

**Application of Plausible Values of Latent Variables
to Analyzing BSI-18 Factors**

Jichuan Wang, Ph.D

Children's National Health System
The George Washington University School of Medicine
Washington, DC

An often encountered challenge in application of SEM: too many variables are involved in a model.

- SEM consists of two components: measurement model and structural model.
- Many items for each factor in the measurement model (e.g., CFA).
- A set of measures for a finite number of latent classes/profiles in the measurement model (e.g., LCA/LPA).

Make model parsimonious:

- Parsimonious theoretical model.
- Replacing the measurement model with estimated values of latent variables → replacing SEM with path analysis:
 - scale total scores.
 - factor scores estimated from CFA.
 - latent class/profile membership estimated from LCA/LPA.
 - plausible values of latent variables (continuous or categorical) .

What are plausible values of latent variable?

A set of generated values of factor scores using multiple imputations (Mislevy, 1991 ; Asparouhov & Muthén, 2010).

Advantages of using plausible values of latent variables:

- The imputed plausible values of a latent variable can be used as observed variable for further analysis.
- In Mplus, plausible values are estimated based on the MCMC Bayesian approach.
- The plausible values estimates and standard errors are more reliable than those obtained by the ML estimator.
- Using factor scores as dependent variables in secondary analysis gives biased slopes. Using plausible values can alleviate the biases.

Two types of applications of plausible values:

- PV-R
 - The Imputation mean of statistic M is:

$$\bar{M}_{IMP} = \frac{\sum_i^K M_i}{K}$$

- Variance of statistic M is expressed as:

$$\hat{V}_{IMP} = \frac{1}{K} \sum_i^K \hat{V}(M_{PVi}) + \left(1 + \frac{1}{K}\right) \left[\frac{1}{K-1} \left(\sum_i^K M_{PVi} - \bar{M}_{PV} \right)^2 \right]$$

- Plausible value data sets are analyzed just like multiple imputation data sets Rubin's method (1987).
 - In Mplus, using the TYPE=IMPUTATION statement in DATA command.
 - A small number (e.g., 5) of imputed data sets is needed.
- PV-W
 - Averaging the plausible values for each individual and then calculating the statistics of interest by using these averages of the plausible values.
 - A large set (e.g., 100-500) of estimates of plausible values is needed.
 - In Mplus, using SAVDATA command.

Example

Data:

Sample: Drug users (N=303) in Changsha, China recruited using RDS, 2012-2013

Outcome measures: The BSI-18

- Somatization: 6 items
- Depression: 6 items
- Anxiety: 6 items

Predictors in MIMIC model:

- Age
- Education
- Marital status
- Employment status
- Meth use in the past 30 days

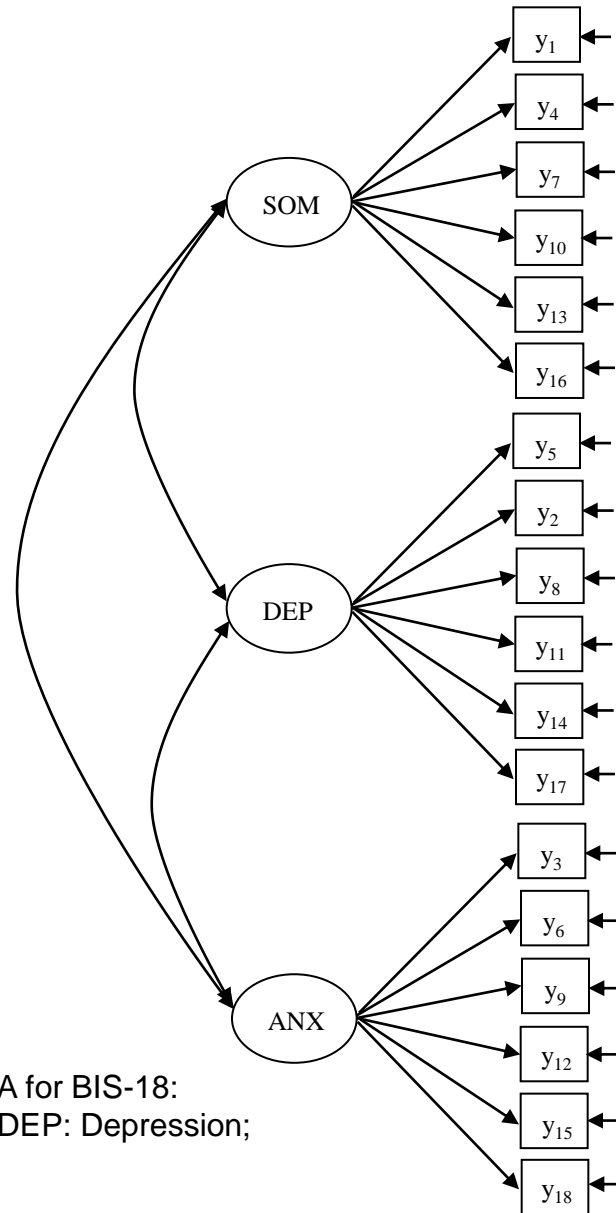


Figure 1. 3-factor CFA for BIS-18:
SOM: Somatization; DEP: Depression;
ANX: Anxiety

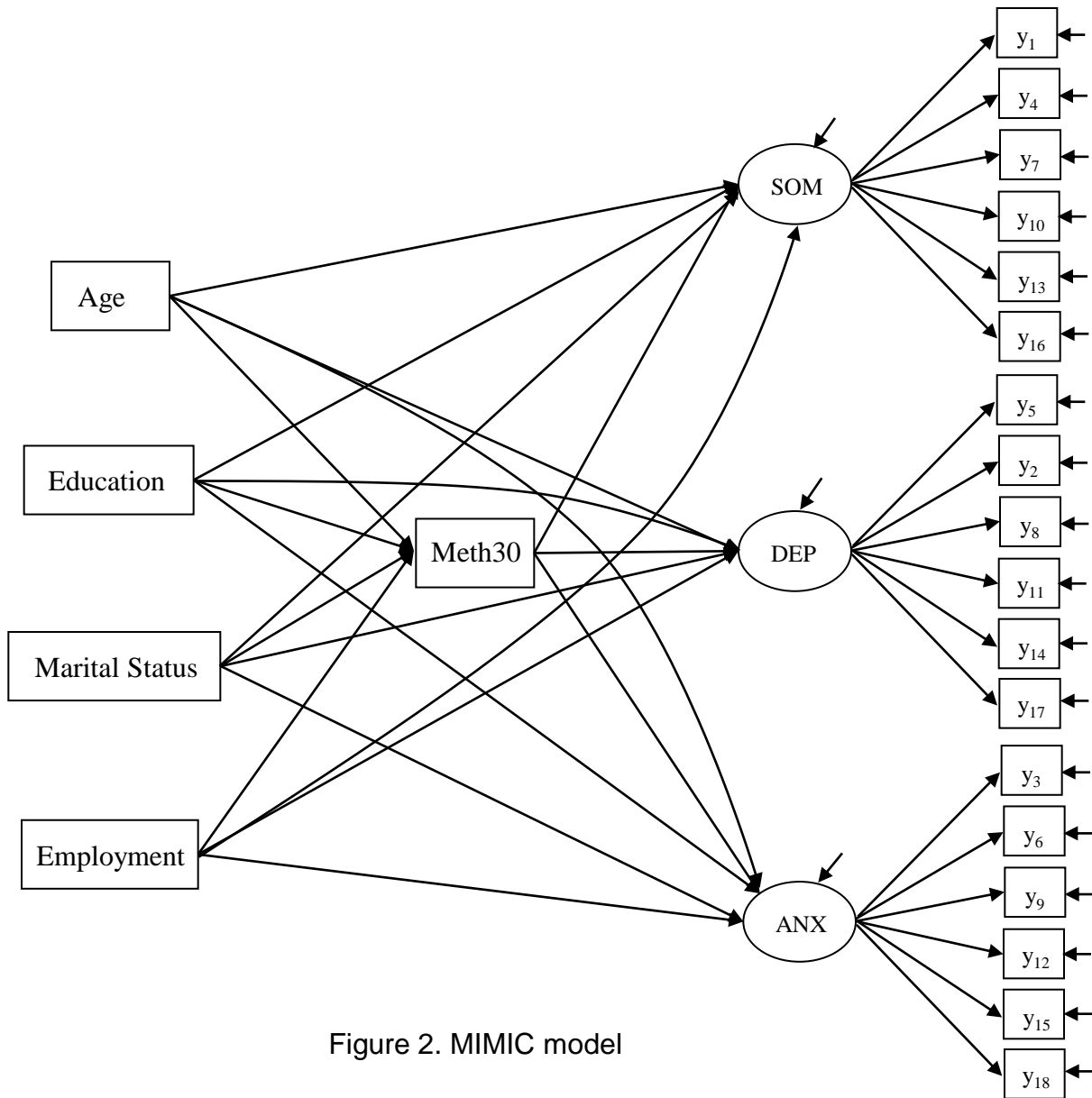


Figure 2. MIMIC model

WARNING: THE RESIDUAL COVARIANCE MATRIX (THETA) IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR AN OBSERVED VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO OBSERVED VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO OBSERVED VARIABLES. CHECK THE RESULTS SECTION FOR MORE INFORMATION. PROBLEM INVOLVING VARIABLE Y10.

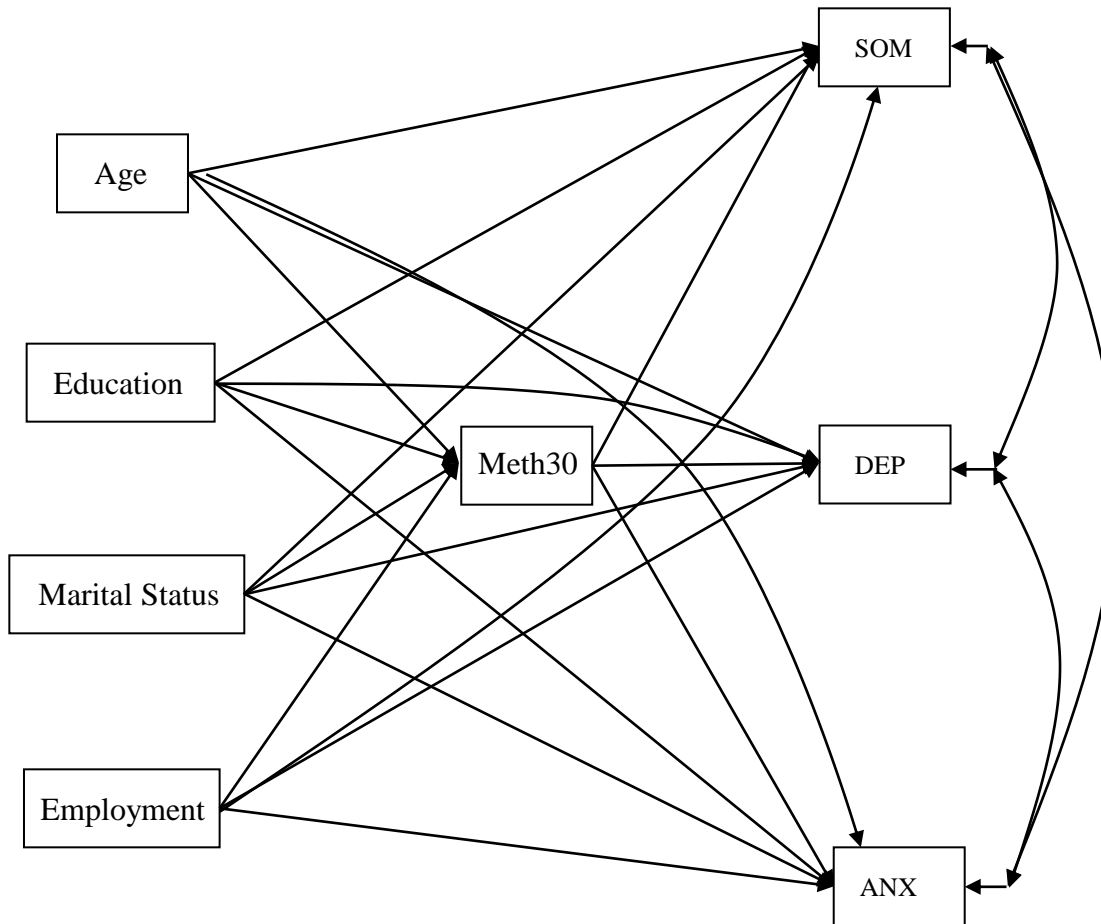


Figure 3. Path analysis model:
 SOM, DEP, and ANX are either total score or traditional point estimates of latent variables in frequentist analysis (e.g., CFA, IRT).

- Plausible values are imputed for each latent variable based on Markov Chain Monte Carlo (MCMC) Bayesian estimation.
- Cross-loadings are specified; error covariance could be specified as well.
- Parameters may be specified as fixed using those estimated from CFA model.
- Non-informative priors using a normal distribution with a mean of zero and a small variance (Muthén and Asparouhov, 2012).
- Model fit: Posterior Predictive P-Value (PPP) = **0.453**

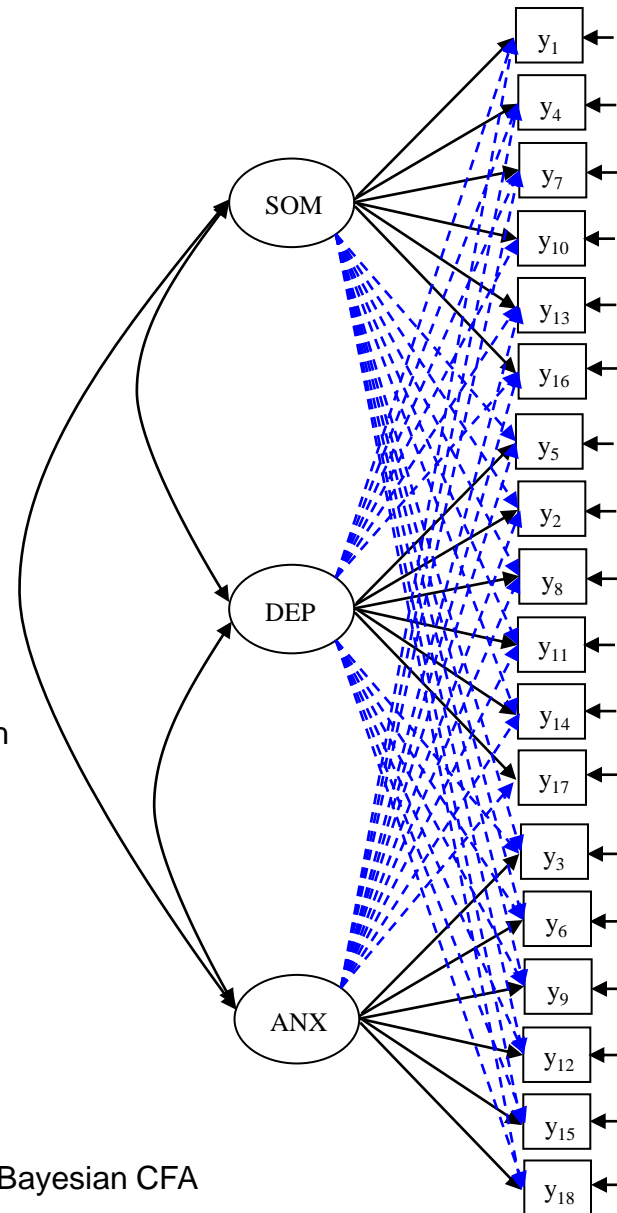


Figure 4. Bayesian CFA

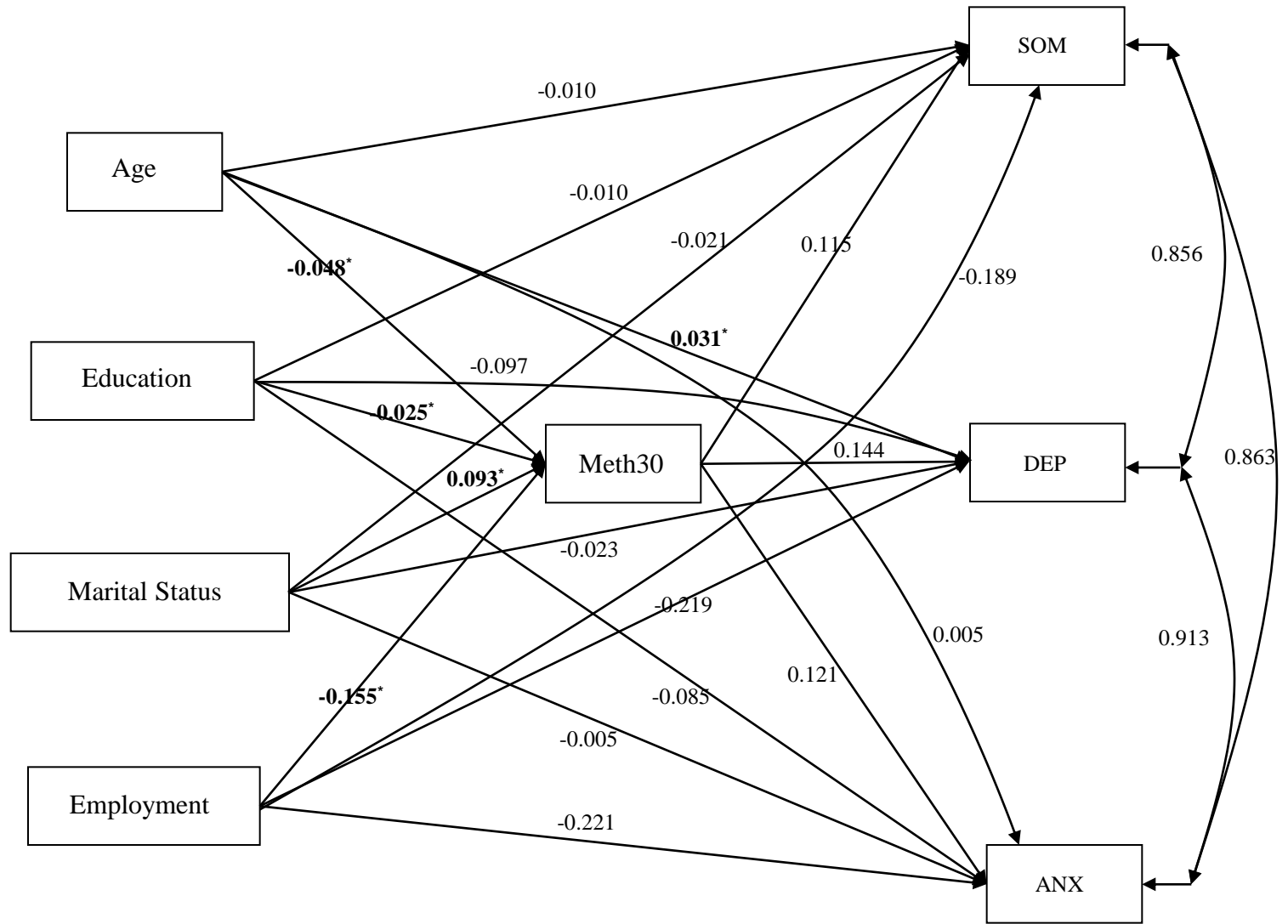


Figure 5. Example of path analysis model with estimated plausible values of latent variables.

Reference

- Asparouhov, T. & Muthén, B. 2010. Plausible values for latent variables using Mplus. Technical Report.
- Mislevy, R. J. 1991. Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177-196.
- Muthen, B. & Asparouhov, T. 2012. Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313-335.
- Rubin, D. 1987. *Multiple Imputation for Nonresponse in Survey*. New York: Wiley.