Establishing Evidence-Based Practice with Structural Equation Modeling

Mike W.-L. Cheung[1]

1 March 2018

[1]Department of Psychology, National University of Singapore (NUS)

A little bit background about me (1)

- PhD: Quantitative psychology, the Chinese University of Hong Kong
- Associate Professor:
  - Department of Psychology, National University of Singapore (NUS)
  - Department of Management & Organisation (courtesy appointment), NUS
- Research areas: Quantitative methods
  - Structural equation modeling, meta-analysis, multilevel model, analysis of missing data, longitudinal data analysis, analysis of non-normal data, etc.

A little bit background about me (2)

- Associate editors:
  - *Research Synthesis Methods*
  - *Neuropsychology Review*
  - *Frontiers in Psychology (Quantitative Psychology and Measurement)*
- Editorial boards:
  - *Psychological Methods*
  - *Psychological Bulletin*
  - *Journal of Management* (Methods task force)
  - *Health Psychology Review* (Research methods and data analysis)

Goals of today's talk

- Introduce the basics of SEM.
- Introduce how to apply and interpret SEM in our work.
- *Note*: We cannot cover how to conduct the analyses in only 2 hours!

What is SEM?

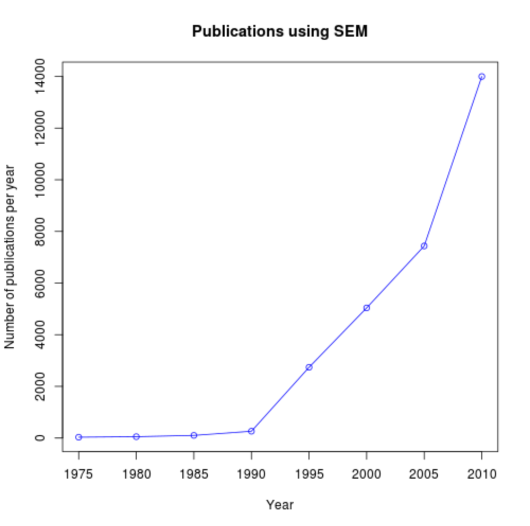
- SEM represents a family of related multivariate techniques.
- It is also known as covariance structural analysis, covariance structure model, analysis of covariance structures, analysis of correlation structure, LISREL model (in the old days), etc.
- It is used to test hypothesized models (theory), which can be used to provide empirical evidence for evidence-based practice/research.

Relationship with other statistical techniques

- Many statistical techniques we have learned before are special cases of SEM, for example, independent and dependent t-tests, ANOVA, ANCOVA, MANOVA, multiple regression, path analysis, confirmatory factor analysis (CFA), item response theory (IRT), multilevel models, and meta-analysis, etc.
- SEM has been extended to combine with other statistical techniques, for example, mixture model, missing data techniques, generalized linear model, categorical data analysis.

Popularity of SEM

- There is no surprise that more and more publications are using SEM as the research tool.

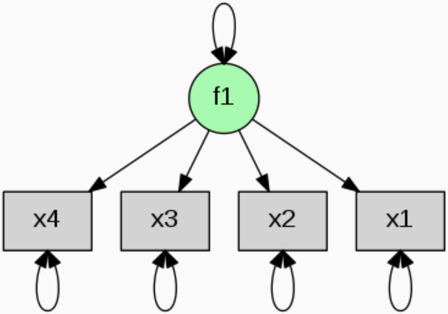


Models on relationship among the constructs

- Most statistical techniques are limited to one dependent variable (DV).
- SEM allows researchers to test models with a complicated relationship.
- Models can be represented by path diagrams.

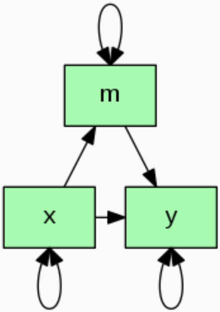
Observed vs. latent variables

- Most constructs in social and behavioral sciences, e.g., psychology, are latent or abstract. They cannot be directly measured or observed.
- **Latent** variables:
  - Abstract and hypothetical constructs.
  - For example, motivation, stress, depression, intelligence, and satisfaction.
- **Observed** or **measured** variables:
  - Indicators of the latent constructs.
  - For example, items to measure depression, test scores of intelligence.


f1
x4
x3
x2
x1

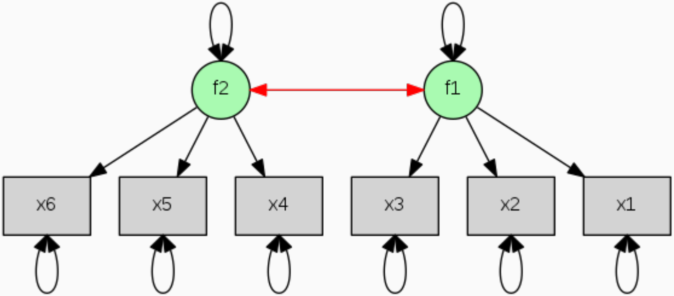
Path analysis

- Linear relationships among observed variables (rectangles).
- No latent variable.
- Multiple regression may also be used to fit this model.
- Research question to answer: What is the mechanism for explaining the dependent variables?

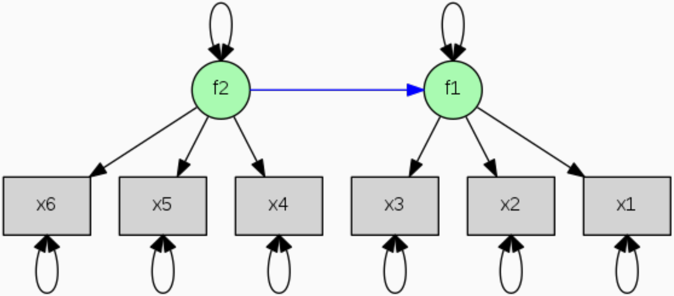


Confirmatory factor analysis (CFA)

- Linear relationships among latent and observed variables.
- No direct effect among the latent variables.
- Research question to answer: What is the construct validity of the psychological constructs?

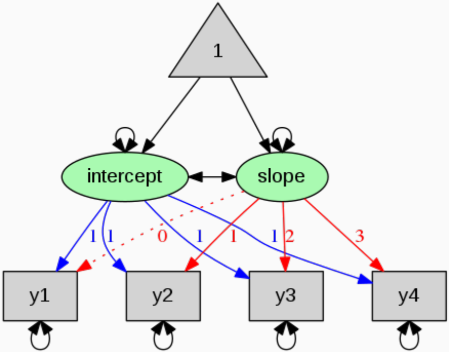


- It combines both CFA and path analysis.
- It may include direct effects among the latent variables.
- Research question to answer: What is the mechanism for explaining the dependent variables in the latent constructs?



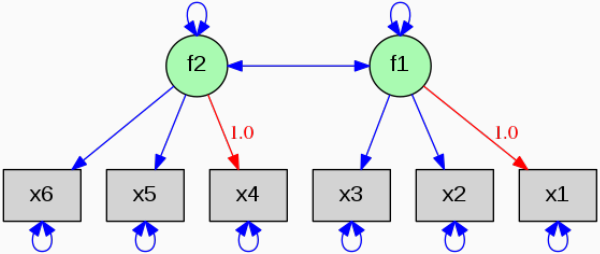
Latent growth model

- What are the growth trajectories of the individuals over time?
- What variables can be used to predict the growth trajectories?



A CFA example

- As an example, we want to fit a two-factor model on intrinsic (f1) and extrinsic (f2) motivation.
- Variable names: x1 to x6 (n=300)
- **Research question**: Does the CFA model fit the data?
- By default, the loading of the first item per factor is fixed at 1.0 in most SEM packages.



Model evaluation in SEM

- There are two major tasks in model evaluation.
- **Overall model fit**: testing whether the proposed model *as a whole* fits the data.
- **Individual parameter estimates**: testing whether the parameter estimates are significant.
- *Note*. If the overall model does not fit the data, we do not test and interpret the parameter estimates.

Chi-square test statistics (1)

- One major difference between SEM and other statistical techniques is how research hypotheses are tested:[1]
  - *Reject-support*: Rejecting the null hypothesis supports the researcher's belief (e.g., t-test, ANOVA, regression analysis, and MANOVA).
  - *Accept-support*: Accepting the null hypothesis supports the researcher's belief (e.g., SEM).
- Based on this rationale, SEM users usually do not want to reject the null hypothesis (the proposed model).

---

[1]Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

Chi-square test statistics (2)

- Chi-square test (also known as the likelihood ratio (LR) test):
    - *If the proposed model is correct*, the test statistic has a chi-square distribution.
    - This is a “badness-of-fit” index: large chi-square statistic indicates a poor fit.
    - The proposed model is rejected at .05 if the test statistic is larger than the critical value.

Issues of chi-square test statistics (1)

- SEM users rarely depend on the chi-square test because of various issues.
- Model misspecification:
    - Are there any "true" models in the world?
    - Most SEM users consider models as approximations of the reality.
    - George Box's favorite quote: "Essentially, all models are wrong, but some are useful."

Issues of chi-square test statistics (2)

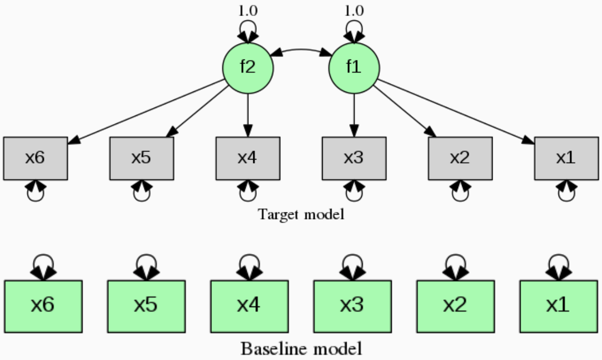
- Violation of underlying assumptions:
  - Data (or residues) are normally distributed.
  - Large samples are required.
  - When data are not normally distributed, especially in clinical studies, or in small sample sizes (e.g., N=100 or 200), the test statistic may not follow a chi-square distribution.
- Sensitive to sample size:
  - All proposed models will be rejected when the sample sizes are large enough.
  - Large samples work against researchers!

Goodness-of-fit indices

- Many SEM users are aware of the problems associated with the chi-square test.
- There are many goodness-of-fit indices developed as alternative measures.
- There are more than 20 goodness-of-fit indices in the market!

Incremental fit indices (1)

- They measure the *relative improvement* in fit by comparing the *target* (or proposed) model against the *baseline* model.
- The baseline model is usually the model stating that all variables are uncorrelated. It is known as the *independence* model. It can be considered as the *worst* model.



Incremental fit indices (2)

- Normed fit index (NFI):
  - $\frac{\chi^2_B - \chi^2_T}{\chi^2_B}$
  - χ^2_T and χ^2_B are the chi-square statistics of the target and the baseline (or null) models.
  - It measures the proportionate reduction in the chi-square values when moving from the baseline model to the hypothesized model.
- Non-normed fit index (NNFI), which is known as Tucker-Lewis index (TLI), is similar to the NFI with an adjustment of the complexity of the model.

Incremental fit indices (3)

- Comparative fit index (CFI): $0 \leq CFI \leq 1$
- What is a well-fitted model?
    - Conventional rule of thumb (without any empirical support): at least > 0.9.
    - The cut-offs are more demanding now (see below).

Residual-based indices (1)

- When the model fits well, the residuals (the difference between the model implied covariance matrix and the sample covariance matrix) should be small.
- Standardized root mean square residual (SRMR)
  - It measures the average value of the standardized residuals.
  - It ranges from zero (perfect fit) to one (very poor fit).
  - Rule of thumb: A well-fitted model $< .05$.

Residual-based indices (2)

- Root mean square error of approximation (RMSEA)[2]
- Similar to SRMR.
- Advantage: Confidence intervals on RMSEA are available on most SEM packages.
- Rules of thumb:
    - Close fit: < 0.05
    - Reasonable fit: 0.05 - 0.08
    - Inadequate fit: > 0.1

[2] Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

What do we need to report in research articles?

- We usually report the chi-square test statistic and it associated *df* and *p*-value, some incremental fit indices and some residual based indices.
- What is a well-fitted model? One popular approach is the combinational rules:[3]
  - NNFI (TLI) or CFI > 0.95 and SRMR $< .09$ *OR* RMSEA $< .05$ and SRMR $< .06$
- Although this recommendation has been widely applied, it is not without criticisms.[4]

[3] Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

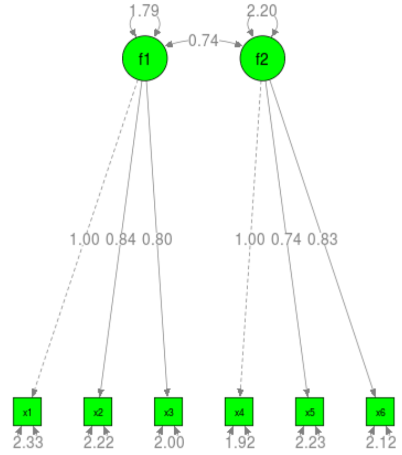
[4] Marsh, H. W., & Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-342.

Our CFA example (intrinsic and extrinsic motivation)

- The proposed model fits the data well with $\chi^2(8) = 2.45, p = .96, CFI = 1.00, RMSEA = 0.000, SRMR = .015$.

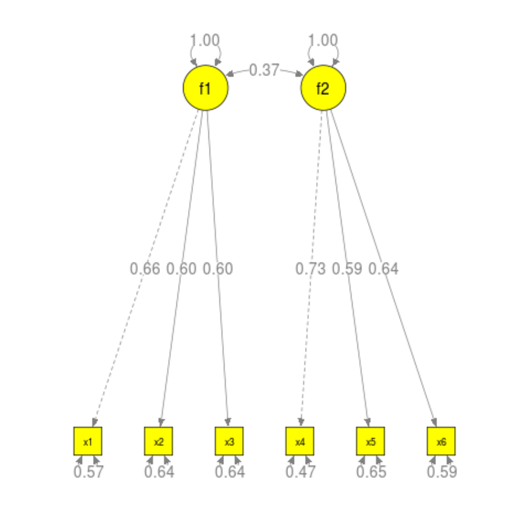
Parameter estimates (1)

- The parameter estimates are *relative* to the fixed loadings.



Parameter estimates (2)

- Sometimes, it is easier to interpret the *standardized* parameter estimates.



Comparing non-nested models

- Sometimes, the models being compared are non-nested. That is, we cannot convert a model into the other by imposing constraints.
- Chi-square difference test is inappropriate.
- Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) measure the parsimonious fit that considers both the model fit and the no. of parameters estimated.
- A smaller value indicates the model is better in compromising between the model fit and the model complexity.
- Choose the model with the smallest AIC or BIC.
- They can be used to compare nested and non-nested models.

Structural equation models

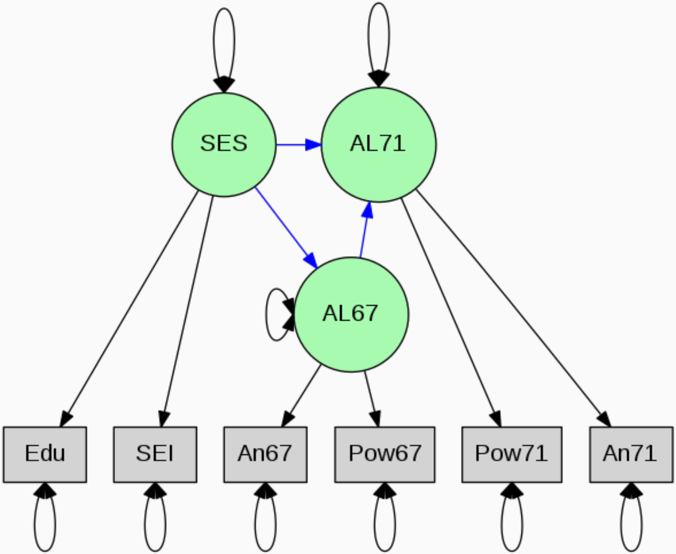
- There are two basic components in SEM.
- **Measurement (CFA) model**:
  - Are the items grouped according to the theory?
  - Assessment of convergent and discriminant validity of measurement.
  - CFA tests construct validity, not reliability.
- **Structural model**:
  - What are the relationships among the latent variables?
  - Assessment of predictive validity.

An example: Stability of alienation

- Wheaton, et al. (1977) studied the stability of attitudes over time (1967 and 1971). These include alienation and the relation to background variables such as education and occupation.[5]
- *Alienation*: Anomia subscale (Anomia), and Powerlessness subscale (Power)
- *Socioeconomic status (SES)*: Duncan's Socioeconomic Index (SEI), and Years of schooling (EDU)

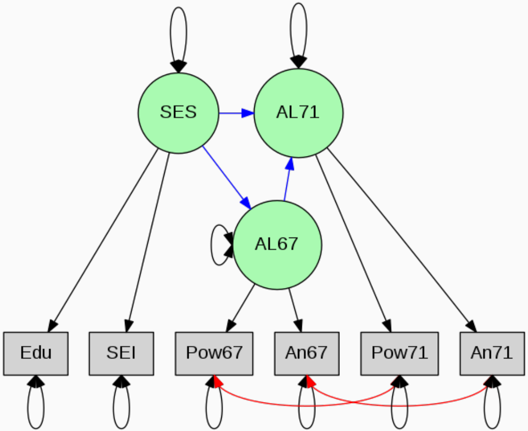
---

[5]Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.): *Sociological Methodology* (pp. 84-136). San Francisco: Jossey-Bass.


SES
AL71
AL67
Edu
SEI
An67
Pow67
Pow71
An71

Correlated residuals

- Since *Anomia* subscale and the *Powerlessness* subscale were measured twice (1967 and 1971), it is reasonable to expect that the measurement errors may be correlated.



Comparing the models with and without correlated residuals

- Results: $\chi^2(4, N = 932) = 4.74, p = .32$; CFI=1.00; TLI=1.00 and RMSEA=0.014. The model fits the data very well.
- Since these two models (with and without correlated errors) are nested, we can use the chi-square difference test to compare them: $\chi^2(2) = 66.81, p < .001$.

Latent growth modeling (LGM)

- Why do researchers want to conduct longitudinal studies?[6]
- To address intra-individual differences:
  - Similar to the within factors in repeated measures ANOVA;
  - Variation over time within individuals.
- To address inter-individual differences:
  - Similar to the between factors or covariates in repeated ANOVA;
  - Variation among individuals;
  - To draw casual inferences.

---

[6] Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*, 501-525.

Advantages of LGM to conventional repeated measures ANOVA

- Each participant has his/her growth curve.
- The number of occasions (incomplete data) can be different for different individuals.
- Time-varying (dynamic) and time-invariant (static) predictors can be handled. Repeated measures ANOVA cannot handle time-varying covariates.
- It can be extended to several levels, e.g., repeated measures of students who are nested within classes and schools.

Types of LGM

- **Unconditional LGM**:
  - There is no predictor.
  - We try to capture the growth patterns of the participants.
- **Conditional LGM**:
  - We try to explain *why* different participants may have different patterns of growth by using subject characteristics as predictors.

An example

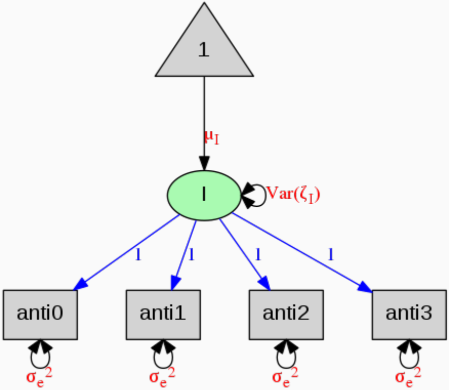
- The sample was drawn from Children of the National Longitudinal Survey of Youth (N=221).
- Time-varying variable:
  - Antisocial behavior: Anti0-Anti3 (time 0 to time 3)
- Time-invariant covariates:
  - Gender: 0: females and 1: males
  - Cog (cognitive support): continuous variable

Main research questions

- What are the growth patterns of antisocial behavior over time (intra-individual differences)?
- What predict the growth patterns (intercepts and slopes) of antisocial behavior over time (inter-individual differences)?

Baseline model

- The intrac class correlation (ICC) = $Var(\zeta_I)/(Var(\zeta_I) + Var(\sigma_I^2))$, which indicates the proportion of between-subject variation to the total variation.

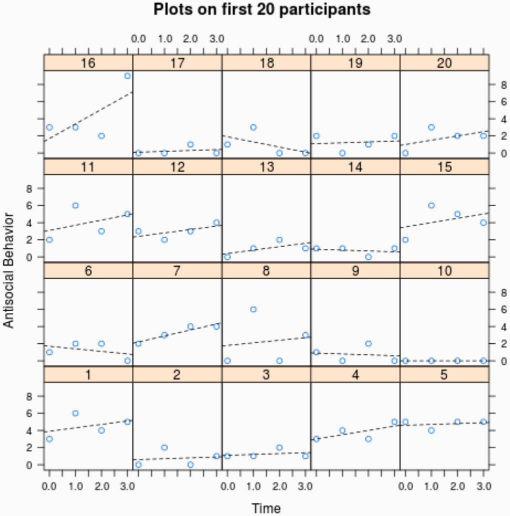


Results:

- $\chi^2(11, N = 221) = 56.416, p < .001$; CFI=0.819; TLI=0.901 and RMSEA=0.137. As expected, the baseline model does not fit the data well.
- ICC=1.579/(1.579+1.741)=.48.
- What if the baseline model fits the data well?

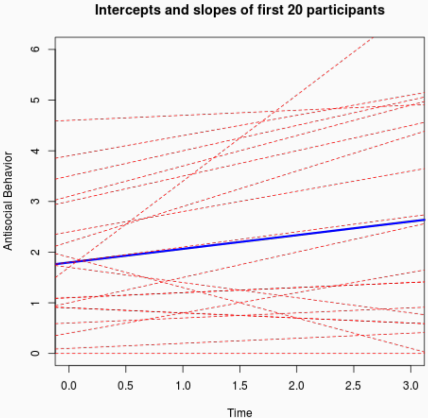
Some observations from the graphical plots (1)

- We may fit a straight line on each child.
- Each child has his/her regression line.



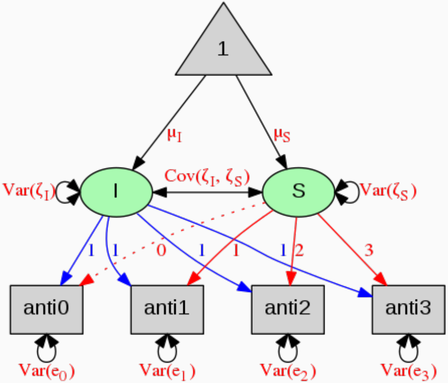
Some observations from the graphical plots (2)

- There are an average intercept and average slope (fixed effects).
- There is a variation on the intercepts and the slopes (random effects).



Linear growth model

- **Fixed effects**: *Average* intercept and *average* linear slope of growth.
- **Random effects**: *Variances* of the intercept and slope and their *covariance*.

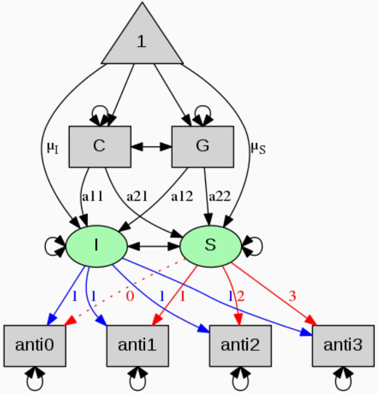


Results

- The linear growth model fits the data well with $\chi^2(df = 5) = 3.16, p = .68$, CFI=1.00, TLI=1.00, RMSEA=0.00 and SRMR=0.02
- The factor loadings are fixed. Thus, there is no estimate and standard error.
- **Fixed effects**:
  - $\hat{\mu}_I = 1.545$, p $<$ .01: the average intercept of antisocial behavior is 1.545.
  - $\hat{\mu_S} = 0.179$, p $<$ .01: the average slope of antisocial behavior is 0.179.
- **Random-effects**:
  - $Var(\hat{\zeta}_I) = 0.991$: the variation on intercepts across subjects.
  - $Var(\hat{\zeta}_S) = 0.10$: the variation on slopes across subjects.

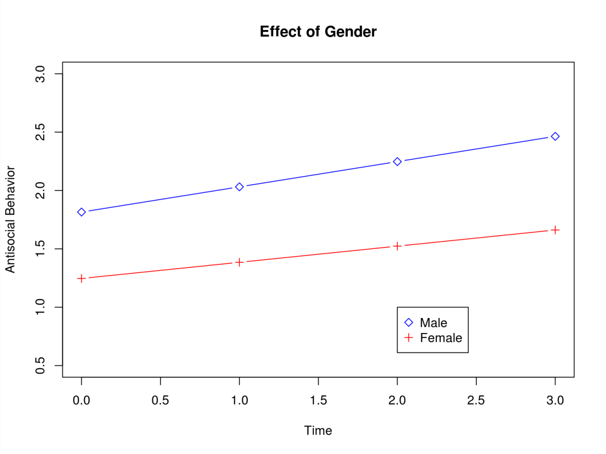
Conditional latent growth model with time-invariant predictors

- It is often of interest to predict why some individuals have larger intercepts or slopes by using cognitive support (C) and gender (G) as covariates.



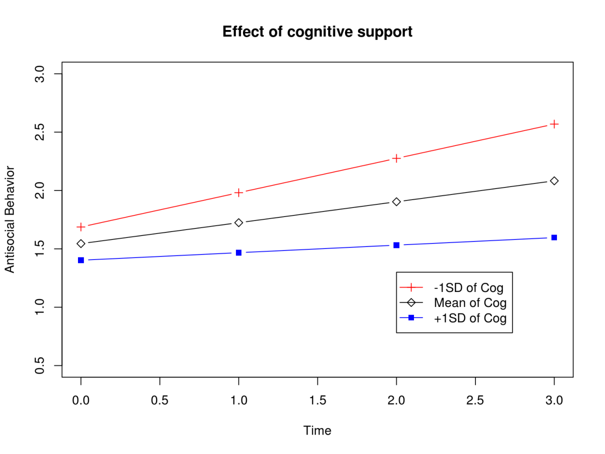
Results: Gender effect

- Males, in general, are more anti-social than females do.
- The growth trend, however, is the same.



Results: Cognitive support effect

- Initially, the level of anti-social behavior is the same.
- Children with less cognitive support from parents have a larger increase in anti-social behaviors.



Conclusion

- SEM is a powerful tool to address research questions in social and behavioral sciences.
- The findings in SEM provide evidence supporting the evidence-based practice/research.
- Other useful topics not discussed in this talk:
  - Handling missing data
  - Handling nonnormal data
  - Handling categorical data

Thank you for your attention!

- Any questions?
- My website: http://mikewlcheung.github.io/
- Source: http://dilbert.com/strip/2012-12-12