STRUCTURAL EQUATION MODELING WITH STATA – TECHNICAL NOTE

Maintained by Tam Nguyen (tamn@email.arizona.edu)

This technical note is based on MISQ Editor's comments on SEM Guidelines (Gefen et al, 2011), practices from the latest published papers in MISQ and ISR (2022 and 2021 citations), and official Stata's manuals. This note focuses more on what to report in the case CB-SEM was used. For PLS-SEM, please refer to Gefen et al (2011) for a list of report items. We should note that the results of each phase will affect later phases.

A. Development of Measures

- Define constructs, observe variables, latent variables, unit of measures.
    - Independent variables (IV) and dependent variables (DV) must be carefully chosen based on the phenomenon of interest. We can perform studying of candidate variables to come up with legitimacy criteria for variable selection (Hsu et al, 2022).
    - Provide articulations for why a latent variable is reflective or formative based on rules like the ones suggested by Jarvis et al. (2003).
    - Provide literature references for each construct and decide if the study is exploratory or confirmatory.
- Scale development (Moore and Benbasat 1991) that may involve more than one round of:
    - Card sorting, expert panel's approval, face validity.
    - Pretesting of wording, the ease with which the questionnaire could be answered, the appropriateness of the question sequence, and consistency in meaning.
    - Pilot testing that involves checks for factorial validity, item-to-construct correlations (Straub et al. 2004), Cronbach's alpha (Nunnally and Bernstein 1994), and average variance extracted – AVE (Barclay et al. 1995). Item-to-construct score should be > 0.6. Cronbach's alpha should be >0.7, and AVE should be >0.5. Robustness check may be needed in certain context (Straub et al., 2004).
    - Must show a table of squared multiple correlation (SMC) or R-squared of constructs. Report significant differences in order of magnitude (if one item is .7 and another is .4) (Gefen et al, 2011).
    - Must include a correlation or covariance table (Gefen et al, 2011).
- Stata menu/commands

    - Statistics > Multivariate analysis > Cronbach's alpha / alpha varlist
    - Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Correlations and covariances
    - Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Pairwise correlations
    - Correlate and pwcorr commands
    - Example on Sagepub
    - Nevo et al. (2021) has a really good appendix on reliability and validity

B. Data Collection

- Choosing the right data collection method
    - Total design method vs Cost-compensation model vs Tailored Design [mail survey methods]
    - Online surveys
    - Carefully articulate the choice of targeted population
- Be transparent about missing data
    - Must explicitly acknowledge missing data
    - Discuss methods used in dealing with missing data (listwise deletion, EM with multiple imputation, maximum likelihood with missing data, etc)
    - Comparing demographics of respondents and nonrespondents for nonresponse bias check
- Identify and control for relationships that can compromise the study (Hsu et al, 20222). If there is a suspected relationship:
    - Divide dataset by a suspected relationship into subgroups (i.e. with/without the relationship)
    - Compare means of IV and DV using t-tests
    - Compare results of fixed effect models
- Independent observations and univariate/multivariate normality should be examined (Arbuckle 2011; Gefen et al. 2011). Data collection process should show that
    - Observations are independent (i.e survey takers can't discuss or exchange answers)
    - Observations meet univariate normality requirements as they are essential for the validity of t tests, F tests, and chi-squared tests. Hsu et al (2022) uses 7-point Likert and have normality requirements as means (3.654~4.733), standard deviations

(0.929~1.834), kurtosis (-0.821~0.052) and skewness (-0.465~0.353). Univariate normality can also be visually checked via (1) q-q plot, (2) histogram plot, or (3) Kolmogorov-Smirnov tests. We note that univariate normality of survey observations is never a perfect distribution.

    - Observations meet acceptable multivariate normality requirements as severe multivariate kurtosis skewness can affect global model fit tests and parameter standard error (Yuan et al. 2005). We can use Mardia's measure (1974) and Z-test.

- Perform nonresponse bias analysis based on Armstrong and Overton (1977) or Sivo et al. (2006)
    - Comparing early and late responses
    - Comparing demographics of response and nonresponse communities
- Common method bias tests
    - Method factor test (Zhou et al, 2021)
    - Market variable test (Zhou et al, 2021)
- Perform factor analysis (How many factors are there? Eigen values greater than 1?)
    - Perform bivariate correlation between indicators and respective constructs for nonsignificant indicators (Cenfetelli & Bassellier, 2009) to decide if nonsignificant indicators should be retained (Zhou et al, 2021).
- Checking for multicollinearity (link). Vif values should be less than 3.3 (Petter et al. 2007).
- Stata menu/commands
    - Basic stats commands
    - Statistics > Summaries, tables, and tests > Distributional plots and tests > Skewness and kurtosis normality test / sktest command
    - Statistics > Summaries, tables, and tests > Distributional plots and tests > Quantile-quantile plot / qqplot command
    - Graphics > Histogram / histogram command
    - Statistics > Nonparametric analysis > Tests of hypotheses > Kolmogorov–Smirnov test / ksmirnov command
    - Statistics > Multivariate analysis > MANOVA, multivariate regression, and related > Multivariate test of means, covariances, and normality / Command for Mardia's multivariate skewness test
    - Statistics > Summaries, tables, and tests > Classical tests of hypotheses > z test (mean-comparison test, known variance / ztest command

- Statistics > Multivariate analysis > Factor and principal component analysis > Factor analysis / Official Stata guide / Quick guide

C. SEM Designing

- Choosing CB or PLS
  - CB-SEM is preferred over PLS-SEM for confirmatory research objectives while PLS may be better for exploratory research objectives (Gefen et al, 2011)
  - CB-SEM may not be a good choice if there are many formative measurements (Gefen et al, 2011). For example, Nevo et al. (2021) used PLS-SEM because the data are not normally distributed and their model has significant formulative latent constructs.
  - PLS is not as robust as CB against bias in estimations (Gefen et al, 2011)
  - Obsolete reasons for choosing CB-SEM: modeling interactions/moderation, distribution assumptions (Gefen et al, 2011)
- Choosing the right method for CB-SEM with the following recommendations from Gefen et al (2011)
  - Maximum likelihood (ML) is the most common
  - Generalize least squares (GLS) may produce counter-intuitive results in highly constrained models
  - Weighted least squares (WLS) only reaches stability at very large sample sizes
- Choosing to do CFA first and then SEM or CFA and SEM at the same time (Gerbing and Anderson 1988)
- Choosing SEM or gSEM (link) – only gSEM supports multilevel models.
- Make sure the realized sample size is sufficient for SEM analysis (Gefen et al., 2011; Wolf et al., 2013). It is around 10 to 20 samples per parameter.
  - For PLS, sample size should be at least 10 times the largest number of predictors for any dependent variable
  - For CB-SEM, there is no specific guidance. It often has a minimum of 200 observations for a moderately complicated structural equation model

D. Sem Analysis

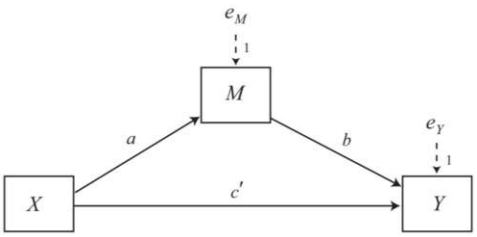
- Include AVE, squared correlations, beta values, p values, t-statistics for each latent construct. Determine if statistics fit the construct type - reflective or formative per - Petter et al. (2007)

- Perform endogeneity check - the Durbin–Wu–Hausman method (Davidson and MacKinnon, 1993; Dong et al.,2017) to make sure exogenous variables are indeed exogenous. What are impacts of control variables on exogenous variables?
- Must include tables of fit indexes as a sign of due diligence. There should be tables before and after each significant modification to observations/model design (Gefen et al, 2011).
    - X^2 , degree of freedom (df), and X^2/df ratio (should be <3).
    - Standardized root mean square residual (SRMR) is the the unweighted average of the squared fitted residuals (Jöreskog and Sörbom 2001) and should be less than 0.05.
    - Root mean square error of approximation (RMSEA) is an estimate of lack of fit per degree of freedom. Good fit has RMSEA <.05, acceptable fit has RMSEA <.08 (Browne and Cudeck 1993).
    - Tucker-Lewis Index (TLI), the comparative fit index (CFI), and relative noncentrality index (RNI) describe fitness relative to a null or baseline model and should be above .90 (Marsh et al. 2004)
    - It is acceptable for the indices to not meet all the rule-of-thumb thresholds (Boudreau et al. 2001)
- Perform post-sem confirmatory factor analysis (CFA)
- Leverage combinatorial rules involving different fit indices from Hu and Bentler's (1998, 1999) because there are biases in fit statistics (Gefen et al, 2011) such as:
    - X^2 fitness is only usable in certain distributions
    - When sample size is large and/or df is large compared to sample size, biases will happen with X^2/df, goodness of fit index (GFI), and Ajusted GFI (AGFI).
- Second-order formative construct validity test: Nevo et al (2021) measured the weights of the first-order constructs (beta >0.5, $p < 0.01$)
- Compare theoretical model with saturated model.
    - Build saturated model (aka measurement model - models with all paths among all pairs of latent variables) and report statistics. Make sure the model has reliability, content validity, convergent validity (AVE), and discriminant validity (AVE).
    - Build theoretical model (aka structural model) and report statistics. Make sure the model has reliability, content validity, convergent validity (AVE), and discriminant validity (AVE).

    - We can join certain pair of latent variables and test for discriminant validity by monitor changes in X^2. Significant discriminant validity is reflected through unsignificant changes of X^2.
- Perform robustness check by computing model statistic with:
    - Averaged data (from study's collected pool)
    - Randomly sampled data (from study's collected pool)
    - Vce (robust) option (assuming there are heterogeneity and autocorrelation)
    - Competing model(s)
- Post-SEM analyses:
    - Chen et al (2021) performed post-SEM ANOVA
- Stata menu/commands
    - Official Stata manual on post-estimation for SEM (link)
    - Postestimation statistics for survey data (link)
    - Post estimation variance inflation factors for the independent variables – estat vif
    - Statistics > Postestimation > Reports and statistics / goodness of fit manual /
    - Statistics > SEM (structural equation modeling) > Goodness of fit > Overall goodness of fit / command
    - Confirmatory Factor Analysis (link)
    - Test of endogeneity – estat endogenous command
    - Stata vcs (robust) option (link)
    - Postestimation ANOVA and related tests (link)
    - Correction for measurement errors with Stata (link)
    - Additional resource on postestimation commands and regression (link)

E. Mediation Test

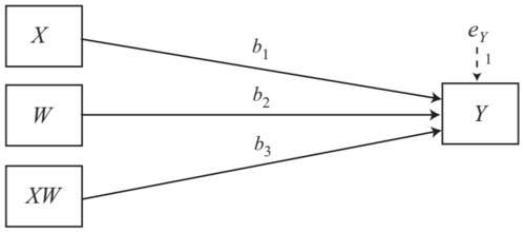
- Mediation is a causal explanation where M is causally located between X and Y. X causes M which in turn causes Y.



- Provide a table of paths, total effect, direct effect, indirect effect, and conclusion on Mediation (Nevo et al, 2021)
- Provide tables of ab test, c' test (Chen et al, 2021)
- Provide table of Ordinary Least Squares (OLS) results (Zhou et al, 2021)
- Indirect paths: simple mediation (one), parallel mediation (multiple)
    - One mediator: multiplication of the path coefficients (Nevo et al, 2021)
    - Multiple mediators: sum of effects across all mediators (Nevo et al, 2021)
    - Sobel test (Zhou et al, 2021)
- The bias corrected confidence interval (CI)
    - CIs were constructed using bootstrapping (Hayes, 2022)
    - Estimating indirect effects of each mediator while accounting for covariance between mediators
- Perform hierarchical regression analysis (Zhou et al, 2021)
- Stata menu/commands
    - Bootstrap sampling and estimation (link)

Moderation Test

- "The effect of X on some variable Y is moderated by W if its size, sign, or strength depends on or can be predicted by W. In that case, W is said to be a moderator of X's effect on Y, or that W and X interact in their influence on Y."
  (Hayes,2022)



- a

References

Arbuckle JL (2011) IBM SPSS Amos 20 User's Guide (SPSS Inc, Chicago, IL)

Armstrong, J .S., and Overton, T. S. 1977. " Estimating Non-Response Bias in Mail Surveys," Journal of Marketing Research (14), pp. 396-402.

Boudreau, M., Gefen, D., and Straub, D. W. 2001. "Validation in IS Research: A State-of-the-Art Assessment," MIS Quarterly (25:1), pp.1-16.

Browne, M. W., and Cudeck, R. 1993. "Alternative Ways of Assessing Model Fit," in Testing Structural Equation Models, K. A. Bollen and J. S. Long (eds.), Newbury Park, CA: Sage Publications.

Chen, Y., Galletta, D. F., Lowry, P. B., Luo, X., Moody, G. D., & Willison, R. (2021). Understanding inconsistent employee compliance with information security policies through the lens of the extended parallel process model. Information Systems Research, 32(3), 1043-1065.

Gefen D, Straub DW, Rigdon EE (2011) An update and extension to SEM guidelines for administrative and social science research. MIS Quart. 35(2):3-14.

Gerbing, D. W., and Anderson, J. C. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment," Journal of Marketing Research (25:May), pp. 186-192.

Hsu, C., Lee, J. N., Fang, Y., Straub, D. W., Su, N., & Ryu, H. S. (2022). The role of vendor legitimacy in IT outsourcing performance: theory and evidence. Information Systems Research, 33(1), 337-361.

Hu, L.-T., and Bentler, P. M. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification," Psychological Bulletin (3), pp. 424-453.

Hu, L.-T., and Bentler, P. M. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structural Analysis: Conventional Criteria Versus New Alternatives," Structured Equation Modeling (6:1), pp. 1-55.

Jarvis CB, Marchenzie SB, Podsakoff PM (2003) A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J. Consumer Res. 30(2):199-218

Jöreskog, K. G., and Sörbom, D. 2001. LISREL 8: User's Reference Guide, Chicago: Scientific Software International.

Mardia KV (1974) Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhya: The Indian J. of Statistics 36(Series B):115–128.

Marsh, H. W., Hau, K.-T., and Wen, Z. 2004. "In Search of the Golder Rules: Comment on the Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings," Structured Equation Modeling (11:3), pp. 320-341.

Moore G, Benbasat I (1991) Development of an instrument to measure the perceptions of adopting an information technology innovation. Inform. Systems Res. 2(3):192-222.

Nevo, S., Nevo, D., & Pinsonneault, A. (2021). Personal Achievement Goals, Learning Strategies, and Perceived IT Affordances. Information Systems Research, 32(4), 1298-1322.

Nunnally JC, Bernstein IH (1994) Psychometric Theory, 3rd ed. (McGraw-Hill, NY)

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. MIS quarterly, 623-656.

Sivo, S. A., Saunders, C., Chang, Q., and Jiang, J. J. 2006. "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research," Journal of the AIS (7:6), pp. 361-414.

Straub DW, Boudreau MC, Gefen D (2004) Validation guidelines for IS positivist research. Commun. AIS, 13(24):380-427.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. Educational and Psychological Measurement, 73(6), 913-934. doi: 10.1177/0013164413495237

Yuan KH, Bentler PM, Zhang W (2005) The effect of skewness and kurtosis on mean and covariance structure analysis. Sociol. Methods and Res. 34(2):240−258.

Zhou, J., Fang, Y., & Grover, V. (2020). Managing Collective Enterprise Information Systems Compliance–A Social and Performance Management Context Perspective. MIS Quarterly.