Unit 4
II. Panel Attrition

PS2701-2019

Longitudinal Analysis

Week 13

Professor Steven Finkel

Panel Attrition

- Can use the causal inference and multilevel framework from past weeks to think about and propose corrections for the problem of **panel attrition**, i.e., units dropping out of the data set over time
- Could occur for reasons related to the individual unit, to his/her motivation to continue in the study, to substantive reasons related to important variables, to more mundane issues related to the data collection process, to individuals moving or being unable to be located, or other reasons
- If attrition not scattered randomly throughout the data set, need to consider corrections
- Big problem: If attrition is related to treatment group status *and* to potential outcomes, huge problem (e.g. if aggressive control individuals are more likely to drop out than aggressive treatment individuals, you may overestimate the effect of an anti-aggression treatment)

Possible Corrections

- Listwise deletion of cases (i.e., do nothing)
  - Similar to missing data in other kinds of analyses – cases are dropped and only "complete cases" are included in model
  - With wide data, drop-outs will be lost to the analysis
  - With long data, drop-outs will contribute only to those waves where data are present on the variables of interest
  - OK so long as drop-outs are Missing Completely at Random (MCAR, i.e., unrelated to any other variable), but inefficient and possibly worse if drop-outs are Missing at Random (MAR, i.e., related to observed independent variables) or Missing Not at Random (MNAR, or related to "unobservables" that may also influence the outcome)
  - Lose lots of cases even if MCAR and affects standard errors

- Maximum Likelihood Estimation with "Ignorable Likelihood Function" (IML)
  - Used in STATA SEM under "method (mlmv)"
  - Each case contributes to the likelihood function only for the waves that he/she participates and for the variables with valid data
  - If MCAR or MAR, this uses available data more efficiently
  - Doesn't correct for MNAR problem
- Multiple Imputation (MI)
  - "fills in" missing data with plausible estimates based on matching of missing and non-missing cases on covariates provided by researcher
  - Multiple data sets imputed that incorporate uncertainty surrounding point estimates for missing cases
  - Analyses conducted on each of data sets and results aggregated to provide best estimate
  - Implemented in STATA with "MI" routine
  - Doesn't correct for MNAR problem

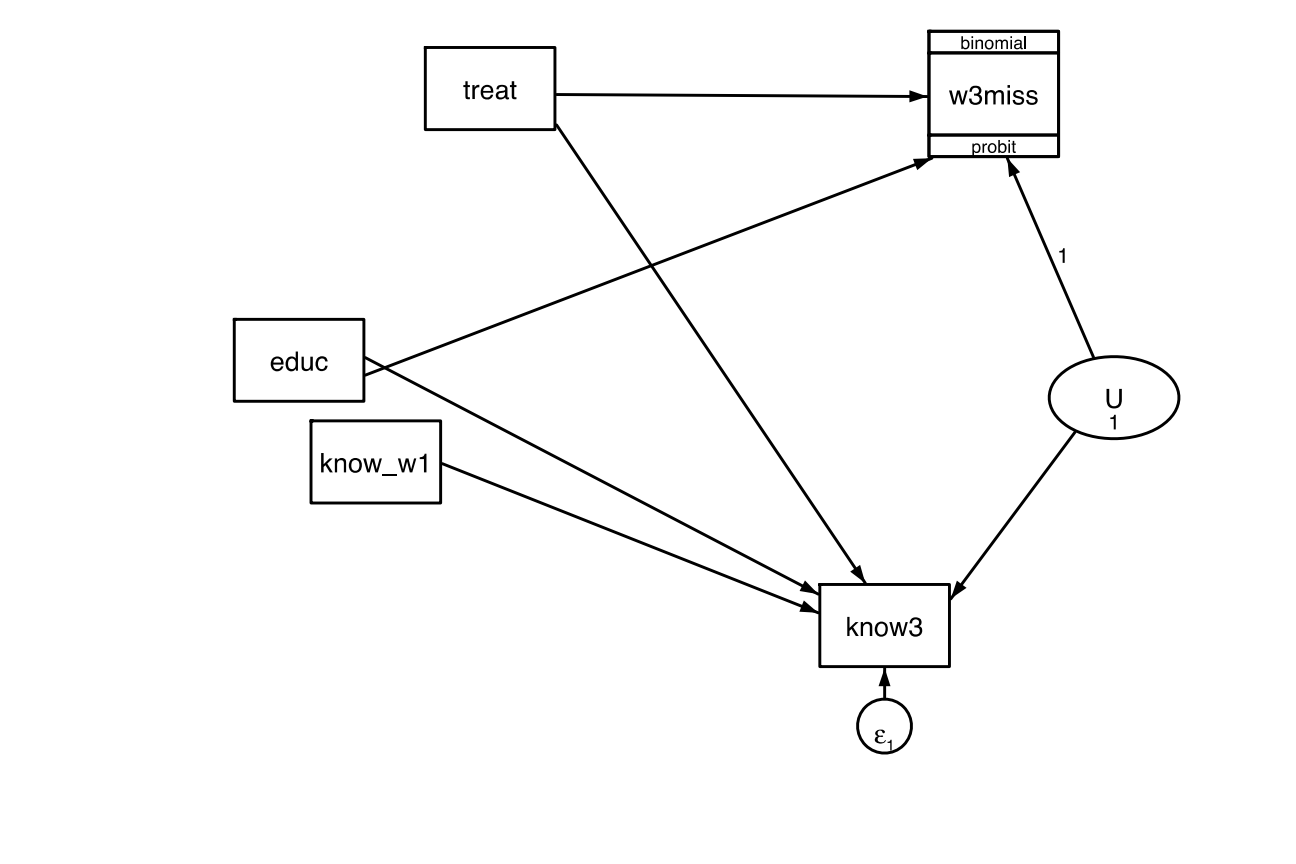
- Inverse Probability Weighting (IPW)
  - Similar logic to IPW in propensity score matching models
  - Builds model of the probability of non-attrition based on observed covariates, then weight the data by the inverse of this probability, so that completed cases with similar characteristics to cases that dropped-out are given greater weight in the analysis. Competed cases with similar characteristics to the kinds of cases that did not drop out are given less weight in the analysis.
  - Can be combined with MI methods or IML methods
  - Some recommend using IPW to handle total panel attrition and MI/IML to handle missing data on covariates
  - Very sensitive to misspecification of the selection equation (as in propensity score matching models)
  - Does not correct for MNAR

- MNAR Panel Selection Models
  - Primary method in Political Science to handle MNAR
  - Treat outcome and selection (attrition) as two processes, with either prior DVs or unobservable unit-level characteristics potentially affecting both outcome and probability of attrition.
  - Estimates outcomes, controlling for the possibility that unobservable characteristics that lead to drop-out may also affect level of the dependent variable
  - Sometimes difficult to estimate in practice
    - Ideally need auxiliary variables that predict drop-out but do not directly affect outcomes (i.e. instruments).
    - Sensitive to normal distribution and functional form assumptions of model

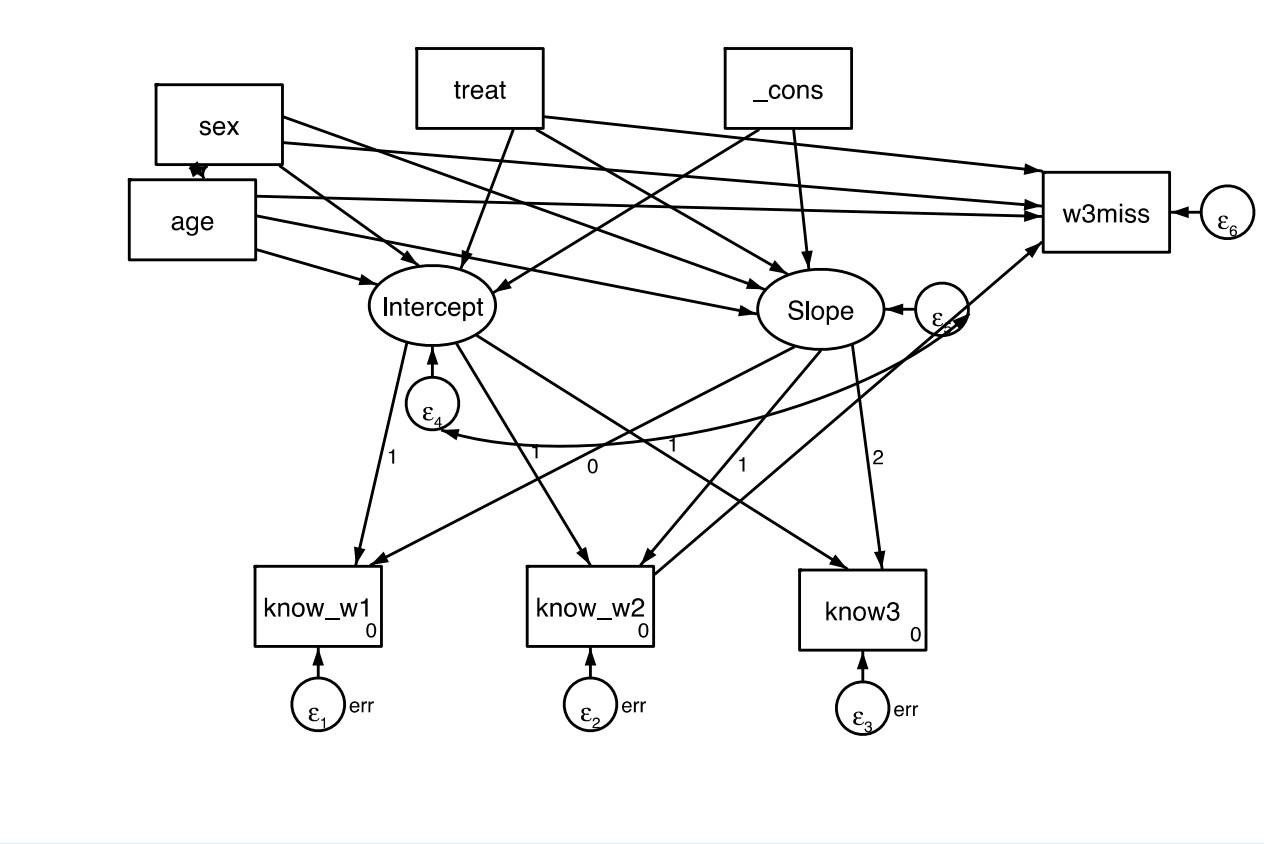
Example of MNAR Models

- Heckman Selection Model: correlated error terms of outcome and selection (drop-out) equations
- Diggle-Kenward model: outcomes at one point in time predict dropout at another point in time
- Shared Parameter Model: intercept/slope of latent growth model predict drop-out

Two Wave Heckman Selection Model for MNAR Data

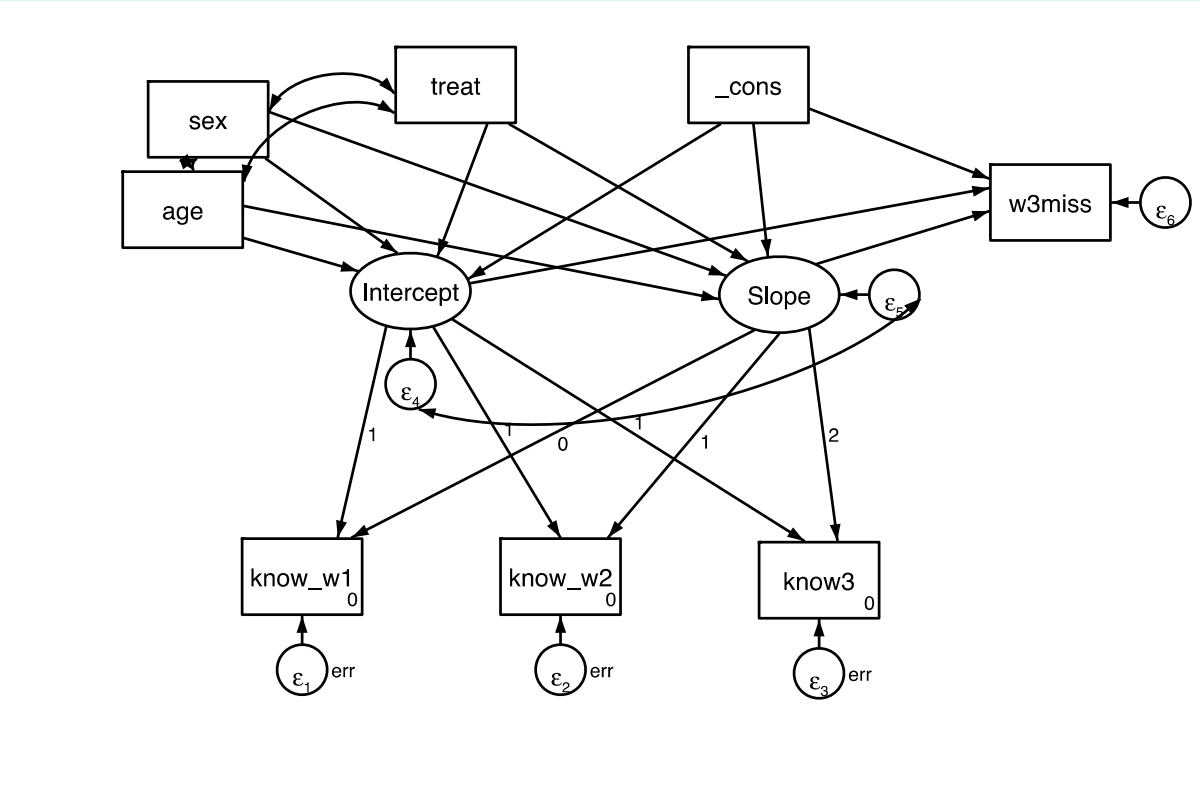


Diggle-Kenward Selection Model for MNAR Data



Exogenous Variables and the Outcome at time t-1 predicts Subsequent Drop-out

"Shared Parameter" Selection Model for MNAR Data



The Parameters of the Growth Model predict Drop-out