Empirical Likelihood Approach to Non-Response

Moshe Feder
& Danny Pfeffermann

Southampton Statistical Sciences Research Institute (S3RI)
University of Southampton
m.feder@soton.ac.uk; moshe.feder@gmail.com

BLS, 12 Nov 2015

Outline of Talk

- Informative non-response

Outline of Talk

- Informative non-response and the respondents distribution

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results
- Testing the Model

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results
- Testing the Model
- Role of the constraints

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results
- Testing the Model
- Role of the constraints
- Imputation

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results
- Testing the Model
- Role of the constraints
- Imputation

NOTE:

- Both informative sampling & informative non-response

Outline of Talk

- Informative non-response and the respondents distribution
- Approaches to non-response
- EL-based approaches to non-response
- Non-response Model
- Estimation
- Simulation set-up & results
- Testing the Model
- Role of the constraints
- Imputation

NOTE:

- Both informative sampling & informative non-response
- Variables observed only for responding units + population means

Informative Non-Response

Informative Non Response

- Informative nonresponse (INR): Response propensity depends on the variable of interest

Informative Non-Response

Informative Non Response

- Informative nonresponse (INR): Response propensity depends on the variable of interest $\Pr(R = 1|y, z) \neq \Pr(R = 1|z)$
  (i.e., response propensity *not* conditionally independent of y given z)

EL-Based Methods for Informative Non-Response

- Qin-Leung-Shao (2002)
- Fang-Hong-Shao (2009)
- Maximization of constrained conditional empirical likelihood: $f_{\mathcal{R}}(y|x) \stackrel{\text{def}}{=} f(y|x, R = 1)$ — our proposed method

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,
  - or even non-parametric model (example to come)

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,
  - or even non-parametric model (example to come)
- The Sample model $f_s(y|x) \stackrel{\text{def}}{=} f(y|x, S = 1)$, and

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,
  - or even non-parametric model (example to come)
- The Sample model $f_s(y|x) \overset{\text{def}}{=} f(y|x, S = 1)$, and
- the respondents' model $f_r(y|x) \overset{\text{def}}{=} f(y|x, S = 1, R = 1)$

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,
  - or even non-parametric model (example to come)
- The Sample model $f_s(y|x) \stackrel{\text{def}}{=} f(y|x, S = 1)$, and
- the respondents' model $f_r(y|x) \stackrel{\text{def}}{=} f(y|x, S = 1, R = 1)$
- ...are usually *different* from the population model $f(y|x)$

The Sample and the Respondents Distributions

- Target model $f(y|x)$ (holding in the population) for example
  - linear regression: $y = \beta' x + \varepsilon$, $E(\varepsilon) = 0$,
  - logistic regression $\Pr(y = 1) = (1 + \exp((-\beta' x)))^{-1}$,
  - or even non-parametric model (example to come)
- The Sample model $f_s(y|x) \stackrel{\text{def}}{=} f(y|x, S = 1)$, and
- the respondents' model $f_r(y|x) \stackrel{\text{def}}{=} f(y|x, S = 1, R = 1)$
- ...are usually *different* from the population model $f(y|x)$ as we'll see next

By Bayes Rule

The Sample Distribution

$$f_s(y|x) = \frac{\Pr(S=1|y,x)}{\Pr(S=1|x)} \cdot f(y|x)$$

By Bayes Rule

The Sample Distribution

$$f_s(y|x) = \frac{\Pr(S=1|y,x)}{\Pr(S=1|x)} \cdot f(y|x)$$

The Respondents Distribution

$$f_r(y|x) = \frac{\Pr(R=1|y,x,S=1)\Pr(S=1|y,x)}{\Pr(R=1|x,S=1)\Pr(S=1|x)} \cdot f(y|x)$$

By Bayes Rule

Informative Non Response
EL
Estimating equations
Estimation
Simulation Results
Testing
Role of Constraints
Imputation
Internet Surveys

The Sample Distribution

$$f_s(y|x) = \frac{\Pr(S = 1|y, x)}{\Pr(S = 1|x)} \cdot f(y|x)$$

The Respondents Distribution

$$\begin{aligned} f_r(y|x) &= \frac{\Pr(R = 1|y, x, S = 1)\Pr(S = 1|y, x)}{\Pr(R = 1|x, S = 1)\Pr(S = 1|x)} \cdot f(y|x) \\ &= \frac{\Pr(R = 1|y, x, S = 1)}{\Pr(R = 1|x, S = 1)} \cdot f_s(y|x) \end{aligned}$$

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1 | y_i, \mathbf{x}_i) = E_p(S_i | y_i, \mathbf{x}_i)$

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1 | y_i, x_i) = E_p(S_i | y_i, x_i)$
  $= E_s(\pi_i^{-1} | y_i, x_i)^{-1}$ (Pfeffermann & Sverchkov 1999)

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1 | y_i, x_i) = E_p(S_i | y_i, x_i)$
  $= E_s(\pi_i^{-1} | y_i, x_i)^{-1}$ (Pfeffermann & Sverchkov 1999)
  $= E_s(w_i | y_i, x_i)^{-1}$

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1|y_i, x_i) = E_p(S_i|y_i, x_i)$
  $= E_s(\pi_i^{-1}|y_i, x_i)^{-1}$ (Pfeffermann & Sverchkov 1999)
  $= E_s(w_i|y_i, x_i)^{-1}$
- So we can estimate $\Pr(S_i = 1|y_i, x_i)$ from the sampling weights of the sampled units

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $$\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1|y_i, x_i) = E_p(S_i|y_i, x_i)$$
  $$= E_s(\pi_i^{-1}|y_i, x_i)^{-1} \text{ (Pfeffermann \& Sverchkov 1999)}$$
  $$= E_s(w_i|y_i, x_i)^{-1}$$
- So we can estimate $\Pr(S_i = 1|y_i, x_i)$ from the sampling weights of the sampled units
- $E_s(w_i|y_i, x_i)$ can be estimated by regressing w_i on (y_i, x_i)

Dealing with informative sampling

The inclusion probability $\pi = \Pr(S = 1)$ and the sampling weight $w = 1/\pi$ may depend on both available and unavailable variables.

- $\tau_i \stackrel{\text{def}}{=} \Pr(S_i = 1|y_i, \boldsymbol{x}_i) = E_p(S_i|y_i, \boldsymbol{x}_i)$
  $= E_s(\pi_i^{-1}|y_i, \boldsymbol{x}_i)^{-1}$ (Pfeffermann & Sverchkov 1999)
  $= E_s(w_i|y_i, \boldsymbol{x}_i)^{-1}$
- So we can estimate $\Pr(S_i = 1|y_i, \boldsymbol{x}_i)$ from the sampling weights of the sampled units
- $E_s(w_i|y_i, \boldsymbol{x}_i)$ can be estimated by regressing w_i on $(y_i, \boldsymbol{x}_i)$
- In the simulation study, we obtained estimates of τ_i, by applying kernel regression of w_i on (y_i, x_i) and their interaction, using the function `npreg` from the R package `np` at its default setting

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

So we need to do something about $\Pr(R_i = 1 | y_i, x_i, S_i = 1)$

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

So we need to do something about $\Pr(R_i = 1 | y_i, \boldsymbol{x}_i, S_i = 1)$ model it:

$$\rho_i = \Pr(R_i = 1 | \boldsymbol{v}_i; \boldsymbol{\gamma}) = g(\boldsymbol{v}_i; \boldsymbol{\gamma})$$

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

So we need to do something about $\Pr(R_i = 1 | y_i, \mathbf{x}_i, S_i = 1)$ model it:

$$\rho_i = \Pr(R_i = 1 | \mathbf{v}_i; \boldsymbol{\gamma}) = g(\mathbf{v}_i; \boldsymbol{\gamma})$$

Informative non-response: $\Pr(R_i = 1 | y_i, \mathbf{x}_i)$ (could be more variables affecting the response probability)

$$\rho_i = \rho(y_i, \mathbf{x}_i; \boldsymbol{\gamma})$$

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

So we need to do something about $\Pr(R_i = 1|y_i, \boldsymbol{x}_i, S_i = 1)$ model it:

$$\rho_i = \Pr(R_i = 1|\boldsymbol{v}_i; \boldsymbol{\gamma}) = g(\boldsymbol{v}_i; \boldsymbol{\gamma})$$

Informative non-response: $\Pr(R_i = 1|y_i, \boldsymbol{x}_i)$ (could be more variables affecting the response probability)

$$\rho_i = \rho(y_i, \boldsymbol{x}_i; \boldsymbol{\gamma})$$

For example

$$\rho_i = \text{logit}^{-1}(\gamma_0 + \gamma_x x_i + \gamma_y y_i)$$

Dealing with informative non-response

In contrast with sampling, non-response is unplanned, and we don't have the analogue of weights...

So we need to do something about $\Pr(R_i = 1 | y_i, \boldsymbol{x}_i, S_i = 1)$ model it:

$$\rho_i = \Pr(R_i = 1 | \boldsymbol{v}_i; \boldsymbol{\gamma}) = g(\boldsymbol{v}_i; \boldsymbol{\gamma})$$

Informative non-response: $\Pr(R_i = 1 | y_i, \boldsymbol{x}_i)$ (could be more variables affecting the response probability)

$$\rho_i = \rho(y_i, \boldsymbol{x}_i; \boldsymbol{\gamma})$$

For example

$$\rho_i = \text{logit}^{-1}(\gamma_0 + \gamma_x x_i + \gamma_y y_i)$$

We will address the issue of *testing* such models later on.

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$
- $y \sim \text{Multinomial}(p_1, \ldots, p_n)$

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$
- $y \sim \text{Multinomial}(p_1, \ldots, p_n)$
- Constraints on $p_1, \ldots, p_n$ may be imposed (estimating equations, using additional information).

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$
- $y \sim \text{Multinomial}(p_1, \ldots, p_n)$
- Constraints on $p_1, \ldots, p_n$ may be imposed (estimating equations, using additional information).
- Empirical likelihood: $\mathscr{L} = \prod_{i=1}^{n} p_i$

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$
- $y \sim \text{Multinomial}(p_1, \ldots, p_n)$
- Constraints on $p_1, \ldots, p_n$ may be imposed (estimating equations, using additional information).
- Empirical likelihood: $\mathscr{L} = \prod_{i=1}^{n} p_i$
- The unconstrained maximum is at $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$

Empirical Likelihood

Basic Idea

- Observed data: $y_1, \ldots, y_n$
- Only observed values present in the population
- Population frequencies: $p_i = \Pr(y = y_i)$, $(\sum_i p_i = 1, p_i \geqslant 0)$
- $y \sim \text{Multinomial}(p_1, \ldots, p_n)$
- Constraints on $p_1, \ldots, p_n$ may be imposed (estimating equations, using additional information).
- Empirical likelihood: $\mathscr{L} = \prod_{i=1}^{n} p_i$
- The unconstrained maximum is at $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$ unless there is additional information (estimating equations) — imposing constraints on the p_i's

Advantages of EL

- No need to specify full distribution — first moments suffice

Advantages of EL

- No need to specify full distribution — first moments suffice
- No need to integrate over $(-\infty, \infty)$ — numerically easier than parametric likelihood

Estimating Equations

Recall that $f_r(y|x) = \dfrac{\Pr(R=1|y,x,S=1)\Pr(S=1|y,x)}{\Pr(R=1|x,S=1)\Pr(S=1|x)} \cdot f(y|x)$

which implies $p_i \propto \dfrac{p_i^{(r)}}{\pi_i \rho_i}$.

If $E(z) = \sum_i p_i z_i = \bar{z}$ is known, $\sum_i p_i(z_i - \bar{z}) = 0$, leading to

$$\sum_i \frac{p_i^{(r)}}{\pi_i \rho_i}(z_i - \bar{z}) = 0$$

and we get our first set of estimating equations

$$\sum_i \frac{p_i^{(r)}}{\pi_i \rho(y_i, x_i; \gamma)}(z_i - \bar{z}) = 0$$

Estimating Equations (cont'd)

The response rate provides additional information.

Let

$\xi_i = \pi_i \rho_i = \Pr(S_i = R_i = 1)$

E_ξ = expectation w.r.t. combined sampling & response

$E_\xi(r) = \sum_{j \in U} \tau_j \rho_j = N \sum_{i \in \mathcal{R}} p_i \tau_i \rho_i = N \sum_{i \in \mathcal{R}} p_i \xi_i = N \bar{\xi}_u$

Thus,

$r \approx N\bar{\xi}_u \quad \Longrightarrow \quad r = N\bar{\xi}_u$

With some algebra

$$\sum_{i \in \mathcal{R}} p_i^{(r)} \left(1 - r/(N\xi_i)\right) = 0$$

Which is equivalent to

$$\sum_{i \in \mathcal{R}} p_i^{(r)} \tau_i^{-1} \rho_i^{-1} = N/r$$

Likelihood

The parameters:

$$p_i^{(r)}, \quad \gamma$$

Likelihood

The parameters:

$$p_i^{(r)}, \quad \gamma$$

and $\quad \beta = \beta(p_1, \ldots, p_r)$ (parametric f_p case)

Likelihood

The parameters:

$$p_i^{(r)}, \quad \gamma$$

and $\quad \beta = \beta(p_1, \ldots, p_r)$ (parametric f_p case)

For ease of notation, denote $q_i = p_i^{(r)}$.

Likelihood

The parameters:

$$p_i^{(r)}, \quad \boldsymbol{\gamma}$$

and $\quad \boldsymbol{\beta} = \boldsymbol{\beta}(p_1, \ldots, p_r)$ (parametric f_p case)
For ease of notation, denote $q_i = p_i^{(r)}$.
The likelihood:

$$\mathscr{L} = \mathscr{L}(\mathbf{q}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} q_i$$

Likelihood

The parameters:

$$p_i^{(r)}, \quad \gamma$$

and $\quad \boldsymbol{\beta} = \boldsymbol{\beta}(p_1, \ldots, p_r)$ (parametric f_p case)
For ease of notation, denote $q_i = p_i^{(r)}$.
The likelihood:

$$\mathscr{L} = \mathscr{L}(\mathbf{q}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} q_i$$

subject to the constraints

$$\sum_i q_i \frac{\mathbf{z}_i - \bar{\mathbf{z}}}{\pi_i \rho(y_i, \mathbf{x}_i; \boldsymbol{\gamma})} = 0$$

$$\sum_{i \in \mathcal{R}} \frac{q_i}{\tau_i \rho(y_i, \mathbf{x}_i; \boldsymbol{\gamma})} = N/r$$

Estimation of the response model

Maximize the likelihood $\mathscr{L}$ subject to the constraints

Estimation of the response model

Maximize the likelihood $\mathscr{L}$ subject to the constraints

$$\max_{\mathbf{q},\boldsymbol{\gamma}} \mathscr{L}(\mathbf{q},\boldsymbol{\gamma}), \quad \text{s.t.} \quad A(\boldsymbol{\gamma})\mathbf{q} = 0$$

Estimation of the response model

Maximize the likelihood $\mathscr{L}$ subject to the constraints

$$\max_{\mathbf{q},\boldsymbol{\gamma}} \mathscr{L}(\mathbf{q},\boldsymbol{\gamma}), \quad \text{s.t.} \quad A(\boldsymbol{\gamma})\mathbf{q} = 0$$

This maximization problem is equivalent to the maximization $\max_{\boldsymbol{\gamma}} \mathrm{G}(\boldsymbol{\gamma})$ where $\mathrm{G}(\boldsymbol{\gamma})$ is the *profile likelihood* of $\boldsymbol{\gamma}$, defined as

$$\mathrm{G}(\boldsymbol{\gamma}) = \max\{\Pi(\mathbf{q}) : A(\boldsymbol{\gamma})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

This maximization can be done using the Owen (2013) algorithm.

Estimation of the response model

Maximize the likelihood $\mathscr{L}$ subject to the constraints

$$\max_{\mathbf{q},\boldsymbol{\gamma}} \mathscr{L}(\mathbf{q},\boldsymbol{\gamma}), \quad \text{s.t.} \quad A(\boldsymbol{\gamma})\mathbf{q} = 0$$

This maximization problem is equivalent to the maximization $\max_{\boldsymbol{\gamma}} \mathrm{G}(\boldsymbol{\gamma})$ where $\mathrm{G}(\boldsymbol{\gamma})$ is the *profile likelihood* of $\boldsymbol{\gamma}$, defined as

$$\mathrm{G}(\boldsymbol{\gamma}) = \max\{\Pi(\mathbf{q}) : A(\boldsymbol{\gamma})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

This maximization can be done using the Owen (2013) algorithm.

Variance estimation of $\hat{\boldsymbol{\gamma}}$:

Inverse Hessian of the *profile* likelihood G

Estimation of the response model

Maximize the likelihood $\mathscr{L}$ subject to the constraints

$$\max_{\mathbf{q},\boldsymbol{\gamma}} \mathscr{L}(\mathbf{q},\boldsymbol{\gamma}), \quad \text{s.t.} \quad A(\boldsymbol{\gamma})\mathbf{q} = 0$$

This maximization problem is equivalent to the maximization $\max_{\boldsymbol{\gamma}} \mathrm{G}(\boldsymbol{\gamma})$ where $\mathrm{G}(\boldsymbol{\gamma})$ is the *profile likelihood* of $\boldsymbol{\gamma}$, defined as

$$\mathrm{G}(\boldsymbol{\gamma}) = \max\{\Pi(\mathbf{q}) : A(\boldsymbol{\gamma})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

This maximization can be done using the Owen (2013) algorithm.

Variance estimation of $\hat{\boldsymbol{\gamma}}$:

Inverse Hessian of the *profile* likelihood G

or parametric bootstrap

Estimation of the target model

Once we have estimates $\hat{\boldsymbol{\gamma}}$, we solve the maximization problem

$$\hat{\mathbf{q}} = \arg\max\{\Pi(\mathbf{q}) : A(\hat{\boldsymbol{\gamma}})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

Recall that $p_i \propto \frac{p_i^{(r)}}{\pi_i \rho_i} = \frac{q_i}{\pi_i \rho_i}$, so we get $\hat{p}_1, \ldots, \hat{p}_r$.

Estimation of the target model

Once we have estimates $\hat{\boldsymbol{\gamma}}$, we solve the maximization problem

$$\hat{\mathbf{q}} = \arg\max\{\Pi(\mathbf{q}) : A(\hat{\boldsymbol{\gamma}})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

Recall that $p_i \propto \frac{p_i^{(r)}}{\pi_i \rho_i} = \frac{q_i}{\pi_i \rho_i}$, so we get $\hat{p}_1, \ldots, \hat{p}_r$.
The regression coefficient $\boldsymbol{\beta} = \boldsymbol{\beta}(\mathbf{p})$ is the solution of the appropriate estimating equation

Estimation of the target model

Once we have estimates $\hat{\gamma}$, we solve the maximization problem

$$\hat{\mathbf{q}} = \arg\max\{\Pi(\mathbf{q}) : A(\hat{\gamma})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

Recall that $p_i \propto \frac{p_i^{(r)}}{\pi_i \rho_i} = \frac{q_i}{\pi_i \rho_i}$, so we get $\hat{p}_1, \ldots, \hat{p}_r$.
The regression coefficient $\beta = \beta(\mathbf{p})$ is the solution of the appropriate estimating equation

Examples:
$\hat{\beta} = (X'D_\mathbf{p}X)^{-1}X'D_\mathbf{p}\mathbf{y}$ (linear regression) where $D_\mathbf{p} = \text{diag}(\mathbf{p})$ or $X'D_\mathbf{p}(Y - \mu) = 0$ (logistic regression)

Estimation of the target model

Once we have estimates $\hat{\boldsymbol{\gamma}}$, we solve the maximization problem

$$\hat{\mathbf{q}} = \arg\max\{\Pi(\mathbf{q}) : A(\hat{\boldsymbol{\gamma}})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

Recall that $p_i \propto \frac{p_i^{(r)}}{\pi_i \rho_i} = \frac{q_i}{\pi_i \rho_i}$, so we get $\hat{p}_1, \ldots, \hat{p}_r$.
The regression coefficient $\boldsymbol{\beta} = \boldsymbol{\beta}(\mathbf{p})$ is the solution of the appropriate estimating equation

Examples:
$\hat{\boldsymbol{\beta}} = (X'D_{\mathbf{p}}X)^{-1}X'D_{\mathbf{p}}\mathbf{y}$ (linear regression) where $D_{\mathbf{p}} = \text{diag}(\mathbf{p})$
or $X'D_{\mathbf{p}}(Y - \boldsymbol{\mu}) = 0$ (logistic regression)
So,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\hat{p}_1, \ldots, \hat{p}_r)$$

Estimation of the target model

Informative Non Response
EL
Estimating equations
Estimation
Simulation Results
Testing
Role of Constraints
Imputation
Internet Surveys

Once we have estimates $\hat{\gamma}$, we solve the maximization problem

$$\hat{\mathbf{q}} = \arg\max\{\Pi(\mathbf{q}) : A(\hat{\gamma})\mathbf{q} = 0 \quad \& \quad \mathbf{q} \in \Omega\}$$

Recall that $p_i \propto \frac{p_i^{(r)}}{\pi_i \rho_i} = \frac{q_i}{\pi_i \rho_i}$, so we get $\hat{p}_1, \ldots, \hat{p}_r$.
The regression coefficient $\beta = \beta(\mathbf{p})$ is the solution of the appropriate estimating equation

Examples:
$\hat{\beta} = (X'D_{\mathbf{p}}X)^{-1}X'D_{\mathbf{p}}\mathbf{y}$ (linear regression) where $D_{\mathbf{p}} = \text{diag}(\mathbf{p})$
or $X'D_{\mathbf{p}}(Y - \mu) = 0$ (logistic regression)
So,

$$\hat{\beta} = \beta(\hat{p}_1, \ldots, \hat{p}_r)$$

or we can estimate $f_p(y|\mathbf{x}$ non-parametrically (example below)

Non-parametric estimation of the population model

The proposed approach does not require any specification of a model for $f_p(y|\mathbf{x})$.

Non-parametric estimation of the population model

The proposed approach does not require any specification of a model for $f_p(y|\mathbf{x})$.
In fact, once estimates $\hat{p}_i$ are obtained, and thus an estimate $\hat{F}$ of the population distribution is available, non-parametric estimation of $f_p(y|\mathbf{x})$ can be made, for example using smooth polynomial spline.

Non-parametric estimation of the population model

The proposed approach does not require any specification of a model for $f_p(y|\mathbf{x})$.
In fact, once estimates $\hat{p}_i$ are obtained, and thus an estimate $\hat{F}$ of the population distribution is available, non-parametric estimation of $f_p(y|\mathbf{x})$ can be made, for example using smooth polynomial spline.

Example: $y = \eta + \varepsilon$

where

$\eta = 0.2 + 0.03x + 0.4x^2$, restricted to $[-0.1, 0.9]$

and $\varepsilon \overset{\text{iid}}{\sim} N(0, 0.25)$

i.e.

$$y = \max(\min(0.2 + 0.03x + 0.4x^2, 0.9), -0.1) + \varepsilon$$

Example: Non-parametric estimation of f_p

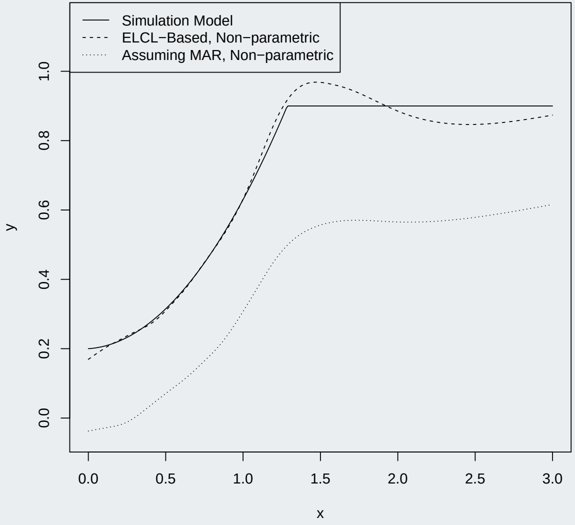


Figure : Results using a smooth cubic spline (average over 10 samples)

Simulation

Target model:
$\text{logit}[\Pr(y = 1)] = \beta_0 + \beta_1 x$, $(\beta_0 = -0.8, \beta_1 = 0.8)$
Response model:
$\text{logit}[\Pr(R = 1)] = \gamma_0 + \gamma_x x + \gamma_y y, (\gamma_0 = 0.7, \gamma_x = 0.5, \gamma_y = -1.5)$

Simulation

Target model:
$\text{logit}[\Pr(y=1)] = \beta_0 + \beta_1 x$, $(\beta_0 = -0.8, \beta_1 = 0.8)$
Response model:
$\text{logit}[\Pr(R=1)] = \gamma_0 + \gamma_x x + \gamma_y y, (\gamma_0 = 0.7, \gamma_x = 0.5, \gamma_y = -1.5)$

500 samples
$N = 10000$
n: $3395 - 3625$; r: $2227 - 2455$
Response rate: $0.64 - 0.69$

Estimates—Response Model

Point Estimates:

	γ_0	γ_x	γ_y
True	0.700	0.500	−1.50
Mean estimate	0.736	0.499	−1.53

Variance Estimates:

	γ_0	γ_x	γ_y
Empirical STD	0.214	0.212	0.319
SQRT Mean variance estimate	0.220	0.212	0.339

Variance estimation — using the inverse Hessian of the profile likelihood.

Estimates—Target Model

Method	Mean Est.		Empirical STD		SQRT Mean Var. Est.	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
TRUE	−0.800	0.800				
MAR UW	−2.665	0.966	0.105	0.093	0.111	0.095
MAR PW	−1.559	0.962	0.106	0.093	0.113	0.097
CREL	−0.797	0.799	0.178	0.104	0.188	0.108

MAR UW = ignoring response mechanism, unweighted
MAR PW = ignoring response mechanism, probability weighted
CREL = proposed method
Variance estimation — parametric Bootstrap (60 samples)

Hosmer-Lemeshow-Type Test Statistic

Hosmer and Lemeshow (1980, 2000): test statistic for the case of logistic regression.
Sample partitioned into G groups of approximately equal size, based on the predicted probability of 'success.' Test statistic:

$$\hat{C} = \sum_{k=1}^{G} \frac{(o_k - n_k \bar{\mu}_k)^2}{n_k \bar{\mu}_k (1 - \bar{\mu}_k)},$$

o_k = number of observed 'successes' in group k, n_k = size of the group, $\bar{\mu}_k$ = the mean number of the estimated probabilities of success, $\bar{\mu}_k = \sum_{i \in G_k} \hat{\mu}_i / n_k$, where G_k is the kth group, and where $\mu_i = \Pr(y_i = 1, I_i = 1, R_i = 1 | x_i)$

Distribution of the Test Statistic

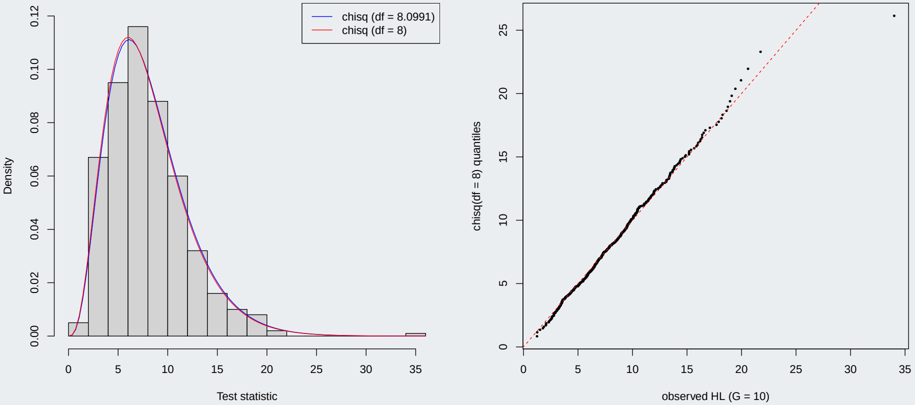
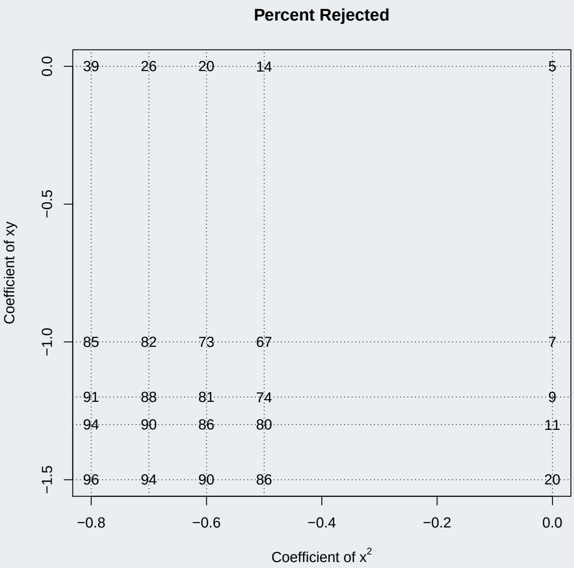


Figure : Distribution of $X_{HL,G=10}$

Power


Percent Rejected
Coefficient of xy
Coefficient of x2
0.0
−0.5
−1.0
−1.5
−0.8
−0.6
−0.4
−0.2
0.0
39 26 20 14 5
85 82 73 67 7
91 88 81 74 9
94 90 86 80 11
96 94 90 86 20

Constraints: What matters

The population mean of the constraining variables $\mathbf{c}$ need to be known. We use proxy variables for the model variables $y, \mathbf{x}$.

A limited study of: (1) Best choice of variables for which the auxiliary variables are proxy, (2) how close should the auxiliary variables be to the variables they are proxy for, and (3) how many auxiliary variables to choose.

The correlation with the target model's variables is more important than the 'noise'.

Even just two auxiliary variables may be enough.

Constraints: What matters (cont'd)

$$\mathbf{c}_i = (1, x_i, y_i, x_i y_i, x_i^2, x_i^2 y_i)' + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_c^2 I_6)$$

		β_0	β_1	γ_0	γ_x	γ_y
Simulation value		-0.800	0.800	0.700	0.500	-1.500
$\sigma_c = 0.5$	c_2, c_3	-0.796	0.797	0.699	0.516	-1.501
$\sigma_c = 1.0$	c_0, c_1, c_4	-1.098	0.761	1.759	0.314	-1.256
$\sigma_c = 9.0$	c_2, c_3	-0.800	0.671	1.526	1.278	-2.112
$\sigma_c = 9.0$	c_2, c_3, c_5	-0.764	0.628	1.656	0.778	-2.023
Six uncorrelated		-1.051	0.753	1.143	2.280	-1.310

Imputation

- Two scenarios: (1) auxiliary variables x and the sampling weights w are available for the non-respondents. (2) x, w not available
- Goal: impute observations for each non-respondent i in such a way that the distribution of $(y, \mathbf{x}, w)'$ in the combined data is the same as in that in the original sample, including the unobserved data.

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled.

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Case 1: Subjects respond no more than once (if at all)

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Case 1: Subjects respond no more than once (if at all)

Case 2: Subjects may respond multiple times and multiple responses from same subject cannot be identified as such

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Case 1: Subjects respond no more than once (if at all)

Case 2: Subjects may respond multiple times and multiple responses from same subject cannot be identified as such

Case 3: Subjects may respond multiple times and multiple responses from same subject *can* be identified as such

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Case 1: Subjects respond no more than once (if at all)

Case 2: Subjects may respond multiple times and multiple responses from same subject cannot be identified as such

Case 3: Subjects may respond multiple times and multiple responses from same subject *can* be identified as such

In Case 1, model $\xi_i \stackrel{\text{def}}{=} \Pr(i \in S) = \xi(y_i, x_i; \gamma)$. Rest is similar to the 'usual' case.

Internet Surveys

Our approach can be applied to Internet survey data where the selection probability is unknown and therefore must be modelled. Need population means for some auxiliary variables

Case 1: Subjects respond no more than once (if at all)

Case 2: Subjects may respond multiple times and multiple responses from same subject cannot be identified as such

Case 3: Subjects may respond multiple times and multiple responses from same subject *can* be identified as such

In Case 1, model $\xi_i \stackrel{\text{def}}{=} \Pr(i \in S) = \xi(y_i, x_i; \gamma)$. Rest is similar to the 'usual' case.
Cases 2 & 3 are more involved, and will not be discussed here

Acknowledgements

- Joint work with Danny Pfeffermann
- Sanjay Chaudhuri for R code
- Funded by the Economic and Social Research Council (ESRC) of the United Kingdom

Acknowledgements

- Joint work with Danny Pfeffermann
- Sanjay Chaudhuri for R code
- Funded by the Economic and Social Research Council (ESRC) of the United Kingdom