Including auxiliary variables in models with missing data using full-information maximum likelihood estimation

Rose Anne Medeiros

Department of Sociology
Rice University

2013 Stata Conference

Outline

Outline

Outline

Outline

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on x is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of x (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of x conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on x is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of x (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of x conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on x is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of x (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of x conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on *x* is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of *x* (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of *x* conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on x is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of x (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of x conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on *x* is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of *x* (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of *x* conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Full-information maximum likelihood (FIML) estimation

- The likelihood function is adjusted so that incomplete observations are used in estimation.
- Implemented in Stata's `sem` command with the `method(mlmv)` option.
- Assumes that missingness on x is either:
  - Unrelated to other observed variables and unrelated to the unobserved values of x (missing completely at random–MCAR)
  - Related to other observed variables, but unrelated to the unobserved value of x conditional on observed variables (missing at random–MAR)
- For the MAR assumption to hold, the predictors of missingness must be modeled

Auxiliary variables

- Variables that are:
  - Correlated with missingness on x, and/or
  - Correlated with the observed values of x
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for x if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on *x*, and/or
  - Correlated with the observed values of *x*
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for *x* if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Auxiliary variables

- Variables that are:
  - Correlated with missingness on x, and/or
  - Correlated with the observed values of x
- While not part of the substantive model they can improve the performance of FIML by:
  - Making the MAR assumption more reasonable
  - Acting as proxies for x if MAR is violated
  - Increase efficiency by reducing uncertainty due to missingness
  - See Collins, Schafer, and Kam (2001).

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
  - Underlying constructs that are approximated by observed variables
  - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
  - Underlying constructs that are approximated by observed variables
  - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
  - Underlying constructs that are approximated by observed variables
  - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
    - Underlying constructs that are approximated by observed variables
    - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
  - Underlying constructs that are approximated by observed variables
  - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
    - Underlying constructs that are approximated by observed variables
    - Error terms

Speaking SEM: Types of variables

Endogenous vs. exogenous

- Exogenous variables are variables not predicted by any other variables in the model (a.k.a. predictor variables)
- Endogenous variables are those that are predicted by other variables in the model (a.k.a. outcome variables)

Observed vs. latent

- Observed variables are variables that have been measured, e.g. *age*, *sex*
- Latent variables are variables that are not observed
  - Underlying constructs that are approximated by observed variables
  - Error terms

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
  - Preferred for models with observed variables
- Saturated correlates approach (SCA)
  - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

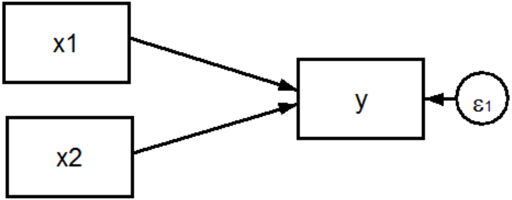
Approaches to including auxiliary variables with FIML

- Use auxiliary variables as extra predictors in the model
- Include auxiliary variables as extra dependent variables (DVs)
    - Preferred for models with observed variables
- Saturated correlates approach (SCA)
    - Preferred for models with latent variables

Note: Both the saturated correlates approach and extra DV models can be applied to models with all observed or latent variables.
See Graham and Coffman (2012) and Graham (2003).

Outline

A simple model



Using complete case analysis:

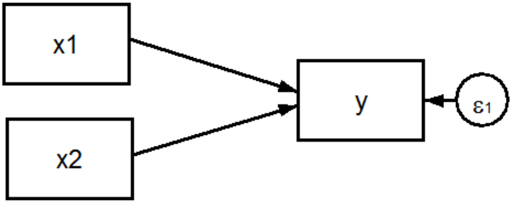
```
regress y x1 x2
```

```
sem (y <- x1 x2)
```

Using FIML (without auxilary variables):

```
sem (y <- x1 x2), method(mlmv)
```

A simple model



Using complete case analysis:

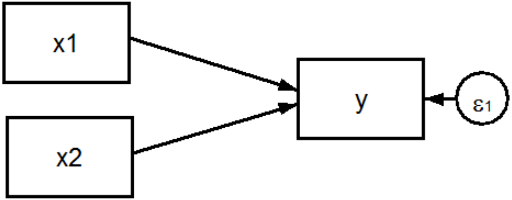
```
regress y x1 x2
```

```
sem (y <- x1 x2)
```

Using FIML (without auxilary variables):

```
sem (y <- x1 x2), method(mlmv)
```

A simple model



Using complete case analysis:

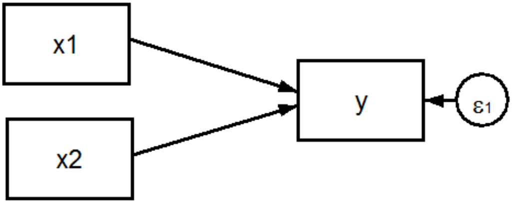
```
regress y x1 x2
```

```
sem (y <- x1 x2)
```

Using FIML (without auxilary variables):

```
sem (y <- x1 x2), method(mlmv)
```

A simple model



Using complete case analysis:

```
regress y x1 x2
```

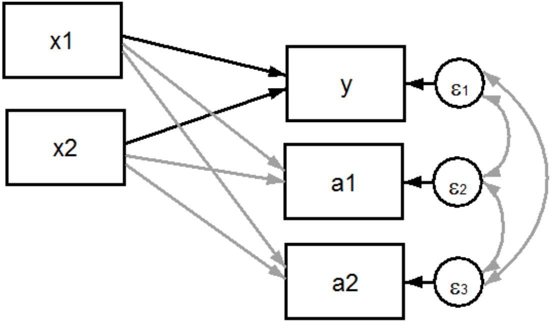
```
sem (y <- x1 x2)
```

Using FIML (without auxilary variables):

```
sem (y <- x1 x2), method(mlmv)
```

The extra DV model with observed variables

- Auxiliary variables are predicted by all predictor variables.
- Residual terms for model dependent variables and the auxiliary variables are correlated.



Syntax for the extra DV model

```
sem (y a1 a2 <- x1 x2), ///
    cov(e.y*e.a1 e.y*e.a2) /// auxiliary with DV
    cov(e.a1*e.a2) /// auxiliary with auxiliary
    method(mlmv)
```

The same model in a more compact form:

```
sem (y a1 a2 <- x1 x2), cov(e.y*e.a1-a2 e.a1*e.a2) meth(mlmv)
```

Syntax for the extra DV model

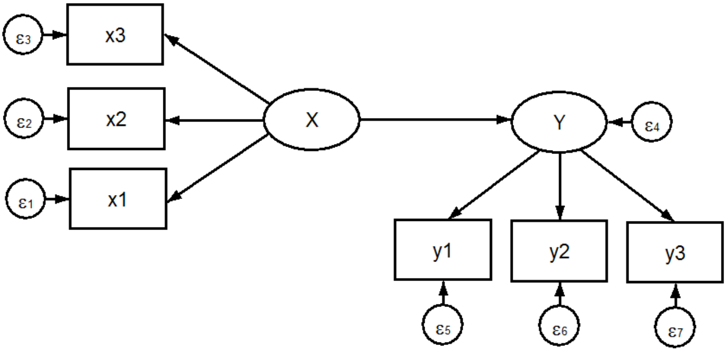
```
sem (y a1 a2 <- x1 x2), ///
    cov(e.y*e.a1 e.y*e.a2) /// auxiliary with DV
    cov(e.a1*e.a2) /// auxiliary with auxiliary
    method(mlmv)
```

The same model in a more compact form:

```
sem (y a1 a2 <- x1 x2), cov(e.y*e.a1-a2 e.a1*e.a2) meth(mlmv)
```

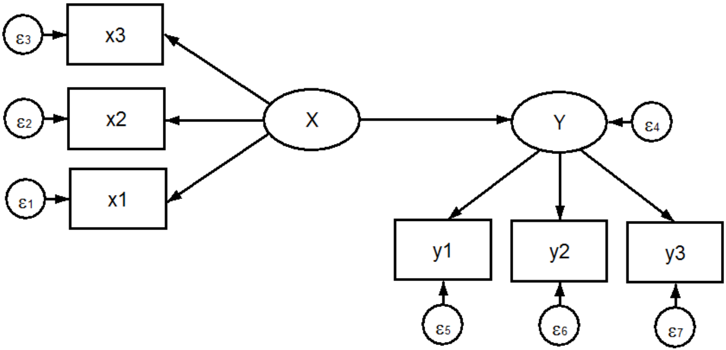
Outline

A simple SEM model




```
sem (x1 x2 x3 <- X) ///
(y1 y2 y3 <- Y) ///
(Y <-X)
```

A simple SEM model




```
sem (x1 x2 x3 <- X) ///
(y1 y2 y3 <- Y) ///
(Y <-X)
```

Including auxiliary variables using the SCA

Auxiliary variables are correlated with:

- All other auxiliary variables
- Any completely exogenous observed variables
- Residual for all endogenous observed variables

Including auxiliary variables using the SCA

Auxiliary variables are correlated with:

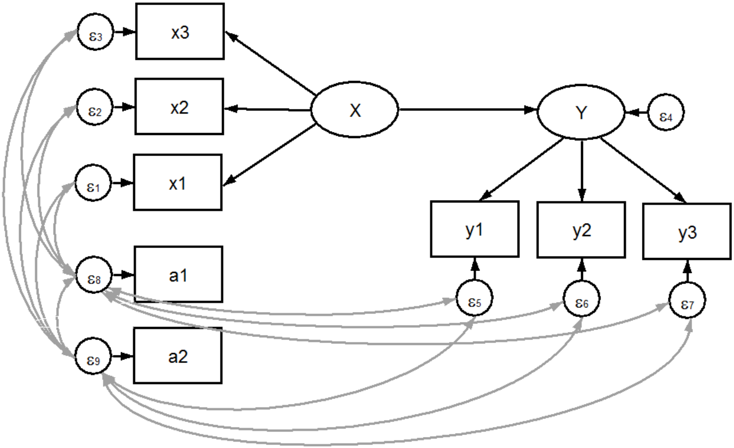
- All other auxiliary variables
- Any completely exogenous observed variables
- Residual for all endogenous observed variables

Including auxiliary variables using the SCA

Auxiliary variables are correlated with:

- All other auxiliary variables
- Any completely exogenous observed variables
- Residual for all endogenous observed variables

Including auxiliary variables using the SCA



Specifying a model using the SCA

```
sem (x1 x2 x3 <- X) ///
(y1 y2 y3 <- Y) ///
(Y <-X) ///
(a1 a2 <- _cons), ///
cov(e.x1-x3*e.a1-a2) /// X observed with auxilary
cov(e.y1-y3*e.a1-a2) /// Y observed with auxilary
cov(e.a1*e.a2) /// auxilary with auxilary
method(mlmv)
```

Measures of model fit with the SCA

- Without auxiliary variables, we can use `estat gof, stats(indices)` to obtain the Comparative fit index (CFI) and Tucker-Lewis index (TLI) after running an over-identified `sem` model
  - These and other incremental fit indices compare the fitted model to a baseline (or null) model
  - With auxiliary variables, the default baseline model does not produce the desired comparison
  - We can specify the desired baseline model and calculate these fit indices by hand
- Graham and Coffman (2012) point out issues in the calculation of RMSEA in models with auxiliary variables and suggest using a utility rho.exe which can be requested from Graham.

Measures of model fit with the SCA

- Without auxiliary variables, we can use `estat gof, stats(indices)` to obtain the Comparative fit index (CFI) and Tucker-Lewis index (TLI) after running an over-identified `sem` model
  - These and other incremental fit indices compare the fitted model to a baseline (or null) model
  - With auxiliary variables, the default baseline model does not produce the desired comparison
  - We can specify the desired baseline model and calculate these fit indices by hand
- Graham and Coffman (2012) point out issues in the calculation of RMSEA in models with auxiliary variables and suggest using a utility rho.exe which can be requested from Graham.

Measures of model fit with the SCA

- Without auxiliary variables, we can use `estat gof, stats(indices)` to obtain the Comparative fit index (CFI) and Tucker-Lewis index (TLI) after running an over-identified `sem` model
  - These and other incremental fit indices compare the fitted model to a baseline (or null) model
  - With auxiliary variables, the default baseline model does not produce the desired comparison
  - We can specify the desired baseline model and calculate these fit indices by hand
- Graham and Coffman (2012) point out issues in the calculation of RMSEA in models with auxiliary variables and suggest using a utility rho.exe which can be requested from Graham.

Measures of model fit with the SCA

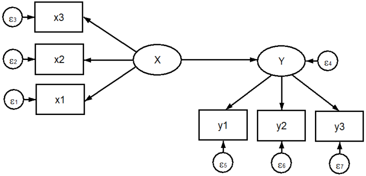
- Without auxiliary variables, we can use `estat gof, stats(indices)` to obtain the Comparative fit index (CFI) and Tucker-Lewis index (TLI) after running an over-identified `sem` model
  - These and other incremental fit indices compare the fitted model to a baseline (or null) model
  - With auxiliary variables, the default baseline model does not produce the desired comparison
  - We can specify the desired baseline model and calculate these fit indices by hand
- Graham and Coffman (2012) point out issues in the calculation of RMSEA in models with auxiliary variables and suggest using a utility rho.exe which can be requested from Graham.

Measures of model fit with the SCA

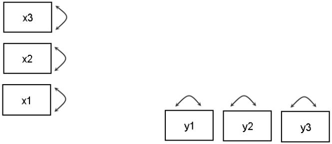
- Without auxiliary variables, we can use `estat gof, stats(indices)` to obtain the Comparative fit index (CFI) and Tucker-Lewis index (TLI) after running an over-identified `sem` model
  - These and other incremental fit indices compare the fitted model to a baseline (or null) model
  - With auxiliary variables, the default baseline model does not produce the desired comparison
  - We can specify the desired baseline model and calculate these fit indices by hand
- Graham and Coffman (2012) point out issues in the calculation of RMSEA in models with auxiliary variables and suggest using a utility rho.exe which can be requested from Graham.

Without auxiliary variables

Fitted model



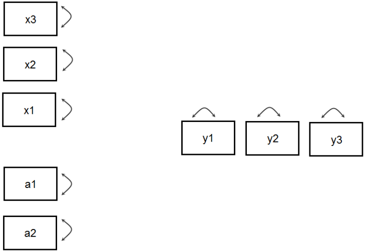
Baseline model



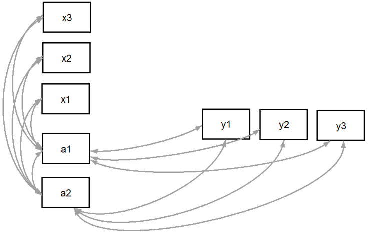
We want to make this same comparison when estimating models using the SCA.

Baseline model with auxiliary variables

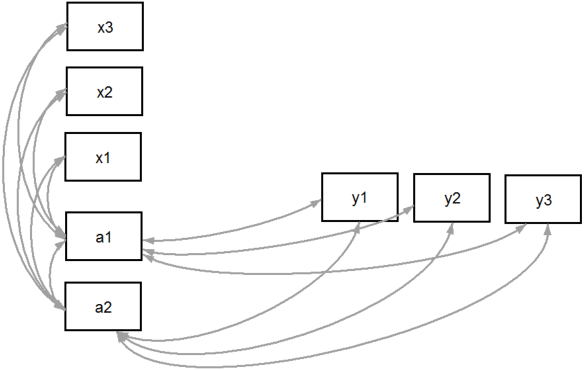
The default (incorrect) baseline model



The correct baseline model with SCA

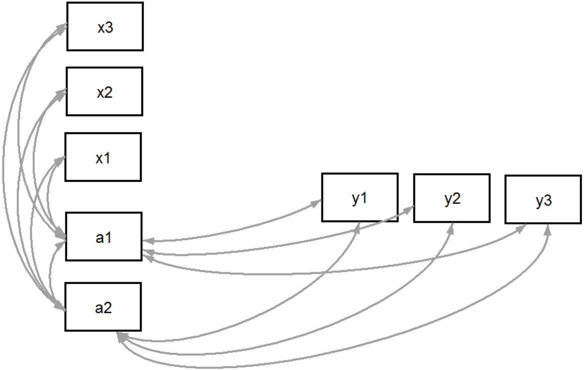


Correct baseline model with SCA




```
sem (x1 x2 x3 y1 y2 y3 a1 a2 <- _cons), ///
  cov(e.x1-x3*e.a1-a2) /// X observed with auxiliary
  cov(e.y1-y3*e.a1-a2) /// Y observed with auxiliary
  cov(e.a1*e.a2) /// auxiliary with auxiliary
  method(mlmv)
```

Correct baseline model with SCA




```
sem (x1 x2 x3 y1 y2 y3 a1 a2 <- _cons), ///
  cov(e.x1-x3*e.a1-a2) /// X observed with auxiliary
  cov(e.y1-y3*e.a1-a2) /// Y observed with auxiliary
  cov(e.a1*e.a2) /// auxiliary with auxiliary
  method(mlmv)
```

χ^2 and df for the baseline and fitted models

Output from the correct baseline model with SCA:

```
LR test of model vs. saturated: chi2(15)  =   1210.92, Prob > chi2 = 0.0000
```

$\chi^2_b = 1210.92$ and $\text{df}_b = 15$

Output from the full (fitted) model including auxiliary variables using SCA:

```
LR test of model vs. saturated: chi2(8)   =      4.19, Prob > chi2 = 0.8397
```

$\chi^2_m = 4.19$ and $\text{df}_b = 8$

χ^2 and df for the baseline and fitted models

Output from the correct baseline model with SCA:

```
LR test of model vs. saturated: chi2(15)  =   1210.92, Prob > chi2 = 0.0000
```

$\chi^2_b = 1210.92$ and $\text{df}_b = 15$

Output from the full (fitted) model including auxiliary variables using SCA:

```
LR test of model vs. saturated: chi2(8)   =      4.19, Prob > chi2 = 0.8397
```

$\chi^2_m = 4.19$ and $\text{df}_b = 8$

Formulae for fit indices

$$\text{CFI} = 1 - \frac{\chi^2_m - \text{df}_m}{\max\{(\chi^2_b - \text{df}_b), (\chi^2_m - \text{df}_m)\}}$$

$$\text{TLI} = \frac{(\chi^2_b/\text{df}_b) - (\chi^2_m/\text{df}_m)}{\chi^2_b/\text{df}_b - 1}$$

Source: [SEM] Methods and formulas for sem

Identifying auxiliary variables

- Theory
- *t*-tests, correlations, etc.
- Planning
- Auxiliary variables may have missing values

Identifying auxiliary variables

- Theory
- *t*-tests, correlations, etc.
- Planning
- Auxiliary variables may have missing values

Identifying auxiliary variables

- Theory
- *t*-tests, correlations, etc.
- Planning
- Auxiliary variables may have missing values

Identifying auxiliary variables

- Theory
- *t*-tests, correlations, etc.
- Planning
- Auxiliary variables may have missing values

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Final thoughts

- Using auxiliary variables has a larger impact when:
  - Auxiliary variables are highly correlated with both missingness and variables of interest
  - High proportion of missing values
- Limitations of auxiliary variables generally
  - Auxiliary variables with high levels of missingness are less helpful
  - Violation of normality assumption is more problematic with missing data
- Limitations of the saturated correlates model
  - Use of saturated correlates model may increase convergence problems
  - May result in non-positive definite variance-covariance matrices

Works cited

Collins, Linda M, Joseph L Schafer, and Chi-Ming Kam. 2001. "A comparison of inclusive and restrictive strategies in modern missing data procedures." Psychological Methods 6(4):330.

Enders, Craig K. 2001. "The impact of non-normality on full-information maximum-likelihood estimation for structural equation models with missing data" *Psychological Methods* 6(4):352-70.

Graham, John W., and Donna L. Coffman. 2012. "Structural Equation Modeling with Missing Data." Pp. 277-94 in *Handbook of Structural Equation Modeling*, edited by Rick H. Hoyle. New York, NY: The Guilford Press.

Little, R. J. A., & D. B. Rubin. 2002. Statistical analysis with missing data. Hoboken, N.J: Wiley.

Rubin, Donald B. 1976. "Inference and missing data." *Biometrika* 63(3):581-92.