Statistical Analysis with Missing Data

Module 10

Missing Not at Random

UNIVERSITY OF MICHIGAN

Likelihood methods for missing not at random (MNAR) models

- Statistical model + incomplete data $\Rightarrow$ Likelihood
- For mnar missing data, need a model for the missing-data mechanism as well as a model for the data
- ML estimates, large sample standard errors
  - tools like factored likelihood, EM apply to these models too
- Bayes posterior distribution
  - stochastic simulation methods apply
- But beware of unidentified or poorly-identified models, leading to numerical problems (e.g. lack of convergence).

(Little and Rubin 2019, chapt. 15)

Models for Y and M

- Let (y_i, m_i) denote the complete-data vector and missing-data indicator for the ith unit, and assume independence across units. Two generic modeling approaches are:

Selection models, which factor:

$$f(y_i, m_i | \theta, \psi) = f(y_i | \theta) \times f(m_i | y_i, \psi)$$

$$\boxed{\text{complete-data model}} \times \boxed{\text{model for md mechanism}}$$

Pattern-mixture models, which factor:

$$f(y_i, m_i | \phi, \gamma) = f(y_i | m_i, \phi) \times f(m_i | \gamma)$$

$$\boxed{\text{model for } y\text{'s within pattern } m_i} \times \boxed{\text{probability of pattern } m_i}$$

EM for Selection Models

E-Step: given current estimates $\left(\theta^{(t)},\psi^{(t)}\right)$, compute

$$Q\left(\theta,\psi\middle|\theta^{(t)},\psi^{(t)}\right)=$$

$$\int \ell\left(\theta,\psi\middle|X,Y_{(0)},Y_{(1)},M\right)f\left(Y_{(1)}\middle|X,Y_{(0)},M,\theta=\theta^{(t)},\psi=\psi^{(t)}\right)dY_{(1)}$$

where $\ell\left(\theta,\psi\middle|X,Y_{(0)},Y_{(1)},M\right)$ is the complete-data loglikelihood

M-Step: find new estimates $\left(\theta^{(t+1)},\psi^{(t+1)}\right)$ to maximize Q:

$$Q\left(\theta^{(t+1)},\psi^{(t+1)}\middle|\theta^{(t)},\psi^{(t)}\right)\geq Q\left(\theta,\psi\middle|\theta^{(t)},\psi^{(t)}\right)\quad \text{for all } \theta,\psi.$$

Won't converge unless parameters are identified.

Selection or Pattern-Mixture Models?

- Selection models are:
  - more natural substantive formulation of model, if inference concerns the entire population
  - more common approach in literature
  - sensitive to specification of the form of the missing-data mechanism, which is often not well understood
- Pattern-mixture models are:
  - More natural when interest is in population strata defined by missing-data pattern
  - closer to the form of the data, often easier to understand and simpler to fit
  - Some models avoid specifying the form of the md mechanism, which is incorporated indirectly via *parameter restrictions*.

Known missingness mechanisms

- There are examples of mechanisms where missingness depends on missing data, but are known, in the sense that they don't involve unknown parameters
  - E.g. grouped or rounded data (Section 15.2 of LR)

Grouped normal regression with covariates (Example 15.4)

$$y_i \sim_{\text{iid}} N(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}, \sigma^2)$$

y_i is observed for $i = 1, \ldots, r$

$y_i \in$ group j if $y_i \in (a_j, b_j)$ $(j = 1, \ldots, J)$

Complete-data sufficient statistics are

$$\sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i x_{ik} (k = 1, \ldots, p)$$

Grouped regression data

E-step:

$$E\left(\sum_{i=1}^{n} y_i \,\middle|\, Y_{(0)}, M, \theta = \theta^{(t)}\right) = \sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} \hat{y}_i^{(t)},$$

$$E\left(\sum_{i=1}^{n} y_i x_{ik} \,\middle|\, Y_{(0)}, M, \theta = \theta^{(t)}\right) = \sum_{i=1}^{r} y_i x_{ik} + \sum_{i=r+1}^{n} \hat{y}_i^{(t)} x_{ik}, \quad k = 1, 2, \ldots, p,$$

$$E\left(\sum_{i=1}^{n} y_i^2 \,\middle|\, Y_{(0)}, M, \theta = \theta^{(t)}\right) = \sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{n} \hat{y}_i^{(t)2} + \hat{s}_i^{(t)2},$$

$$\hat{y}_i^{(t)} = \mu_i^{(t)} + \sigma^{(t)} \delta_i^{(t)}, \hat{s}_i^{(t)2} = \sigma^{(t)2}\left(1 - \gamma_i^{(t)}\right), \mu_i^{(t)} = \beta_0^{(t)} + \sum\nolimits_{k=1}^{p} \beta_k^{(t)} x_{ik}$$

$$\delta_i^{(t)} = -\frac{\phi\left(d_i^{(t)}\right) - \phi\left(c_i^{(t)}\right)}{\Phi\left(d_i^{(t)}\right) - \Phi\left(c_i^{(t)}\right)}, \gamma_i^{(t)} = \delta_i^{(t)2} + \frac{d_i^{(t)} \phi\left(d_i^{(t)}\right) - c_i^{(t)} \phi\left(c_i^{(t)}\right)}{\Phi\left(d_i^{(t)}\right) - \Phi\left(c_i^{(t)}\right)},$$

$\phi()$ = standard normal density, $\Phi()$ = standard normal cdf

Grouped regression data

M-step:

Compute new $\theta^{(t+1)}$ using estimated complete-data sufficient statistics from E-Step

An application is when people unwilling to give exact amount of income are asked to report income in known categories -- apply model with Y = log(income)

Setting $J = 1, a_1 = -\infty, b_1 = 0$ yields Tobit model in economics

A more complex example is Example 15.6 -- coarsened data from Health and Retirement Survey

MNAR models with unknown parameters

- MNAR models involving unknown parameters are more difficult to deal with, because of lack of identifiability of the parameters
- See for example regression with missing outcomes:

Ex 15.7: Regression with missing outcomes

Heckman Selection model:

$(y_i \mid x_i) \sim_{\text{ind}} N(\beta_0 + \beta_1 x_i, \sigma^2)$

$(m_i \mid x_i, y_i) \sim_{\text{ind}} \text{BERN}\left(\Phi(\psi_0 + \psi x_i + \lambda y_i\right)$

MAR if $\lambda = 0$

No real information about λ in the sample,

because y_i is only observed when $m_i = 0$

Ex 15.8: Regression with missing outcomes

Pattern-mixture model

$(y_i \mid m_i = m, x_i) \sim_{\text{ind}} N(\beta_0^{(m)} + \beta^{(m)} x_i, \sigma^{(m)2})$

$(m_i \mid x_i) \sim_{\text{ind}} \text{BERN}\left(\Phi(\psi_0 + \psi x_i\right)$

$\left(\beta_0^{(1)}, \beta^{(1)}\right)$ not estimable because y_i is missing when $m_i = 1$

Or: $(y_i \mid m_i = m, x_i) \sim_{\text{ind}} N(\beta_0^{(m)} + \beta x_i, \sigma^2)$

Then $\delta = \beta_0^{(1)} - \beta_0^{(0)}$ characterizes difference between respondents and nonrespondents

5 approaches to lack of estimability

(a) Follow up a sample of nonrespondents, and incorporate this information into the main analysis.
(b) Adopt a Bayesian approach, assigning the parameters prior distributions. Inferences tend to be sensitive to the choice of prior distribution.
(c) Impose additional restrictions on model parameters, such as on the regression coefficients in the Heckman model.
(d) Conduct analysis to assess sensitivity of inferences for quantities of interest to different choices of the values of parameters poorly estimated from the data.
(e) Selectively discard data to avoid modeling the missingness mechanism.

15.3.2. Following up a Subsample of Nonrespondents.

- One way to reduce sensitivity of inference to MNAR nonresponse is to follow up at least some nonrespondents to obtain the desired information.
- The subsample of nonrespondents can then be weighted up to all nonrespondents, or used to multiply impute the remaining nonrespondents
- Even if only a few nonrespondents are followed up, these can be exceedingly helpful in reducing sensitivity of inference
- See simulation experiment described in Example 15.9

15.3.3 Bayesian Approach

Rubin (1977) proposes the following model for scalar x_i:

$$(y_i \mid x_i, M_i = m) \sim_{\text{ind}} N\left(\phi_{(m)} + \beta_{(m)}(x_i - \bar{x}_0), \sigma^2\right)$$

with priors:

$$p(\phi_{(0)}, \beta_{(0)}, \log\sigma^2) \propto \text{const.},$$

$$p(\beta_{(1)} \mid \phi_{(0)}, \beta_{(0)}, \sigma^2) \sim N_q(\beta_{(0)}, \psi_1^2 \beta_{(0)} \beta_{(0)}^T),$$

$$p(\phi_{(1)} \mid \phi_{(0)}, \beta_{(0)}, \beta_{(1)}, \sigma^2) \sim N(\phi_{(0)}, \psi_2^2 \phi_{(0)}^2)$$

width of 95% posterior credibility interval is

$$3.92\bar{y}_{(0)}\left(\psi_1^2 h_1^2 + \psi_2^2 h_2^2 + h_3^2\right)^{1/2}$$

Bayesian Approach

Posterior distribution of $\bar{y}$ is Normal with mean

$\bar{y}_{(0)} + \hat{\beta}_{(0)}^{T}(\bar{x} - \bar{x}_{(0)})$ and variance $\bar{y}_{(0)}^{2}\left(\psi_1^2 h_1^2 + \psi_2^2 h_2^2 + h_3^2\right)$

$$h_1^2 = (\sigma^2 / \bar{y}_{(0)}^2)\left[\left(\hat{\beta}_{(0)}^{T}(\bar{x} - \bar{x}_{(0)})\right)^2 / \sigma^2 + \left(\bar{x} - \bar{x}_{(0)}\right)^T S_{xx}^{-1}\left(\bar{x} - \bar{x}_{(0)}\right)\right]$$

$$h_2^2 = p^2\left\{1 + \sigma^2 \big/ \left(r\bar{y}_{(0)}\right)\right\}$$

$$h_3^2 = \left(\sigma^2 \big/ \bar{y}_{(0)}^2\right)\left((p/r) + \left(\bar{x} - \bar{x}_{(0)}\right)^T S_{xx}^{-1}\left(\bar{x} - \bar{x}_{(0)}\right)\right)$$

Application (Ex 15.10)

- Rubin (1977) illustrates the method of Example 15.10 with data from a survey of 660 schools, 472 of which filled out a compensatory reading questionnaire consisting of 80 items.
- Twenty-one dependent variables (Y's) and 35 background variables (X's) describing the school and the socioeconomic status and achievement of the students were considered.
- The dependent variables in the study measure characteristics of compensatory reading in the form of frequency with which they were present, and were scaled to lie between zero (never) and one (always).

Table 15.4 Example 15.10, Widths of Subjective 95 Percent Intervals of Finite Population Means $\bar{y}$, as Percentages of Observed Sample Mean, $\bar{y}_{(0)}$.

	$\psi_1 = 0$				$\psi_1 = 0.4$			
Variable	$\psi_2 = 0$	$\psi_2 = 0.1$	$\psi_2 = 0.2$	$\psi_2 = 0.4$	$\psi_2 = 0$	$\psi_2 = 0.1$	$\psi_2 = 0.2$	$\psi_2 = 0.4$
17B	5.6	8.0	12.7	23.7	6.0	8.3	12.9	23.6
18A	7.9	9.8	13.9	24.2	8.1	9.9	14.0	24.3
18B	15.4	16.5	19.3	27.8	16.6	17.6	20.2	28.5
23A	2.1	6.1	11.6	22.9	2.3	6.1	11.6	22.9
23C	2.0	6.0	11.6	22.9	2.0	6.1	11.6	22.9
32A	1.2	5.8	11.5	22.8	1.2	5.8	11.5	22.8
32D	1.1	5.8	11.4	22.8	1.1	5.8	11.4	22.8

Description of Outcome Variables:

17B: Compensatory reading carried out during school hours released from other classwork
18A: Compensatory reading carried out during time released from social studies, science, and/or foreign language
18B: Compensatory reading carried out during time released from mathematics
23A: Frequency of organizing compensatory reading class into groups by reading grade level
23C: Frequency of organizing compensatory reading class into groups by shared interests
32A: Compensatory reading teaches textbooks other than basal readers
32D: Compensatory reading teaches teacher-prepared materials

Quantity ψ_2 has major impact on interval widths, exhibiting sensitivity of Bayesian approach to choice of prior

5 approaches to lack of estimability

(a) Follow up a sample of nonrespondents, and incorporate this information into the main analysis.
(b) Adopt a Bayesian approach, assigning the parameters prior distributions. Inferences tend to be sensitive to the choice of prior distribution.
(c) Impose additional restrictions on model parameters, such as on the regression coefficients in the Heckman model.
(d) Conduct analysis to assess sensitivity of inferences for quantities of interest to different choices of the values of parameters poorly estimated from the data.
(e) Selectively discard data to avoid modeling the missingness mechanism.

5 approaches to lack of estimability

(a) Follow up a sample of nonrespondents, and incorporate this information into the main analysis.

(b) Adopt a Bayesian approach, assigning the parameters prior distributions. Inferences tend to be sensitive to the choice of prior distribution.

(c) Impose additional restrictions on model parameters, such as on the regression coefficients in the Heckman model.

(d) Conduct analysis to assess sensitivity of inferences for quantities of interest to different choices of the values of parameters poorly estimated from the data.

(e) Selectively discard data to avoid modeling the missingness mechanism.

More on Heckman Selection Model

$Y_1 =$ outcome variable, with missing values,

M = missing data indicator for Y_1

$Y_2 =$ latent variable, not observed

X_1, X_2 = covariates, fully observed

$$y_{1i} = x_{1i}\beta_1 + u_{1i}$$

$$y_{2i} = x_{2i}\beta_2 + u_{2i} \text{ (latent -- not observed)}$$

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix}\right] \text{ (MAR if } \rho = 0\text{)}$$

$$m_i = \begin{cases} 0, y_{2i} > 0 \text{ -- } y_{1i} \text{ observed} \\ 1, y_{2i} \le 0 \text{ -- } y_{1i} \text{ missing} \end{cases}$$

Heckman Selection Model

- In Heckman's application

  $Y_1 = \text{wife's wages, if working } (M = 0)$

  $Y_2 = \text{ wife's wages - "reservation wages"}$

- Normality of Y_2 implies a probit model for M given X and Y_1 – hence the model is the bivariate normal probit selection model with added covariates
- Parameters estimated by maximum likelihood or two-stage least squares (see Chapter 15 of LR)
- Not well identified unless the set of predictor variables in the models for Y_2 and Y_1 differ. –often a strong assumption.

Application to Current Population Survey Income nonresponse

- Missing (Y_1): Earnings Variables (Box-Cox transformed)
- Observed (X): Education (6 categories), Years of market experience (4 linear splines, Exp 0-5, Exp 5- 10, Exp 10-20, Exp 20 +), Pr(first year of market experience (Prob 1), Region (South or other), Child of household head (1 = yes, 0 = no), Other relative of household head (yes, no), Personal interview (1 = yes, 0 = no), and Year in survey (1 or 2).
- Last four variables omitted from the earnings equation; that is, their coefficients in the vector β were set equal to zero. The variables education, years of market experience, and region were omitted from the response equation.

Table 15.5 Example 15.11, Estimates for the Regression of ln (Earnings) on Covariates: 1980 Current Population Survey.

Variable	OLS on 27,909 Respondents	Coefficient (β_1) from ML for Selection Model on 32,879 Survey Units
Constant	9.5013 (.0039)	9.6816 (.0051)
Sch 8	.2954 (.0245)	.2661 (.0202)
Sch 9–11	.3870 (.0206)	.3692 (.0169)
Sch 12	.6881 (.0188)	.6516 (.0158)
Sch 13–15	.7986 (.0201)	.7694 (.0176)
Sch 16 +	1.0519 (.0199)	1.0445 (.0178)
Exp 0–5	– .0225 (.0119)	– .0294 (.0111)
Exp 5–10	.0534 (.0038)	.0557 (.0039)
Exp 10–20	.0024 (.0016)	.0240 (.0016)
Exp 20 +	– .0052 (.0008)	– .0036 (.0008)
Prob 1	–1.8136 (.1075)	–1.7301 (.0945)
South	– .0654 (.0087)	– .0649 (.0085)
$\rho = \lambda\sigma / \sqrt{1+\lambda^2\sigma^2}$	0	-.6842

reflects sensitivity of the correction to skewness in the transformed-income respondent residuals

Table 15.6 Example 15.11, The Maximized Loglikelihood as a Function of γ, with Associated Values of $\hat{\rho}$.

γ	Maximized Loglikelihood	$\hat{\rho}$
0	–300,613.4	–0.6812
0.45	–298,169.7	–0.6524
1.0	–300,563.1	0.8569

Source: Lillard, Smith, and Welch (1982).

$\hat{\gamma}$
γ

CPS Example

- Lillard, Smith, and Welch's best-fitting model predicts income amounts for nonrespondents that are 73% larger on average than imputations supplied by the Census Bureau, which uses a hot deck method that assumes ignorable nonresponse. However, this large adjustment is founded on the normal assumption for the population residuals from the fitted model.
- It is quite plausible that nonresponse is MAR and the unrestricted residuals follow the same (skewed) distribution as that in the respondent sample.
- Comparisons of Census Bureau imputations with IRS income amounts from matched CPS/IRS files do not indicate substantial underestimation from the CPS hot-deck imputations (David et al., 1986).

Ex. 15.12. AIDS Incidence from a Demographic Survey with Randomly-Assigned Interviewers

- Survey estimates of HIV prevalence may be biased because of nonresponse, with refusal to participate in the HIV test plausibly related to HIV status, even after controlling for other observed variables.
- Janssens et al. (2014) applied the Heckman model to data on HIV prevalence from a survey of 1,992 individuals in urban Namibia that included an HIV test. The specification was justified by random assignment of interviewers to households in the sample.
- Specifically, their model took the form:

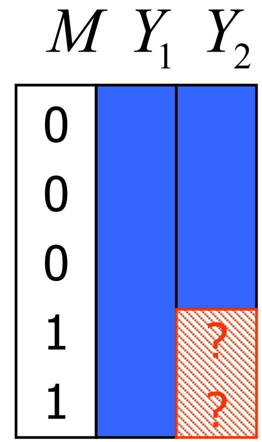
$$\Pr(y_i = 1 \mid x_i, z_i, \beta, \psi) = \Phi(\beta_0 + \beta_1 x_i)$$

$$\Pr(m_i = 1 \mid y_i, x_i, z_i, \beta, \psi) = \Phi(\psi_0 + \psi_1 x_i + \psi_2 z_i + \psi_3 y_i),$$

- z_i = identity codes of the nurses who administered the HIV test. These variables were omitted in model for HIV, because these nurses can influence the missingness rate, but not the outcome of the HIV test directly, as they are randomly assigned to households.

Normal Pattern-Mixture Model

- In pattern-mixture models, lack of information about the mechanism is reflected in unidentified parameters for the incomplete patterns
  - e.g. consider the bivariate normal pattern-mixture model for monotone pattern in figure:



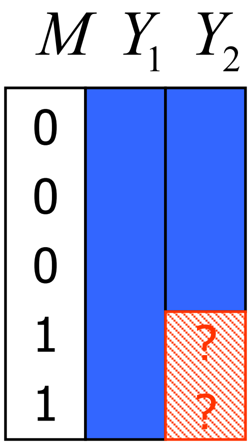
$$\left(y_i \mid m_i = j\right) \sim_{\text{iid}} N(\mu^{(j)}, \Sigma^{(j)}),\ j = 0,1$$

$$\Pr(m_i = j) = \text{BERN}(\pi)$$

$$(\mu^{(0)}, \Sigma^{(0)}, \mu_1^{(1)}, \sigma_{11}^{(1)}) \text{ are identified;}$$

$$\text{3 parameters, } (\mu_2^{(1)}, \sigma_{22}^{(1)}, \sigma_{12}^{(1)}), \text{ are not identified}$$

Normal Pattern-Mixture Model



Assumptions about the mechanism yield restrictions on the parameters that may identify the model

For example, if missingness of Y_2 depends on Y_2 but not on Y_1, that is:

$\Pr(m_{i2} = 1 \mid y_{i1}, y_{i2}) = g(y_{i2})$ for any function g

Then the regression of y_{i1} on y_{i2} is same for complete $(m_i = 0)$ and incomplete $(m_i = 1)$ cases

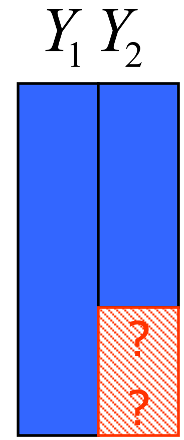
$\Rightarrow \beta_{10\cdot 2}^{(0)} = \beta_{10\cdot 2}^{(1)},\ \beta_{12\cdot 2}^{(0)} = \beta_{12\cdot 2}^{(1)},\ \sigma_{11\cdot 2}^{(0)} = \sigma_{11\cdot 2}^{(1)}$

These 3 restrictions just identify the model parameters

ML for Normal Pattern-Mixture Model

- More generally, if

$$\Pr(m_{i2} = 1 \mid y_{i1}, y_{i2}) = g(y_{i1} + \lambda y_{i2})$$



for known λ, the ML estimate of the marginal mean of Y_2 is:

$$\hat{\mu}_2 = \bar{y}_2 + \hat{\beta}_{21\cdot1}^{(\lambda)}(\hat{\mu}_1 - \bar{y}_1), \hat{\beta}_{21\cdot1}^{(\lambda)} = \frac{s_{12} + \lambda s_{22}}{s_{11} + \lambda s_{12}}$$

where (s_{jk}) is the sample cov. matrix from complete cases. Other parameters have similarly simple expressions.

Extends the regression estimator ($\lambda = 0$) to MNAR mechanisms

- No information in the data about λ. We can assess sensitivity of inferences to different values of this parameter

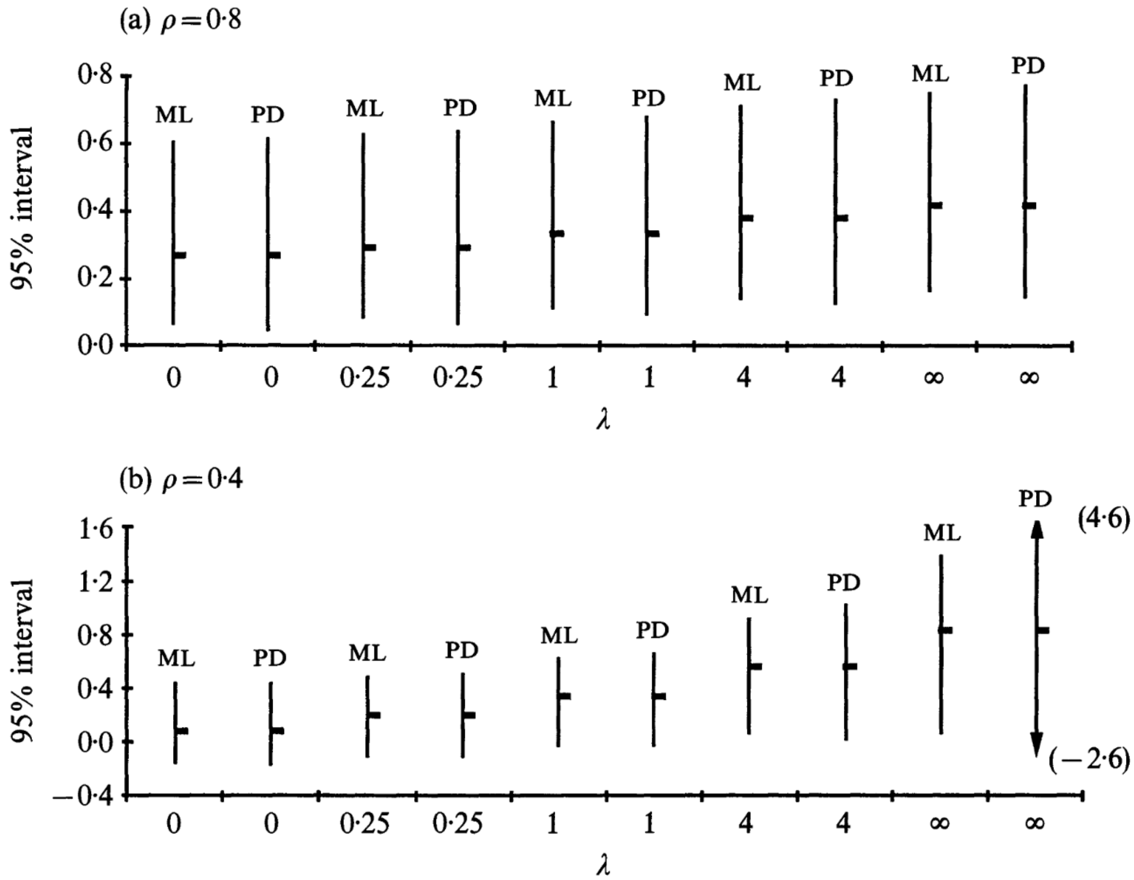
…

5 approaches to lack of estimability

(a) Follow up a sample of nonrespondents, and incorporate this information into the main analysis.
(b) Adopt a Bayesian approach, assigning the parameters prior distributions. Inferences tend to be sensitive to the choice of prior distribution.
(c) Impose additional restrictions on model parameters, such as on the regression coefficients in the Heckman model.
(d) Conduct analysis to assess sensitivity of inferences for quantities of interest to different choices of the values of parameters poorly estimated from the data.
(e) Selectively discard data to avoid modeling the missingness mechanism.

Sensitivity analysis for bivariate normal pattern-mixture model

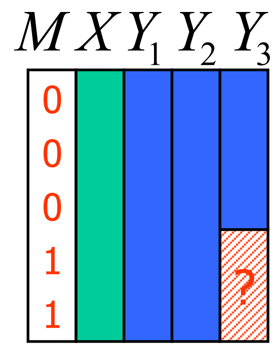
Figure 15.1 Example 15.15, 95% Intervals for Generated Data with ρ =0.8 and 0.4.



Example. Haloperidol Study

- Clinical trial to compare three alternative dose regimens of haloperidol for schizophrenia (Little and Wang 96).
  - 65 patients assigned to receive 5, 10 or 20 mg./day of haloperidol for 4 weeks. X = 3 indicators for drug dose (or equivalently two indicators and a constant term).
  - Y = Brief Psychiatric Rating Scale Schizophrenia (BPRSS) factor, measured at baseline, week 1 and week 4. Main interest in mean change in BPRSS between baseline and week 4 for each dose group.
  - 29 patients dropped out of the study before the 4th week for a variety of reasons, including side effects of the drug. Proportions dropping out varied across dose groups, suggesting that missingness was related to dose.

Pattern-mixture model for this example



- Little and Wang (1996) assume

$$(m_i \mid x_i) \sim_{\text{ind}} \text{Bernoulli}\left(p_i(x_i)\right);$$

$$\operatorname{logit}\left(p_i(x_i)\right) = \pi^T x_i$$

- This logistic regression of m on x is assumed saturated, that is, the fraction of incomplete cases is estimated independently for each treatment group.

$$(y_i \mid x_i, m_i = k) \sim_{\text{ind}} N_3(\mathrm{B}^{(k)} x_i, \Sigma^{(k)})$$

- Parameters $B^{(k)}, \Sigma^{(k)}$ are identified by alternative assumptions about the missing-data mechanism...

Four alternative md mechanisms

- The effect of nonignorable nonresponse is assessed by computing estimates under a range of assumptions about the missing-data mechanism. Specifically, missingness is assumed to be an arbitrary function of treatment group and one of the following four linear combinations of *Y*:

$$A.\ y_A^* = 0.4Y_1 + 0.4Y_2 + 0.2Y_3$$

$$B.\ y_B^* = 0.3Y_1 + 0.3Y_2 + 0.4Y_3$$

$$C.\ y_C^* = 0.1Y_1 + 0.1Y_2 + 0.8Y_3$$

$$D.\ y_D^* = Y_3$$

A is closest to ignorable, where coefficient of $Y_3 = 0$;
D is the most extreme departure from ignorable nonresponse

Comparison of Estimates

- Following table shows estimated difference in mean BPRSS between baseline and week 4 for the 3 dosage groups, for:
  - Complete-Case (CC) analysis.
  - Ignorable ML, where missingness is assumed to depend on the BPRSS scores at baseline and week 1. Standard errors are the sds of estimates from 1000 bootstrap samples.
  - (a) ML for the four pattern-mixture models. Asymptotic standard errors were computed using the SEM algorithm.
  - (b) Bayes for the four pattern-mixture models. Mean and variance of the posterior distributions were simulated via Gibbs sampling. The prior was

$$p(\pi,\phi) \propto |\Sigma_{11.2}|^{-(p_1+1)/2} |\Sigma_{22}^{(0)}|^{-1/2} |\Sigma_{22}^{(1)}|^{-1/2}$$

  - ML estimates for the probit selection model.

Estimates for Haliperidol Data

	Treatment Group		
	Dose 5	**Dose 10**	**Dose 20**
Sample Size	15	34	16
Fraction missing	.33	.41	.63
CC Analysis	3.700 (1.027)	4.350 (0.726)	5.667 (1.326)
Ignorable ML	3.291 (0.897)	4.087 (0.618)	6.463 (1.044)
Pattern-Mixture ML			
Mechanism A	3.276 (0.898)	4.139 (0.621)	6.528 (1.048)
Mechanism B	3.251 (0.909)	4.184 (0.631)	6.610 (1.072)
Mechanism C	3.181 (0.945)	4.249 (0.663)	6.808 (1.155)
Mechanism D	3.140 (0.968)	4.268 (0.684)	6.913 (1.208)
Probit Selection ML	3.345 (1.027)	4.155 (0.560)	6.586 (1.194)

Estimates for Haliperidol Data (contd)

	Treatment Group		
	Dose 5	**Dose 10**	**Dose 20**
Sample Size	15	34	16
Fraction missing	.33	.41	.63
CC Analysis	3.700 (1.027)	4.350 (0.726)	5.667 (1.326)
Ignorable ML	3.291 (0.897)	4.087 (0.618)	6.463 (1.044)
Pattern-Mixture Bayes			
Mechanism A	3.229 (0.985)	4.070 (0.710)	6.464 (1.185)
Mechanism B	3.212 (1.016)	4.133 (0.717)	6.559 (1.221)
Mechanism C	3.133 (1.117)	4.226 (0.772)	6.812 (1.393)
Mechanism D	3.075 (1.187)	4.258 (0.820)	6.964 (1.526)
Probit Selection ML	3.345 (1.027)	4.155 (0.560)	6.586 (1.194)

Haloperidol Data Findings

- Complete-case estimates deviate noticeably from estimates from the other methods
- Estimates for the probit selection model are similar to those for the ignorable selection model and pattern mixture model with mechanism (A).
  - Bootstrap distributions of the estimated coefficients for selection were very dispersed, confirming that the ability to estimate these parameters simultaneously from the data is very limited.
  - This approach is not recommended

Haloperidol Findings continued

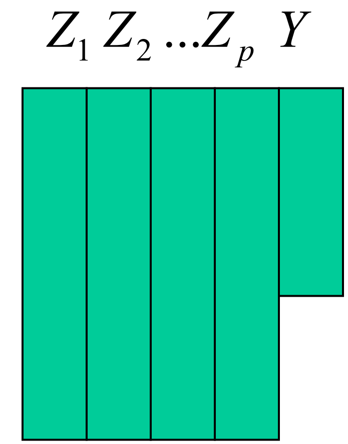
- ML and Bayes estimates for pattern-mixture models are similar.
- Size of treatment effects is only moderately sensitive to the choice of pattern-mixture model (ignorable, A-D):
  - As missingness becomes increasingly dependent on the missing Week 4 BPRSS value, differentials in treatment effects by size of dose increase slightly
- Asymptotic se's are a bit smaller than the posterior standard errors; the latter are superior in that they do not rely on quadratic approximations to the log likelihood
- Se's of the pattern-mixture model estimates increase from models A though D, reflecting a loss of information with increasing degree of non-ignorable nonresponse

Proxy Pattern-Mixture Analysis for Survey Nonresponse

- Three major components to consider in evaluating nonresponse:
  1. Amount of missingness (measures of nonresponse rate)
  2. Differences between respondents and nonrespondents on auxiliary data
  3. Relationship between auxiliary covariates and survey outcome
- Previous approaches have focused on a subset of these components
- Proxy Pattern-Mixture Analysis is a simple method of nonresponse bias measurement and adjustment that integrates all three aspects

Proxy variables for outcome

- Setting: univariate nonresponse
- Y = survey outcome
- Z = auxiliary covariate information
- Goal: nonresponse adjustment of mean of Y
  - (non-MAR as well as MAR)



Create $X =$ single best proxy for Y based on $Z = (Z_1, \ldots, Z_p)$

Compute by regression of Y on $Z_1, \ldots, Z_p$ using complete cases

$\rho = \text{Corr}(X, Y) > 0$

Call X a <u>strong proxy</u> if ρ is high, a <u>weak proxy</u> if ρ is low

Proxy pattern-mixture model

Transform $Z \rightarrow (X, V)$,

$X = Z^T \alpha$ = best predictor of Y, V = other covariates

$[Y, X, V, M, \alpha] = [Y, X \mid M, \alpha][M][\alpha][V \mid Y, X, M, \alpha]$

$\left((X, Y) \mid M = m\right) \sim N_2\left((\mu_x^{(m)}, \mu_y^{(m)}), \Sigma^{(m)}\right)$

$M \sim Bernoulli(\pi)$

Unspecified

$$\Sigma^{(m)} = \begin{pmatrix} \sigma_{xx}^{(m)} & \rho^{(m)}\sqrt{\sigma_{xx}^{(m)}\sigma_{yy}^{(m)}} \\ \rho^{(m)}\sqrt{\sigma_{xx}^{(m)}\sigma_{yy}^{(m)}} & \sigma_{yy}^{(m)} \end{pmatrix}$$

$\Pr(M = 1 \mid X, Y) = f(Y^*), Y^* = X\sqrt{\sigma_{yy}^{(0)} / \sigma_{xx}^{(0)}} + \lambda Y, \lambda \geq 0$

rescaling X aids interpretation of λ

PPMA ML estimate

ML estimate of mean of Y is

$$\hat{\mu}(\lambda) = \overline{y}_R + g(\hat{\rho})\sqrt{\left(s_{yy}^{(0)} / s_{xx}^{(0)}\right)}(\overline{y}_{NR}^* - \overline{y}_R^*), \quad g(\hat{\rho}) = \left(\frac{\hat{\rho} + \lambda}{1 + \hat{\rho}\lambda}\right)$$

$\lambda \geq 0$ is a sensitivity parameter,

determined by assumed missing data mechanism

Propose sensitivity analysis with three values of λ:

λ=0, $g(\hat{\rho}) = \hat{\rho}$ (MAR, usual regression estimator)

λ=1, $g(\hat{\rho}) = 1$ (NMAR, carries over bias adjustment from proxy)

λ=∞, $g(\hat{\rho}) = 1 / \hat{\rho}$ (NMAR, inverse regression estimator)

Note: $g(\hat{\rho})$ varies between $\hat{\rho}$ and $1 / \hat{\rho}$, reduced sensitivity as $\hat{\rho} \uparrow 0$

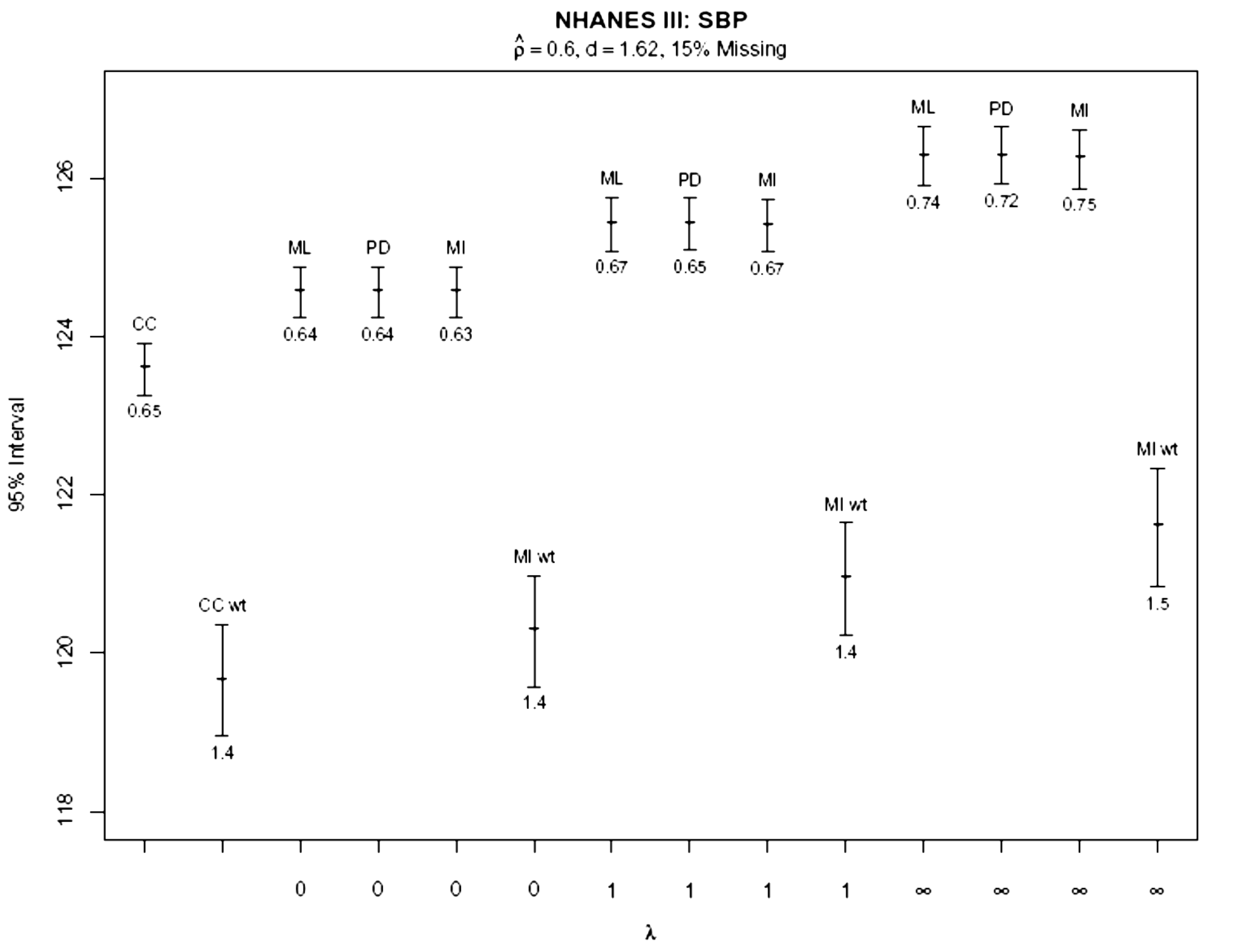
Estimation methods

- 1. Maximum Likelihood
  - Doesn't incorporate uncertainly in regression parameters used to create the proxy
  - Large-sample variances by Taylor series calculations
- 2. Bayesian, non-informative priors
  - Proxy recreated at each draw of regression parameters, so uncertainty is incorporated
  - Easy to implement, non-iterative
- 3. Multiple Imputation of missing Y's
  - Allows complex design features to be incorporated in the within-imputation component of variance
  - Easy to implement

Ex: NHANES III data

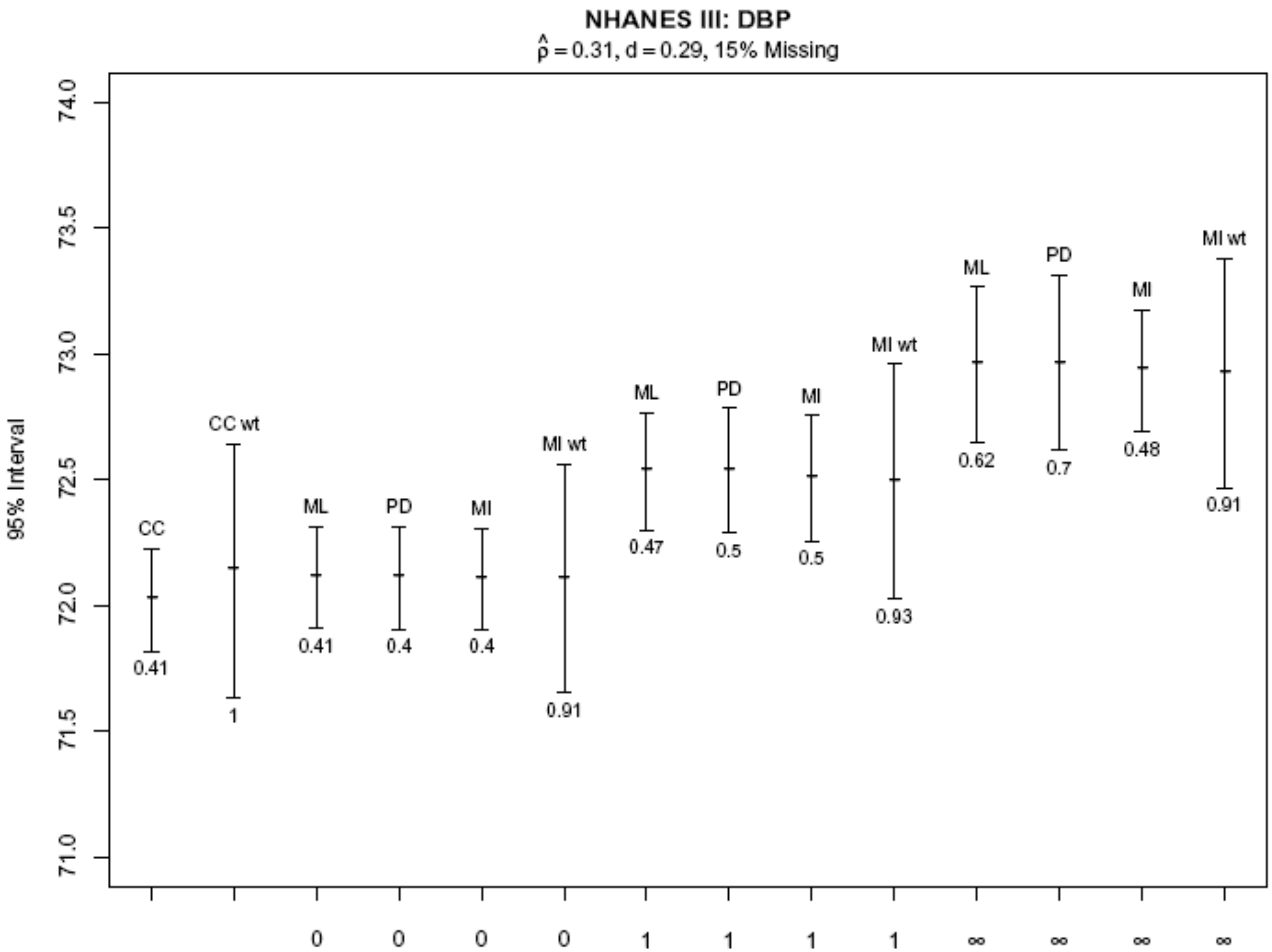
- Nonresponse to Medical Exam, using covariates from the personal interviw
- CC: Complete cases analysis (SRS)
- ML: maximum likelihood (SRS)
- PD: posterior prob interval (SRS)
- MI: MI inference with data imputed using PPMA model
- CC wt, MI wt: as CC, MI, with complete data inference incorporating design weights, clustering

NHANES III examples



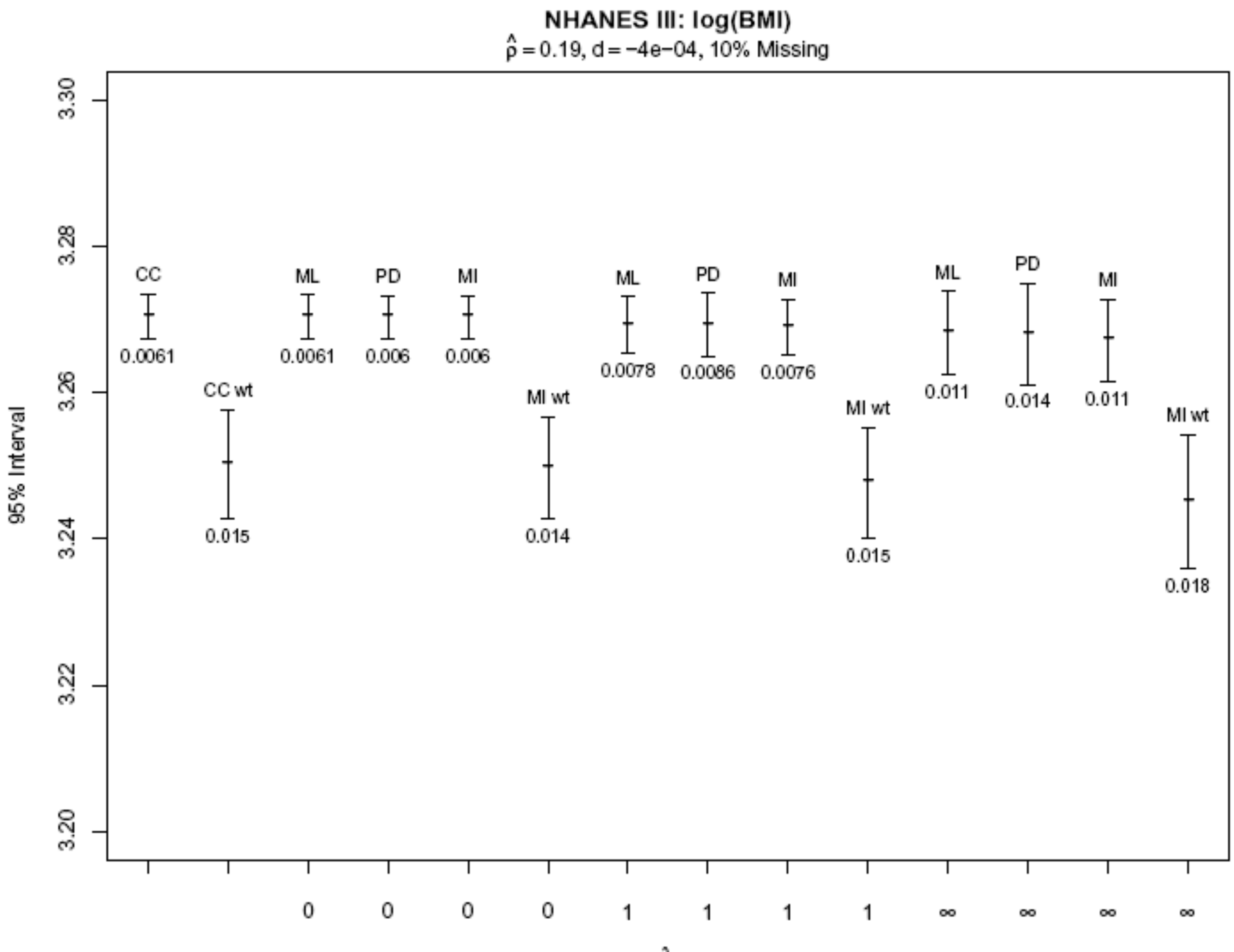
missing not at random

NHANES III Examples



missing not at random

NHANES III Examples



Attractive features of PPMA

- Integrates various components of nonresponse into a single sensitivity analysis reflecting the hierarchy of evidence about bias in the mean
- Easy to implement
- Includes but does not assume MAR; sensitivity analysis is preferred method of assessing MNAR nonresponse
- Gives appropriate credit to the existence of good predictors of the observed outcomes
  - Reinforces that emphasis should be on collecting strong auxiliary data, not solely on obtaining the highest possible response rate

Potential "Disadvantages"

- Reducing covariates to single best predictor limits the interpretability of λ
- Analysis needs to be repeated on each of the key outcomes -- no single measure is readily available
- PPMA is for continuous outcomes where normality is reasonable
  - Extensions to categorical variables via probit models, other generalized linear model

Offsets: a simple way to modify Chained Equation methods for deviations from MAR

Chained Equation methods (IVEware, MICE) assume MAR:
what if one variable (say Y_1) is MNAR? Run program, but when imputing Y_1 given other variables $Y_2,...,Y_p$, add an offset δ to all the imputations. This assumes that the conditional mean of Y_1 given $Y_2,...,Y_p$ for nonrespondents deviates from respondents by δ

Sensitivity analysis: vary size of δ.

Possible choices: $\delta = \lambda s_{1\cdot 23...p}, \lambda = 0.2, 0.5$

(Giusti and Little 2011 J Official Statistics 27, 211-219)

Conclusions

- MNAR mechanisms can be included in a missing-data analysis, but this is a difficult modeling problem
- Often little is known about the missingness mechanism, and results may be sensitive to formulation
- Parameters are often unidentified or weakly-identified from the data …
- As a result, it may be more appropriate to do a sensitivity analysis, fixing weakly identified parameters at different values.
- Software for fitting MNAR models is not widely available
- Design to MNAR missing data is preferable if possible

Subsample ignorable likelihood for regression with missing covariates
(throwing data away can sometimes pay!)

Subsample Ignorable likelihood (SSIL)

- Ignorable likelihood methods: methods based on likelihood, ignoring missingness mechanism
  - Maximum Likelihood, Bayes, Multiple imputation
- Subsample ignorable likelihood methods: ignorable likelihood methods applied to a subsample of the data
- Little, R. J. and Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 60, 4, 591–605.

Options for missing data in regression

- <u>Discard</u> all incomplete cases: valid (though perhaps inefficient) when missingness depends on covariates but not outcomes
- <u>Apply an IL method to all the data</u> (e.g. multiple imputation): valid when missingness depends on observed variables but not on missing variables
- <u>Subsample ignorable likelihood</u> (SSIL) methods are a hybrid of these approaches: – discard some incomplete cases and apply IL method to others
  - Choice of which cases to discard depends on assumptions about missingness mechanism

Basic idea of SSIL

- Outcomes Y, predictors (Z,X,W) where:
- Z fully observed
- Suppose missingness in X is covariate-dependent: it can depend on covariates but not on outcomes Y
- Suppose missingness in W is MAR in the subset of cases with X fully observed: it can depend on Z, X, Y_{obs}, W_{obs} but not Y_{mis}, W_{mis}
- SSIL: discard cases where X has missing values, and apply ignorable likelihood method (ML, MI, Bayes) to remaining cases

Unweighted CC analysis

- Drops incomplete cases
- Hence inefficient if there is substantial information in these cases
- Loss of information depends on pattern and estimand
- E.g. Figure 1: for mean of *Y* the incomplete cases have substantial information, when *X*'s are predictive
- For regression of *Y* on *X*, incomplete cases have no information, under MAR
- But there is info under MNAR

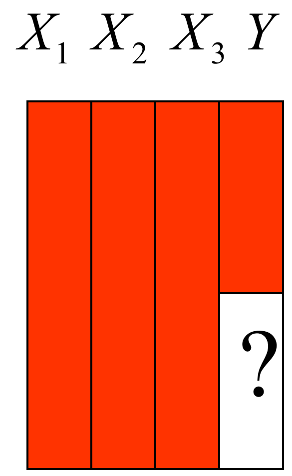
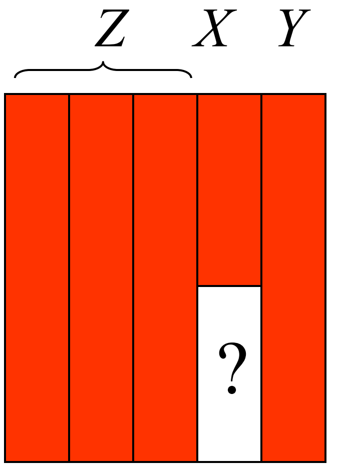


Figure 1

Missing data in X

Target: regression of Y on X, Z; missing data on X



Ignorable Likelihood (IL) methods include information for the regression in the incomplete cases (particularly for intercept and coefficients of Z) and are valid assuming MAR:

$\Pr(X \text{ missing}) = g(Z, Y)$

BUT: if $\Pr(X \text{ missing}) = g(Z, X)$

CC analysis is consistent, but IL methods (or weighted CC) are inconsistent since mechanism is not MAR

Simulations favoring IL often generate data under MAR, <u>hence are biased against CC</u>

(A) Missing data in X

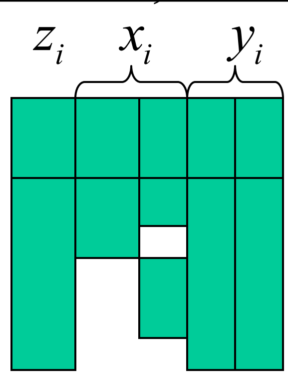
Could be vector

Pattern	Observation, i	z_i	x_i	y_i	R_{x_i}
P1	$i = 1,\ldots,m$	√	√	√	$u_x = (1,\ldots,1)$
P2	$i = m+1,\ldots,n$	√	?	√	$\bar{u}_x$

<u>Key: √ denotes observed, ? denotes observed or missing</u>

P1: complete cases

P2: incomplete X



XMAR $\Rightarrow$ Ignorable Likelihood

Target: Parameters ϕ of regression of Y on X, Z

Full Model: $p(x_i, y_i \mid z_i, \theta) \times p(R_{x_i} \mid z_i, x_i, y_i, \psi); \phi = \phi(\theta)$

If we assume XMAR:

$p(R_{x_i} \mid z_i, x_i, y_i, \psi) = p(R_{x_i} \mid z_i, x_{\text{obs},i}, y_i, \psi)$ for all $x_{\text{mis},i}$

Then $L_{\text{full}}(\theta, \psi) = L_{\text{ign}}(\theta) \times L_{\text{md}}(\psi)$, can base inference on

$$L_{\text{ign}}(\theta) = \text{const.} \times \prod_{i=1}^{n} p(x_{\text{obs},i}, y_i \mid z_i, \theta)$$

XMAR $\Rightarrow$ Ignorable Likelihood

Target: $\phi = \phi(\theta)$ = parameters of regression of Y on X, Z

ML: $\hat{\phi} = \phi(\hat{\theta})$

Bayes: draw $\phi^{(d)} = \phi(\theta^{(d)})$

Multiple imputation: draw $X_{\text{mis}}^{(d)} \sim P(X_{\text{mis}} \mid \text{data})$,

apply MI combining rules to estimates of ϕ

XCOV $\Rightarrow$ Complete-Case Analysis

Assume XCOV:completeness of X depends on covariates, not outcomes:

$$p\left(R_{x_i} = u_x \mid z_i, x_i, y_i, \psi\right) = p\left(R_{x_i} = u_x \mid z_i, x_i, \psi\right) \text{ for all } y_i \text{ (MNAR)}$$

$$L_{\text{full}}(\theta,\psi) = \prod_{i=1}^{m} p(R_{x_i} = u_x, x_i, y_i \mid z_i, \theta, \psi) \prod_{i=m+1}^{n} p(R_{x_i}, x_{\text{obs},i}, y_i \mid z_i, \theta, \psi)$$

$$= \prod_{i=1}^{m} p(y_i \mid x_i, R_{x_i} = u_x, z_i, \theta, \psi) p(R_{x_i} = u_x, x_i \mid z_i, \theta, \psi)$$

By XCOV

$$\times \prod_{i=m+1}^{n} p(R_{x_i}, x_{\text{obs},i}, y_i \mid z_i, \theta, \psi)$$

$$= \prod_{i=1}^{m} p(y_i \mid x_i, z_i, \phi) \times p(R_{x_i} = u_x, x_i \mid z_i, \theta, \psi) \prod_{i=m+1}^{n} p(R_{x_i}, x_{\text{obs},i}, y_i \mid z_i, \theta, \psi)$$

$$= L_{\text{cc}}(\phi) \times L_{\text{rest}}(\theta,\psi)$$

Maximizing $L_{cc}(\phi)$ is valid, but info in $L_{\text{rest}}(\theta,\psi)$ except in special cases

(B) Missing data on *X* and *Y*

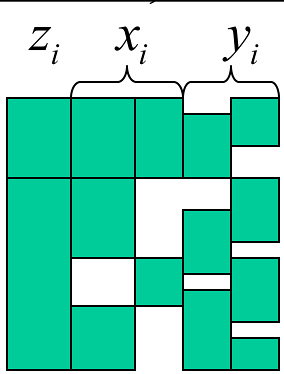
Could be vector

Pattern	Observation, i	z_i	x_i	y_i	R_{x_i}
P1	$i = 1,\ldots,m$	√	√	?	$u_x = (1,\ldots,1)$
P2	$i = m+1,\ldots,n$	√	?	?	$\bar{u}_x$

<u>Key: √ denotes observed, ? denotes observed or missing</u>

P1: covariates complete

P2: x incomplete



XYMAR $\Rightarrow$ Ignorable Likelihood

Target: Parameters ϕ of regression of Y on X, Z

Model: $p(x_i, y_i \mid z_i, \theta)$

Assume XYMAR:

$$p(R_{x_i}, R_{y_i} \mid z_i, x_i, y_i, \psi) = p(R_{x_i}, R_{y_i} \mid z_i, x_{\text{obs},i}, y_{\text{obs},i}, \psi)$$

for all $x_{\text{mis},i}, y_{\text{mis},i}$

Then $L_{\text{full}}(\theta,\psi) = L_{\text{ign}}(\theta) \times L_{\text{md}}(\psi)$, can base inference on

$$L_{\text{ign}}(\theta) = \text{const.} \times \prod_{i=1}^{n} p(x_{\text{obs},i}, y_{\text{obs},i} \mid z_i, \theta)$$

IL Inference about $\phi(\theta)$, as before

XCOV, YSMAR $\Rightarrow$ IL on cases with X observed

Target: Parameters ϕ of regression of Y on X, Z

Assume:

XCOV: completeness of X depends on covariates, not outcomes:

$$p\left(R_{x_i} = u_x \mid z_i, x_i, y_i, \psi)\right)$$

$$= p\left(R_{x_i} = u_x \mid z_i, x_i, \psi)\right) \text{ for all } y_i$$

YSMAR: Y is MAR in subsample with X observed:

$$p(R_{y_i} \mid R_{x_i} = u_x, z_i, x_i, y_i, \psi)$$

$$= p(R_{y_i} \mid R_{x_i} = u_x, z_i, x_i, y_{\text{obs},i}, \psi) \text{ for all } y_i$$

MNAR

SSIL: Apply IL to subsample with X fully observed (P1)

(B) Missing data on X and Y

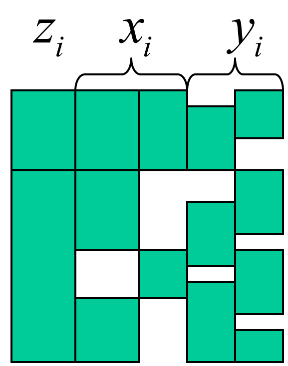
Could be vector

Pattern	Observation, i	z_i	x_i	y_i	R_{x_i}
P1	$i = 1,\ldots,m$	√	√	?	$u_x = (1,\ldots,1)$
P2	$i = m+1,\ldots,n$	√	?	?	$\bar{u}_x$

Key: √ denotes observed, ? denotes observed or missing

P1: covariates complete

P2: x incomplete



SSIL likelihood under XCOV, YSMAR

$$L_{\text{full}}(\theta,\psi)=\prod_{i=1}^{n} p(R_{x_i}, x_{\text{obs},i}, R_{y_i}, y_{\text{obs},i} \mid z_i,\theta,\psi)$$

$$=\prod_{i=1}^{m} p(R_{x_i}=u_x, x_i, R_{y_i}, y_{\text{obs}.i} \mid z_i,\theta,\psi)\times L_{\text{rest}}(\theta,\psi)$$

$$=L_{\text{rest}}(\theta,\psi)\times\prod_{i=1}^{m}\Big(p(R_{x_i}=u_x, x_i \mid z_i,\theta,\psi)\Big) \longleftarrow L^*_{\text{rest}}(\theta,\psi)$$

$$\times\prod_{i=1}^{m}\Big(\int \underbrace{p(y_i \mid x_i, R_{x_i}=u_x, z_i,\theta,\psi)}_{\text{XCOV}\,\downarrow}\, \underbrace{p(R_{y_i} \mid y_i, x_i, R_{x_i}=u_x, z_i,\theta,\psi)}_{\text{YSMAR}\,\downarrow}\, dy_{\text{mis}}\Big)$$

$$=L^*_{\text{rest}}(\theta,\psi)\times\prod_{i=1}^{m}\int p(y_i \mid x_i, z_i,\phi)\, p(R_{y_i} \mid y_{\text{obs},i}, x_i, R_{x_i}=u_x, z_i,\psi)\, dy_{\text{mis}}$$

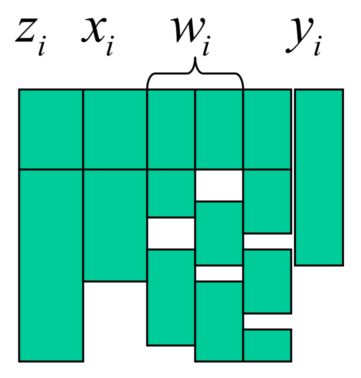
$$=L^*_{\text{rest}}(\theta,\psi)\times\prod_{i=1}^{m} p(y_{\text{obs},i} \mid x_i, z_i,\phi)\prod_{i=1}^{m} p(R_{y_i} \mid y_{\text{obs},i}, x_i, R_{x_i}=u_x, z_i,\psi)$$

SSIL maximizes this

Two covariates *X*, *W* with different mechanisms

Pattern	Observation, i	z_i	x_i	w_i	y_i	R_{x_i}	R_{w_i}
P1	$i = 1,\ldots,m$	√	√	√	?	u_x	u_w
P2	$i = m+1,\ldots,m+r$	√	√	?	?	u_x	$\bar{u}_w$
P3	$i = m+r+1,\ldots,n$	√	?	?	?	$\bar{u}_x$	$\bar{u}_w$

P1: covariates complete
P2: x obs, w, y may be mis
P3: x mis, w, y may be mis



XCOV, WYSMAR $\Rightarrow$ IL on cases with X observed

- Target: regression of Y on Z, X, and W
- Assume:

(XCOV) Completeness of X can depend on covariates but not Y:

$$p\left(R_{x_i} = u_x \mid z_i, x_i, w_i, y_i, \psi_x)\right) = p\left(R_{x_i} = u_x \mid z_i, x_i, w_i, \psi_x)\right) \text{ for all } y_i$$

(WYMAR) Missingness of (W, Y) is MAR within subsample of cases with X observed:

$$p(R_{(w_i, y_i)} \mid z_i, x_i, w_i, y_i, R_{x_i} = u_x; \psi_{wy \cdot x}) =$$
$$p(R_{(w_i, y_i)} \mid z_i, x_i, w_{\text{obs},i}, y_{\text{obs},i}, R_{x_i} = u_x; \psi_{wy \cdot x}) \quad \text{for all } w_{\text{mis},i}, y_{\text{mis},i}$$

- SSIL: apply IL method (e.g. ML) to the subsample of cases for which X is observed
- Proof of consistency: similar to previous case, treating W and Y as block

Two covariates X, W with different mechanisms

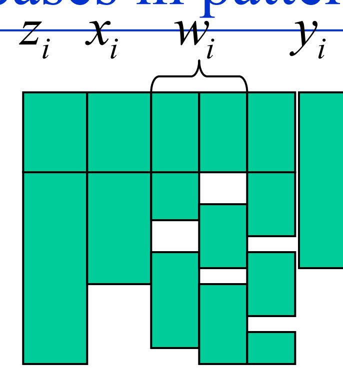
Pattern	Observation, i	z_i	x_i	w_i	y_i	R_{x_i}	R_{w_i}
P1	$i = 1,\ldots,m$	√	√	√	?	u_x	u_w
P2	$i = m+1,\ldots,m+r$	√	√	?	?	u_x	$\bar{u}_w$
P3	$i = m+r+1,\ldots,n$	√	?	?	?	$\bar{u}_x$	$\bar{u}_w$

SSIL: analyze cases in patterns 1 and 2

P1: covariates complete

P2: x obs, w, y may be mis

P3: x mis, w, y may be mis



Simulation Study

- For each of 1000 replications, 5000 observations *Z*, *W*, *X* and *Y* generated as:

$$\left(y_i \mid z_i, w_i, x_i\right) \sim_{\text{ind}} N(1+z_i+w_i+x_i, 1)$$

$$(z_i, w_i, x_i) \sim_{\text{ind}} N(0, \Sigma), \ \Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

- 20-35% of missing values of *W* and *X* generated by four mechanisms

Simulation: missing data mechanisms

Mechanisms	$\alpha_0^{(w)}$	$\alpha_z^{(w)}$	$\alpha_w^{(w)}$	$\alpha_x^{(w)}$	$\alpha_y^{(w)}$	$\alpha_0^{(x)}$	$\alpha_z^{(x)}$	$\alpha_w^{(x)}$	$\alpha_x^{(x)}$	$\alpha_y^{(w)}$
I: All valid	-1	1	0	0	0	-1	1	0	0	0
II: CC valid	-1	1	1	1	0	-1	1	1	1	0
III: IML valid	-2	1	0	0	1	-2	1	1	0	1
IV: SSIML valid	-1	1	1	1	0	-2	1	1	0	1

$$\operatorname{logit}\left(P(R_{w_i}=0 \mid z_i, w_i, x_i, y_i)\right) = \alpha_0^{(w)} + \alpha_z^{(w)} z_i + \alpha_w^{(w)} w_i + \alpha_x^{(w)} x_i + \alpha_y^{(w)} y_i$$

$$\operatorname{logit}\left(P(R_{x_i}=0 \mid R_{w_i}=1, z_i, w_i, x_i, y_i)\right) = \alpha_0^{(x)} + \alpha_z^{(x)} z_i + \alpha_w^{(x)} w_i + \alpha_x^{(x)} x_i + \alpha_y^{(x)} y_i$$

RMSEs*1000 of Estimated Regression Coefficients for Before Deletion (BD), Complete Cases (CC), Ignorable Maximum Likelihood (IML) and Subsample Ignorable Maximum Likelihood (SSIML), under Four Missing Data Mechanisms.

	$\rho = 0$				$\rho = 0.8$			
	I*	II	III	IV	I	II	III	IV
BD	27	28	28	27	50	46	50	46
CC	45	44	553	322	86	71	426	246
IML	37	231	36	116	58	96	53	90
SSIML	42	133	360	49	70	80	319	69
Valid:	ALL	CC	IML	SSIML	ALL	CC	IML	SSIML

Missing Covariates in Survival Analysis

$\{t_1,...,t_k\}$ distinct survival times, j = unit that fails at time t_j (no ties);

$R_j =$ risk set at time t_j, $z_j, x_j, w_j =$ covariates, as before.

Complete data: contribution of data at time t_j to partial likelihood is

$$L_j = \frac{\lambda(y = t_j \mid z_j, x_j, w_j, \beta)}{\sum_{k \in R_j} \lambda(y = t_k \mid z_k, x_k, w_k, \beta)}, \lambda(y = t_j \mid z_j, x_j, w_j, \beta) = \text{ hazard}$$

With z_j, w_j fully observed, x_j covariate-dependent complete, i.e.:

$$\Pr(R_{x_j} = u_x \mid y_j, z_j, x_j, w_j) = \Pr(R_{x_j} = u_x \mid z_j, x_j, w_j)$$

then $\lambda(y = t_j \mid R_{x_j} = u_x, z_j, x_j, w_j, \beta) = \lambda(y = t_j \mid z_j, x_j, w_j, \beta)$

That is, conditioning on $R_{x_j} = u_x$ for each risk set

gives a valid partial likelihood

also OK for time-varying x_j

SSIL for Survival Analysis

SSIL for partial likelihood: Assume

XCOV: x_j is covariate-dependent missing:

$$\Pr(R_{x_j} = u_x \mid y_j, z_j, x_j, w_j) = \Pr(R_{x_j} = u_x \mid z_j, x_j, w_j)$$

WSMAR: missing values of w_j are MAR

in subsample with x_j observed:

$$\Pr(R_{w_j} \mid R_{x_j} = u_x, y_j, z_j, x_j, w_j) = \Pr(R_{w_j} \mid R_{x_j} = u_x, y_{\text{obs},j}, z_j, x_j, w_{\text{obs},j})$$

Then can apply SSIL methods to partial likelihood

in subsample with w_j observed. (Zhang and Little 2014)

How to choose *X*, *W*

- Choice requires understanding of the mechanism:
- Variables that are missing based on their underlying values belong in *X*
- Variables that are SMAR belong in *W*
- Collecting data about why variables are missing is obviously useful to get the model right
- But this applies to all missing data adjustments…

Other questions and points

- How much is lost from SSIL relative to full likelihood model of data and missing data mechanism?
  - In some special cases, SSIL is efficient for a pattern-mixture model
  - In other cases, there is a trade-off between additional specification of mechanism and loss of efficiency from conditional likelihood
- MAR analysis applied to the subset does not have to be likelihood-based
  - E.g. weighted GEE, AIPWEE

Conclusions

- Sometimes discarding data is useful!
- SSIL: selectively discards data based on assumed missing-data mechanism
- More efficient than CC
- Valid for mechanisms where IL, CC are inconsistent