Unbiased, consistent estimation and size α inference for longitudinal and multilevel models with missing data

BM Ringham, JD Hoffman, D Dabelea, KE Muller, DH Glueck

1. Introduction

Why are we writing this paper? --Current approaches have convergence problems with typical longitudinal data and unstructured covariance matrices. Analytic and simulation results (see Gurka edwards Muller, include table here and from Ringham et al) provide convincing evidence that underfitting may substantially inflate Type I error. We provide a solution that works for a useful class of models.

Good points

a. Can write down estimators.

b. We can write down approximate and exact distributional theory where the approximate theory will give the exact results in certain useful cases.

c. one does not need to specify the variance structure before fitting the model.

d. Although different method of estimation, should give asymptotically same results as usual ML mixed model estimation.

Bad points

a. will fail for certain sample sizes and missing data processes

b. can't have repeated covariates or heteroscedasticity

Say we are interested in GLH

Check out Sandwich Estimator

Exhibit 1: This describes current state of the art (USE tables from failed doubly robust submission.

		Convergence	
		Yes	No
Simple Σ	Yes	YES	X
	No		X

Exhibit 2: This describes current state of the art (USE tables from failed doubly robust submission.

		Correct Type I error	
		Yes	No
Simple Σ	Yes		
	No		

2. General Notation

For a set $\mathcal{L}$, write the cardinality of the set as $\#(\mathcal{L})$.

Let $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\}$be a set of p vectors of dimension $(N \times 1)$. Write the $(N \times p)$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_p]$. Let $\text{vec}(\boldsymbol{A})$ be the $(Np \times 1)$ vector such that $\text{vec}(\boldsymbol{A}) = [\boldsymbol{a}_1' \quad \boldsymbol{a}_2' \quad \cdots \quad \boldsymbol{a}_p]'$. Write the expected value of a matrix $\boldsymbol{A}$ as $\mathcal{E}(\boldsymbol{A})$. With $\boldsymbol{a}$ a $(p \times 1)$ random vector, write the $(p \times p)$ variance-covariance matrix of $\boldsymbol{a}$ as $\boldsymbol{\mathcal{V}}(\boldsymbol{a})$.

As in Arnold, (Theorem 17.2, p. 312, 1981), indicate that the $(q \times p)$ matrix $\boldsymbol{A}$ follows the (possibly singular) matrix normal distribution with $(q \times p)$ expected value matrix $\boldsymbol{M}$, $(q \times q)$ column covariance matrix $\boldsymbol{U}$ and $(p \times p)$ row covariance matrix $\boldsymbol{V}$ by

$$\boldsymbol{A} \sim (\mathcal{S})\mathcal{N}_{q,p}(\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V}). \tag{1}$$

Under that definition, we have

$$\text{vec}(\boldsymbol{A}) \sim (\mathcal{S})\mathcal{N}_{qp}(\boldsymbol{M}, \boldsymbol{V} \otimes \boldsymbol{U}) \tag{2}$$

Let $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k\}$ be a set of matrices of arbitrary dimensionality. Define $\max\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k\}$ to be the largest element in any matrix $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k\}$. Let $\mathbf{1}_{N,p}$ indicate the $(N \times p)$ matrix with every element equal to 1.

Other things go here. Dear Brandy, Grampa Keith would like us to bold the $\mathcal{V}$ in the variance operator when it yields a matrix result. Also, we have no idea why we needed the max notation. Let's drop it if we don't want it.

3. Planned data and missing data assumptions

4.1 Define planned outcome data

Suppose one planned to observe n independent sampling units, with n a positive integer. The plan was that each independent sampling unit would yield p observations. For $i \in \{1, 2, \ldots, n\}$, and $j \in \{1, 2, \ldots, p\}$, and $y_{ij} \in \Re$, let $\boldsymbol{Y} = \{y_{ij}\}$ be the $(n \times p)$ planned outcome matrix, with $\boldsymbol{Y}_i$ the $(1 \times p)$ matrix of planned outcomes for the i^{th} independent sampling unit. We assume that $p \ll n$.

4.2 Distribution of planned outcome data

Define $\boldsymbol{B}$ a $(q \times p)$ matrix, and $\boldsymbol{\Sigma}$ a $(p \times p)$ symmetric and positive definite matrix, both consisting of unknown parameters. Let

$$\boldsymbol{E} \sim N_{n,p}(\mathbf{0}, \boldsymbol{I}_n, \boldsymbol{\Sigma}) \tag{3}$$

be an $(n \times p)$ matrix of residual errors.

Let $\boldsymbol{X}$ be the $(n \times q)$ planned design matrix of predictors. Futhermore, assume no repeated covariates, and no randomness in $\boldsymbol{X}$.

Please explain what repeated covariates and randomness mean

If

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E} \tag{4}$$

then

$$\boldsymbol{Y} \sim N_{n,p}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Sigma}). \tag{5}$$

Write the density function of $\boldsymbol{Y}$ as $f_{\boldsymbol{Y}}\left(\overset{\clubsuit}{\boldsymbol{Y}}; \boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Sigma}\right)$. For convenience, we occasionally write it with the parameters suppressed as $f_{\boldsymbol{Y}}\left(\overset{\clubsuit}{\boldsymbol{Y}}\right)$.

4.3 Missing data process

Let p be a positive integer. Let $\mathcal{X} = \{1, 2, \ldots, p\}$ and let $d \in \{1, 2, \ldots, 2^p\}$. Let R_d be a potentially empty subset of $\mathcal{X}$, of cardinality $C(R_d) = p_d$, with $0 \le p_d \le p$. Assume without loss of generality that d is assigned so that the subsets R_d are sorted in quasi-lexicographic order.

Exhibit C. An example with $p = 3$, $R_d \subseteq \mathcal{X} = \{1, 2, 3\}$ and $d \in \{1, 2, \ldots, 8\}$.

$$
\begin{array}{lll}
R_1 = \emptyset & \mu_m(R_1) = \pi_{m,1} & \boldsymbol{z} = [1, 0, \ldots, 0] \\
R_2 = \{1\} & \mu_m(R_2) = \pi_{m,2} & \boldsymbol{z} = [0, 1, \ldots, 0] \\
R_3 = \{2\} & \mu_m(R_3) = \pi_{m,3} & \boldsymbol{z} = [0, 0, \ldots, 0] \\
R_4 = \{3\} & \mu_m(R_4) = \pi_{m,4} & \boldsymbol{z} = [0, 0, \ldots, 0] \\
R_5 = \{1, 2\} & \mu_m(R_5) = \pi_{m,5} & \boldsymbol{z} = [0, 0, \ldots, 0] \\
R_6 = \{1, 3\} & \mu_m(R_6) = \pi_{m,6} & \boldsymbol{z} = [0, 0, \ldots, 0] \\
R_7 = \{2, 3\} & \mu_m(R_7) = \pi_{m,7} & \boldsymbol{z} = [0, 0, \ldots, 0] \\
R_8 = \{1, 2, 3\} & \mu_m(R_8) = \pi_{m,8} & \boldsymbol{z} = [0, 0, \ldots, 1] \\
\hline
 & \multicolumn{2}{c}{1}
\end{array}
$$

Define $\mathcal{A} = \{R_1, R_2, \ldots, R_{2^p}\}$. Note $\mathcal{A}$ is a sigma-field of subsets of $\mathcal{X}$, since, with $\emptyset = R_1$, and $\mathcal{X} = R_{2^p}$, $R_1 \in \mathcal{A}$, and $R_{2^p} \in \mathcal{A}$. Further, $\mathcal{A}$ is closed under union (since $\mathcal{A}$ is the set of all subsets of $\mathcal{X}$, and similarly, under complementation).

Indicate a probability measure on the sigma-field $\mathcal{A}$ by μ_m, where m indexes a specific missing data mechanism m, by $\mu_m : \mathcal{A} \to [0, 1]$, with $\mu_m(R_d) = \pi_{m,d}$ for all $d \in \mathcal{A}$, and $\mu_m(R_d \cap R_{d'}) = 0$, for all $d \neq d'$. We therefore have $\sum_{d=1}^{2^p} \pi_{m,d} = 1$. Write $\boldsymbol{\pi}_m = [\pi_{m,1}, \pi_{m,2}, \ldots, \pi_{m,2^p}]$.

Let $\mathcal{B}$ be the standard basis for $\Re^{2^p}$. That is, let $\mathcal{B}$ be a set of all vectors of length 2^p which have exactly one entry with value one and zeroes elsewhere. Define the random $(2^p \times 1)$ vector $\boldsymbol{z} = \{Z_d\} = [Z_1, Z_2, \ldots, Z_{2^p}]$ such that $\boldsymbol{z} : \mathcal{A} \to \mathcal{B}$. Define the elements of $\boldsymbol{z}$ by letting $Z_d = 1$ if R_d occurs and 0 otherwise. Then Z_d is Bernoulli$(\pi_{m,d})$, and $\boldsymbol{z}$ is multinomial$(1, \boldsymbol{\pi}_m)$ if (8) holds, and not necessarily otherwise.

Now consider a random sample of size n of the random vector $\boldsymbol{z}$. Indicate each sample element by $\boldsymbol{z}_i$, for $i \in \{1, 2, \ldots n\}$. Define the random vector $\boldsymbol{n} = \sum_{i=1}^{n} \boldsymbol{z}_i = [N_1, N_2, \ldots N_{2^p}]$. Note that each N_d is a random variable.

We have $0 \leq N_d \leq n$. Note $\boldsymbol{n}'\mathbf{1}_{2^p} = n$. The vector $\boldsymbol{n}$ is multinomial$(n, \boldsymbol{\pi}_m)$, if (8) holds, and not necessarily otherwise. Let $F(n, \boldsymbol{n}, \boldsymbol{\pi}_m) \equiv F \leq \min(n, 2^p)$ be the random number of non-zero entries in the vector $\boldsymbol{n}$, and write the $(F \times 1)$ random vector of non-zero entries as $[N_{i_1}, N_{i_2}, \ldots, N_{i_F}]$ for $1 \leq i_1 \leq \cdots \leq i_F \leq 2^p$.

This is some tutorial rambing. It cannot actually appear here in the paper because it says the forbidden words independent sampling unit, and number of times things occur. However, it's still here because we will never understand this paper without reading this text.

Intuitively, N_1 is the number of times that R_1 occurs, and so on. If $i_1 = 1$, then we have observed the empty set N_{i_1} times. If $i_1 = 2$, the empty set was not observed. That is, there are no independent sampling units with completely missing data.

The missing data pattern for the i^{th} independent sampling unit is represented as $\boldsymbol{z}_i$. The random sample of size n of the vectors $\boldsymbol{z}_i$ corresponds to the missing data patterns for an entire data set of n independent sampling units. The number of independent sampling units with all observations missing is N_1. The number of independent sampling units with no observations missing is N_{2^p}.

Exhibit D. Frequency of deletion classes

$$p = 3$$
$$\mathcal{X} = \{1, 2, 3\}$$
$$d \in \{1, 2, \ldots, 8\}$$
$$R_d \subseteq \{1, 2, 3\}$$
$$N = 9$$
$$i_1 = 5;\ i_2 = 8; F = 2$$
$$n_5 = 5; n_8 = 4$$

Deletion pattern	Deletion class	Frequency of deletion class
$R_1 = \emptyset$	$\boldsymbol{z} = [1, 0, \ldots, 0]$	0
$R_2 = \{1\}$	$\boldsymbol{z} = [0, 1, \ldots, 0]$	0
$R_3 = \{2\}$	$\boldsymbol{z} = [0, 0, \ldots, 0]$	0
$R_4 = \{3\}$	$\boldsymbol{z} = [0, 0, \ldots, 0]$	0
$R_5 = \{1, 2\}$	$\boldsymbol{z} = [0, 0, \ldots, 0]$	5
$R_6 = \{1, 3\}$	$\boldsymbol{z} = [0, 0, \ldots, 0]$	0
$R_7 = \{2, 3\}$	$\boldsymbol{z} = [0, 0, \ldots, 0]$	0
$R_8 = \{1, 2, 3\}$	$\boldsymbol{z} = [0, 0, \ldots, 1]$	4
		9

4.4 Joint distribution of data matrix and data indicator matrix

With $\overset{\clubsuit}{\boldsymbol{Y}}$ the realization of $\boldsymbol{Y}$ and $\overset{\clubsuit}{\boldsymbol{n}} = [N_1 = n_1, N_2 = n_2, \ldots, N_{2^p} = n_{2^p}]$ the realization of the random vector $\boldsymbol{n}$, define the joint multivariate density function of the $(n \times p)$ outcome matrix $\boldsymbol{Y}$ and the $(2^p \times 1)$ vector $\boldsymbol{n}$ by

$$f_{\boldsymbol{Y},\boldsymbol{n}}\left(\overset{\clubsuit}{\boldsymbol{Y}}, \overset{\clubsuit}{\boldsymbol{n}}\right). \tag{6}$$

Further define the marginal density function of $\boldsymbol{n}$ so that

$$f_{\boldsymbol{n}}\left(\overset{\clubsuit}{\boldsymbol{n}}; n, \boldsymbol{\pi}\right), \tag{7}$$

written as $f_{\boldsymbol{n}}\left(\overset{\clubsuit}{\boldsymbol{n}}\right)$ with the parameters n and $\boldsymbol{\pi}$ suppressed for convenience. Assume that

$$f_{\boldsymbol{Y},\boldsymbol{n}}\left(\overset{\clubsuit}{\boldsymbol{Y}}, \overset{\clubsuit}{\boldsymbol{n}}\right) = f_{\boldsymbol{n}}\left(\overset{\clubsuit}{\boldsymbol{n}}\right) f_{\boldsymbol{Y}}\left(\overset{\clubsuit}{\boldsymbol{Y}}\right). \tag{8}$$

MNAR

MAR

MCAR

Assumptions

Expectation

$\mathcal{E}[g(D)|Y]$ or perhaps $\mathcal{E}[g(D)|Y^*]$

$\mathcal{E}[g(Y)|D]$ or perhaps $\mathcal{E}[g(Y^*)|D]$

These expectations will have different results depending on the assumptions.

Equation 4 may move here.

4. Wackers

Dear team. Here, we introduce three sorts of wacking matrices. These wacking matrices are defined by the missing process. We believe that they can eventually be unified by using an appropriate operator, and that they may all be the same, but we will clean this up later by deriving a grand unified wacking theory. For now, it will be complex. And we will love it.

4.1 Row deletion matrix $\boldsymbol{G}_{i_f}$

For $f = 1$, let the row deletion matrix $\boldsymbol{G}_{i_1}$, be the $(N_{i_1} \times n)$ submatrix of the $(n \times n)$ identity matrix $\boldsymbol{I}_n$ formed by keeping rows numbered 1 to N_{i_1} inclusive. For $f \in \{2, 3, \ldots, F\}$, let the row deletion matrix $\boldsymbol{G}_{i_f}$, be the $\left(N_{i_f} \times n\right)$ submatrix of the $(n \times n)$ identity matrix $\boldsymbol{I}_N$ formed by keeping rows numbered $\left(\sum_{g=1}^{f-1} N_{i_g} + 1\right)$ to $\sum_{g=1}^{f} N_{i_g}$, inclusive.

4.2 Column deletion matrix $\boldsymbol{D}_d$

For every $d > 1$, and for p_d the cardinality of R_d let the deletion matrix $\boldsymbol{D}_d$, be the $(p \times p_d)$ submatrix of the $(p \times p)$ identity matrix $\boldsymbol{I}_p$ formed by keeping each column k of

$\boldsymbol{I}_p$ such that $k \in R_d$. For example, if $p = 3$ and $R_d = \{1, 3\}$ then $p_d = 2$, and

$$\boldsymbol{D}_d = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{9}$$

When $d = i_f$, we write $\boldsymbol{D}_d = \boldsymbol{D}_{i_f}$

We note here that the previous line is completely unclear. We may wish to clean up this notation. Why were we using d instead of i_f? Brandy says it is because there is a deletion class corresponding to each of our $R_d s$, but we may or may not observe each one of them. They may not fit our criteria for having enough rows, so we had to do a secondary indexing to pick out the ones that fulfill our criteria. We should make a global statement ant define $\boldsymbol{D}_{i_f}$and $\overset{\bigstar}{\boldsymbol{Y}}_{i_f}$ **and** $\boldsymbol{Y}_{i_f}$ and $\boldsymbol{G}_{i_f}$ to explain this point. Later.

4.3 Yet another deletion matrix $\boldsymbol{T}(\boldsymbol{A})$

Write the $(w \times m)$ matrix $\boldsymbol{A} = [\,\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_m\,]$. Let

$$\Omega(\boldsymbol{a}_i) = \begin{cases} 0 & \max\{\boldsymbol{a}_i\} = \min\{\boldsymbol{a}_i\} = 0 \\ 1 & \text{else} \end{cases}. \tag{10}$$

Define the $(m \times 1)$ vector $\boldsymbol{\Omega}(\boldsymbol{A})$ as $\boldsymbol{\Omega}(\boldsymbol{A}) = [\Omega(\boldsymbol{a}_1)\ \Omega(\boldsymbol{a}_2) \ldots \Omega(\boldsymbol{a}_m)]'$. Let b be the total number of entries that are 1 in $\Omega(\boldsymbol{A})$. There are $2^m - 1$ possible $\boldsymbol{\Omega}(\boldsymbol{A})$ matrices (omitting the empty set). Indicate which one we have by $\boldsymbol{\Omega}_w(\boldsymbol{A})$, $w \in \{1, 2, \ldots, 2^m - 1\}$. We might as well sort them in quasi-lexicographic order. Let the deletion matrix $\boldsymbol{T}(\boldsymbol{A})$ be the $(m \times b)$ submatrix of the $(m \times m)$ identity matrix $\boldsymbol{I}_m$ formed by keeping each column i of $\boldsymbol{I}_m$ such that row i of $\boldsymbol{\Omega}(\boldsymbol{A})$ is non-zero.

We have boys and girls. They get their blood pressure measured 3 times. Sometimes, the boys and girls do not show up for their measurement and so we do not observe the measurement. A supreme being puts all their measurements into a matrix called Y. The supreme being knows what the measurements would be regardless of whether we lowly mortals actually observed them or not. Sort the matrix into sub-matrices depending on the pattern of observed measurements. If we assume that boys and girls are equally likely to show up to be measured, then we can calculate the expected number of boys (or girls) in each sub-matrix as number of rows in the sub-matrix times the probability that a boy (or girl) shows up to be measured. We can also still probably calculate the expected number of boys (or girls) in each sub-matrix if they show up at different probabilities. It is stil a binomial process, just with different p's.

Brandy would really like a common operator here. Actually, she wants $\boldsymbol{\Omega}(\boldsymbol{X}_{i_f})$ to return the row labels, not 1's or zero's. Actually, she wants a common operator so that this deletion matrix construction task parallels that from section 4.3. Which is reasonable, I think. Brandy objects to using functional notation since we do not use it for the other two deletion matrices.

5. Formation and distribution of the reduced data

5.1 Overview

The matrices $\boldsymbol{Y}$ and $\boldsymbol{X}$ may have missing data. We will describe the missing data pattern by subdividing the rows of $\boldsymbol{Y}$ and $\boldsymbol{X}$ by deletion class. Each subdivision will have rows with the same pattern of missing data. We will eventually form estimates of the mean and variance for each of those deletion classes. Then, we will remove the columns of $\boldsymbol{Y}$ that are all zeroes, which corresponds to removing repeated measures for which those rows have no observed data. We will also remove columns of $\boldsymbol{X}$ that are all zeroes, which corresponds to removing levels of the predictor variables that have no observed data.

5.2 Row whackery

Subdivide $\boldsymbol{Y}$ and $\boldsymbol{X}$ into the $F \leq \min(n, 2^p)$ submatrices $\boldsymbol{Y}_{i_1}, \boldsymbol{Y}_{i_2}, \ldots \boldsymbol{Y}_{i_F}$, each with N_{i_f} rows and p columns, respectively, for $f \in \{1, 2, \ldots, F\}$, by defining

$$\boldsymbol{Y}_{i_f} = \boldsymbol{G}_{i_f}\boldsymbol{Y}. \tag{11}$$

Similarly, for $f \in \{1, 2, \ldots, F\}$, write

$$\boldsymbol{X}_{i_f} = \boldsymbol{G}_{i_f}\boldsymbol{X}, \tag{12}$$

where $\boldsymbol{X}_{i_f}$ is a $\left(N_{i_f} \times q\right)$ matrix. Note that $\operatorname{rank}\left(\boldsymbol{X}_{i_f}\right) \leq \operatorname{rank}(\boldsymbol{X})$. Let $\operatorname{rank}\left(\boldsymbol{X}_{i_f}\right) = q_{i_f} \leq q$.

Exhibit E. $\boldsymbol{Y}$ row deletion for $i_1 > 1$

$$N_{i_1} = 5$$
$$n = 9$$

$$\underset{(5\times 9)}{\boldsymbol{G}_{i_1}} = \begin{bmatrix} \boldsymbol{I}_5 & \underset{(5\times 4)}{\mathbf{0}} \end{bmatrix} \qquad \underset{(9\times 3)}{\boldsymbol{Y}} \qquad \underset{(5\times 3)}{\boldsymbol{Y}_{i_1}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{13} \\ y_{31} & y_{32} & y_{13} \\ y_{41} & y_{42} & y_{13} \\ y_{51} & y_{52} & y_{13} \\ y_{61} & y_{62} & y_{63} \\ y_{71} & y_{72} & y_{73} \\ y_{81} & y_{82} & y_{83} \\ y_{91} & y_{92} & y_{93} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \\ y_{41} & y_{42} & y_{43} \\ y_{51} & y_{52} & y_{53} \end{bmatrix}$$

Exhibit F. $\boldsymbol{X}$ row deletion for $i_1 > 1$

$$N_{i_1} = 5$$
$$n = 9$$

$$\underset{(5\times 9)}{\boldsymbol{G}_{i_1}} = \begin{bmatrix} \boldsymbol{I}_5 & \underset{(5\times 4)}{\mathbf{0}} \end{bmatrix} \qquad \underset{(9\times 2)}{\boldsymbol{Y}} \qquad \underset{(5\times 2)}{\boldsymbol{X}_{i_1}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{13} \\ y_{31} & y_{32} & y_{13} \\ y_{41} & y_{42} & y_{13} \\ y_{51} & y_{52} & y_{13} \\ y_{61} & y_{62} & y_{63} \\ y_{71} & y_{72} & y_{73} \\ y_{81} & y_{82} & y_{83} \\ y_{91} & y_{92} & y_{93} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \\ x_{51} & x_{52} \end{bmatrix}$$

Note that $\boldsymbol{Y}_{i_1}$ is of dimension (5×3), and is a submatrix of $\boldsymbol{Y}$, and $\boldsymbol{X}_{i_1}$ is of dimension (5×2) and is a submatrix of $\boldsymbol{X}$.

5.3 Column deletion of $\boldsymbol{Y}_{i_f}$ and $\boldsymbol{X}_{i_f}$

For $f \in \{1, 2, \ldots, F\}$, define the $\left(N_{i_f} \times p_{i_f}\right)$ matrix $\overset{\bigstar}{\boldsymbol{Y}}_{i_f}$ such that

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} = \underset{(N_{i_F} \times n \times p \times p_{i_f})}{\boldsymbol{G}_{i_f} \boldsymbol{Y} \boldsymbol{D}_{i_f}} = \boldsymbol{Y}_{i_f} \boldsymbol{D}_{i_f}. \tag{13}$$

Nonempty Y big stars

Then the set of random matrices that represent the non-missing data is given by

$$\mathcal{Y} = \begin{cases} \left\{ \overset{\bigstar}{\boldsymbol{Y}}_{i_2}, \ldots, \overset{\bigstar}{\boldsymbol{Y}}_{i_F} \right\} & i_1 = 1 \\ \left\{ \overset{\bigstar}{\boldsymbol{Y}}_{i_1}, \overset{\bigstar}{\boldsymbol{Y}}_{i_2}, \ldots, \overset{\bigstar}{\boldsymbol{Y}}_{i_F} \right\} & i_1 > 1. \end{cases} \tag{14}$$

Note that $\text{rank}(\boldsymbol{X}_{i_1})$ may be less than or equal to $\text{rank}(\boldsymbol{X})$. For example, if

$$\underset{(9 \times 2)}{\boldsymbol{X}} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}, \tag{15}$$

and

$$\underset{(5 \times 2)}{\boldsymbol{X}_{i_1}} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{16}$$

then $\text{rank}(\boldsymbol{X}) = 2$ and $\text{rank}\left(\boldsymbol{X}_{i_f}\right) = 1$.

We cannot perform estimation if $\text{rank}(\boldsymbol{X}_{i_1})$ is less than full rank. We define a transformation to ensure that $\boldsymbol{X}_{i_1}$ is full rank. Note that $\boldsymbol{\Omega}\left(\boldsymbol{X}_{i_f}\right) = [0 \quad 1]'$, yielding $\boldsymbol{T}_{i_1} = [0 \quad 1]'$. Now $\overset{\bigstar}{\boldsymbol{X}}_{i_f} = \boldsymbol{X}_{i_f} \boldsymbol{T}_{i_f}$ is full rank, as shown in Equation 17.

$$\underset{(5\times 1)}{\overset{\bigstar}{\boldsymbol{X}}_{i_f}} = \boldsymbol{X}_{i_f}\boldsymbol{T}_{i_f} \tag{17}$$

$$= \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Find $\boldsymbol{\Omega}(\boldsymbol{X}_{i_1})$ and the $\left(q \times q_{i_f}\right)$ matrix $\boldsymbol{T}\left(\boldsymbol{X}_{i_f}\right) = \boldsymbol{T}_{i_f}$. Construct a full column rank $\left(N_{i_f} \times q_{i_f}\right)$ matrix $\overset{\bigstar}{\boldsymbol{X}}_{i_f}$ of rank q_{i_f} as

$$\overset{\bigstar}{\boldsymbol{X}}_{i_f} = \underset{N_{i_F} \times n \times q \times q_{i_f}}{\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{T}_{i_f}}. \tag{18}$$

$\overset{\bigstar}{\boldsymbol{X}}_{i_f}$ is the observed predictor matrix in the missing data universe for deletion class i_f

5.4 Distribution of the reduced data for $\boldsymbol{X}_{i_f}$ full rank

This section, and the following section may wish to combine, if we get fancy, and the math looks the same. We did it seperately, because we don't know until we do it what the answer is. Also, we are not sure if the constraint $N_{i_f} > q$ is correct.

Here, we assume that $N_{i_f} > q$. When $\operatorname{rank}\left(\boldsymbol{X}_{i_f}\right) = q$ for all $f \in \{1, 2, \ldots, F\}$, the distribution theory is as follows. By Arnold, 1981, p. 312, Theorem 17.2 (the reproductive property of normals theorem), since

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} = \boldsymbol{G}_{i_f}\boldsymbol{Y}\boldsymbol{D}_{i_f}, \tag{19}$$

we have

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} \sim N_{N_{i_f}, p_{i_f}}\left(\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}_{i_f}, \boldsymbol{G}_{i_f}\boldsymbol{I}_n\boldsymbol{G}'_{i_f}, \boldsymbol{D}'_{i_f}\boldsymbol{\Sigma}\boldsymbol{D}_{i_f}\right). \tag{20}$$

Note that because

$$\begin{aligned}\boldsymbol{G}_{i_f}\boldsymbol{I}_n\boldsymbol{G}'_{i_f} &= \underset{(N_{i_1}\times n)}{\boldsymbol{G}_{i_f}}\underset{(n\times N_{i_1})}{\boldsymbol{G}'_{i_f}} \\ &= \left(\boldsymbol{G}_{i_f}\boldsymbol{G}'_{i_f}\right)' \\ &= \boldsymbol{I}_{N_{i_f}},\end{aligned} \tag{21}$$

it implies

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} \sim N_{N_{i_f},p_{i_f}}\left(\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}_{i_f}, \boldsymbol{I}_{N_{i_f}}, \boldsymbol{D}'_{i_f}\boldsymbol{\Sigma}\boldsymbol{D}_{i_f}\right). \tag{22}$$

Define the $\left(p_{i_f}\times p_{i_f}\right)$ matrix $\boldsymbol{\Sigma}_{i_f}$ by

$$\boldsymbol{\Sigma}_{i_f} = \boldsymbol{D}'_{i_f}\boldsymbol{\Sigma}\boldsymbol{D}_{i_f}, \tag{23}$$

yielding

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} \sim N_{N_{i_f},p_{i_f}}\left(\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}_{i_f}, \boldsymbol{I}_{N_{i_f}}, \boldsymbol{\Sigma}_{i_f}\right). \tag{24}$$

The deletion operation preserves row independence, albeit now with a reduced number of rows. That is, we now have N_{i_f} rather than n rows.

5.5 Distribution of the reduced data for $\boldsymbol{X}_{i_f}$ potentially less than full rank

Here, we assume that $N_{i_f} > q_{i_f}$. When rank$\left(\boldsymbol{X}_{i_f}\right) = q_{i_f} \leq q$ for all $f \in \{1,2,\ldots,F\}$, the distribution theory is as follows. Furthermore, $\boldsymbol{B}_{i_f} =$ For this case, we have $N_{i_F}\times n\times q\times q_{i_f}$

$$\underset{N_{i_f}\times p_{i_f}}{\overset{\bigstar}{\boldsymbol{Y}}_{i_f}} = \underset{N_{i_F}\times n\times q\times q_{i_f}\times q\times p\times p_{i_f}}{\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{T}_{i_f}\boldsymbol{T}'_{i_f}\boldsymbol{B}\boldsymbol{D}_{i_f}}. \tag{25}$$

We guess that

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} \sim N_{N_{i_f},p_{i_f}}\left(\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{T}_{i_f}\boldsymbol{T}'_{i_f}\boldsymbol{B}\boldsymbol{D}_{i_f}, \boldsymbol{I}_{N_{i_f}},\boldsymbol{\Sigma}_{i_f}\right). \tag{26}$$

Note that $\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{T}_{i_f}$ is the only predictor matrix that Deb and Brandy in the missing data universe can see, and $\boldsymbol{T}_{i_f}'\boldsymbol{B}\boldsymbol{D}_{i_f}$ is the only beta that the Deb and Brandy in the missing data universe can calculate.

$OCTOPUS$

6. Estimation

6.1 Estimation within deletion classes

Suppose $T \in \{0,1,\ldots,F\}$ is the number of $\overset{\bigstar}{\boldsymbol{Y}}_{i_f}$ such that $i_f > 1$ and $\left(N_{i_f} - 1\right) > p_{i_f}$.

The above constraints may not be entirely correct. Please fix later. We are especially concerned about $\left(N_{i_f} - 1\right) > p_{i_f}$. Also, Keith would like to put on further side conditions/ fix this one later. Goals: HLT df > 0 and invertibility.

Here, T is a random variable. Let $\{k_1, k_2, \ldots, k_T\} \subseteq \{i_1, i_2, \ldots, i_F\}$ index the elements of $\mathcal{A}$ that satisfy the following two conditions : 1) $i_f > 1$ and 2)$\left(N_{i_f} - 1\right) > p_{i_f}$. Estimation is not possible for $i_f = 1$ as that corresponds to the condition where all data are missing.

Because we are dividing the data into subsets, a bad thing might happen. X has levels. We might not get 1 participant per level of X in a subset. This creates an X columnwise deletion for that subgroup (subgroups are the deletion classes). Previously, we defined Y deletion classes. We now need to define X deletion classes. In addition to defining X deletion classes, we need to describe the rank of X in the resulting deletion classes. So we will end up with more subsets when we cross classify the Y and X deletion classes. We will need to redo the weighting on beta and sigma and redo proofs of unbiasedness and consistency for those new weighted estimators.

This brings to light the need to talk about efficiency and whether and under what circumstances our estimators achieve the Cramer-Rao lower bound. Under the assumption that we understand the missing process, we can define indicator variables that are 1 if a cross-classification of an X deletion and a Y deletion has enough rows in it and 0 otherwise. We can compute the probability that that indicator variable is 1. We can also compute the variance of that indicator variable, which will allow us to calculate the expected relative asymptotic efficiency with a confidence interval around it. Note that the formula for the expected asymptotic relative efficiency will depend on the missing process. The missing process and design features (1:1 randomization or 2:1 randomization for example) is what defines the expected value of the number of rows in a specific Y deletion class and the number of rows in a specific X deletion class.

KEVIN CODE NOTE: Kevin coded a base10 order for the deletion classes. We need to go back to his dataset and instead code a variable to order the deletion classes in quasi lexicographic order.

NOTE: Brandy will go back and correct the definition of $\overset{\bigstar}{\boldsymbol{Y}}_{k_t}$ so that it does not use i_f. Here is the sentence she will fix: "When $T = 0$, the parameters are non-estimable. When $T > 0$, and with $i_f \in \{k_1, k_2, \ldots, k_T\}$, consider T models of the form". In addition, she will change all the subsequent notation to k_t.

$$\underset{(N_{i_f} \times p_{i_f})}{\overset{\bigstar}{\boldsymbol{Y}}_{i_f}} = \underset{(N_{i_f} \times N \times q \times p \times p_{i_f})}{\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}_{i_f}} + \boldsymbol{G}_{i_f}\boldsymbol{E}\boldsymbol{D}_{i_f}. \tag{27}$$

Estimate the $(q \times p_{i_f})$ matrix $\boldsymbol{B}_{i_f} = \boldsymbol{B}\boldsymbol{D}_{i_f}$ by

$$\underset{(q \times p_{i_f})}{\widehat{\boldsymbol{B}}_{i_f}} = \left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{G}'_{i_f}\overset{\bigstar}{\boldsymbol{Y}}_{i_f}, \tag{28}$$

and estimate the $(p_{i_f} \times p_{i_f})$ matrix $\boldsymbol{\Sigma}_{i_f}$ by

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{i_f} &= \left(N_{i_f} - q_{i_f}\right)^{-1} \overset{\bigstar}{\boldsymbol{Y}}'_{i_f} \left(\boldsymbol{I}_{N_{i_f}} - \boldsymbol{H}_{i_f}\right) \overset{\bigstar}{\boldsymbol{Y}}_{i_f}, \\
&= \left(N_{i_f} - q_{i_f}\right)^{-1} \widehat{\boldsymbol{S}}_{e,N_{i_f}}
\end{aligned}
\tag{29}
$$

where

$$
\boldsymbol{H}_{i_f} = \boldsymbol{G}_{i_f} \boldsymbol{X} \left(\boldsymbol{X}' \boldsymbol{G}'_{i_f} \boldsymbol{G}_{i_f} \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \boldsymbol{G}'_{i_f}. \tag{30}
$$

6.2 Estimation of effective sample sizes

The *total estimable sample size* is given by

$$
N_* = \sum_{t=1}^{T} N_{k_t}. \tag{31}
$$

Let $\mathcal{L}_j = \{d : j \in R_d\}$. Define the *column-wise sample size*, totalled over all the subsets R_d that contain j, by

$$
\widetilde{N}_j = \sum_{k_t \in \mathcal{L}_j} N_{k_t}. \tag{32}
$$

An example is shown in Exhibit G.

Exhibit G. Example $\mathcal{L}_j$ and $\widetilde{N}_j$ for $j = 3$.

$$
\begin{aligned}
p &= 3 \\
\mathcal{X} &= \{1, 2, 3\} \\
d &\in \{1, 2, \ldots, 8\} \\
R_d &\subseteq \{1, 2, 3\} \\
N &= 9 \\
\mathcal{L}_3 &= \{4, 6, 7, 8\} \\
\widetilde{N}_3 &= N_4 + N_6 + N_7 + N_8 = 4
\end{aligned}
$$

Deletion pattern	Contain 3	Frequency of deletion class
$R_1 = \emptyset$	No	$N_1 = 0$
$R_2 = \{1\}$	No	$N_2 = 0$
$R_3 = \{2\}$	No	$N_3 = 0$
$R_4 = \{3\}$	Yes	$N_4 = 0$
$R_5 = \{1, 2\}$	No	$N_5 = 5$
$R_6 = \{1, 3\}$	Yes	$N_6 = 0$
$R_7 = \{2, 3\}$	Yes	$N_7 = 0$
$R_8 = \{1, 2, 3\}$	Yes	$N_8 = 4$
		$N = 9$

Similarly, for j and $j' \in \{1, 2, \ldots, p\}$, $j \neq j'$, let $\mathcal{L}_{jj'} = \{d : j \text{ and } j' \in R_d\}$, and define the *paired column-wise sample size* as

$$\widetilde{N}_{jj'} = \sum_{k_t \in \mathcal{L}_{jj'}} N_{k_t}. \tag{33}$$

An example is shown in Exhibit H.

Exhibit H. Example $\mathcal{L}_{jj'}$ and $\widetilde{N}_{jj'}$ for $j = 1$ and $j' = 3$.

$$\begin{aligned} p &= 3 \\ \mathcal{X} &= \{1, 2, 3\} \\ d &\in \{1, 2, \ldots, 8\} \\ R_d &\subseteq \{1, 2, 3\} \\ N &= 9 \\ \mathcal{L}_{jj'} &= \{6, 8\} \\ \widetilde{N}_{jj'} &= N_6 + N_8 = 4 \end{aligned}$$

Deletion pattern	Contain 1 and 3	Frequency of deletion class
$R_1 = \emptyset$	No	$N_1 = 0$
$R_2 = \{1\}$	No	$N_2 = 0$
$R_3 = \{2\}$	No	$N_3 = 0$
$R_4 = \{3\}$	No	$N_4 = 0$
$R_5 = \{1, 2\}$	No	$N_5 = 5$
$R_6 = \{1, 3\}$	Yes	$N_6 = 0$
$R_7 = \{2, 3\}$	No	$N_7 = 0$
$R_8 = \{1, 2, 3\}$	Yes	$N_8 = 4$
		$N = 9$

6.3 *Estimation of* $\boldsymbol{B}$

When $\#(\mathcal{L}_j) \geq 1$ for all $j \in \{1, 2, \ldots, p\}$, $\boldsymbol{B}$ is estimable. Note we can only check that $\#(\mathcal{L}_j) \geq 1$ for all $j \in \{1, 2, \ldots, p\}$ if the missingness process has been observed. Define $\widehat{\boldsymbol{B}}_W$, an estimator of $\boldsymbol{B}$, by

$$\widehat{\boldsymbol{B}}_W = \sum_{t=1}^{T} \frac{N_{k_t}}{N_*} \underset{(q \times q_{k_t} \times p_{k_t} \times p)}{\boldsymbol{T}_{k_t} \widehat{\boldsymbol{B}}_{k_t} \boldsymbol{D}'_{k_t}}. \tag{34}$$

6.4 *Estimation of* $\boldsymbol{\Sigma}$

When $\#(\mathcal{L}_j) \geq 1$ for all $j \in \{1, 2, \ldots, p\}$ and $\#(\mathcal{L}_{jj'}) \geq 1$ for all $j \neq j'$, and $j, j' \in \{1, 2, \ldots, p\}$, $\boldsymbol{\Sigma}$ is estimable. Let $\boldsymbol{\delta}_j$ be an $(p \times 1)$ vector with a value of 1 in the j^{th} row, and a value of 0 everywhere else.

$$\widehat{\sigma}_{jj}^2 = \sum_{k_t \in \mathcal{L}_j} \frac{N_{k_t}}{\widetilde{N}_j} \boldsymbol{\delta}_j' \boldsymbol{D}_{k_t} \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}_{k_t}' \boldsymbol{\delta}_j. \tag{35}$$

$$\widehat{\rho}_{jj'} \widehat{\sigma}_{jj} \widehat{\sigma}_{j'j'} = \sum_{k_t \in \mathcal{L}_{jj'}} \frac{N_{k_t}}{\widetilde{N}_{\ jj'}} \boldsymbol{\delta}_j' \boldsymbol{D}_{k_t} \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}_{k_t}' \boldsymbol{\delta}_{j'}. \tag{36}$$

$$\widehat{\boldsymbol{\Sigma}}_W = \begin{bmatrix} \widehat{\sigma}_{11}^2 & \widehat{\rho}_{12}\widehat{\sigma}_{11}\widehat{\sigma}_{22} & \cdots & \widehat{\rho}_{1p}\widehat{\sigma}_{11}\widehat{\sigma}_{pp} \\ \widehat{\rho}_{12}\widehat{\sigma}_{11}\widehat{\sigma}_{22} & \widehat{\sigma}_{22}^2 & \cdots & \widehat{\rho}_{2p}\widehat{\sigma}_{22}\widehat{\sigma}_{pp} \\ & & \ddots & \\ \widehat{\rho}_{1p}\widehat{\sigma}_{11}\widehat{\sigma}_{pp} & \widehat{\rho}_{2p}\widehat{\sigma}_{22}\widehat{\sigma}_{pp} & \cdots & \widehat{\sigma}_{pp}^2 \end{bmatrix}. \tag{37}$$

6.5 Chance of estimation success

If you do not observe data, you cannot estimate things.

Define the probability that sigma jj estimation will fail, that sigma jj' will fail and that betaij estimation will fail. This is basically having too small a sample size for a deletion class. So $\mathcal{L}$j can be empty or it can be non-empty but the sample size is insufficient for the Rkt in $\mathcal{L}$j.

Cool figures that are functions of sample size and missingness go here.

7. Distributional results

7.1 Exact distribution of $\widehat{\boldsymbol{B}}_W$

From Equation 38,

$$\overset{\bigstar}{\boldsymbol{Y}}_{i_f} \sim N_{N_{i_f}, p_{i_f}} \left(\boldsymbol{G}_{i_f} \boldsymbol{X} \boldsymbol{B} \boldsymbol{D}_{i_f}, \boldsymbol{I}_{N_{i_f}}, \boldsymbol{\Sigma}_{i_f} \right). \tag{38}$$

Recall that the $\left(q \times p_{i_f}\right)$ matrix $\boldsymbol{B}_{i_f} = \boldsymbol{B}\boldsymbol{D}_{i_f}$ is defined by

$$\underset{\left(q \times p_{i_f}\right)}{\widehat{\boldsymbol{B}}_{i_f}} = \left(\boldsymbol{X}' \boldsymbol{G}_{i_f}' \boldsymbol{G}_{i_f} \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \boldsymbol{G}_{i_f}' \overset{\bigstar}{\boldsymbol{Y}}_{i_f}. \tag{39}$$

Using Theorem 17.2 on the reproductive property of matrix normals, (Arnold, 1981, p. 312),

$$\widehat{\boldsymbol{B}}_{i_f} \sim N_{q,p_{i_f}}\left[\boldsymbol{B}\boldsymbol{D}_{i_f}, \left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}, \boldsymbol{\Sigma}_{i_f}\right] \tag{40}$$

since

$$\begin{aligned}\left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{G}'_{i_f}\left(\boldsymbol{G}_{i_f}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}_{i_f}\right) &= \left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)\boldsymbol{B}\boldsymbol{D}_{i_f} \\ &= \boldsymbol{B}\boldsymbol{D}_{i_f}\end{aligned}$$

and

$$\begin{aligned}&\left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{I}_{N_{i_f}}\boldsymbol{G}_{i_f}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1} \\ &\quad= \left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1} \\ &\quad= \left(\boldsymbol{X}'\boldsymbol{G}'_{i_f}\boldsymbol{G}_{i_f}\boldsymbol{X}\right)^{-1}.\end{aligned}$$

Recall that $\{k_1, k_2, \ldots, k_t\} \subseteq \{i_1, i_2, \ldots, i_F\}$. Thus distributional results for $\widehat{\boldsymbol{B}}_{i_f}$, such as appear in Equation 0, apply to $\widehat{\boldsymbol{B}}_{k_t}$.

Again invoking Theorem 17.2 (Arnold, 1981, p. 312)

$$\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t} \sim (S)\mathcal{N}_{q,p}\left[\boldsymbol{B}\boldsymbol{D}_{k_t}\boldsymbol{D}'_{k_t}, \left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}, \boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t}\right] \tag{41}$$

Note that the column covariance term, $\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t}$, may contain zeroes and thus be less than full rank, potentially causing the matrix normal to be singular. Recall that

$$\widehat{\boldsymbol{B}}_W = \sum_{t=1}^{T}\frac{N_{k_t}}{N_*}\underset{(q\,\times\,p_{k_t}\,\times\,p)}{\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t}} \tag{42}$$

and then,

$$\begin{aligned}\mathcal{E}\left(\widehat{\boldsymbol{B}}_W \middle| \boldsymbol{n}\right) &= \mathcal{E}\left(\sum_{t=1}^{T}\frac{N_{k_t}}{N_*}\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t}\middle|\boldsymbol{n}\right) \\ &= \sum_{t=1}^{T}\mathcal{E}\left(\frac{N_{k_t}}{N_*}\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t}\middle|\boldsymbol{n}\right) \\ &= \sum_{t=1}^{T}\left(\frac{N_{k_t}}{N_*}\right)\boldsymbol{B}\boldsymbol{D}_{k_t}\boldsymbol{D}'_{k_t} \\ &= \boldsymbol{B}\end{aligned} \tag{43}$$

Thus, $\widehat{\boldsymbol{B}}_W$ is unbiased.

We now show that $\widehat{\boldsymbol{B}}_W$ is consistent. Note that if

$$\boldsymbol{X} \sim \mathcal{N}_{q,p}(\boldsymbol{M},\boldsymbol{U},\boldsymbol{V}) \tag{44}$$

then

$$\mathrm{vec}(\boldsymbol{X}) \sim \mathcal{N}_{qp,1}[\mathrm{vec}(\boldsymbol{M}),1,\boldsymbol{V}\otimes\boldsymbol{U}]. \tag{45}$$

$$\begin{aligned}\boldsymbol{\mathcal{V}}\left[\mathrm{vec}\left(\widehat{\boldsymbol{B}}_W\right)\middle|\boldsymbol{n}\right] &= \boldsymbol{\mathcal{V}}\left[\mathrm{vec}\left(\sum_{t=1}^{T}\frac{N_{k_t}}{N_*}\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t}\right)\middle|\boldsymbol{n}\right] \\ &= \sum_{t=1}^{T}\boldsymbol{\mathcal{V}}\left[\frac{N_{k_t}}{N_*}\mathrm{vec}\left(\widehat{\boldsymbol{B}}_{k_t}\boldsymbol{D}'_{k_t}\right)\middle|\boldsymbol{n}\right] \\ &= \sum_{t=1}^{T}\left(\frac{N_{k_t}^2}{N_*^2}\right)\left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t}\otimes\left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right].\end{aligned} \tag{46}$$

Note that if $\max(\boldsymbol{\Sigma}_{k_t},\boldsymbol{X}) < \infty$, then

$$\lim_{N\to\infty}\max\left|\left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t}\otimes\left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right]\right| = 0. \tag{47}$$

Proof: Since $\max(\boldsymbol{\Sigma}_{k_t},\boldsymbol{X}) < \infty$, $\exists$ some number $M \in \Re$ such that $\max|(\boldsymbol{\Sigma}_{k_t},\boldsymbol{X})| < M$. Define $\boldsymbol{M} = M\mathbf{1}_{N,p}$. Then

$$
\begin{aligned}
\lim_{N\to\infty} & \left\{ \max\left| \left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t} \otimes \left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right]\right|\right\} \\
& \leq \lim_{N\to\infty} \max\left|\left[\boldsymbol{\Sigma}_{k_t} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right]\right| \\
& \leq \lim_{N\to\infty} \max\left|\left[\boldsymbol{\Sigma}_{k_t} \otimes (\boldsymbol{M}'\boldsymbol{M})^{-1}\right]\right| \\
& = \lim_{N\to\infty} \max\left| M^2\left[\boldsymbol{\Sigma}_{k_t} \otimes \left(\mathbf{1}'_{N,p}\mathbf{1}_{N,p}\right)^{-1}\right]\right| \\
& = \lim_{N\to\infty} \max\left| M^2 N^{-1}\left[\boldsymbol{\Sigma}_{k_t} \otimes (\mathbf{1}_{p,p})^{-1}\right]\right| \\
& = 0
\end{aligned}
$$

Since by Equation 31, we have $N_{k_t} \leq N_*$, and since $T < \infty$,

$$
\lim_{N\to\infty} \sum_{t=1}^{T} \left(\frac{N_{k_t}^2}{N_*^2}\right) \left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t} \otimes \left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right] = 0
$$

So $\widehat{\boldsymbol{B}}_W$ is consistent.

Must show that $\sum_{t=1}^{T} \left(\frac{N_{k_t}^2}{N_*^2}\right) \left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t} \otimes \left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right]$ is positive definite, and hence $\widehat{\boldsymbol{B}}_W \sim$ N(something, something), not $\widehat{\boldsymbol{B}}_W \sim (\mathcal{S})$N(something, something)

To show it is positive definite, for $t \in \{1, 2, \ldots, T\}$, let $\boldsymbol{\mathcal{A}}_t = \left(\frac{N_{k_t}^2}{N_*^2}\right) \left[\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t} \otimes \left(\boldsymbol{X}'\boldsymbol{G}'_{k_t}\boldsymbol{G}_{k_t}\boldsymbol{X}\right)^{-1}\right]$, and define a $Tp \times Tp$ supermatrix by

$$
\boldsymbol{\mathcal{S}} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{A}}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{A}}_T \end{bmatrix}. \tag{49}
$$

Then

$$\boldsymbol{\mathcal{V}}\left[\text{vec}\left(\widehat{\boldsymbol{B}}_W\right)\middle|\boldsymbol{n}\right] =$$

7.2 Approximate distribution of $\widehat{\boldsymbol{\Sigma}}_W$

Claim: For the complete case (no missing data) $\widehat{\boldsymbol{\Sigma}}_W$ has parameter (n). If you delete the nth person, the resulting distribution of $\widehat{\boldsymbol{\Sigma}}_W$ has parameter $(n-1)$. Thus partial deletions of the nth person, such as removing measures 1, 3 and 7 must have parameter $(n-1) \leq n^* \leq n$.

Further claim: For the complete case (no missing data) $\widehat{\boldsymbol{\Sigma}}_W$ has dimension p. If you delete the pth measure, the resulting distribution of $\widehat{\boldsymbol{\Sigma}}_W$ has dimension $p-1$ (and appropriately deleted $\boldsymbol{\Sigma}$ matrix. Thus for partial deletions of measure p, such as removing measure p for persons 2, 4 and 6, $\widehat{\boldsymbol{\Sigma}}_W$ must have something reduced- but not the dimension. Brandy suspects it will be the n parameter.

Show result is positive definite with probability 1.

Find expected value and variance of the result.

$$
\begin{aligned}
\mathcal{E}\left(\,\widehat{\sigma}^2_{jj}|\boldsymbol{n}\right) &= \mathcal{E}\left(\sum_{k_t\in\mathcal{L}_j}\frac{N_{k_t}}{\widetilde{N}_j}\boldsymbol{\delta}'_j\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}'_{k_t}\boldsymbol{\delta}_j|\boldsymbol{n}\right) && 1 \\
&= \sum_{k_t\in\mathcal{L}_j}\mathcal{E}\left(\frac{N_{k_t}}{\widetilde{N}_j}\boldsymbol{\delta}'_j\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}'_{k_t}\boldsymbol{\delta}_j|\boldsymbol{n}\right) && 2 \\
&= \sum_{k_t\in\mathcal{L}_j}\frac{N_{k_t}}{\widetilde{N}_j}\boldsymbol{\delta}'_j\boldsymbol{D}_{k_t}\left[\mathcal{E}\left(\widehat{\boldsymbol{\Sigma}}_{k_t}|\boldsymbol{n}\right)\right]\boldsymbol{D}'_{k_t}\boldsymbol{\delta}_j && 3 \\
&= \sum_{k_t\in\mathcal{L}_j}\frac{N_{k_t}}{\widetilde{N}_j}\boldsymbol{\delta}'_j\boldsymbol{D}_{k_t}\left[\mathcal{E}\left(\widehat{\boldsymbol{\Sigma}}_{k_t}\right)\right]\boldsymbol{D}'_{k_t}\boldsymbol{\delta}_j && 4 \\
&= \sum_{k_t\in\mathcal{L}_j}\frac{N_{k_t}}{\widetilde{N}_j}\boldsymbol{\delta}'_j\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}'_{k_t}\boldsymbol{\delta}_j && 5 \\
&= \sum_{k_t\in\mathcal{L}_j}\frac{N_{k_t}}{\widetilde{N}_j}\;\sigma^2_{jj} && 6 \\
&= \frac{\sigma^2_{jj}}{\widetilde{N}_j}\sum_{k_t\in\mathcal{L}_j}N_{k_t} && 7 \\
&= \sigma^2_{jj}\frac{\widetilde{N}_j}{\widetilde{N}_j} && 8 \\
&= \sigma^2_{jj}. && 9
\end{aligned}
\tag{50}
$$

where the second equality holds because $\mathcal{L}_j$ is a fixed set which depends only the choice of j, and not on $\boldsymbol{n}$. The the fourth equality holds because

$$\mathcal{E}\left(\widehat{\boldsymbol{\Sigma}}_{k_t}|\boldsymbol{n}\right) = \mathcal{E}\left(\widehat{\boldsymbol{\Sigma}}_{k_t}\right), \tag{51}$$

under the assumption that the data is independent of the missing process. The fifth equality holds because (Muller and Stewart, 2006, p. 316, Equations 16.15 and 16.18) gives

$$\widehat{\boldsymbol{\Sigma}}_{i_f} \equiv \left(N_{i_f} - q_{i_f}\right)^{-1}\widehat{\boldsymbol{S}}_{e,N_{i_f}}. \tag{52}$$

which yields

$$\widehat{\boldsymbol{S}}_{e,N_{i_f}} \sim W_{p_{i_f}}\left(N_{i_f} - q_{i_f}, \boldsymbol{\Sigma}_{i_f}\right) \tag{53}$$

and

$$\mathcal{E}\left(\widehat{\boldsymbol{S}}_{e,N_{i_f}}\right) = \left(N_{i_f} - q_{i_f}\right)\boldsymbol{\Sigma}_{i_f} \tag{54}$$

and thus

$$\mathcal{E}\left[\widehat{\boldsymbol{\Sigma}}_{i_f}\right] = \boldsymbol{\Sigma}_{i_f}. \tag{55}$$

Matrix algebra gives

$$\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j = \sigma_{jj}^2. \tag{56}$$

Show result is positive definite with probability 1.

This does not have to be blue. Note that

$$\begin{aligned}\mathcal{E}_{\boldsymbol{n}}\left[\mathcal{E}\left(\,\widehat{\sigma}_{jj}^2|\boldsymbol{n}\right)\right] &= \mathcal{E}_{\boldsymbol{n}}\left[\sigma_{jj}^2\right] \\ &= \sigma_{jj}^2,\end{aligned}$$

since σ_{jj}^2 is constant.

To give four examples of the result in Equation 56, it may be helpful to consider Example I.

Exhibit I. Example $\mathcal{L}_j$ and $\{\boldsymbol{D}_{k_t}\}$ for $j = 3$.

$$
\begin{aligned}
p &= 3 \\
\mathcal{X} &= \{1,2,3\} \\
d &\in \{1,2,\ldots,8\} \\
R_d &\subseteq \{1,2,3\} \\
N &= 9 \\
\mathcal{L}_3 &= \{4,6,7,8\} \\
\widetilde{N}_3 &= N_4 + N_6 + N_7 + N_8 = 4
\end{aligned}
$$

Deletion pattern	Contain 3	p_d	$\boldsymbol{D}_{k_t}$
$R_1 = \emptyset$	No		
$R_2 = \{1\}$	No		
$R_3 = \{2\}$	No		
$R_4 = \{3\}$	Yes	1	$\boldsymbol{D}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
$R_5 = \{1,2\}$	No		
$R_6 = \{1,3\}$	Yes	2	$\boldsymbol{D}_6 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$
$R_7 = \{2,3\}$	Yes	2	$\boldsymbol{D}_7 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$
$R_8 = \{1,2,3\}$	Yes	3	$\boldsymbol{D}_8 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$$
\begin{aligned}
\boldsymbol{\delta}_3' \boldsymbol{D}_4 \boldsymbol{\Sigma}_4 \boldsymbol{D}_4' \boldsymbol{\delta}_3 &= \boldsymbol{\delta}_3' \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \sigma_{33}^2 [\,0 \quad 0 \quad 1\,] \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} 0 \\ 0 \\ \sigma_{33}^2 \end{bmatrix} [\,0 \quad 0 \quad 1\,] \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= [\,0 \quad 0 \quad 1\,] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \sigma_{33}^2
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\delta}_3' \boldsymbol{D}_6 \boldsymbol{\Sigma}_6 \boldsymbol{D}_6' \boldsymbol{\delta}_3 &= \boldsymbol{\delta}_3' \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 \\ \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} 0 & \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 & 0 \\ 0 & \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 & 0 \\ 0 & \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \sigma_{33}^2
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\delta}_3' \boldsymbol{D}_7 \boldsymbol{\Sigma}_7 \boldsymbol{D}_7' \boldsymbol{\delta}_3 &= \boldsymbol{\delta}_3' \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33} \\ \rho_{32}\sigma_{22}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} 0 & 0 \\ \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33} \\ \rho_{32}\sigma_{22}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_3' \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33} \\ 0 & \rho_{32}\sigma_{22}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33} \\ 0 & \rho_{32}\sigma_{22}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \sigma_{33}^2
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\delta}_3'\boldsymbol{D}_8\boldsymbol{\Sigma}_8\boldsymbol{D}_8'\boldsymbol{\delta}_3 &= \boldsymbol{\delta}_3'\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}\sigma_{11}^2&\rho_{12}\sigma_{11}\sigma_{22}&\rho_{13}\sigma_{11}\sigma_{33}\\\rho_{12}\sigma_{11}\sigma_{22}&\sigma_{22}^2&\rho_{23}\sigma_{22}\sigma_{33}\\\rho_{13}\sigma_{11}\sigma_{33}&\rho_{23}\sigma_{22}\sigma_{33}&\sigma_{33}^2\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\boldsymbol{\delta}_3\\
&= \boldsymbol{\delta}_3'\begin{bmatrix}\sigma_{11}^2&\rho_{12}\sigma_{11}\sigma_{22}&\rho_{13}\sigma_{11}\sigma_{33}\\\rho_{12}\sigma_{11}\sigma_{22}&\sigma_{22}^2&\rho_{23}\sigma_{22}\sigma_{33}\\\rho_{13}\sigma_{11}\sigma_{33}&\rho_{23}\sigma_{22}\sigma_{33}&\sigma_{33}^2\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\boldsymbol{\delta}_3\\
&= \boldsymbol{\delta}_3'\begin{bmatrix}\sigma_{11}^2&\rho_{12}\sigma_{11}\sigma_{22}&\rho_{13}\sigma_{11}\sigma_{33}\\\rho_{12}\sigma_{11}\sigma_{22}&\sigma_{22}^2&\rho_{23}\sigma_{22}\sigma_{33}\\\rho_{13}\sigma_{11}\sigma_{33}&\rho_{23}\sigma_{22}\sigma_{33}&\sigma_{33}^2\end{bmatrix}\boldsymbol{\delta}_3\\
&= \begin{bmatrix}0&0&1\end{bmatrix}\begin{bmatrix}\sigma_{11}^2&\rho_{12}\sigma_{11}\sigma_{22}&\rho_{13}\sigma_{11}\sigma_{33}\\\rho_{12}\sigma_{11}\sigma_{22}&\sigma_{22}^2&\rho_{23}\sigma_{22}\sigma_{33}\\\rho_{13}\sigma_{11}\sigma_{33}&\rho_{23}\sigma_{22}\sigma_{33}&\sigma_{33}^2\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\\
&= \sigma_{33}
\end{aligned}
$$

7.2.1 Distribution of $\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j$ conditional on $\boldsymbol{n}$.

Define the $(1 \times p_{k_t})$ matrix

$$\boldsymbol{A} = \boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}. \tag{57}$$

By Theorem 17.6d, p. 315, Arnold, 1981, we have

$$
\begin{aligned}
(N_{k_t} - q_{k_t})\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j &= (N_{k_t} - q_{k_t})\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{A}'\\
&= \boldsymbol{A}\widehat{\boldsymbol{S}}_{e,N_{k_t}}\boldsymbol{A}'
\end{aligned}
$$

$$
\begin{aligned}
(n_{k_t} - q_{k_t})\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\Big|\boldsymbol{n} &\sim W_1(n_{k_t} - q_{k_t}, \boldsymbol{A}\boldsymbol{\Sigma}_{k_t}\boldsymbol{A}')\\
&\sim W_1\big(n_{k_t} - q_{k_t}, \sigma_{jj}^2\big).
\end{aligned}
$$

This gives

$$\mathcal{E}\left[(n_{k_t} - q_{k_t})\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\Big|\boldsymbol{n}\right] = (n_{k_t} - q_{k_t})\sigma_{jj}^2 \tag{58}$$

and thus

$$
\begin{aligned}
\mathcal{E}\left(\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\Big|\boldsymbol{n}\right) &= \mathcal{E}\big(\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\big)\\
&= \sigma_{jj}^2
\end{aligned} \tag{59}
$$

and also

$$\mathcal{V}\left[(n_{k_t}-q_{k_t})\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\middle|\boldsymbol{n}\right]=2(n_{k_t}-q_{k_t})\sigma_{jj}^2 \tag{60}$$

which gives

$$\begin{aligned}\mathcal{V}\left[\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\middle|\boldsymbol{n}\right]&=\frac{2(n_{k_t}-q_{k_t})\sigma_{jj}^2}{(n_{k_t}-q_{k_t})^2}\\&=\frac{2\sigma_{jj}^2}{(n_{k_t}-q_{k_t})}\end{aligned} \tag{61}$$

Now, finding the variance of $\widehat{\sigma}_{jj}^2$

$$\begin{aligned}\mathcal{V}\left(\,\widehat{\sigma}_{jj}^2|\boldsymbol{n}\right)&=\mathcal{V}\left(\sum_{k_t\in\mathcal{L}_j}\frac{n_{k_t}}{\widetilde{n}_j}\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\middle|\boldsymbol{n}\right) && 10\\&=\sum_{k_t\in\mathcal{L}_j}\mathcal{V}\left(\frac{n_{k_t}}{\widetilde{n}_j}\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\middle|\boldsymbol{n}\right) && 11\\&=\sum_{k_t\in\mathcal{L}_j}\left(\frac{n_{k_t}}{\widetilde{n}_j}\right)^2\mathcal{V}\left(\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_j\middle|\boldsymbol{n}\right) && 12\\&=\sum_{k_t\in\mathcal{L}_j}\left(\frac{n_{k_t}}{\widetilde{n}_j}\right)^2\left[\frac{2\sigma_{jj}^2}{(n_{k_t}-q_{k_t})}\right] && 13\\&=2\sigma_{jj}^2\sum_{k_t\in\mathcal{L}_j}\left[\frac{(n_{k_t})^2}{(n_{k_t}-q_{k_t})(\,\widetilde{n}_j\,)^2}\right] && 14\end{aligned} \tag{62}$$

The first equality follows since the independent sampling units of any two deletion classes are independent, and hence the variance estimates for any two deletion classes are independent.

Asymptotic result: Assume, for $0\leq\pi\leq1$

$$\lim_{n\to\infty}\mathcal{E}\left(\frac{N_{k_t}}{\widetilde{N}_j}\right)=\pi.$$

Then

$$\mathcal{V}\big(\,\widehat{\sigma}^2_{jj}\big) \longrightarrow 0$$

Another pseudo result. $\widehat{\boldsymbol{\Sigma}}_W$ is probably approximately wisharty, with

$$N\heartsuit = \sum_{k_t \in \mathcal{L}_j} \left[\frac{(n_{k_t})^2}{(n_{k_t} - q_{k_t})(\,\widetilde{n}_j\,)^2} \right]$$

and variance = $\boldsymbol{\Sigma}$

We need to examine asymptotic relative efficiency and determine if it achieves the Kramer-Rao lower bound.

Now, finding the expectation of $\widehat{\sigma}^2_{jj'}$

$$\widehat{\sigma}^2_{jj'} = \sum_{k_t \in \mathcal{L}_{jj'}} \frac{N_{k_t}}{\widetilde{N}_{\ jj'}} \boldsymbol{\delta}'_j \boldsymbol{D}_{k_t} \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}'_{k_t} \boldsymbol{\delta}_{j'}. \tag{63}$$

$$\begin{aligned}
\mathcal{E}\big(\widehat{\sigma}^2_{jj'} | \boldsymbol{n}\big) &= \mathcal{E}\left(\sum_{k_t \in \mathcal{L}_{jj'}} \frac{n_{k_t}}{\widetilde{n}_{\ jj'}} \boldsymbol{\delta}'_j \boldsymbol{D}_{k_t} \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}'_{k_t} \boldsymbol{\delta}_{j'} \middle| \boldsymbol{n} \right) && 15 \\
&= \sum_{k_t \in \mathcal{L}_{jj'}} \mathcal{E}\left(\frac{n_{k_t}}{\widetilde{n}_{\ jj'}} \boldsymbol{\delta}'_j \boldsymbol{D}_{k_t} \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}'_{k_t} \boldsymbol{\delta}_{j'} \middle| \boldsymbol{n} \right) && 16 \\
&= \sum_{k_t \in \mathcal{L}_{jj'}} \frac{n_{k_t}}{\widetilde{n}_{\ jj'}} \sigma^2_{jj'} && 17 \\
&= \frac{\sigma^2_{jj'}}{\widetilde{n}_{\ jj'}} \sum_{k_t \in \mathcal{L}_{jj'}} n_{k_t} && 18 \\
&= \frac{\sigma^2_{jj'}}{\widetilde{n}_{\ jj'}} \widetilde{n}_{\ jj'} && 19 \\
&= \sigma^2_{jj'}, && 20
\end{aligned} \tag{64}$$

where the second equality holds because $\mathcal{L}_{jj'}$ is a fixed set which depends only the choice of j and j', and not on $\boldsymbol{n}$, the the fourth equality holds because

$$\mathcal{E}\big(\widehat{\boldsymbol{\Sigma}}_{k_t}|\boldsymbol{n}\big) = \mathcal{E}\big(\widehat{\boldsymbol{\Sigma}}_{k_t}\big), \tag{65}$$

under the assumption that the data is independent of the missing process, and the fifth equality holds because

$$\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\boldsymbol{\Sigma}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_{j'} = \sigma_{jj'}^2. \tag{66}$$

To give two examples of the result in Equation 33, it may be helpful to consider Example J.

Exhibit J. Example $\mathcal{L}_{jj'}$ and $\widetilde{N}_{jj'}$ for $j = 1$ and $j' = 3$.

$$p = 3$$
$$\mathcal{X} = \{1, 2, 3\}$$
$$d \in \{1, 2, \ldots, 8\}$$
$$R_d \subseteq \{1, 2, 3\}$$
$$N = 9$$
$$\mathcal{L}_{jj'} = \{6, 8\}$$
$$\widetilde{N}_{jj'} = N_6 + N_8 = 4$$

Deletion pattern	Contain 1 and 3	Frequency of deletion class
$R_1 = \emptyset$	No	$N_1 = 0$
$R_2 = \{1\}$	No	$N_2 = 0$
$R_3 = \{2\}$	No	$N_3 = 0$
$R_4 = \{3\}$	No	$N_4 = 0$
$R_5 = \{1, 2\}$	No	$N_5 = 5$
$R_6 = \{1, 3\}$	Yes	$N_6 = 0$
$R_7 = \{2, 3\}$	No	$N_7 = 0$
$R_8 = \{1, 2, 3\}$	Yes	$N_8 = 4$
		$N = 9$

$$\begin{aligned}
\boldsymbol{\delta}_1' \boldsymbol{D}_6 \boldsymbol{\Sigma}_6 \boldsymbol{D}_6' \boldsymbol{\delta}_3 &= \boldsymbol{\delta}_1' \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_1' \begin{bmatrix} \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 \\ \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \boldsymbol{\delta}_1' \begin{bmatrix} 0 & \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 & 0 \\ 0 & \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \boldsymbol{\delta}_3 \\
&= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \sigma_{11}^2 & \rho_{13}\sigma_{11}\sigma_{33} \\ 0 & 0 & 0 \\ 0 & \rho_{13}\sigma_{11}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \rho_{13}\sigma_{11}\sigma_{33}
\end{aligned}$$

$$\begin{aligned}\boldsymbol{\delta}_1'\boldsymbol{D}_8\boldsymbol{\Sigma}_8\boldsymbol{D}_8'\boldsymbol{\delta}_3 &= \boldsymbol{\delta}_1'\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}\sigma_{11}^2 & \rho_{12}\sigma_{11}\sigma_{22} & \rho_{13}\sigma_{11}\sigma_{33}\\ \rho_{12}\sigma_{11}\sigma_{22} & \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33}\\ \rho_{13}\sigma_{11}\sigma_{33} & \rho_{23}\sigma_{22}\sigma_{33} & \sigma_{33}^2\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\boldsymbol{\delta}_3\\ &= \boldsymbol{\delta}_1'\begin{bmatrix}\sigma_{11}^2 & \rho_{12}\sigma_{11}\sigma_{22} & \rho_{13}\sigma_{11}\sigma_{33}\\ \rho_{12}\sigma_{11}\sigma_{22} & \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33}\\ \rho_{13}\sigma_{11}\sigma_{33} & \rho_{23}\sigma_{22}\sigma_{33} & \sigma_{33}^2\end{bmatrix}\boldsymbol{\delta}_3\\ &= \rho_{13}\sigma_{11}\sigma_{33}.\end{aligned}$$

Now, finding the variance of $\widehat{\sigma}_{jj'}^2$

$$\widehat{\sigma}_{jj'}^2 = \sum_{k_t\in\mathcal{L}_{jj'}} \frac{N_{k_t}}{\widetilde{N}_{jj'}}\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_{j'}. \tag{67}$$

$$\begin{aligned}\mathcal{V}(\widehat{\sigma}_{jj'}|\boldsymbol{n}) &= \mathcal{V}\left(\sum_{k_t\in\mathcal{L}_{jj'}} \frac{n_{k_t}}{\widetilde{n}_{jj'}}\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_{j'}\middle|\boldsymbol{n}\right) && 1\\ &= \sum_{k_t\in\mathcal{L}_{jj'}}\mathcal{V}\left(\frac{n_{k_t}}{\widetilde{n}_{jj'}}\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_{j'}\middle|\boldsymbol{n}\right) && 2\\ &= \sum_{k_t\in\mathcal{L}_{jj'}}\left(\frac{n_{k_t}}{\widetilde{n}_{jj'}}\right)^2\mathcal{V}\left(\boldsymbol{\delta}_j'\boldsymbol{D}_{k_t}\widehat{\boldsymbol{\Sigma}}_{k_t}\boldsymbol{D}_{k_t}'\boldsymbol{\delta}_{j'}\middle|\boldsymbol{n}\right) && 3\\ &= \sum_{k_t\in\mathcal{L}_{jj'}}\left(\frac{n_{k_t}}{\widetilde{n}_{jj'}}\right)^2\frac{\left(\rho_{jj'}^2\sigma_{jj}^2\sigma_{j'j'}^2+\sigma_{jj}^2\sigma_{j'j'}^2\right)}{(n_{k_t}-q_{k_t})} && 4\end{aligned} \tag{68}$$

where the second equality holds because $\mathcal{L}_{jj'}$ is a fixed set which depends only the choice of j and j', and not on $\boldsymbol{n}$, and Equality 4 holds from Equation 5, p.3, Nydick, 2012.

Find approximate distribution of the weighted thing.

$\widehat{\boldsymbol{\Sigma}}_W \approx \sim$ Wishart(interesting degrees of freedom, funky thing*Σ)

$\widehat{\boldsymbol{\Sigma}}_W$ is pretty close to unbiased

$\widehat{\boldsymbol{\Sigma}}_W$ is almost certainly consistent

7.3 Independence of $\widehat{\boldsymbol{B}}_W$ and $\widehat{\boldsymbol{\Sigma}}_W$

Show independence

7.4 Wald statistic approximator

Put these things here:

1. General linear hypothesis here

2. Define Wald statistic

3. Define Wald statistic approximator

7.5 Approximate distribution of the null case Wald statistic approximator

7.6 Approximate distribution of the alternative case Wald statistic approximator

8. Deviations from normality

Show that the convergence targets for global $\widehat{\boldsymbol{B}}$, and $\widehat{\boldsymbol{\Sigma}}$, are the same as the convergence targets for the Zeger and Liang GEE estimates (http://links.jstor.org/sici?sici=0006-341X%28198812%2944%3A4%3C1049%3AMFLDAG%3E2.0.CO%3B2-R), and calculate asymptotic relative efficiency of the Kreidler and Z+L estimates. If available, do the same thing for the ML estimators for the mixed model (see Harville), for which you need the following ingredients for the proof.

a. Explicit ML estimator for the mixed model (beta is unbiased, but sigma is asymptotically biased).

b. Explicit ML convergence target for the mixed model.

c. OR- real analysis theorem that allows me to demonstrate asymptotic convergence without 1 or 2.

9. Special cases

9.1 Complete cases (no missing data)

Show that for the special multivariate case, the result reduces to the standard estimator of $\widehat{\boldsymbol{\Sigma}}_W$

Under the assumptions of the multivariate model, no data is missing, and, as long as $n > p - 1$,

$$\widetilde{N}_{jj'} = n, \tag{69}$$

$$N_{k_t} = n, \tag{70}$$

$$\widetilde{N}_j = n, \tag{71}$$

$$\boldsymbol{D}_{k_t} = \boldsymbol{I}_p, \tag{72}$$

$$\boldsymbol{\delta}_{j'} = \boldsymbol{\delta}_{j'}, \tag{73}$$

$$k_t = 2^p, \tag{74}$$

$$\mathcal{L}_j = \{2^p\}, \tag{75}$$

and

$$\boldsymbol{\delta}_j = \boldsymbol{\delta}_j. \tag{76}$$

Then

$$\begin{aligned} \mathcal{E}(\widehat{\sigma}_{jj}) &= \sum_{k_t \in \mathcal{L}_j} \frac{N_{k_t}}{\widetilde{N}_j} \boldsymbol{\delta}_j' \boldsymbol{D}_{k_t}' \boldsymbol{\Sigma}_{k_t} \boldsymbol{D}_{k_t} \boldsymbol{\delta}_j \\ &= \boldsymbol{\delta}_j' \boldsymbol{\Sigma}_{k_t} \boldsymbol{\delta}_j \\ &= \sigma_{jj}. \end{aligned} \tag{77}$$

It follows that

$$\mathcal{E}\left(\widehat{\boldsymbol{\Sigma}}_W\right) = \boldsymbol{\Sigma} \tag{78}$$

Exhibit K. Example outcome matrix sorted by deletion classes,
under no missing data assumption.

$$\begin{aligned} p &= p \\ \mathcal{X} &= \{1, 2, \ldots, p\} \\ d &\in \{1, 2, \ldots, 2^p\} \\ R_d &\subseteq \{1, 2, \ldots, p\} \\ N &= N \\ i_F &= i_1 = 2^p; F = 1 \\ N_1 &= N \end{aligned}$$

$$\widehat{\sigma}_{jj'} = \sum_{k_t \in \mathcal{L}_{jj'}} \frac{N_{k_t}}{\widetilde{N}_{jj'}} \boldsymbol{\delta}_j' \boldsymbol{D}_{k_t}' \widehat{\boldsymbol{\Sigma}}_{k_t} \boldsymbol{D}_{k_t} \boldsymbol{\delta}_{j'}. \tag{79}$$

$$\widehat{\boldsymbol{\Sigma}}_W = \{\widehat{\sigma}_{jj'}\}. \tag{80}$$

9.2 Cluster (which may have missing data, and hence differently sized clusters)

A thing Keith said to consider here

1. The Sarah paper fails for cluster. But this give a new estimate for cluster (surprise, it provised the analytic rationale for why Catellier works so well).

10. Numerical evaluations

Do simulations for accuracy and maybe other stuff.

10.1 Methods

10.2 Results

Exhibit 12- convergence graph (in folder)

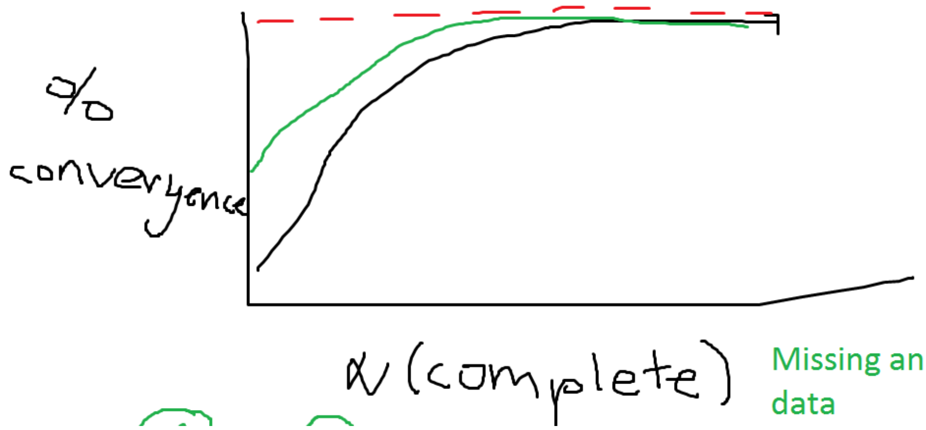


Exhibit 13- Bias of standard and novel mixed model estimators for $\boldsymbol{B}$
as a function of sample size, percent missing, missing process, sharks, balloons

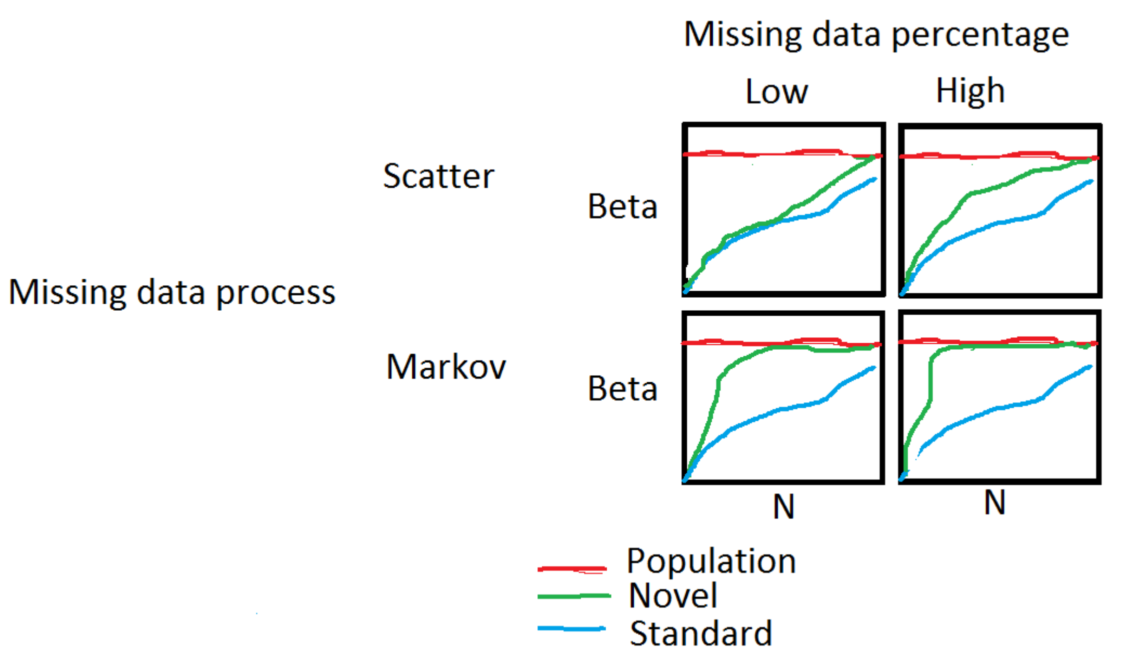


Exhibit 14- Consistency of standard and novel mixed model estimators for Σ as a function of sample size, percent missing, missing process, sharks, balloons

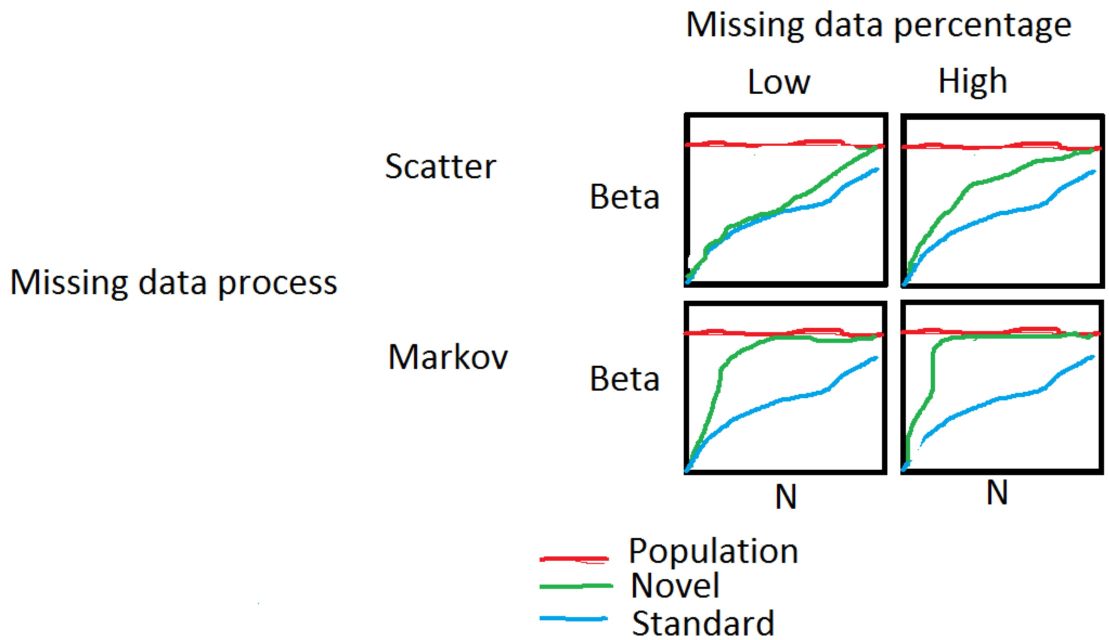
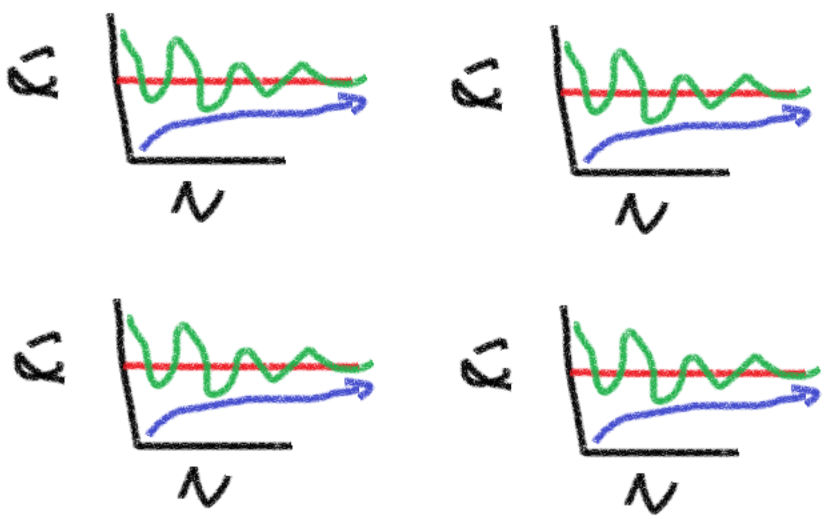


Exhibit 15- Type I error of standard and novel mixed model hypothesis tests as a function of sample size, percent missing, missing process, sharks, balloons



11. Example power analysis

Problem:

1. Two measurements of maternal nutrient intake, one measurement of infant fat mass or maternal status (gestational diabetes, or not, for example).

2. Hypothesis: no association of infant fat mass and two measurements, or no association between maternal gestational diabetes and two nutrient measurements.

3. Missing data pattern: 30% missing measurement 1, 30% missing measurement 2, 60% have all the data.

Assuming that all people have two measurements, and those measurements are equally spaced, we can write the model and hypotheses as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{e} \quad (1)$$

$$\boldsymbol{X} = \boldsymbol{I}_2 \otimes \mathbf{1}_{N/2} \quad (2)$$

$$H_0 : \boldsymbol{C}\boldsymbol{\beta}\boldsymbol{U} = \mathbf{0} \quad (3)$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad (4)$$

$$\boldsymbol{U} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (5)$$

Now assuming that all people have six repeated measurements, and those measurements are equally spaced, we can write the model and hypotheses as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{e} \quad (6)$$

$$\boldsymbol{X} = \boldsymbol{I}_2 \otimes \mathbf{1}_{N/2} \quad (7)$$

$$H_0 : \boldsymbol{C}\boldsymbol{\beta}\boldsymbol{U} = \mathbf{0} \quad (8)$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad (9)$$

$$\boldsymbol{U} = \begin{bmatrix} \text{linear} & \text{quad} & \text{cubic} & \text{quartic} & \text{quintic} \end{bmatrix} \quad (10)$$

Two approaches

1. Fit Kreidler multivariate method to estimate 15 of 21 parameters.

2. Fit lear to estimate remaining 6 parameters, which are functions of the observed data and have distributions.

3. Cobble together inference.

4. Keith thinks we can fit the fifth order polynomial and inherently non-linearizable models.....

12. Discussion

Limitations

1. The observations for each independent sampling unit occur at p planned times, or p planned locations

2. $p \ll N$.

3. No repeated covariates

4. No randomness in X. (not sure that's true).

5. $G > 0$

6. $\left(N_{i_f} - 1\right) > p_{i_f}$ (this should be the general constraint for a multivariate model, which Keith knows, and may not be this exactly).

7. Future research: Also can get KR analogue, and Chi-Muller missing data analogue and Catellier and Muller data analogue.

8. Suggests big data sampling algorithms, so you sample from each deletion class with defined error, and then use epsilon delta to bound the overall error (using the consistency argument).

9. Future Research 1. Think about whether if the B_{I_f} differed by deletion class whether we would have a diagnostic test for MNAR.

Acknowledgments

The research presented in this paper was supported by grant number NIGMS 9R01GM121081 awarded to the Colorado School of Public Health, D. H. Glueck, D. Dabelea and K. E. Muller, Principal Investigators. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Science nor the National Institutes of Health.

References

Appendix