Estimating and Testing Latent Interactions: Advancements in Theories and Practical Applications

Herbert W. Marsh
Oxford University
Zhonglin Wen
South China Normal University
Hong Kong Examinations Authority
Benjamin Nagengast
Oxford University
Kit-Tai Hau
The Chinese University of Hong Kong

International Congress of Applied Psychology, Melbourne 2010

What is in the Book: Outline

- Multiple regression: between observed variables
- Latent-Variable Approaches
  - Multiple SEM
  - SEM interaction between latent variables
  - Main issues related to latent interaction model
- Strategies for creating product indicator
  - Types of product indicators
  - Strategies for matching indicators
- Parameter Constrained and Unconstrained approaches
  - Constrained approach
  - Partially constrained
  - Unconstrained approach

Outline (Cont)

- Centering of indicators and mean structure of latent models
  - Raw indicators
  - Mean-centered indicators
  - Orthogonalized product indicators
  - Double-mean-centered indicators
- Appropriate standard solutions and scale-free properties
  - Appropriate std solutions
  - Scale free propertie
  - Calculation of SE through Bootstrap samples
- An example with appropriate std solution
- Robustness to violations of normality assumption

Outline (cont)

- Distribution-analytic Approaches
  - Latent-moderated structural equation (LMS)
  - Quasi-Maximum Likelihood (QML)
  - Comparison to Product-Indicator Approaches
- Bayesian Method
- Summary
- Limitations and Directions for Future research
  - Quadratic Effects: confounding nonlinear and interaction
  - Test of measurement invariance
  - Multilevel Designs and Clustered Sample

Introduction

- Examples
  - Ed Psych: effects of an instructional technique interact with students' characteristics
  - Dev Psych: effects of a variable interact with age
  - Soc Psych: effects of individual characteristic depends on Group
  - Organizational Psych: employee characteristics × workplace characteristics
  - Moderator: variable affects direction and/or strength of relation between indep var and dep var, typically defined as $X_1 \times X_2$

Introduction
Traditional (nonlatent) Approaches

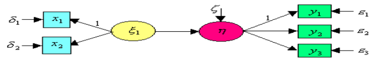
- Interaction between two manifest variables (X_1, X_2) on outcome (Y)
  - X_1, X_2 small number of categories: ANOVA
  - **X_1, X_2 cont., regression to estimate main and int'n**
    $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + e$$
  - Helpful to graph if interaction is significant
  - Empirical interactions typically small, non-sig, substantial measurement error reduces power of sig test
  - Latent interaction controls for measurement error, increase power, provide more defensible interpretation of interaction

Latent Variable Approaches

- **Two Broad categories**
  - **at least one variable involved is categorical with few categories (e.g., male/female) → multiple group SEM**
  - **both variables are continuous and latent → various approaches and best practice still evolving**

Latent Variable Approach
Multiple Group Analysis

- latent variable (ξ_1) × observed categorical variable (X_2) → latent variable (η)

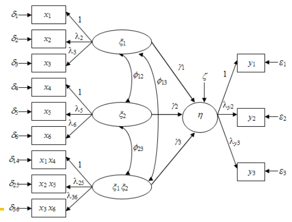


- X_2 small number of naturally existing categories, as grouping var
- test: invariance of ξ_1 → η effects over multiple groups; decline in goodness of fit with invariance constraint
- easily implemented in most SEM software
- problems: limitation in interpretation of the interaction, reduce power (small N), ignore measurement error categorizing var
- Not recommended, unless it is a true catergorial var with small number of categories with at least moderate sample sizes

Latent Variable Approaches
Full Latent (variable) Approach

- **Kenny & Judd (1984) proposed an ingenious heuristic model by constraining of loadings/variances of the product term**

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$$



Latent Variable Approach
Main Issues

- different ways to form the product indicator; How many product indicators? How to form best set ?
- many constraints on parameters make the method tedious /difficulty, are they absolutely necessary ?
- even if both ξ_1 ξ_2 have mean of zero, product term ξ_1 ξ_2 mean is not zero; mean structure complicates the application, is it really necessary?
- typical software do not provide appropriate SE for std effects, more serious with interaction model, how to obtain appropriate std solution?

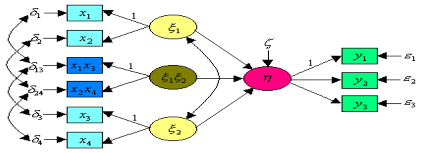
Strategies for Creating Product Indicators

- 2 guidelines
  - use all the information (all multiple indicators should be used in forming product indictors)
  - do NOT reuse information: each indicator used once in forming product indicators to avoid artificially created correlated residuals (variance/covariance matrix of errors becomes diagonal)
- Other possible strategies
  - Use the better indicators (when cannot use all indicators)
  - Use parcels (average of indicators) when there are too many indicators in a certain indep var

Parameter Constrained & UnconstrainedApproaches
– Constrained Approach

- Kenny & Judd (1984) proposed an ingenious heuristic model by constraining of loadings/variances of the product term

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$$



$x_1 = \xi_1 + \delta_1 \quad x_2 = \lambda_2\xi_1 + \delta_2 \quad x_3 = \xi_2 + \delta_3 \quad x_4 = \lambda_4\xi_2 + \delta_4$

Parameter Constrained & UnconstrainedApproaches
constrained approach (cont)

- Judd suggested using x_1x_3, x_1x_4, x_2x_3, x_2x_4 as indicators of the interaction $\xi_1\xi_2$ and imposed many constraints on loadings and variances, e.g.

  $x_2x_4 = \lambda_2\lambda_4\xi_1\xi_2 + \lambda_2\xi_1\delta_4 + \lambda_4\xi_2\delta_2 + \delta_2\delta_4$

  (i) loading of x_2x_4 on $\xi_1\xi_2$ (i.e., λ_{24}) constrained to be $\lambda_2\lambda_4$
- (ii) $\text{var}(\delta_{24}) = \lambda_2^2\,\text{var}(\xi_1)\,\text{var}(\delta_4) + \lambda_4^2\,\text{var}(\xi_2)\,\text{var}(\delta_2) + \text{var}(\delta_2)\,\text{var}(\delta_4)$
- Generally 2 constraints for each additional product indicator one for the loading, one for the measurement variance,
- **Specification of these constraints so tedious, prone to error, thus method seldom used in applied research**

Parameter Constrained & UnconstrainedApproaches
constrained approach (cont)

- Jöreskog & Yang (1996) proposed a general model. When observed var are not mean-centered, measurement eqn with intercept terms are used

  $x_1 = \tau_1 + \xi_1 + \delta_1 \quad x_2 = \tau_2 + \lambda_2\xi_1 + \delta_2$

  $x_3 = \tau_3 + \xi_2 + \delta_3 \quad x_4 = \tau_4 + \lambda_4\xi_2 + \delta_4$
- additional intercepts → (i) not only involve specification of mean structure, (ii) but also many nonlinear constraints (generally 5 constraints for one product indicator

Parameter Constrained & UnconstrainedApproaches
constrained approach (cont)

- Algina & Moulder (2001) revised Jöreskog-Yang model so that observed var are mean-centered as in Jaccard & Wan and a mean structure as in Joreskog-Yang
  - model was more likely to converge
  - even when all models converge, simulation results favor this revised model
  - Moulder & Algina (2002) compared 6 methods and concluded that their method was most effective with less bias, better control of Type I error rate, and higher power
  - thus we recommend this among all constrained approaches, and refer this as "constrained approach"

Parameter Constrained & Unconstrained Approaches
Partially Constrained

- the assumption in constrained approach that ξ_1, ξ_2, ζ and all measurement errors are normally with mean of zero, distributed is untenable, i.e., $\phi_{31} = \text{cov}(\xi_1\xi_2, \xi_1) = 0$ is typically false, so are $\phi_{32} = 0$ and $\phi_{33} = \phi_{11}\phi_{22} + \phi_{21}^2$
- applying constrained approach to non-normal data led to systematically biased estimates of interaction
- Wall & Amemiya (2001) proposed a generalized appended product indicator (partially constrained) procedure that did not constrain on Φ, but keeping all other constraints

  $$\Phi = \begin{bmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}$$

Parameter Constrained & Unconstrained Approaches
Partially Constrained (Cont)

- Advantage: relaxes the assumption that ξ_1, ξ_2 are normally distributed
- Disadvantages: specification of constraints still complicated,

Parameter Constrained & Unconstrained Approaches
Unconstrained

- Marsh, Wen, Hau (2002) evaluated an unconstrained approach
- **Similar to constrained approach: product of observed variables used to form indicators of latent term, however, without imposed complicated nonlinear constraints**
- Unconstrained model is identified when there are at least 2 product indicators

Parameter Constrained & Unconstrained Approaches
Unconstrained

- Marsh et al.(2002) simulation showed unconstrained approach::
  - Comparable goodness of fit, proportions of proper solutions, bias in estimation for first-order and interaction effects, precision as the partially constrained approach
  - Importantly, it is much easier to implement (no constrained needed)
  - However, when N is small, normality assumptions are met, the precision is somewhat lower than the constrained approach
- **Summary: unconstrained approach is recommended for its ease in implementation and acceptable bias /precision**

Centering of Indicators & Mean Structure of Latent Interaction Models

- Raw Indicators: Mean structures are always necessary for structural and measurement equations
- Mean-centering Indicators: centering x-indicators simplifies model considerably → intercepts terms of measurement eqn of x- & product-indicators no longer necessary; intercepts of measurement eqn of y necessary (even if y's are centered)
- Even if ξ_1, ξ_2 are, $E(\xi_1, \xi_2) = \mathrm{cov}(\xi_1, \xi_2)$ typically not zero, hence constant term of $\xi_1 \xi_2$ is necessary
- Thus, intercept for y-indicators, and mean structure for latent interaction model needed

Centering of Indicators & Mean Structure of Latent Int'n Models (cont)

- Orthogonalized Product Indicators –
  - alternative to mean-centering, orthogonalize interaction term by regressing on both ξ_1, ξ_2 :), regress x_1x_3 on x_1, x_2, x_3, x_4, and regress x_2x_4 on x_1, x_2, x_3, x_4,
  - treat these two residuals as indicators of the latent construct in a corresponding latent interaction model that does not require a mean structure
  - Cumbersome 2 steps procedure, non-random bias when (ξ_1, ξ_2) is not bivariate normal

Centering of Indicators & Mean Structure of Latent Int'n Models (cont)

- Double-Mean-Centered Indicators
  - Lin, Wen, Marsh, Lin proposed to double center the indicators, let the centered x-indicator be: $x_1^C = \xi_1 + \delta_1$
  - then matched product indicators $x_1^C x_3^C, x_2^C x_4^C$ are cenered again denoted by $(x_1^C x_3^C)^C, (x_2^C x_4^C)^C$

$$(x_2^C x_4^C)^C = x_2^C x_4^C - E\left(x_2^C x_4^C\right)$$

  - it can be shown when x, y mean-centered, product indicators double-mean-centered, mean structure is unnecessary
- **Summary: mean-center all x, y indicators, create product indicator, fit model without mean structure (because software routinely centers them again)**

An Appropriate Standardized Solution and Its Scale-free Properties (cont)

- appropriate std solution of interaction model not directly provided by usual commercial software
- **Wen, Marsh, Hau (2010) derived appropriate std solution for latent interaction, which are scale free, SE and t-values are also scale free**
- **Let usual std coefficients be γ_1', γ_2', γ_3', appropriate std coefficients γ_1'', γ_2'', γ_3'' are obtained:**

$$\gamma_1'' = \gamma_1' \quad \gamma_2'' = \gamma_2' \quad \gamma_3'' = \gamma_3' \frac{\sqrt{\phi_{11}\phi_{22}}}{\sqrt{\phi_{33}}}$$

where $\phi_{11} = \mathrm{var}(\xi_1)$ $\phi_{22} = \mathrm{var}(\xi_2)$ $\phi_{33} = \mathrm{var}(\xi_1\xi_2)$
are from the original solutions

An Appropriate Standardized Solution and Its Scale-free Properties (cont)

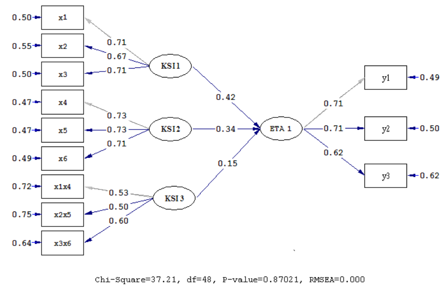
- Scale-free properties of std solution
  - Wen, Marsh, Hau (2010) proved that the appropriate std estimates have the scale-free properties → invariant when calcualted from either the centered or std data
- Calculation of SE of appropriate std coef through Bootstrap samples (similar to original estimates) → t-values of original estimates can be used to test the significance of the appropriate std estimates, if close to cutoff point use bootstrap method

Unconstrained Approaches: Examples

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3[\xi_1\xi_2 - E(\xi_1\xi_2)] + \zeta$$

- Each latent variable has 3 indicators
- Assume η is math achievement, ξ_1 is prior math ability, ξ_2 is math motivation, $\xi_1\xi_2$ is the interaction of prior math ability and math motivation
- y_1 to y_3, $x_1 \ldots x_6$ centered, product indicators x_1x_4, x_2x_5, x_3x_6 are created, but not re-standardized
- γ_1=0.425, γ_2=0.331, γ_3 =0.197; ϕ_{11}=0.501, ϕ_{22}=0.529, ϕ_{33} =0.308; and the completely standardized estimates: γ_1' =0.423, γ_2' =0.338 and γ_3' =0.153. By using Formula 27, γ_1'' =0.423, γ_2''=0.338, and γ_3'' = 0.142

Unconstrained Approaches: Examples (cont)



el

Unconstrained Approaches: Examples (cont)

- SE and t-values from bootstrap resampling of the original N=500, a total of 800 bootstrap samples generated from PRELIS 2.72, minor differences in t-values

  SY='bs0.psf'

  OU MA=CM RA=bs1.dat XM WI=11 ND=6 IX=111 BS=800 SF=100
- Significant interaction shows:
  - +ve effect of Math ability is more substantial for highly motivated students, or equivalently
  - +ve effect of math motivation is more substantial for students with higher level of prior ability

Robustness to Normality in Unconstrained Approach

- Considerations when normality is violated:
  - ML typically used is based on assumption of normality, however, this is a common problem to all CFA, SEM research (not specific to interaction/quadratic analyses)
  - even when ξ_1,ξ_2 are normal, the product are non-normal, constrained, partially constrained, unconstrained all suffer when ML estimation is used
  - **Fortunately, ML tends to be robust to violation of normality in parameter estimates, though ML likelihood ratio test is too large, standard errors are too small under nonnormality**

Robustness in relation to Violation of Normality (cont)

- In most conditions, ML still outperforms alternative estimators (e.g., Arbitrary distribution function, ADF; weighted least square) that do not assume normality
- Nevertheless, still appropriate to adjust/correct SE and χ^2
- Specific to constrained approach (but not to partially constrained & unconstrained), the constraints are set on the assumption that ξ_1 ,ξ_2 are normal, interaction estimates are not robust to violation of this assumption, size of bias does not decrease with increasing N; in contrast, both the partially constrained /unconstrained approaches provide relatively unbiased estimates under varying degree of nonnormality and this bias became smaller as N increased
- thus constrained approach not recommended

Distribution-analytic Approaches

- **Whereas they have many desirable features, they are computationally demanding, and not available in widely accessible SEM softwares**
  - **Latent Moderated Structural Equation (LMS, Klein & Moosbrugger, 2000) implemented in Mplus**
  - **Quasi-Maximum Likelihood(QML, Klein & Muthén, 2002) – available from author, not available in software yet**
- QML (Klein & Muthen, 2002) was developed for more efficient estimation than LMS
- Both estimate parameters in $\eta = \alpha + \Gamma\xi + \xi'\Omega\xi + \zeta$
- LMS and QML differ in the distributional assumptions about the latent dependent variable η and its indicators

Distribution-analytic Approaches (cont)

- LMS assumes ***x*** of the latent predictor, structural disturbance term ζ and all residuals in measurement model are assumed to be normal; can become computationally demanding and becomes unfeasible with a large number of nonlinear effects
- QML theoretically more robust against violations of normal distribution of indicators and residuals, but less efficient if distributional assumptions of LMS are fulfilled
- Computationally LMS is more efficient and can be used to fit models with a larger number of nonlinear effects and interactions

Distribution-analytic Approaches (cont)

- Comparison to product-indicator approaches
- In LMS/QML not necessary to construct product indicators as the product of latent variables, non-normal distribution of the latent outcome variables (and its indicators y) are modeled directly
- Product indicator approaches usually assume normality of latent variables and indicators, which are violated in models with latent interaction, distribution-analytic approaches maximize special fitting functions taking into account non-normality of indicators and dependent latent variable explicitly
- Due to lack of properly defined null model, no general fit statistics is provided (only nested models can be compared), unable to obtain appropriate std solutions

Bayesian Method

- Lee et al. (2007) developed a Bayesian approach fundamentally different from likelihood based approaches
- Assume all parameters are random and model their distribution conditionally on prior information and data
- Similar to distribution-analytic approaches, Bayesian models do not require product-indicators
- Good performance, especially in small samples, however, require sound statistical knowledge and careful thinking about the distributions of all model parameters and their priors

Summary

- **One of predictor variables is a manifest grouping variable with small number of categories → multiple group SEM, but not recommended when all predictors are continuous or based on multiple indicators**
- **Product-indicator dominated latent interaction research, still evolving, unconstrained approach – ease of implementation and robustness**
- **More recently, LMS/QML hold considerable promise over product-indicator approach**
- **Many issues not appropriately dealt with and applied research is limited**

Limitation and Directions for Future Research

- **Quadratic effects: special case of nonlinearity effect**
  - **Strong quadratic effect may appear as spurious interaction effect and hard to distinguish**
  - **Similar complicated issues arise when higher-order interaction involving more than two latent variables**
- **Tests of Measurement Invariance**
  - **Often ignored is the test of latent mean differences across multiple group without ensuring whether variables have same meaning in different groups (DIF): configural (pattern) , weak (loading), strong (intercept+loading), strict (+ unique) invariance, need at least strong invariance**

Limitation and Directions for Future Research (cont)

  - Marsh, Tracey, Craven (2006) proposed a hybrid approach using MIMIC (multiple-indicator-multiple-cause) and multiple group approach
- **Multilevel Design and Clustered Sample**
  - **Special type of interactions with data pointed related as clusters**
  - **Historically HLM tends to work with manifest variables, while SEM works with latent variables, inevitably, the integration will lead to more sophisticated analyses**

Thank You