

Multiple imputation of incomplete non-linear terms, when sample size is extremely small

Friday 25 July 2025 09:45 (15 minutes)

Oral presentation

Multiple imputation of incomplete non-linear terms, when sample size is extremely small

Author

Kristian Kleinke

Affiliation

University of Siegen, Department of Psychology

Abstract

Handling incomplete non-linear terms in regression models like interaction terms (to test for moderating effects) is not trivial (Kleinke, 2021). One state-of-the-art approach to do so is factorized regression (e.g. Keller, 2022), also known under the name “substantive model compatible multiple imputation”. Imputations are generated under a model, which is fully compatible to the analyst’s model and to the assumed data generating process. The approach is implemented in various software packages including the Blimp software (Keller & Enders, 2023). The focus of the present study is to evaluate the factorized regression approach implemented in Blimp in a variety of scenarios where sample size was extremely small, ranging from 20 to 200 participants. I evaluated bias in point estimates and in measures of uncertainty of the incomplete non-linear effect in a 10 (sample size) by 11 (strength of the moderating effect) by 2 (missing data percentage) factorial experiment. Results show that point estimates were usually unbiased unless sample size was very small. Confidence interval (CI) widths increased with decreasing sample size. CI widths also increased with increasing missing data percentage, reflecting the increasing estimation uncertainty due to missing data. Especially for smaller samples and for smaller effects, this could mean power problems. Confidence interval coverage was widely accurate and centered around the nominal level. For very small samples, some over-coverage was observed, indicating that the obtained standard errors were too conservative (i.e. too large), which could result in type-II errors. If possible, applied researchers should try to obtain sample sizes of at least around 50, when the interaction effect of interest involves two continuous variables.

Keller, B. T. (2022). An introduction to factored regression models with Blimp. *Psych*, 4(1), 10-37. <https://doi.org/10.3390/psych4010002>

Keller, B. T., & Enders, C. K. (2023). Blimp user’s guide (Version 3). Retrieved from www.appliedmissingdata.com/blimp

Kleinke, K. (2021). Estimation of partially observed non-linear terms in a multilevel model: An evaluation of the robustness of ad hoc and state-of-the-art missing data methods. *Psychological Test and Assessment Modeling*, 63(3), 432–455.

Keywords

multiple imputation, factorized regression

Primary author: KLEINKE, Kristian (Universität Siegen)

Session Classification: Session 5 : “Bayesian methods in Psychology and Statistics”

Track Classification: Statistical analyses: Statistical analyses