A Full Information Maximum Likelihood (FIML) Approach to Compensating for Missing Data in Matrix Sampling

Paul Biemer

RTI International and

University of North Carolina

Content of this talk

- Simple matrix sampling for two questionnaires
- Presents basic idea of FIML for matrix sampling
- Some results based upon simulation
- Implications for future work

Advantages and Disadvantages of FIML

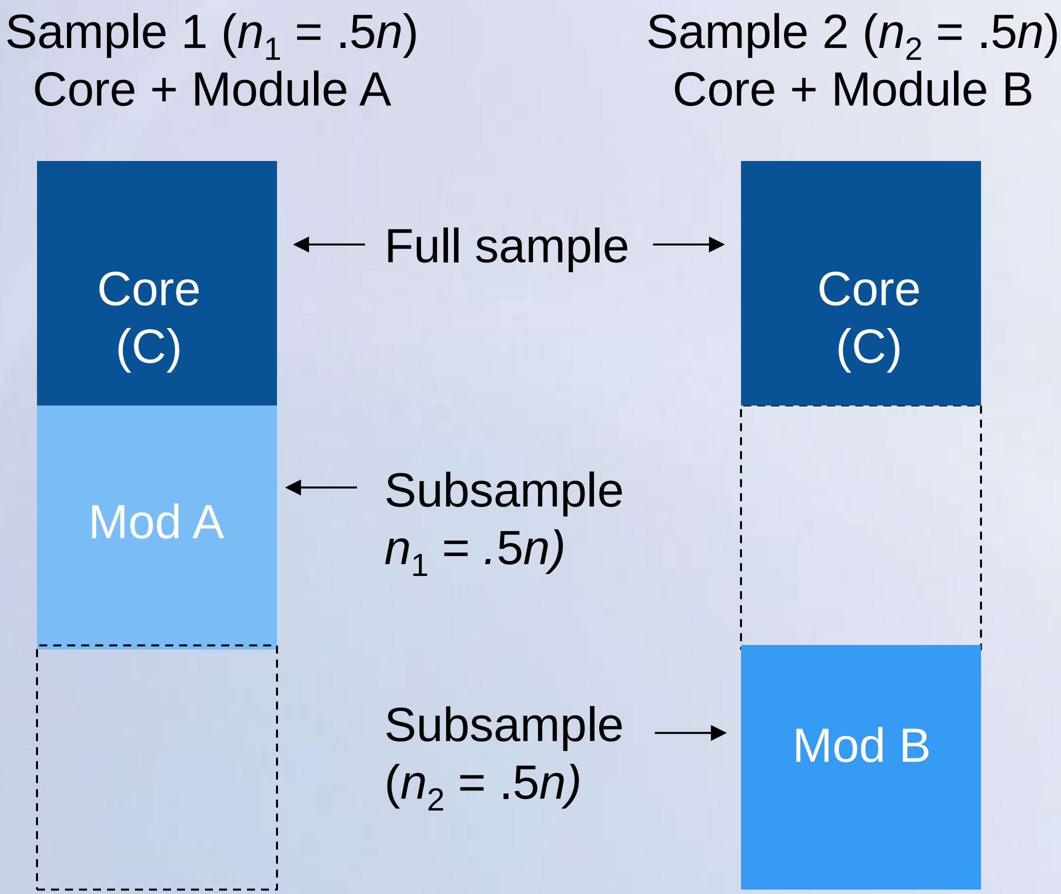
Advantages

- More efficient that MI
- Easier to use that MI
- Uses full information
  - Unlike case-wise deletion, for example
- Useful for simulating various matrix sampling scenarios

Disadvantage

- Requires special software such as Mplus or Latent Gold
- Modeling complexity

Simple Matrix Sampling Design



Notation for Analyzing C, A and B

Subtable for C × A

	C=1		C=2
A=1	$n_{a=1,c=1}$		$n_{a=1,c=2}$
A=2	$n_{a=2,c=1}$		$n_{a=2,c=2}$

Subtable for C × B

	C=1		C=2
B=1	$n_{b=1,c=1}$		$n_{b=1,c=2}$
B=2	$n_{b=2,c=1}$		$n_{b=2,c=2}$

Define Response Indicators

Notation:

- *R*, *S* are response indicators for *A*, *B*, respectively
  - E.g., *RS* = 12 denotes table CA, *RS* = 21 denotes CB
  - Note: *RS* = 11 (CAB) and *RS* = 22 (C) are not observed
- Use log-linear path models specify relationships among *C*,*A*,*B*,*R*, *S*
- Both ignorable and nonignorable response mechanisms can be estimated
- Matrix sampling is primarily concerned with ignorable (MAR) response mechanisms

Likelihood Assuming Multinomial Sampling

- Incomplete data likelihood

$$\log \mathcal{L}_{(\pi)} = \sum_{cab} n_{cab} \log \pi_{cab}\pi_{11|cab} + \sum_{ca} n_{ca} \log \sum_{b} \pi_{cab}\pi_{12|cab}$$
$$+ \sum_{cb} n_{abd} \log \sum_{a} \pi_{cab}\pi_{21|cab} + \sum_{c} n_{c} \log \sum_{ab} \pi_{cab}\pi_{22|cab}$$

where

$$\pi_{cab} = \Pr(C = c, A = a, B = a)$$
$$\pi_{rs|cab} = \Pr(R = r, S = s \mid C = c, A = a, B = a)$$

Possible Logit Models for *R* and *S*

MCAR:

$$\pi_{rs|cab} = \pi_{rs} = \frac{\exp(u_r^R + u_s^S + u_{rs}^{RS})}{\sum_{rs} \exp(u_r^R + u_s^S + u_{rs}^{RS})}$$

MAR:

$$\pi_{rs|cab} = \frac{\exp(u_r^R + u_s^S + u_{rs}^{RS} + u_{rc}^{RC} + u_{rb}^{RB} + u_{sc}^{SC} + u_{sa}^{SA})}{\sum_{rs} \exp(u_r^R + u_s^S + u_{rs}^{RS} + u_{rc}^{RC} + u_{rb}^{RB} + u_{sc}^{SC} + u_{sa}^{SA})}$$

$$\pi_{rs|cab} = \frac{\exp(u_r^R + u_s^S + u_{rs}^{RS} + u_{rc}^{RC} + u_{sc}^{SC})}{\sum_{rs} \exp(u_r^R + u_s^S + u_{rs}^{RS} + u_{rc}^{RC} + u_{sc}^{SC})}$$

How is the precision of estimates affected by matrix sampling?

- When C, A and B are uncorrelated?
- When C and A or C and B are correlated?
- When C and A or C and B or A and B are correlated?

Illustration of a Simulation to Investigate the Effect of Correlation on Precision

- FIML employed to estimate the proportion positives for A (or B); i.e. $\pi_{a=1}$ or $\pi_{b=1}$
- Simulation setup

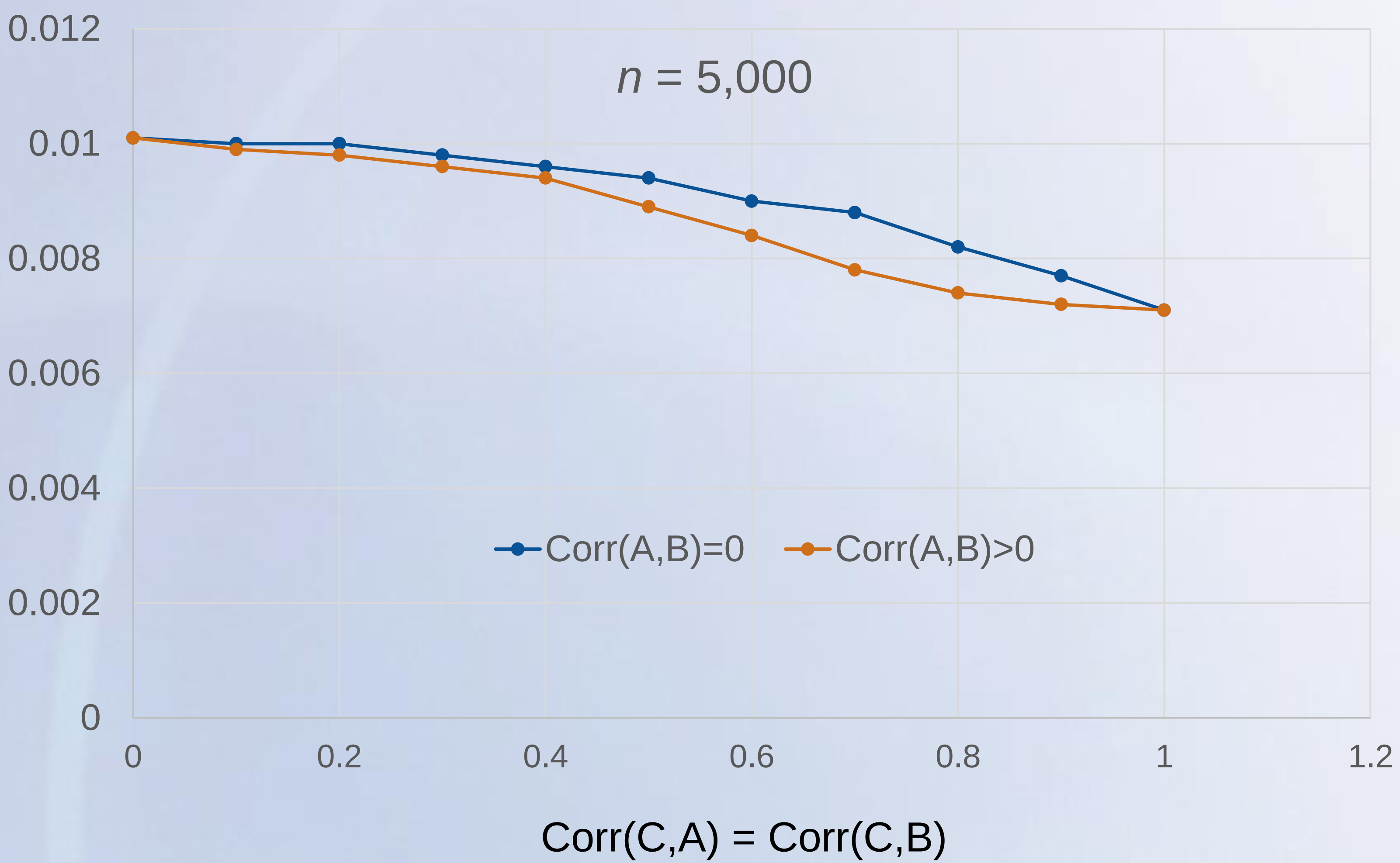
$$\pi_{c=1} = \pi_{a=1} = \pi_{b=1} = 0.5$$

- Simulation 1:

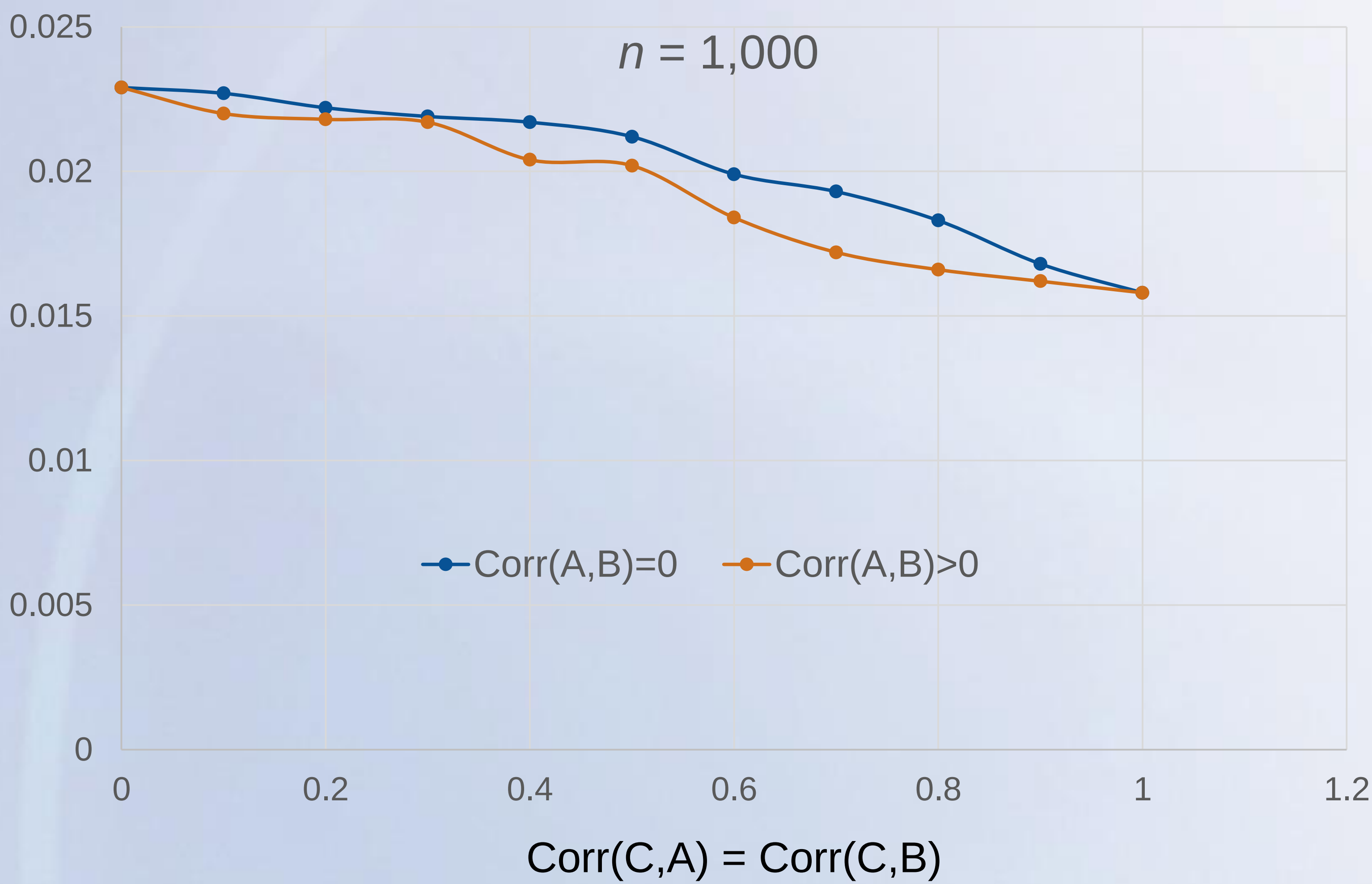
  Corr(C,A) = Corr(C,B) varied between 0.0 and 1.0
- Simulation 2:

  Corr(C,A) = Corr(C,B) = Corr(A,B) varied between 0.0 and 1.0
- n = 5000 and n = 1000

Standard Error of Proportion A (or B) as a Function of Corr(C,A) = Corr(C,B)



Standard Error of Proportion A (or B) as a Function of Corr(C,A) = Corr(C,B)



Remarks

- FIML is a viable approach for point, interval and model estimation in matrix sampling
- FIML standard errors equivalent to MI standard errors with $m = \infty$
- S.E's can be improved by incorporating correlates within and across disjoint subsamples
- FIML with response indicators makes this quite straight-forward