Lecture 17

Multiple Imputation

How many times to impute?

Relative efficiency of $m$ multiple imputations versus $m = \infty$

$$= \left(1 + \frac{1}{m}\right)^{-1} \text{ where } \lambda = \text{fraction missing}$$

$$= \left(1 + \left(\frac{5}{5}\right)\right)^{-1} \lambda = 0.5, \ m = 5$$

$$= 0.909$$

in terms of variance of $\bar{\theta}$

or 0.95 in terms of standard deviation

$m = 5$ often okay, $m = 20$ almost always enough.
Data augmentation

Recall we wanted $p(\theta | y)$

We suppose $y$ functionally related to latent variables $y^*$

Then we use the augmented posterior $p(\theta, y^* | y)$

Use Gibbs samplers with alternating draws

$y^*(s)$ from $p(y^* | \theta^{(s-1)}, y)$

$\theta^{(s)}$ from $p(\theta | y^*(s), y)$

We get draws from $p(\theta, y^* | y)$

We just keep the draws of $\theta$. 
- Apply data augmentation to imputation
- Change of notation

Fully observed data are $Y_{obs}$ (this was $y$)
Latent data are $Y_{m3}$ if missing (this was $y^*$)
Parameter $\bar{y}$ are for the assumed dag for $Y_{m3}$ (this was $\Theta$)

- Gibbs sampler alternates draws

$Y_{m3}^{(s)}$ from $p(Y_{m3}^{(s)} | \bar{y}^{(s-1)}, Y_{obs})$ imputation step

$\bar{y}^{(s)}$ from $p(\bar{y}^{(s)} | Y_{m3}^{(s)}, Y_{obs})$ prediction step

Converges to draws from $p(Y_{m3}, \bar{y} | Y_{obs})$
and use the $Y_{m3}^{(s)}$ draws, e.g. every one hundredth
Imputation using multivariate normal model

Let \( \mathbf{z}_i \) be \( p \times 1 \) vector of \( p \) variables with incomplete data

\( \sim \)

\( \mathbf{t}_i \) be \( q \times 1 \) " - \( q \) variables with complete data

Model

\[
\mathbf{z}_i \sim \mathcal{N}(\Theta \mathbf{t}_i + \mathbf{\xi}_i, \Sigma)
\]

\( \mathbf{\xi}_i \sim \mathcal{N}(0, \bar{\Sigma}) \)

\( (Y_{\text{mic}, X}, Y_{\text{obs}, \mathbf{t}}, \pi) \Rightarrow (\Theta, \Sigma) \).

Prior

uniform for \( \Theta \)

inverse Wishart for \( \Sigma \)
- Inverse-Wishart \( p(\Sigma) \propto |\Sigma|^{\frac{n+\nu+1}{2}} \exp\left(-\frac{1}{2} \Sigma^\top S \Sigma\right) \)

State default is a uniform which sets \( \nu = -(p+1) \)

\[ \Sigma^{-1} \sim \mathcal{O}_{p \times p} \]

- In imputation step draw \( x_i \sim \mathcal{N}(\cdot, \cdot) \)
- In prediction step draw \( \Sigma \) from an inverse-Wishart
- Draw \( \text{vec}(\Theta) \sim \mathcal{N}(\cdot, \cdot) \)

- May transform eg. logs before MVA impute

Special methods for (0,1) data eg. MI impute logit.
MISSINGNESS TERMINOLOGY

1. MCAR Missing completely at random

\[ \Pr(x \text{ missing} \mid \text{observed, unobserved}) = \Pr(x \text{ missing}) \]

Then no problems as observed data is a random sample of the unobserved data.
Then can use case deletion.
Just an efficiency loss.
(2) MAR (missing at random)

\[ \Pr(x \text{ missing} | \text{ observed, unobserved}) = \Pr(x \text{ missing} | \text{ observed}) \]

Ignorable missing

Selection on observables only

Problems: Observed data is not a random sample
Listwise deletion is no good.
However, can do imputation.
Also inverse-probability weighting.
(3) MNAR Missing not at random
Non-ignorable missing
Missingness depends on unobservables
y_i. Tobit model
\[ y_i = y_i^* \text{ if } y_i^* = x_i' \beta + \varepsilon_i > 0 \]
\[ = 0 \quad \text{otherwise} \]
Need to make strong assumptions to proceed.
Eg. ML
Or could impute.
INFERENCES WITH CLUSTERED ERRORS

- Focus on OLS
  Much generalizes to nonlinear models except in nonlinear models clustered errors may lead to inconsistency.
- \( \hat{\beta} \) remains consistent
  but need different \( \hat{\text{V}}(\hat{\beta}) \)
  and could be more efficient feasible GLS
- Clustered errors are correlated within clusters (or groups)
  uncorrelated across clusters.
\[ y_i = x_i \beta + z_i \]
\[ y = X \beta + z \]

Order date by cluster.

So all of cluster 1, all of cluster 2...

\[ \text{Var}(\epsilon) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \]

\[ n \times n \]

\[ \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix} \]
Two leading examples

1. Individuals in Clusters

How do job injury rates effect wages?

Problem: job injury rate is unknown at the individual level, it is known at a more aggregated level, e.g., occupation.

\[ y_i^g = \alpha + \delta_i g + \theta_i z + u_i^g \]

- \( y_i^g \) is occupation
- \( z \) is job injury rate in occupation \( g \)
Here $t_j$ is perfectly correlated within cluster and $u_{ij}$ is (mildly) correlated within cluster as the model overpredicts for one person in occupation $g$ it is likely to overpredict for others in occupation $g$. 