Missing Data in Panels and Dropout

Jamie Monogan

Washington University in St. Louis

March 18, 2010
Objectives

By the end of this meeting, participants should be able to:

- Describe how missing data mechanisms are defined for panel data.
- Define dropout and the various ways it may occur.
- Explain various ways to handle missing data in panel studies.
- Analyze panel data using multiple random imputation.
Missing Data Mechanisms

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCAR</td>
<td>( P(R_i \mid Y_i^O, Y_i^M, X_i) = P(R_i) )</td>
</tr>
<tr>
<td>Covariate-Dependent</td>
<td>( P(R_i \mid Y_i^O, Y_i^M, X_i) = P(R_i \mid X_i) )</td>
</tr>
<tr>
<td>MAR</td>
<td>( P(R_i \mid Y_i^O, Y_i^M, X_i) = P(R_i \mid Y_i^O, X_i) )</td>
</tr>
<tr>
<td>NI</td>
<td>( P(R_i \mid Y_i^O, Y_i^M, X_i) = P(R_i \mid Y_i^O, Y_i^M, X_i) )</td>
</tr>
</tbody>
</table>

Censoring

Different numbers of waves or times of measurement may arise from the structure of the study. MCAR/Covariate-Dependent is a pretty safe assumption.
Dropout

- Also called “attrition” or “panel mortality.”
- Refers to individuals’ leaving the study permanently.
- MCAR: \( \theta_2 = \theta_3 = 0 \)
- MAR: \( \theta_3 = 0 \)
- NI: \( \theta_3 \neq 0 \)

Model of Dropout Odds

\[
\log \left\{ \frac{P(D_i = k | D_i \geq k, Y_{i1}, \ldots, Y_{ik})}{P(D_i > k | D_i \geq k, Y_{i1}, \ldots, Y_{ik})} \right\} = \theta_1 + \theta_2 Y_{ik-1} + \theta_3 Y_{ik}
\]
Handling Missing Panel Data
With Special Attention to Dropout

- Complete-Case Analysis: Analagous to Case-Wise Deletion (MCAR)
- Available-Data Analysis (MCAR)
- Last Value Carried Forward (Dubious Assumption)
- Multiple Random Imputation (MAR)
  - Regression-based technique.
  - Match the missing observations with observed observations with a similar probability of missingness.
  - Returning individuals: Use later (and earlier) observations to impute middle values.
- Weighting Methods: Give more weight to observations with a high dropout probability. (MAR)
- Refreshment Samples (Application: Bartels 1999)
Key Equations for Multiple Random Imputation

**Standard Error of Averaged Parameter**

\[
\sqrt{\frac{1}{M} \sum_{i=1}^{M} \hat{s}^2_{\theta i} + \frac{M+1}{M} \frac{1}{M-1} \sum_{i=1}^{M} (\hat{\theta}_i - \bar{\theta})^2}
\]

**Degrees of Freedom of Averaged Parameter**

\[
(M - 1) \left(1 + \frac{1}{M + 1} \frac{s^2_{\hat{\theta}W}}{s^2_{\hat{\theta}B}} \right)^2
\]
For Next Time

- Read FLW chapter 15
- Suppose you have the following results on the parameter $\theta$ from five imputed datasets:
  - $\hat{\theta} = \{3, 1, 4, 1, 5\}$
  - $s_{\hat{\theta}} = \{0.9, 0.4, 1.2, 1.0, 0.6\}$
- Tell me:
  - Your averaged estimate of $\hat{\theta}$
  - Your estimate of the standard error of the averaged estimate.
  - The degrees of freedom for testing a hypothesis about $\theta$.
  - The $p-$value when you test:
    
    $H_0: \theta = 0$
    
    $H_A: \theta \neq 0$