Portland State University

A Solution for Including Auxiliary Variables with Categorical Dependent Variable Estimation in SEM

Jason T. Newsom
Portland State University

Mallory R. Kroeck
Portland State University

Brian T. Keller
University of Missouri

Nicholas A. Smith
University of Texas Arlington

Paper Presented at the Modern Modeling Methods Conference, Storrs, CT, June 2024

Objective

- Illustrate and evaluate a method of including auxiliary variables with three estimation methods for structural equation models with binary dependent variables

Objective

- Inclusion of auxiliary variables with missing data estimation reduces bias and increases efficiency
- Auxiliary variables are any variables that contain information relevant to missingness and the variables with missing values, whether a cause of missingness or not.

Objective

- But most estimators and/or model parameterizations for binary dependent variables (i.e., logistic estimation with marginal maximum likelihood, delta parameterization with WLSMV, Bayes estimation) cannot include variables not explicitly part of the hypothesized model the saturated correlations approach

- To extend one method of manual specification for the auxiliary variable to the case with binary dependent variable estimation

Auxiliary Variables

- Two methods of using auxiliary variables with SEM, originally suggested by Graham, Hofer, Donaldson, MacKinnon, and Schafer (1997) and detailed by Graham (2003)

- Saturated correlates (“spider model”) and extra dependent variables (“extra DV”)

Auxiliary Variables

- Saturated correlates (RX is the auxiliary variable)

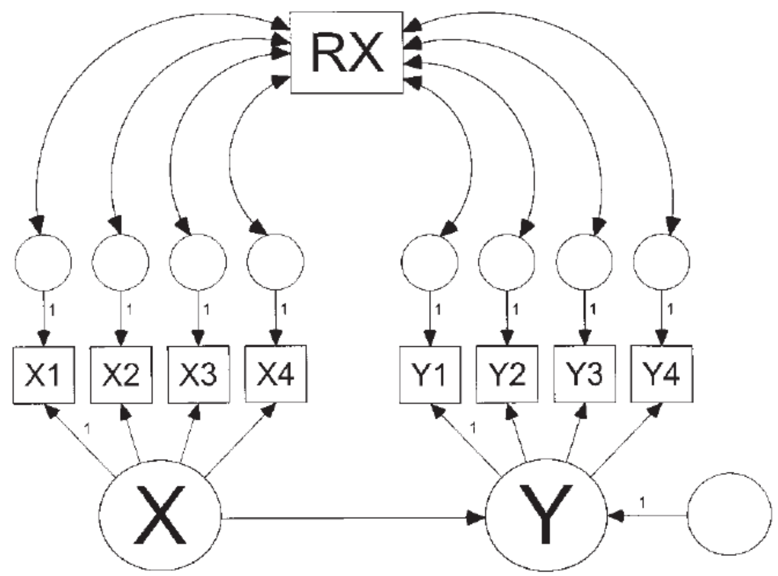


FIGURE 6 Model 3: “saturated correlates” model (latent variable version).

From John W. Graham (2003) Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models, Structural Equation Modeling: A Multidisciplinary Journal, 10:1, 80-100, DOI: 10.1207/S15328007SEM1001_4, p. 84

Auxiliary Variables

Extra Dependent Variable (RX is the auxiliary variable)

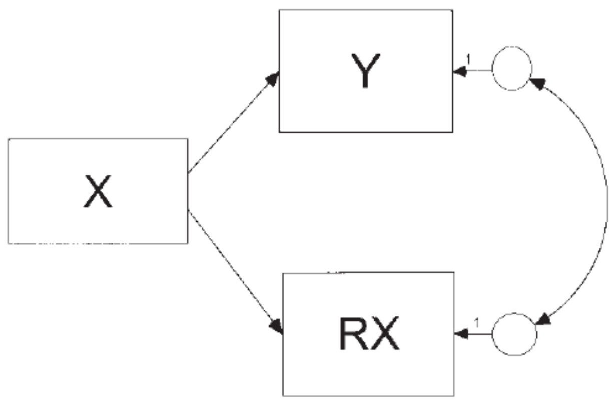


FIGURE 3 Model 2: “extra DV” model (manifest variable version).

From John W. Graham (2003) Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models, Structural Equation Modeling: A Multidisciplinary Journal, 10:1, 80-100, DOI: 10.1207/S15328007SEM1001_4, p. 83

Auxiliary Variables

- Graham (2003) demonstrates that the saturated correlates and extra dv approaches are highly similar in bias and efficiency

- Saturated correlations is implemented automatically in Mplus with AUXILIARY command

- Manually implemented saturated correlations or the AUXILIARY command are not available with marginal maximum likelihood, WLSMV delta, or Bayes estimation for binary variables

- Extra DV option is possible, however

Auxiliary Variables with Full Maximum Likelihood

- FIML as implemented in SEM (as well as MI) provides unbiased estimates and more efficient than traditional missing data handling alternatives if data are The missing at random (MAR)

- MAR assumption can be met if the association between missingness and the values of variable with missing data can be accounted for by other variables

- Collins, Schafer, & Kam (2001) show inclusion of auxiliary variables when data are MNAR reduces bias and may increase efficiency

- Best if missingness > 25%, linear missing data mechanisms, and the magnitude of the relationship to missingness and variables of interest are *moderate* or *high*

Auxiliary Variables with FIML

TABLE 17.2. N_{EFF} Benefits from Adding an Auxiliary Variable

ρ_{YZ}	N_{EFF}	% benefit
.10	502	0.4%
.20	509	1.8%
.30	522	4.4%
.40	542	8.4%
.50	570	14%
.60	608	22%
.70	660	32%
.80	733	47%
.90	839	78%

Note. N_{TOT} = 1,000; N_{CC} = 500; % benefit = $(N_{EFF} - N_{CC})/N_{CC}$.

From Graham & Coffman (2012), p. 285; Collins, Schafer, & Kam, 2001; ρ is the correlation of auxiliary variable with Y

Binary Estimation with Missing Values

- Diagonal weighted least squares (WLSMV from here on) is limited information method with hybrid approach to missing data, with some steps based on FIML and some based on pairwise deletion, and this may be less optimal when values are MAR or MNAR (Asparouhov & Muthén, 2010; 2021 ; Jai & Wu, 2019)

- Marginal maximum likelihood (MLR from here on) should be expected to behave similarly to FIML for continuous variables under various missing data mechanisms (e.g., Asparouhov & Muthén, 2021) [MAR assumption]

- Bayesian estimation takes into account parameters and auxiliary variables in the multivariate posterior distribution [MAR assumption]

Binary Estimation with Missing Values

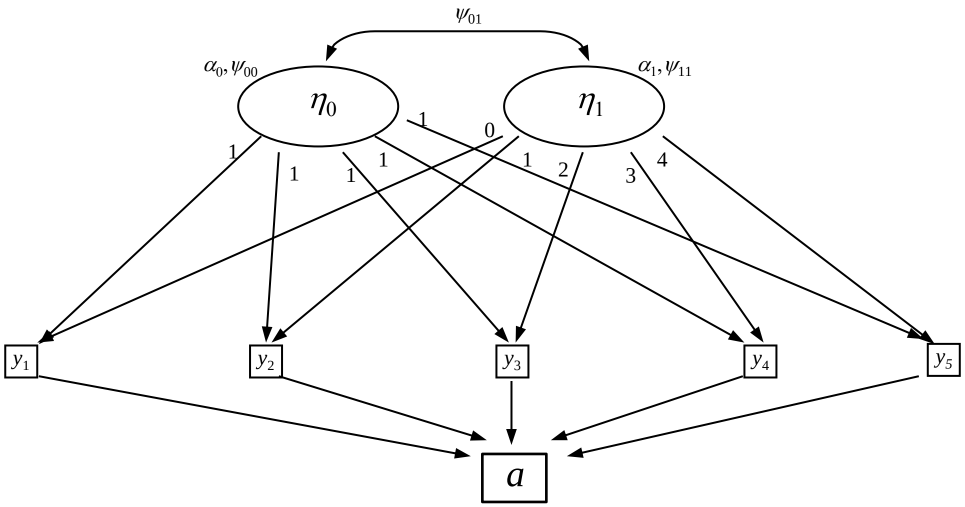
- Asparouhov and Muthén (2021) show biased estimates for WLSMV and unbiased estimates for Bayesian for a covariance between two variables under MAR

- For a growth model with MAR missingness, ML and Bayesian estimates had minimally biased estimates of growth parameter means and variances, whereas WLSMV estimates were biased for slope means and parameter variances (Asparouhov & Muthén, 2021)

- Another simulation comparing ML (non-robust) and Bayesian (informative and noninformative priors) with dropout pattern found ML outperformed Bayesian with noninformative priors (Kim, Huh, Zhou, & Mun, 2020).
  - Missing data mechanism unclear

Auxiliary Variables

Extra DV adaptation for growth curve



Method

- Data were generated in SAS version 9.4 and the RandomMVBinary macro (Wicklin, 2013)
- 5 binary variables with nearly symmetric distributions at baseline (Pr = .45)
- Linear slope increment proportion of .025 per wave, a moderate effect, approximately equal to standardized slope value of .4
- Intercepts and slopes designed to have significant variance

Method

- N = 200 for each sample
- # replications = 1,000
- Average proportion of cases with dropout equal to .2 per wave, based on similar rates for longitudinal studies, such as the Health and Retirement Study (Heeringa & Conner, 1995)
- Percent cases with at least one missing value: 54%

Portland State UNIVERSITY

Method

- Missing values were created to mimic three dropout patterns (missing on y_3, y_4, and y_5; missing on y_3 and y_4; or missing on y_5)

y_1	y_2	y_3	y_4	y_5
				.
			.	.
		.	.	.

- Three patterns combined (i.e., we do not examine the differences among the three patterns)

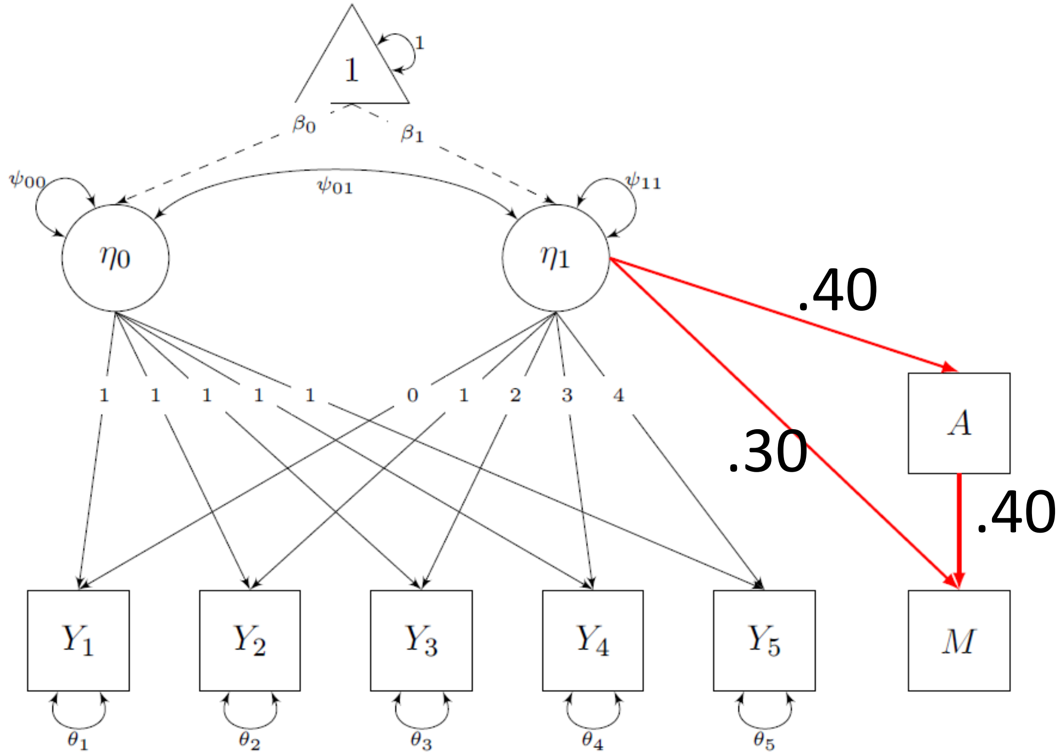
Method

- These analyses focus on a "partial MNAR" mechanism, in which the missingness is related to the value of the auxiliary variable and to the slopes

- Moderate correlation of auxiliary variable with missingness (.4) and with slopes (.4), so that some association of slopes and missingness remained

Method

"Partial MNAR"



Normal FIML

- To demonstrate the equivalence of the “all correlates” and “extra dv” approaches, we first estimated growth curve models using standard FIML

- Normal theory FIML is *inappropriate* given binary outcomes for y at each time point

- Compared five conditions: (1) complete data; (2) no auxiliary variable included in model; (3) auxiliary function in Mplus (same as all correlates); (4) saturated correlates model; (5) extra DV model

Normal FIML

	Complete			No Auxiliary			Auxilary			All Correlates			Extra DV	
	Est	SE		Est	SE		Est	SE		Est	SE		Est	SE
α_I	0.4485	0.0314		0.4624	0.0318		0.4559	0.0319		0.4559	0.0319		0.4559	0.0319
α_S	0.0249	0.0097		0.006	0.0121		0.0147	0.0124		0.0147	0.0124		0.0147	0.0124
ψ_I	0.1104	0.0214		0.1074	0.0224		0.1083	0.0224		0.1083	0.0224		0.1083	0.0224
ψ_S	0.0047	0.0023		0.0041	0.0031		0.0045	0.0031		0.0045	0.0031		0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059		-0.0074	0.0072		-0.0080	0.0072		-0.0080	0.0072		-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Normal FIML

	Complete		No Auxiliary		Auxilary		All Correlates		Extra DV	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_I	0.4485	0.0314	0.4624	0.0318	0.4559	0.0319	0.4559	0.0319	0.4559	0.0319
α_S	0.0249	0.0097	0.006	0.0121	0.0147	0.0124	0.0147	0.0124	0.0147	0.0124
ψ_I	0.1104	0.0214	0.1074	0.0224	0.1083	0.0224	0.1083	0.0224	0.1083	0.0224
ψ_S	0.0047	0.0023	0.0041	0.0031	0.0045	0.0031	0.0045	0.0031	0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059	-0.0074	0.0072	-0.0080	0.0072	-0.0080	0.0072	-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Normal FIML

	Complete		No Auxiliary		Auxilary		All Correlates		Extra DV	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_I	0.4485	0.0314	0.4624	0.0318	0.4559	0.0319	0.4559	0.0319	0.4559	0.0319
α_S	0.0249	0.0097	0.006	0.0121	0.0147	0.0124	0.0147	0.0124	0.0147	0.0124
ψ_I	0.1104	0.0214	0.1074	0.0224	0.1083	0.0224	0.1083	0.0224	0.1083	0.0224
ψ_S	0.0047	0.0023	0.0041	0.0031	0.0045	0.0031	0.0045	0.0031	0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059	-0.0074	0.0072	-0.0080	0.0072	-0.0080	0.0072	-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Normal FIML

	Complete			No Auxiliary			Auxilary			All Correlates			Extra DV	
	Est	SE		Est	SE		Est	SE		Est	SE		Est	SE
α_I	0.4485	0.0314		0.4624	0.0318		0.4559	0.0319		0.4559	0.0319		0.4559	0.0319
α_S	0.0249	0.0097		0.006	0.0121		0.0147	0.0124		0.0147	0.0124		0.0147	0.0124
ψ_I	0.1104	0.0214		0.1074	0.0224		0.1083	0.0224		0.1083	0.0224		0.1083	0.0224
ψ_S	0.0047	0.0023		0.0041	0.0031		0.0045	0.0031		0.0045	0.0031		0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059		-0.0074	0.0072		-0.0080	0.0072		-0.0080	0.0072		-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Normal FIML

	Complete		No Auxiliary		Auxilary		All Correlates		Extra DV	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_I	0.4485	0.0314	0.4624	0.0318	0.4559	0.0319	0.4559	0.0319	0.4559	0.0319
α_S	0.0249	0.0097	0.006	0.0121	0.0147	0.0124	0.0147	0.0124	0.0147	0.0124
ψ_I	0.1104	0.0214	0.1074	0.0224	0.1083	0.0224	0.1083	0.0224	0.1083	0.0224
ψ_S	0.0047	0.0023	0.0041	0.0031	0.0045	0.0031	0.0045	0.0031	0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059	-0.0074	0.0072	-0.0080	0.0072	-0.0080	0.0072	-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Portland State University

Normal FIML

	Complete			No Auxiliary			Auxilary			All Correlates			Extra DV	
	Est	SE		Est	SE		Est	SE		Est	SE		Est	SE
α_I	0.4485	0.0314		0.4624	0.0318		0.4559	0.0319		0.4559	0.0319		0.4559	0.0319
α_S	0.0249	0.0097		0.006	0.0121		0.0147	0.0124		0.0147	0.0124		0.0147	0.0124
ψ_I	0.1104	0.0214		0.1074	0.0224		0.1083	0.0224		0.1083	0.0224		0.1083	0.0224
ψ_S	0.0047	0.0023		0.0041	0.0031		0.0045	0.0031		0.0045	0.0031		0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059		-0.0074	0.0072		-0.0080	0.0072		-0.0080	0.0072		-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Normal FIML

	Complete		No Auxiliary		Auxilary		All Correlates		Extra DV	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_I	0.4485	0.0314	0.4624	0.0318	0.4559	0.0319	0.4559	0.0319	0.4559	0.0319
α_S	0.0249	0.0097	0.006	0.0121	0.0147	0.0124	0.0147	0.0124	0.0147	0.0124
ψ_I	0.1104	0.0214	0.1074	0.0224	0.1083	0.0224	0.1083	0.0224	0.1083	0.0224
ψ_S	0.0047	0.0023	0.0041	0.0031	0.0045	0.0031	0.0045	0.0031	0.0045	0.0031
$\psi_{I,S}$	-0.0092	0.0059	-0.0074	0.0072	-0.0080	0.0072	-0.0080	0.0072	-0.0080	0.0072

N = 200, 1,000 replications. α_I = average intercept, α_S = average slope, ψ_I = intercept variance, ψ_S = slope variance, $\psi_{I,S}$ = intercept-slope covariance

Auxiliary Variables with Binary Estimation

- Compare three estimation approaches:
  - WLSMV
  - MLR (robust marginal maximum likelihood)
  - Bayes (default noninformative priors)

Method

Same data:

- Binary indicators
- 5 time points
- N=200 for each sample
- Replications = 1,000
- “partial NMAR” mechanism

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV				
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

Portland State UNIVERSITY

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622
	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

	WLSMV					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2130	0.1306	-0.1556	0.1308	-0.1556	0.1308
α_S	0.1032	0.0417	0.0389	0.0541	0.0389	0.0541
ψ_I	1.7949	0.5827	1.6751	0.6098	1.6751	0.6098
ψ_S	0.0783	0.0433	0.0756	0.0587	0.0756	0.0587
$\psi_{I,S}$	-0.1526	0.1348	-0.0969	0.1622	-0.0969	0.1622

	MLR (Marginal Maximum Likelihood Robust)					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2073	0.1354	0.1036	0.1092	-0.1653	0.1673
α_S	0.1023	0.0450	0.0945	0.0569	0.0234	0.0556
ψ_I	1.3998	0.4321	0.6519	0.2657	1.1968	0.5714
ψ_S	0.1014	0.0407	0.1789	0.0515	0.0342	0.0482
$\psi_{I,S}$	-0.1338	0.1034	-0.1177	0.0780	-0.0441	0.1595

	Bayes					
	Complete		No Auxiliary		Extra DV	
	Est	SE	Est	SE	Est	SE
α_I	-0.2098	0.1353	-0.1572	0.1390	-0.1816	0.1379
α_S	0.1023	0.0456	0.0367	0.0549	0.0689	0.0550
ψ_I	1.9695	0.5703	1.9882	0.6706	1.9402	0.6553
ψ_S	0.1029	0.0478	0.1218	0.0674	0.1140	0.0644
$\psi_{I,S}$	-0.1806	0.1342	-0.1794	0.1741	-0.1712	0.1690

Summary

- With continuous variables, the extra DV method is essentially equivalent to the saturated correlates method
- However, the utility of including auxiliary variables with models for binary variables depends on the estimation method
- WLSMV with auxiliary variable showed no improvement over not including an auxiliary variable at all[1]

[1]Analyses not shown: saturated correlations or AUXILIARY command with theta parametrization in WLSMV show identical results to extra DV approach and no improvement over omitting auxiliary variable

Summary

- MLR with auxiliary variable resulted in more biased estimates of slopes and slope variance than not including an auxiliary variable
- But generally did considerably better than WLSMV with no auxiliary variable

Summary

- Bayes estimation showed comparable or better estimates without an auxiliary variable than did MLR without an auxiliary variable

- And generally substantially reduced bias when auxiliary variable was included

Discussion

- Attrition common in longitudinal research and it is not uncommon to reach high levels of missing data overall
- Unlikely to meet MAR in this case but many auxiliary variables (e.g., from baseline measures or earlier time points) are available
- Only one auxiliary variable that was modestly related to outcome and missingness
- Yet, biased can be reduced, so use of auxiliary variables can be helpful for reducing attrition biases

Discussion

- Diagonal weighted least squares (WLSMV specification in Mplus and lavaan) is not full information method and involves univariate likelihood initially and so is not inclusive of auxiliary variable in missing data estimation

- MI also viable alternative and may tolerate more auxiliary variables and would be expected to perform similarly to the auxiliary variable approach illustrated (Asparouhov & Muthén, 2021; Jai & Wu, 2019)

- Might fix issues with WLSMV because auxiliary variables can be included at imputation step

Conclusions

- Auxiliary variables can be included in SEM with binary dependent variables through the use of extra DV approach

- Extra DV is perhaps more convenient in many situations than the two step MI approach necessary with SEM, but is *best implemented with a Bayesian SEM approach*

References

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus.* Retrieved from http://www.statmodel.com/download/Imputations7.pdf

Asparouhov, T., & Muthén, B. (2021*). Bayesian analysis of latent variable models using Mplus.* Retrieved from http://www.statmodel.com/download/BayesAdvantages18.pdf

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6(*4), 330.

Graham, J.W. (2003) Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(1), 80-100, DOI: 10.1207/S15328007SEM1001_4

Graham, J. W., & Coffman, D. L. (2012). Structural equation modeling with missing data. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 277–295). The Guilford Press.

Graham, J.W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P.,&Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. Bryant, M.Windle, & S.West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–366).Washington, DC: American Psychological Association.

Heeringa, S. G., & Conner, J. H. (1995, May). *Technical description of the Health and Retirement Survey sample design.* Institute for Social Research: University of Michigan Survey Research Center. Retrieved December 31, 2016 from http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior research methods, 51*, 2337-2355.

Kim, S. Y., Huh, D., Zhou, Z., & Mun, E. Y. (2020). A comparison of Bayesian to maximum likelihood estimation for latent growth models in the presence of a binary outcome. *International journal of behavioral development, 44*(5), 447-457.

Wicklin, R. (2013). *Simulating data with SAS.* SAS Institute.

Thank You

Questions?

Your comments and suggestions are greatly appreciated!

Contact: newsomj@pdx.edu