Notes on Complete- and Available-Case Analyses

Writing Notes

- Add introduction
- Add section with real dataset that illustrates complete-case analysis and shifting-units
- Nomenclature: *complete-case analysis* (CCA) is pretty clearly defined. *Available-case analysis* is a broad term, especially given the application to meta-analysis where *available-case analysis* has become synonymous with *shifting units of analysis*. But, *shifting units of analysis* is kind of burdensome for writing. How about *shifting-case analysis* (SCA)?
- There's some dodgy notation that needs to be fixed.
- Structure for results: General form followed by concrete example.
  - Derive results for distributions, which highlight conditions for unbiasedness.
  - Derive conditional expectations of $T|X, v, R$.
  - Show general matrix-form biases as a function of conditional biases
  - Walk through simple example

Conceptual Notes

- How do we relate coefficients in log-linear selection models to coefficients in meta-regression models?

1 Introduction

[HOLD FOR INTRO]

2 Case Study

[HOLD FOR ILLUSTRATION OF ANALYSIS METHODS WITH REAL DATA]

3 Model and Notation

Let T_i be the estimate of the effect parameter θ_i, and let v_i be the estimation error variance of T_i. Denote a vector of covariates that pertain to T_i as $X_i = [1, X_{i1}, \ldots, X_{i,p-1}]^T$. Note that the first element of X_i is a 1, which corresponds to an intercept term in a meta-regression model, and that model can be expressed as:

$$T_i | X_i, v_i, \eta = X_i^T \beta + u_i + e_i \tag{1}$$

Here, $\beta \in \mathbb{R}^p$ is the vector of regression coefficients. The term e_i is the estimation error for effect i and $V[e_i] = v_i$, and u_i is the random effect such that $u_i \perp e_i$ and $V[u_i] = \tau^2$.

This is the standard random effects meta-regression model, and it is also consistent with subgroup analysis models. The parameter $\eta = [\beta, \tau^2]$ refers to the parameters of model. Under a fixed-effects model, it is assumed that $\tau^2 = 0$, in which case $\eta = \beta$, and $u_i \equiv 0$.

A common assumption in random effects meta-regression is that the random effects u_i are independent and normally distributed with mean zero and variance τ^2:

$$u_i \sim N(0, \tau^2).$$

In that case, the distribution $p(T|X, v, \eta)$ can be written as

$$p(T_i | X_i, v_i, \eta) = \frac{1}{\sqrt{2\pi(\tau^2 + v_i)}} e^{-\frac{(T_i - X_i^T \beta)^2}{2(\tau^2 + v_i)}} \tag{2}$$

Note that this assumes that all covariates are observed, and is referred to as the *complete-data likelihood function*. We note that a meta-regression with no missing data will be accurate if the complete-data model is correctly specified. Thus, to illustrate the properties of incomplete data meta-regression, we assume that the complete-data model is correctly specified.

Let R_i be a vector of response indicators for effect i. Each element R_{ij} of R_i corresponds to a variable in a meta-analytic dataset. The R_{ij} take a value of either 0 or 1: $R_{ij} = 1$ indicates the corresponding variable in the meta-analysis is observed and $R_{ij} = 0$, indicates a that the corresponding variable is not observed. For the data $[T_i, v_i, X_i]$, $R_i \in \{0, 1\}^{p+1}$ is a vector of 0s and 1s of length $p + 1$. If v_i were missing, then $R_{i2} = 0$.

Our focus in this article is on missing covariates, and we assume that T_i and v_i are observed for every effect of interest in a meta-analysis. Thus, we amend the notation so that $R_i \in \{0, 1\}^{p-1}$ and $R_{ij} = 1$ if X_{ij} is observed and $R_{ij} = 0$ if X_{ij} is missing. Note that this omits the intercept term described above.

For instance if $X_i = [1, X_{i1}]$ with $X_{i1} \in \mathbb{R}$, then R_i is a scalar such that $R_i = 1$ if X_{i1} is observed, and $R_i = 0$ if X_{i1} is missing.

Denote $O = \{(i,j) : R_{ij} = 1\}$ as the indices of covariates that are observed and $M = \{(i,j) : R_{ij} = 0\}$ be the set of indices for missing covariates. Then, the complete-data model can be written as

$$p(T_i|X_i, v_i, \eta) = p(T_i|X_{iO}, X_{iM}, v_i, \eta). \tag{3}$$

Note that the complete-data model depends on X_{iM}, which are unobserved.

4 Conditional Incomplete Data Meta-Regression

When meta-analytic datasets are missing covariates, analyses involve incomplete data. In practice, meta-regressions of incomplete data have largely relied on one of two approaches: complete-case analyses and available case analyses. A complete-case analysis (CCA) includes only effects for which all covariates of interest are observed, which means that $R_i = [1, \ldots, 1] = \mathbf{1}$ for all effects included in the analysis. In such cases, inferences are based on the conditional distribution of $T_i|X_i, v_i, R_i = \mathbf{1}$ for some missing data pattern $r \in \{0,1\}^{p-1}$.

Available-case meta-regressions typically amount to including a subset of relevant covariates that are completely observed. For instance, if two covariates X_1 and X_2 are of interest, an available-case analysis often involves regressing effects on X_1 (excluding X_2) and then on X_2 (excluding X_1). Analyses of this sort have been referred to as *shifting units of analysis* in the meta-analytic literature, and in this article we refer to them as *shifting-cases analyses* (SUA).

A shifting-cases analysis is equivalent to including observations for which R_i fall into some set of missingness patterns $\mathcal{R}_j$. For instance, including effects for which X_1 is observed and omitting X_2 from the model means that the analysis include effects for which both X_1 and X_2 are observed (i.e., $R = [1,1]$), as well as effects for which X_1 is observed, but X_2 is missing (i.e., $R = [1,0]$). In this formulation, available-case inferences are based on the distribution of $T_i|X_i, v_i, R_i \in \mathcal{R}_j$ for some $\mathcal{R}_j \subset \mathcal{R}$.

Note that both complete- and available-case analyses condition on the value of R_i. In that sense, we can see both of these approaches as *conditional on missingness*. Because both models above are conditional, they are not necessarily identical to the complete-data model, which is the model of interest, because the complete-data model does not condition on R_i. Yet, the analytic approaches of complete- and available-case analyses proceed as if the complete-data and conditional models are equivalent. Doing so ignores the sources and impacts of missingness, and can lead to inaccurate results.

The complete-data model can be related to the conditional models through the distribution of missingness R_i. This approach is referred to as a *selection model* in the missing data literature. Let ψ parameterize the distribution of R given the observed and unobserved data. We can write a selection model as:

$$p(T_i|X_i, v_i, R_i = r, \eta, \psi) = \frac{p(R_i = r|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)}{p(R_i = r|X_i, v_i, \psi, \eta)} \tag{4}$$

This describes the conditional model as a function of the complete-data model $p(T|X, v, \eta)$ and a selection model $p(R_i = r|T_i, X_i, v_i, \psi)$ that gives the probability that a given effect and its covariates are used in the analysis. The denominator on the right hand side of (4) is the probability of a missingness pattern r given the estimation error variance v_i and the observed and unobserved covariates in the vector X_i, and can be written as

$$p(R_i = r|X_i, v_i, \psi, \eta) = \int p(R_i = r|T_i, X_i, v_i, \psi)p(T_i|X_i, v_i, \eta)dT \tag{5}$$

Note that R_i can take one of many values of $r \in \mathcal{R} \equiv \{0, 1\}^{p-1}$.

A standard approach for modelling missingness in covariates is to assume R_i follows some log-linear distribution.

$$\text{logit}P[R_i = r_j|T_i, X_i, v_i] = \sum_{i=0}^{n} \psi_{ij} f_{ij}(T_i, X_i, v_i) \tag{6}$$

where $f_{0j}(T_i, X_i, v_i) = 1$, so that ψ_{0j} would be the intercept term for the logit model for missingness pattern r_j. While log-linear models are not the only applicable or appropriate model for missingness, we make this assumption at points throughout this article in order to demonstrate conditions under which conditional meta-regressions are inaccurate, and how inaccurate they can be.

5 Missingness Mechanisms

[HOLD FOR DISCUSSION OF MCAR, MAR, MNAR, AND IGNORABILITY]

6 Issues with Conditional Inference on Incomplete Data

Both complete- and shifting-case ignore information, which can lead to biased meta-regression estimates. Complete-case analyses omit missing data. Shifting-case analyses omit covariates *and* missing data. Neither typically make any adjustments for the information they exclude, which is an approach known to induce bias in a variety of statistical analyses.

The bias induced by complete- or shifting-case analyses will depend on the conditional expectation $E[T_i|X_i, v_i, R_i = r]$, and how it differs from $E[T_i|X_i, v_i] = X_i^T\beta$. It is possible to derive an approximation

for $E[T_i|X_i, v_i, R_i = r]$ when $R_i|T_i, X_i, v_i$ follows a log-linear distribution as described in the previous section.

Proposition: Suppose $p(T_i|X_i, v_i)$ is the standard fixed- or random effects meta-regression model in equation (2), and suppose $p(R_i = r_j|T_i, X_i, v_i) = G_j(T_i, X_i, v_i)$ follows the log-linear model in (6). If we denote $\mathcal{D}_j = \{k : f_{kj} \text{ depends on } T_i\}$, then

$$E[T_i|X_i, v_i, R_i = r] \approx X_i^T\beta + (1 - G_j(X_i^T\beta, X_i, v_i))(\tau^2 + v_i)\sum_{k\in\mathcal{D}} \psi_{kj} f'_{kj}(X_i^T\beta, X_i, v) \tag{7}$$

Proof:

For $i \in \mathcal{D}_j$, we can write $f_{ij}(T, X, v)$ and we can write $f_{ij}(T, X, v) = f_{ij}(X, v)$ for $i \notin \mathcal{D}$. Further, denote

$$G_j(X^T\beta, X, v) \equiv G_j(X, v) = P[R = r_j|T, X, v]|_{T=X^T\beta}$$

Then an approximation for $E[T_i|X_i, v_i, R_i = r_j]$ is as follows:

omitting subscripts, Taylor series for exponents in $P[R = r_j|T, X, v]$ at $T = X^T\beta$

$$\begin{aligned}
E[T|X, v, R = r_j] &= \int \frac{T \exp\left\{-\frac{(T-X\beta)^2}{2(\tau^2+v)} + \sum_i \psi_{ij} f_{ij}(T, X, v)\right\}}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}\log\left(1 + e^{\sum_i \psi_{ij} f_{ij}(T,X,v)}\right)} dT \\
&= \int \frac{T \exp\left\{-\frac{(T-X\beta)^2}{2(\tau^2+v)} + \sum_i \psi_{ij} f_{ij}(X^T\beta, X, v) + \sum_i \psi_{ij} f'_{ij}(X^T\beta, X, v)(T - X^T\beta)\right.}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} \\
&\qquad \frac{\left. - \log\left(1 + e^{\sum_i \psi_{ij} f_{ij}(X^T\beta,X,v)}\right) - G_j(X, v)(\sum_i \psi_{ij} f'_{ij}(X^T\beta, X, v))(T - X\beta) + O(T^2)\right\}}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} dT \\
&\approx \int \frac{T \exp\left\{-\frac{1}{2(\tau^2+v)}\left(T^2 - 2TX\beta - 2T(\tau^2 + v)\sum_{i\in\mathcal{D}} \psi_{ij} f'_{ij}(X^T\beta, X, v)\right.\right.}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} \\
&\qquad \frac{\left.\left. + 2T(\tau^2 + v)G_j(X, v)(\sum_{i\in\mathcal{D}} \psi_{ij} f'_{ij}(X^T\beta, X, v)) + \ldots\right)\right\}}{g_j(X, v)\sqrt{2\pi(\tau^2 + v)}} dT \\
&= X\beta + (1 - G_j(X, v))(\tau^2 + v)\sum_{i\in\mathcal{D}} \psi_{ij} f'_{ij}(X^T\beta, X, v) \qquad \blacksquare
\end{aligned}$$

Note that this uses a first order Taylor expansion of the exponentiated terms of the log-linear model, and thus assumes the f_j are differentiable. This approximation will be more accurate if the f_j are linear in T. In that case only a Taylor expansion of the denominator of the log-linear model is required.

7 Complete-Case Analyses

A common approach in meta-regression with missing covariates is to use a complete-case analysis. There are conditions under which the complete case analysis will lead to unbiased estimates. First, if the covariates are MCAR, so that $R \perp T, X, v$, then

$$P(R|T, X, v, \psi) = \psi$$

and hence

$$\begin{aligned} p(T|X, v, R = r, \eta) &= \frac{p(R = r|T, X, v, \psi)p(T|X, v, \eta)}{p(R = r|X, v, \psi)} \\ &= \frac{\psi p(T|x, v, \eta)}{\psi} \\ &= p(T|x, v, \eta) \end{aligned}$$

Thus, likelihood-based estimation should be consistent, assuming it can be done when X is MCAR. This is consistent with broader results on analyses of MCAR data.

However, it would appear that a complete-case analysis is also valid under slightly less restrictive assumptions. Suppose that $R \perp (X, T)|v$, then

$$\begin{aligned} p(T|X, v, R = r) &= \frac{p(R = r|T, X, v, \psi)p(T|X, v, \eta)}{p(R = r|X, v, \psi)} \\ &= \frac{p(R = r|v, \psi)p(T|x, v, \eta)}{p(R = r|v, \psi)} \\ &= p(T|x, v, \eta) \end{aligned}$$

The assumption that $R \perp (X, T)|v$ implies that if missingness only depends on the estimation error variances, then a complete case analysis may be unbiased. This is a weaker assumption than MCAR, which requires $R \perp (T, X, v)$. For most effect size indices, variances v are functions of the sample sizes within studies n. Some effect sizes, such as the z-transformed correlation coefficient, have variances v that depend entirely on the sample size of a study, while for other effect sizes this is approximately true, such as the standardized mean difference. For such effect sizes, this assumption implies that missingness depends only on the sample size of the study. This may be true, for instance, if smaller studies are less likely to report more fine-grained demographic information regarding their sample out of concern for the privacy of the subjects who participated in the study (and that no other factors affect missingness).

- Note about reduction in precision.

However, when R is not independent of X or T (given v), then analyses can be biased. Precisely how biased will depend on the distribution of R and its relationship to effect estimates T and their covariates X. A general result based on the approximation in the previous section follows from the fact that a complete-case analysis involves $R = \mathbf{1}$. Further, denote $\mathbf{X}_R$ as the matrix of covariates such $R_i = \mathbf{1}$. If we denote $\mathbf{W}_R = diag(v_i + \tau^2) : R_i = \mathbf{1}$, the complete-case analysis will estimate coefficients as:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}_R^T W_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R E[T|X, v, R = r] \\
&= (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[X_i^T \beta + \left((1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f_j'(X_i^T \beta, X_i, v_i) \right) \right] \\
&= \beta + (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[(1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f_j'(X_i^T \beta, X_i, v_i) \right]
\end{aligned}$$

Thus, the bias of the regression coefficients is given by

$$(\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \left[(1 - G(X_i, v_i))(\tau^2 + v_i) \sum_{j \in \mathcal{D}} \psi_j f_j'(X_i^T \beta, X_i, v_i) \right]$$

- Refine for matrix notation. Let $\mathbf{I}$ be the identity matrix, $\mathbf{G} = diag(G(X_i, v_i))$ and $Y = [\sum_{j \in \mathcal{D}} \psi_j f_j'(X_i^T \beta, X_i, v_i), i = 1, \ldots]^T$. Then we can write:

$$(\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T \mathbf{W}_R \mathbf{W}_R^{-1} (\mathbf{I} - \mathbf{G}) Y = (\mathbf{X}_R^T \mathbf{W}_R \mathbf{X}_R)^{-1} \mathbf{X}_R^T (\mathbf{I} - \mathbf{G}) Y$$

- If the likelihood for conditional inference can be written free of R then conditional inference may be appropriate.

- If we're worried about bias, then the PMM or selection model can help us understand bias. PMM may be most useful for this. Show formulas.

- If we're worried about uncertainty, selection models can help unpack that.

7.1 Example: Complete-Case Analysis with a Single Binary Covariate

Suppose there is only one covariate $X_{i1} \equiv X_i \in \mathbb{R}$, so that the complete data model is

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i$$

where β_0, β_1 are the regression coefficients of interest. Assume that the model for missingness is log-linear:

$$p(R = 1|T, X, v) \propto \frac{\exp\{\psi_0 + \psi_1 X + \psi_2 T + \psi_3 TX\}}{1 + exp\{\psi_0 + \psi_1 X + \psi_2 T + \psi_3 TX\}}$$

Note that $G(X, v) \equiv G(X)$. Further, $\mathcal{D} = \{2, 3\}$ and $f_j(T, X, v) = Tf_j(X, v)$, and hence $f'_j(T, X, v) = f_j(X, v)$ for $j \in \mathcal{D}$. Thus,

$$E[T_i|X_i, v_i, R_i = 1] = \beta_0 + \beta_1 X_i + (1 - G(X_i))(v_i + \tau^2)(\psi_2 + \psi_3 X_i)$$

Suppose X is a binary variable, so that $X \in \{0, 1\}^{p-1}$, and that X is observed for $M \leq k$ effects. Denote m_0 as the number of observed $X_i = 0$ and m_1 be the observed $X_i = 1$ so that $M = m_0 + m_1$. Further, assume that $X_i = 0$ (but $R_i = 1$) for $i = 1, \ldots m_0$ and $X_i = 1$ (and $R_i = 1$) for $m_0 + 1, \ldots, M$. Then then the ML and least squares estimator for β_0 is given by

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{m_0} T_i/(v_i + \tau^2)}{\sum_{i=1}^{m_0} 1/(v_i + \tau^2)}$$

This would imply that the complete-case estimator of β_0 under the missing data model specified would have expectation:

$$\begin{aligned} E[\hat{\beta}_0] &= \beta_0 + \frac{(1 - G(0))\, m_0}{\sum_{i=1}^{m_0} 1/(v_i + \tau^2)} \psi_2 \\ w_{j\cdot} &= \left[\sum_{i: X_i = j, R_i = 1} 1/(v_i + \tau^2) \right] / m_j \\ E[\hat{\beta}_0] &= \beta_0 + \frac{1 - G(0)}{w_{0\cdot}} \psi_2 \end{aligned}$$

Note that the second term on the right hand side constitutes the bias of $\hat{\beta}_0$. The bias depends on a variety of quantities. First, it depends on the selection model, solely through ψ_2 When $\psi_2 > 0$ and hence when larger effect estimates are more likely to be missing covariates, the bias is negative. The bias is positive when $\psi_0 < 0$, which occurs when larger effect estimates are less likely to be missing covariates.

The ML estimator of β_1 is given by

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=m_0+1}^{M} T_i/(v_i+\tau^2)}{\sum_{i=m_0+1}^{M} 1/(v_i+\tau^2)} - \frac{\sum_{i=1}^{m_0} T_i/(v_i+\tau^2)}{\sum_{i=1}^{m_0} 1/(v_i+\tau^2)} \\ &= \frac{\sum_{i=m_0+1}^{M} T_i/(v_i+\tau^2)}{\sum_{i=m_0+1}^{M} 1/(v_i+\tau^2)} - \hat{\beta}_0\end{aligned}$$

The expectation of this estimator is approximately

$$\begin{aligned}E[\hat{\beta}_1] &= \beta_0 + \beta_1 + \frac{(1-G(1))m_1}{\sum_{i=m_0+1}^{M} 1/(v_i+\tau^2)}(\psi_2+\psi_3) - \beta_0 - \frac{(1-G(0))\,m_0}{\sum_{i=1}^{m_0} 1/(v_i+\tau^2)}\psi_2 \\ &= \beta_1 + \frac{(1-G(1))m_1}{\sum_{i=m_0+1}^{M} 1/(v_i+\tau^2)}(\psi_2+\psi_3) - \frac{(1-G(0))\,m_0}{\sum_{i=1}^{m_0} 1/(v_i+\tau^2)}\psi_2 \\ &= \beta_1 + \left[\frac{1-G(1)}{w_{1\cdot}} - \frac{1-G(0)}{w_{0\cdot}}\right]\psi_2 + \frac{1-G(1)}{w_{1\cdot}}\psi_3\end{aligned}$$

- Plots?
- Continuous covariates?

For a single continuous covariate, note that

$$\hat{\beta}_1 = \frac{\sum w_i(T_i - \bar{T}_\cdot)(X_i - \bar{X}_\cdot)}{\sum w_i(X_i - \bar{X}_\cdot)^2}$$

Given the selection model above, the bias of $\hat{\beta}_1$ can be expressed as:

$$\psi_2 \frac{\sum(1-G(X_i))(X_i - \bar{X}_\cdot)}{\sum w_i(X_i - \bar{X}_\cdot)^2} + \psi_3 \frac{\sum(1-G(X_i))X_i(X_i - \bar{X}_\cdot)}{\sum w_i(X_i - \bar{X}_\cdot)^2}$$

It is not immediately clear that these simplify. Perhaps we leave this result out?

8 Available Cases and Shifting Units of Analysis

When there are multiple covariates of interest, each of which has some missing data, there may only be a few effects for which all covariates of interest are observed. When that happens, a complete case analysis can be unfeasible. A common solution to this in meta-analysis is to use an available-case analysis.

In meta-analysis, an *available-case analysis* typically takes the form of fitting several meta-regression models, each including a subset of the covariates of interest. Sometimes this even takes the form of regressing effect estimates on one covariate at a time. Referred to as "sifting units of analysis" (SUA), this approach inherently conditions on a set of missingness patterns: $R \in \{\tilde{r}_1, \ldots, \tilde{r}_n\}$. To see this, note

that SUA amounts to a complete-case analysis on a subset of covariates. Thus $R_j = 1$ for those covariates X_j that are observed and included in the analyses, but $R_k \in \{0, 1\}$ for covariates X_k that are excluded; that is, excluded covariates may be observed or unobserved.

Let $U = \{j : X_j \text{ included in model}\}$ denote the covariates included in a SUA model. Let E be the complement of U so that it indexes all of the covariates excluded from the SUA model. Further, let $\mathcal{R} = \{R : R_j = 1, \forall j \in U\}$. Note that $\mathcal{R}$ contains missingness patterns R such that all the included covariates are observed, but any excluded covariates may be either observed or unobserved. Then, the SUA model can be written as:

$$p(T|X_U, v, R \in \mathcal{R}) = \frac{P[R \in \mathcal{R}|T, X_U, v]p(T|X_U, v)}{P[R \in \mathcal{R}|X_U, v]}$$

The model above is slightly different from the models in the previous sections. Note that all of the functions involved condition only on the included covariates X_U. Thus, the function $p(T|X_U, v)$ is analogous to the complete-data likelihood $p(T|X, v)$, however it is important to note that the two functions are not necessarily equivalent because the former conditions only on X_U and not the full set of covariates X.

There are two sources of bias that arise in SUA models. The first is due to the fact that $p(T|X_U, v)$ need not be equivalent to $p(T|X, v)$. These models would only be equivalent of $T \perp X_E|X_U, v$. That is, unless the excluded covariates are completely unrelated to effect size (given the covariates included in the SUA model), then even if none of the included covariates had any missingness, there would be bias in an SUA model.

The second source arises from the fact that SUA approaches omit effects for which any of the included covariates are missing. That is, if any X_{ij} is missing for $j \in U$, then SUA ignores the effect T_i and its associated covariates.

Taken together, these two facets of the conditional SUA model in the expression above suggest a very strict set of conditions for which SUA analyses are unbiased. First, $R \perp (T, X_U)|v$, so that missingness must be independent of effect size estimates and any included covariates conditional on the (completely observed) estimation error variances. This is a similar, though slightly weaker assumption as that made for unbiased complete-case analyses. Second, $T \perp X_E|X_U, v$, which means that any excluded covariates must be completely irrelevant given the included covariates. This amounts to $\beta_j = 0$ for all $j \in E$. Alternatively, we may write $(T, X_U) \perp X_E|v$, which would imply that the complete data likelihood involves no interactions between X_U, X_E and that X_U and X_E are completely orthogonal.

When assumptions about both the missingness mechanism and the relevance of X_E doe not hold, then SUA will be biased. Just how biased will depend on a number of factors, including the amount of missingness, the missingness mechanism, and the relevance of any excluded covariates.

The matrix representation of the bias when no data are missing is:

$$\begin{aligned} E[(\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{T}] &= (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X} \beta \\ &= (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} (\mathbf{X}_U \beta_U + \mathbf{X}_E \beta_E) \\ &= \beta_U + (\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X}_E \beta_E \end{aligned}$$

Thus, the bias of excluding covariates is given by:

$$(\mathbf{X}_U^T W \mathbf{X}_U)^{-1} \mathbf{X}_U^T \mathbf{W} \mathbf{X}_E \beta_E$$

This bias arises even if no data are missing. When data are missing, there is a second source of bias that can arise due to missingness. This bias will depend on $E[T|X, v, R \in \mathcal{R}_j]$. Note that we can write:

$$\begin{aligned} p(T|X, v, R \in \mathcal{R}_j) &= \frac{\sum_{r \in \mathcal{R}_j} p(R = r|T, X, v) p(T|X, v)}{\sum_{r \in \mathcal{R}_j} p(R = r|X, v)} \\ &= p(T|X, v) \frac{\sum_{r \in \mathcal{R}_j} p(R = r|T, X, v)}{\sum_{r \in \mathcal{R}_j} p(R = r|X, v)} \end{aligned}$$

Using the approximation from the proposition above, we can write:

$$E[T|X, v, R \in \mathcal{R}_j] = \frac{\sum_{r \in \mathcal{R}_j} p(R = r|X, v) \left[X^T \beta + (1 - G_j(X, v)) \right]}{\sum_{r \in \mathcal{R}_j} p(R = r|X, v)}$$

8.1 Example: Shifting-Cases Analysis with Two Binary Covariates

Suppose $X = [1, X_1, X_2]$ and that there is missingness in both X_1 and X_2. Then $R \in \{0, 1\}^2$ so that $R = [1, 1]$ indicates both covariates are observed, and $R = [1, 0]$ indicates only X_1 is observed.

If missingness is such that $R = [1, 1]$ for very few effect estimates, then a shifting units analysis might involve regressing T on the observed values of X_1 and then on the observed values of X_2.

The first regression would take only rows for which X_1 is observed. It would use the following estimates:

$$\hat{\beta}_0 = \frac{\sum_{X_i=0} w_i T_i}{\sum_{X_i=0}} = \bar{T}_{10}, \qquad \hat{\beta}_1 = \frac{\sum_{X_i=1} w_i T_i}{\sum_{X_i=1}} - \hat{\beta}_0 = \bar{T}_{11} - \bar{T}_{10}$$

Note that if no X_1 were missing then the bias of these estimates is given by [

]

However, suppose that some of the X_1 and X_2 are missing, and that R follows a log-linear model:

[

]

Then the bias is given by [

]

9 Discussion