The Mixed Effects Structural Equations Model

Lynne Steuerle Schofield

Swarthmore College

PIAAC Methodological Seminar: Paris, France
June 14, 2019

Motivation

Researchers across diverse disciplines in the social sciences rely on latent variables as predictors of an outcome of interest.

- Neal and Johnson (1996) show the large effect of human capital as measured by the the Armed Forces Qualifying Test (AFQT) in explaining the Black-white wage gap in the US
- Heckman, Stixrud and Urzua, (2006) demonstrate the role of cognitive abilities and non-cognitive personality traits (e.g., motivation and self-esteem), as key to later-life outcomes, including labor market, health, and educational decisions.

Motivation

A typical model to study the effect of latent variables on the outcomes of interest is some form of a regression analysis:

$$Y_i = \beta_0 + \beta_1 \theta_i + \beta_2 Z_i + \epsilon_i \tag{1}$$

where

- Y_i = the outcome of interest for individual i (e.g., log wages)
- θ_i = some latent construct(s) (e.g., human capital, cognitive ability)
- Z_i = some other covariates (e.g. race, years of labor force experience)

Non-linear regression analyses such as logistic, probit, and Poisson regressions are also common.

Problems with Errors in Variables

Thus, while most researchers want to estimate (1), they actually estimate:

$$Y_i = \beta_0 + \beta_1 \hat{\theta}_i + \beta_2 Z_i + \epsilon_i \tag{2}$$

Ignoring the measurement error leads to biased results (Fuller, 2006, Stefanski, 2000).

The size and direction of the bias depends on the size and type of the measurement error.

The Mixed Effects Structural Equations Model (MESE)

- When θ is observed and not measured with error, our likelihood is

$$f(Y|\beta, \theta, Z)$$

- But when θ is unobserved and X (a proxy test score or a set of item responses) is observed, our likelihood becomes

$$f(Y, X|\beta, Z) \tag{3}$$

The Mixed Effects Structural Equations Model (MESE)

$f(Y, X|\beta, Z)$ is a marginal distribution of a more general model in which the unknown θ is integrated out.

Factoring by the Law of Total Probability,

$$\begin{aligned} f(Y, X|\beta, Z) &= \int f(Y, X, \theta|Z, \beta) d\theta \quad (4) \\ &= \int f(Y|X, \theta, Z, \beta) f(X|\theta, Z, \beta) f(\theta|Z, \beta) d\theta. \quad (5) \end{aligned}$$

implies a form of the Mixed Effects Structural Equations (MESE) Model

Assumptions of MESE

- Assume Y depends only on θ and Z; $Y \perp\!\!\!\perp X|\theta$ such that X provides no additional information about Y once θ is known.
- Assume $\theta \perp\!\!\!\perp \beta|Z$
- Assume $X \perp\!\!\!\perp Z, \beta|\theta$

These are based on good measurement practice and modern psychometric theory.

The Mixed Effects Structural Equations Model (MESE)

The MESE model suggests three general submodels (Richardson and Gilks, 1993):

$$
\begin{aligned}
\text{Structural Model:} \quad & Y_i|Z_i, \theta_i, \beta \;\sim\; f(Y_i|\theta_i, Z_i, \beta) & (6)\\
\text{Measurement Model:} \quad & X_{ij}|\theta_i, \gamma_j \;\sim\; f(X_{ij}|\theta_i, \gamma_j) & (7)\\
\text{Conditioning Model:} \quad & \theta_i|Z_i, \alpha \;\sim\; f(\theta_i|Z_i, \alpha) & (8)
\end{aligned}
$$

where γ_j are the parameters in the measurement model, α are the parameters in the population model for $\theta|Z$, and β, θ, Y, X, and Z are defined as before.

Submodels of MESE: The Measurement Model

- Often is the item response theory (IRT) model underlying the design, construction and scoring of the assessment
- Flexible in using different IRT measurement models for different latent constructs
- Misspecification of the IRT model is relatively robust (see simulation study in Schofield, 2008)

IRT Models

3-PL model (for binary items)

$$P_j(\theta_i) \equiv P[X_{ij} = 1] = c_j + \frac{1 - c_j}{1 + \exp[-a_j(\theta_i - b_j)]} \ . \tag{9}$$

Samejima's (1969) graded response model (GRM) (for Likert-scale survey responses and other ordinal items),

$$P^*_{jk}(\theta_i) \equiv P[X_{ijk} \geq x_{ijk}] = \frac{exp[a_j(\theta_i - b_{jk})]}{1 + \exp[a_j(\theta_i - b_{jk})]} \ . \tag{10}$$

- X_{ij} is the response of individual i to item j,
- a_j is the "discrimination" item parameter,
- b_j is the "difficulty" item parameter,
- c_j is the "guessing" item parameter, and
- P^*_{jk} is the probability of individual i with proficiency θ scoring k or above on item j.

IRT Models: Measurement Error

IRT models provide a direct estimate of the measurement error for $\hat{\theta}$, which is equivalent to the standard error of $\hat{\theta}$. Asymptotically,

$$SE(\theta_i) = \frac{1}{\sqrt{\sum_{j=1}^{J} I_j(\theta_i)}} \tag{11}$$

where $I_j(\theta_i)$ is the Fisher information.

IRT Models: Measurement Error

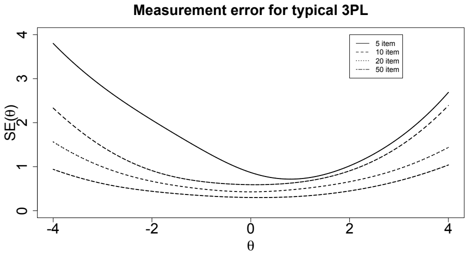


Figure: Measurement error for a typical 3-PL model by θ where $a \sim Unif(0,2)$, $b \sim N(0,1)$ and $c = 0$ for all items.

- $SE(\theta)$ varies for different values of θ: largest for those in the tails of the distribution of θ and smallest for those in the middle.
- $SE(\theta) \to 0$ as $J \to \infty$.
- $SE(\theta)$ unknown for all individuals. Using $SE(\hat{\theta})$ to correct for measurement error leads to bias (Lockwood and McCaffrey, 2014).

Submodels of MESE: The Conditioning Model

- Often assumed to be multivariate normally distributed
- Allows for possible differences in the distribution of θ across subgroups of the sample
- Is flexible in allowing the latent constructs to be associated with one another conditional on the other covariates in the model.
- Mis-specification in shape is robust (See Schofield, 2008 and Dresher, 2006)

Submodels of MESE: The Conditioning Model

Which variables to include?

- Many large scale assessments (e.g, NAEP and PISA) follow Mislevy (1991) and condition on a huge set of background covariates to avoid bias in population statistics estimated from the test.
- Schofield et al. (2014) suggest "goldilocks result" when θ is the independent variable in an analysis:
  - θ must be conditioned on all of the covariates in the structural equation
  - θ may not be conditioned on Y or any other variable associated with Y conditional on θ that is not already in the structural equation, unless there is model congeniality.

Submodels of MESE: The Conditioning Model

Congeniality: Assume the conditioning model on θ (placed by the survey institution) is

$$\theta | Y, Z \sim N(\beta_1 Z + \beta_2 Y, \sigma^2)$$

By the Law of Total Probability and Bayes' Rule, this conditioning model forces:

$$p(Y|\theta, Z) = \frac{p(\theta|Z, Y)p(Y|Z)}{p(\theta|Z)} = \frac{p(\theta|Z, Y) * p(Y|Z))}{\int p(\theta|Z, Y) * p(Y|Z)d\theta}$$

Submodels of MESE: The Conditioning Model

If we assume $p(Y|Z) \sim N(\gamma Z, \tau^2)$ then

$$p(Y|Z,\theta) \propto N(\alpha_1\theta + \alpha_2 Z, \xi^2)$$

where

$$\alpha_1 = \frac{\beta_2\tau^2}{\beta_2^2\tau^2 + \sigma^2}$$

$$\alpha_2 = \frac{\gamma\sigma^2 - \beta_1\beta_2\tau^2}{\beta_2^2\tau^2 + \sigma^2}$$

$$\xi^2 = \frac{\sigma^2\tau^2}{\beta_2^2\tau^2 + \sigma^2}$$

Submodels of MESE: The Conditioning Model

If you believe that the *right* structural model is

$$Y|\theta, Z \sim N(\alpha_1\theta + \alpha_2 Z, \xi^2), \tag{12}$$

then a conditioning model that contains Y will produce unbiased estimates of α_1 and α_2.

But suppose you want to estimate a model like

$$Y|\theta, Z \ N(\alpha_1\theta + \alpha_2\theta^2 + \alpha_3 Z, \xi^2)$$

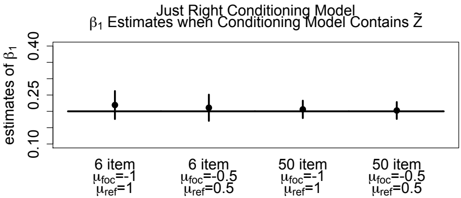
or *any other model* other than (12), then the conditioning model which includes Y doesn't match the structural model.

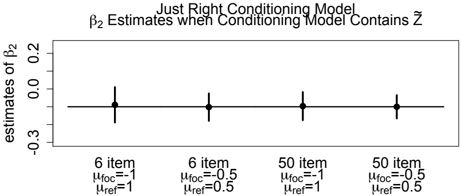
Submodels of MESE: The Conditioning Model

Mis-specification of which covariates are in the conditioning model will cause bias

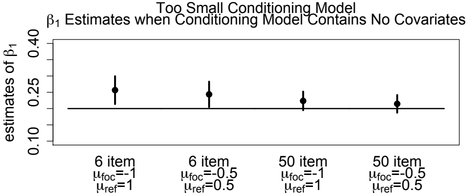
- Size and direction of the bias varies based on
  - the number of items on the test
  - the strength of the correlation between Y, Z, and θ.

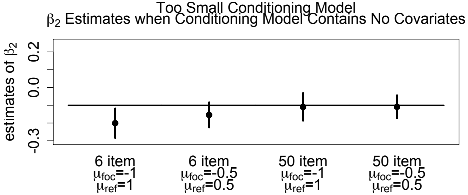
Submodels of MESE: The Conditioning Model



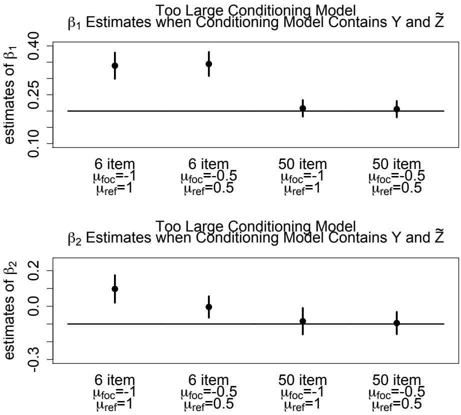


Submodels of MESE: The Conditioning Model

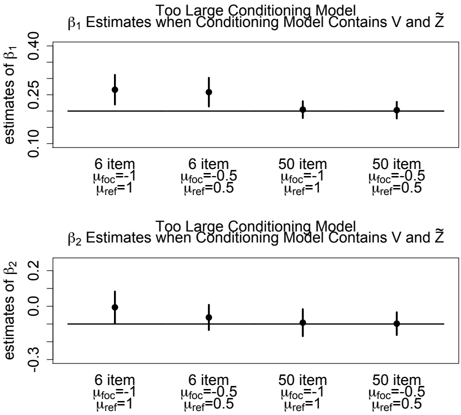




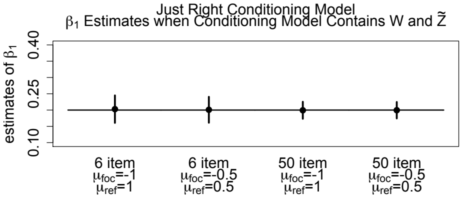
Submodels of MESE: The Conditioning Model

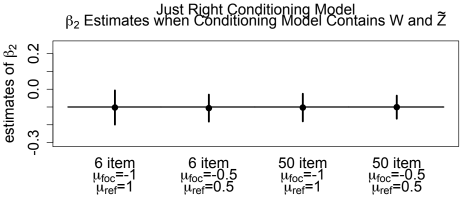


Submodels of MESE: The Conditioning Model



Submodels of MESE: The Conditioning Model





Submodels of MESE: The Structural Model

- Equation of primary interest
- θ is treated as a random variable in a mixed-effects regression
- Functional form depends on the substantive question of interest and the response variable, Y.
- Can accommodate several models; among them, any generalized linear model

MESE In Action: STEM Retention

- Rising concern about the under-representation, and specifically the retention of minorities and women in science, technology, engineering, and mathematics (STEM) disciplines in higher education.
- In 2008, 31.7% of black, 33.1% of Hispanic students versus 43.9% of whites persisted in STEM in the U.S.
- Griffith (2010) found only 37% of women versus 43% of men persisted in STEM in the U.S.

MESE In Action: STEM Retention

Studies control for many latent variables in trying to understand these differentials

- academic achievement (e.g., Maltese and Tai, 2011),
- math and science identity (Chang, et al., 2011),
- interest (Sullins, Hernandez and Fuller, 1995),
- future time perspective (Husman, et al, 2007),
- sense of community (Espinosa, 2011),
- goals (Leslie, McClure and Oaxaca, 1998), or
- personality traits (Korpershoek, Kuyper and van der Werf, 2012).

MESE In Action: Modeling STEM Retention

$$\begin{aligned} Y_i = 1 &\sim \text{Bernoulli}(p_i) \\ \log \frac{p_i}{1 - p_i} &= \beta_0 + \beta_1 \theta_i + \beta_2 Z_i \end{aligned} \tag{13}$$

where

- Y_i is a binary measure of STEM persistence,
- $\theta_i = (\theta_{1i}, \theta_{2i}, \ldots, \theta_{ki})$ is a vector of k latent variables measuring cognitive and non-cognitive traits, and
- Z_i is a vector of demographic variables including indicator variables for underrepresented minorities (URMs) and female gender.

Modeling STEM Retention: The Data

- 1997 National Longitudinal Survey of Youth (NLSY97) $n =$ 8900 youths ages 12 to 16 years old as of December 31, 1996.
- $Y_i =$ a "stayer," someone who persisted in a STEM major; or a "leaver," someone who declared a STEM major but did not persist to graduation.
- Race $=$ 1 when a URM and 0 when not
- Gender $=$ 1 indicating Female gender or 0 when not
- Six latent variables:
  - The PIAT: a measure of cognitive proficiency in mathematics ($J \leq 100$; binary questions)
  - The TIPI: measures of the Big Five: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness. ($J = 2$ per personality characteristic; Likert-type responses)

Modeling STEM Retention: The Data

Table: Sample Characteristics, 1997 National Longitudinal Survey (NLSY97)

	Female	Male	URM	NonURM	Total
N	163	265	133	295	428
Proportion Stayers	0.49	0.67	0.50	0.65	0.60

Notes: Author's calculations, 1997 National Longitudinal Survey of Youth. Sample of only those youth who have completed a two or four-year college degree and declared a STEM major at some point in their college career.

Modeling STEM Retention: The Data

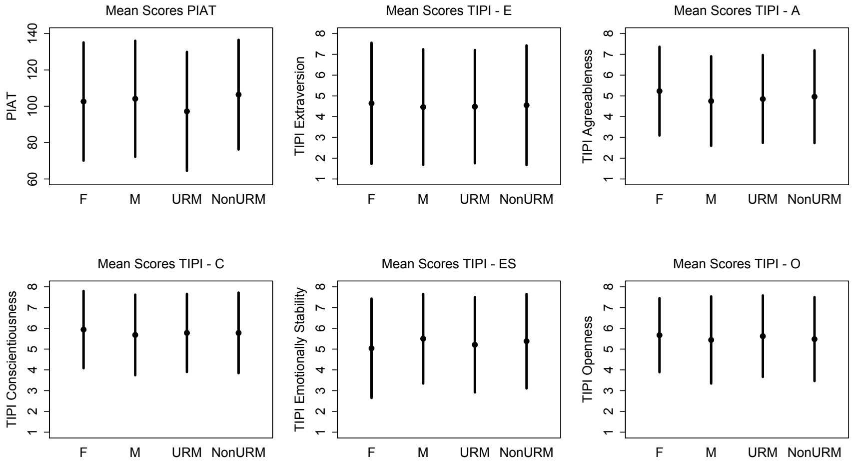


Figure: Average PIAT and TIPI Scores by Race and Gender

Modeling STEM Retention: The Results

Table: Logistic Regression of Persistence in STEM (NLSY97)

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
	Baseline	PIAT		TIPI		TIPI & PIAT	
adjusted for ME?		N	Y	N	Y	N	Y
URM	−0.584* (0.223)	−0.427* (0.217)	−0.375 (0.235)	−0.641* (0.222)	−0.894* (0.474)	−0.472 (0.237)	−0.690 (0.411)
Female	−0.704* (0.203)	−0.717* (0.210)	−0.728* (0.212)	−0.594* (0.226)	−0.128 (0.616)	−0.612* (0.223)	−0.270 (0.526)
PIAT		0.330* (0.109)	0.324* (0.121)			0.341* (0.113)	0.400* (0.200)
TIPI Extraversion				−0.284* (0.111)	−0.625 (0.652)	−0.280* (0.115)	−0.661 (0.546)
TIPI Agreeableness				−0.227 (0.117)	−0.943 (0.745)	−0.242* (0.115)	−0.957* (0.489)
TIPI Conscientiousness				0.082 (0.107)	0.432 (0.327)	0.081 (0.106)	0.444 (0.312)
TIPI Emo. Stability				0.036 (0.114)	0.397 (0.460)	0.017 (0.116)	0.215 (0.388)
TIPI Openness				−0.083 (0.113)	−0.177 (0.948)	−0.083 (0.116)	0.354 (0.857)
N	428	428	428	428	428	428	428
DIC	560	552	554	555	505	547	507
Error Rate	36.9%	35.7%	36.4%	32.7%	23.1%	32.7%	22.7%

Notes: Sample of those youth who have completed a two or four-year college degree who declared a STEM major at some point. All estimates of latent variables have been standardized.

MESE In Action: Intergenerational Transference of Human Capital

- Since the 1970s, inequality in the US has grown remarkably
- Chetty, Hendren, Kline, Saez, and Turner (2014) report that for children born between 1971 and 1986 face a remarkably stable intergenerational correlation, consistently showing rank-rank slopes ranging from 0.30 to 0.35
- There are a host of measurement issues with using income that has been explored quite extensively by Solon and others
- In addition, interpretation of the coefficient is a bit opaque because it includes human capital decisions, labor supply, and location decisions
- Could examine how **human capital** as measured by test scores is transmitted across generations

Ideal Model for Examining Intergenerational Transference

Ideally, we would like to estimate a regression with a random intercept for mother

$$\theta_{(c)i} = \beta_{0p} + \beta_1 \theta_{(m)i} + e_i$$

We will adjust the MESE model to do this:

$$\begin{aligned}
\theta_{(c)i} | \theta_{(m)i}, \beta_0, \beta_1 &\sim N(\beta_{0p} + \beta \theta_{(m)i}, \sigma^2) && (14)\\
\beta_{0p} &\sim N(0, 1) && (15)\\
X_{(c)ij} | \theta_{(c)i}, \gamma_{(c)j} &\sim IRT(X_{(c)ij} | \theta_{(c)i}, \gamma_{(c)j}) && (16)\\
X_{(m)il} | \theta_{(m)i}, \gamma_{(m)l} &\sim IRT(X_{(m)il} | \theta_{(m)i}, \gamma_{(m)l}) && (17)\\
\theta_{(m)i} &\sim N(0, 1) && (18)
\end{aligned}$$

NLSY data

- We use the children of the NLSY79 performances on the PIAT math test as the dependent variable
- For the mother's human capital we use the AFQT, $J = 104$ item subset of the ASVAB test for the1979 Cohort
- We divide our sample of mothers into three groups based on the race and ethnicity (white, black, and Hispanic) reported by the mother in 1979 round
- We report the impact of a one-standard deviation increase in the mother's AFQT score on the number of standard deviation her child's PIAT score for both OLS and MESE

Intergenerational Transference of Human Capital: Results

Table: Results: Intergenerational Transference of Human Capital

	n	OLS	MESE
White	4170	0.410	0.378
Black	2693	0.341	0.267
Hisp	1742	0.371	0.294

Notes: Author's calculations, 1979 National Longitudinal Survey of Youth and their children. Sample of only those for whom we have both an AFQT score for the mom and a PIAT score for the child.

Observations

- Measurement error inherent in latent variables must be modeled when the latent variables are used as predictors in secondary analyses or else bias ensues in both
  - the effect of the latent variable on the outcome of interest
  - the effect of any covariate associated with the latent variables
- The bias increases with shorter tests/surveys
- One way to do account for the measurement error is the MESE model
- MESE results are substantially different than OLS

Observations: STEM Retention Gap

- In STEM retention gaps, OLS substantially underestimates the effect of the latent variables, especially the personality traits.
- Racial gaps become insignificant after controlling for math proficiency <u>and</u> its measurement error
- After adjusting for measurement error, results suggest comparably-skilled and comparably-traited men and women are equally likely to remain in STEM

Observations: Intergenerational Transference of Human Capital

- The results indicate that OLS substantially overstates the degree of correlation, especially for Hispanic and African Americans
- The correlation between generation is the strongest for whites. Given the academic progress made by Hispanic and African American students, this is probably as expected
- The correlation for whites is quite similar from those found in Solon's work and Chetty and his co-authors, but for African Americans and Hispanics it is considerably lower