Using Auxiliary Variables in a Maximum Likelihood Analysis

Inclusive Analysis Strategy and Auxiliary Variables

The literature recommends an inclusive strategy that incorporates auxiliary variables into missing data handling

An auxiliary variable is not of substantive interest but is used to improve power or reduce nonresponse bias

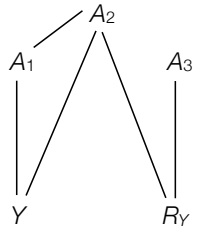
The benefit of an auxiliary variable depends on the pattern and magnitude of its correlations with the analysis variables and missing data indicators

Diagram of Auxiliary Variable Correlations

Conditioning on A_1 improves power but ignoring this variable does not introduce bias

Ignoring A_2 induces an NMAR mechanism and nonresponse bias

A_3 cannot introduce bias nor can it increase power



Motivating Example

Data from a sample of 250 chronic pain patients

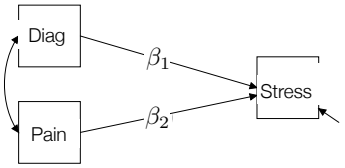
Variables include gender, the number of diagnosed physical ailments, sleep quality, pain ratings, positive and negative affect, and stress

Gender and number of diagnoses are complete, the remaining variables have up to 14% missing data

Analysis Model

The analysis is a multiple regression model that examines the influence of pain on stress, controlling for the number of diagnosed ailments

$$Stress = \beta_0 + \beta_1(Diagnose) + \beta_2(Pain) + e$$



Auxiliary Variables as Covariates

Auxiliary variables can be added as covariates

This approach alters the interpretation of model parameters (e.g., regression slopes partial out auxiliary variables), and is cumbersome in complex models

A better strategy introduces auxiliary variables in a way that maintains the same interpretations as a complete-data analysis

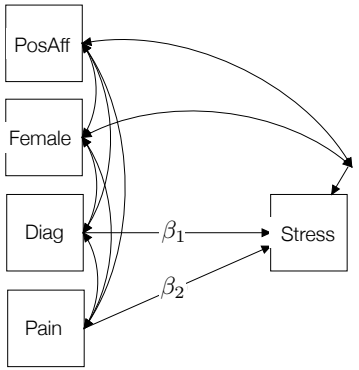
Auxiliary Variable Models

Graham (2003) outlined the saturated correlates (spider) and extra dependent variable models for auxiliary variables

The models transmit information from the auxiliary variables to all analysis variables via correlations or regressions

The models not alter the substantive interpretation of the parameter estimates, nor does it affect model fit

Saturated Correlates Model Diagram



Model Specification Rules

Correlate each auxiliary variable with …

Manifest predictor variables

Other auxiliary variables

The residual terms of all dependent variables

Auxiliary variables never correlate with latent variables

Ex11.1.inp
Mplus Analysis Script

```
DATA:
file = pain.csv;
VARIABLE:
names = id female diagnose sleep pain posaff
  negaff stress;
usevariables = diagnose pain stress;
auxiliary = (m) female posaff;
missing = all(-99);
MODEL:
diagnose pain;
stress on diagnose pain;
OUTPUT:
standardized(stdyx);
```

Mplus Analysis Output

```
SUMMARY OF ANALYSIS

Number of groups                                 1
Number of observations                         250

Number of dependent variables                    1
Number of independent variables                  2
Number of continuous latent variables            0

Observed dependent variables

  Continuous
   STRESS

Observed independent variables
   DIAGNOSE    PAIN

Observed auxiliary variables
   FEMALE      POSAFF
```

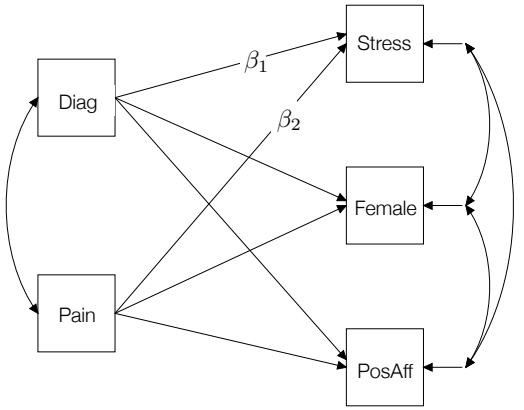
Mplus Analysis Output

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

STRESS   ON
   DIAGNOSE         -0.019      0.059     -0.327      0.744
   PAIN              0.261      0.068      3.828      0.000

PAIN     WITH
   DIAGNOSE          0.550      0.090      6.115      0.000

Means
   DIAGNOSE          3.800      0.074     51.673      0.000
   PAIN              4.558      0.072     63.568      0.000
```

Mplus Analysis Output

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

...

 Intercepts
    STRESS           2.679      0.294      9.121      0.000

 Variances
    DIAGNOSE         1.352      0.121     11.180      0.000
    PAIN             1.161      0.111     10.454      0.000

 Residual Variances
    STRESS           0.765      0.075     10.217      0.000
```

Extra Dependent Variable Model Diagram



Model Specification Rules

Regress each auxiliary variable on manifest predictors

Specify correlated residuals among all auxiliary variables and the auxiliary variables and the outcome variable

Latent variables never predict auxiliary variables

Ex11.2.inp
Mplus Analysis Script

```
DATA:
file = pain.csv;
VARIABLE:
names = id female diagnose sleep pain posaff
  negaff stress;
usevariables = diagnose pain stress female posaff;
missing = all(-99);
MODEL:
diagnose pain;
stress on diagnose pain;
female posaff on diagnose pain;
female with posaff;
OUTPUT:
standardized(stdyx);
```

Mplus Analysis Output

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STRESS ON				
DIAGNOSE	-0.019	0.059	-0.327	0.744
PAIN	0.261	0.068	3.828	0.000
FEMALE ON				
DIAGNOSE	0.122	0.026	4.650	0.000
PAIN	0.150	0.030	5.059	0.000
POSAFF ON				
DIAGNOSE	-0.123	0.059	-2.080	0.038
PAIN	-0.118	0.066	-1.771	0.077

Mplus Analysis Output

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
...				
FEMALE WITH				
POSAFF	0.108	0.027	3.996	0.000
STRESS	-0.019	0.026	-0.718	0.472
POSAFF WITH				
STRESS	-0.210	0.058	-3.615	0.000
PAIN WITH				
DIAGNOSE	0.550	0.090	6.115	0.000

Mplus Analysis Output

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
...				
Means				
DIAGNOSE	3.800	0.074	51.673	0.000
PAIN	4.558	0.072	63.568	0.000
Intercepts				
STRESS	2.679	0.294	9.121	0.000
FEMALE	-0.648	0.129	-5.023	0.000
POSAFF	5.493	0.289	19.014	0.000
Variances				
DIAGNOSE	1.352	0.121	11.180	0.000
PAIN	1.161	0.111	10.454	0.000
Residual Variances				
STRESS	0.765	0.075	10.217	0.000
FEMALE	0.183	0.017	11.017	0.000
POSAFF	0.840	0.079	10.650	0.000