Published in final edited form as:

Multivariate Behav Res. 2013 January ; 48(1): 153–154. doi:10.1080/00273171.2012.748359.

Abstract: Applying Maximum Likelihood Estimation and Multiple Imputation to Moderated Regression Models With Incomplete Predictor Variables

Heining Cham, Amanda N. Baraldi, and Craig K. Enders

Department of Psychology, Arizona State University

Although maximum likelihood (ML) estimation and multiple imputation (MI) suit many common statistical analyses, their applications to moderated regression models with incomplete predictors are not straightforward.

For ML, we follow the logic of the unconstrained product indicator approach by Marsh, Wen, and Hau (2004, 2006) to develop a latent variable interaction model that defines each predictor as the sole manifest indicator of a latent construct (i.e., pseudo-outcome). We outline parameter constraints that yield estimates with interpretations equivalent to those from a centered solution and simple slopes in the ordinary least squares framework (Aiken & West, 1991). Figure 1 shows a path diagram of the moderated regression model with two predictors.

For MI, we outline the following procedure: (a) form a product term by multiplying predictors in raw scores, (b) impute all variables on their raw score metrics, and (c) rescale the predictor variables *and* the product term following imputation. We developed equations to rescale the imputed raw score product term to deviation score format following imputation. These rescaling equations, which are equivalent to the parameter constraints in ML, again, yield estimates with interpretations equivalent to those from a centered solution and simple slopes in the ordinary least squares framework.

We used Monte Carlo simulations to evaluate the performance of the ML and MI approaches. The simulations used a 2-predictor moderated regression model and manipulated the following factors: predictors distributions (normal, kurtotic, skewed and kurtotic), moderation effect sizes (small vs. large R^2 increment), sample sizes (N 's 50–5,000), and missing data mechanisms (missing completely at random [MCAR]; missing at random [MAR]). Under MCAR, parameter estimates exhibited very little bias, with standardized bias values ranging between 1% and 6% of a standard error unit. Under MAR, estimates generally exhibited some bias, but most values were below 40% of a standard error unit.

Acknowledgments

This abstract was supported in part by the National Institute on Drug Abuse Grant DA09757. Heining Cham expresses his thanks and appreciation to his SMEP sponsors, Craig Enders and Stephen West.

References

- Aiken, L.S.; West, S.G. Multiple regression: Testing and interpreting interactions. Newbury, CA: Sage; 1991.
- Marsh HW, Wen Z, Hau KT. Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*. 2004; 9:275–300. [PubMed: 15355150]
- Marsh, HW.; Wen, Z.; Hau, KT. Structural equation models of latent interaction and quadratic effects. In: Hancock, G.; Mueller, RO., editors. *Structural equation modeling: A second course*. Greenwich, CT: Information Age; 2006. p. 225-265.

FIGURE 1. Latent variable interaction model. The triangles denote a unit vector, and the associated directional arrows are regression intercepts. Direct factor loadings are constrained to one, and the error variances of X , Z , and XZ are constrained to zero.