Multiple imputation of incomplete panel data based on a piecewise growth curve model

Kristian Kleinke[1)] & Jost Reinecke[2)]

[1)]*University of Siegen*

[2)]*Bielefeld University*

Keywords: missing data, multiple imputation, zero-inflation, piecewise growth curve models

Missing data are a nuisance. Complex panel data can be affected by various 'sources' of nonresponse including missingness in the target variables, in predictors at level-1, level-2, or higher hierarchical levels, or even missingness in the cluster identifiers. Depending on the type and pattern of missing data, unobserved information poses a threat to the validity of statistical inferences. Modern imputation approaches try to recreate the sample in a way as if ideally no information had gone missing. To this end, the model that is used to create the imputations usually needs to reflect the assumed (and typically unknown) true data generating process, and if necessary, the mechanism that created the missing data patterns, well. If the imputation model is somehow misspecified, bias is to be expected. We focus on complex (nonlinear) relationships of the target variable over time. The purpose of this paper is to elucidate how the choice of the imputation method and model affects substantive model results. We present results of a Monte Carlo Simulation based on empirical data from 12 waves of the Crime in the modern City (CrimoC) study (www.crimoc.org), focussing on the development of juvenile delinquency. The data show the typical age-crime-relationship with an increase in delinquent activities early on in adolescence and a decrease later on. Data were imputed based on a piecewise growth curve model, and based on proxy methods with either a close fit (a growth curve model with a linear and quadratic time trend), or based on relatively robust allround methods such as semi-parametric predictive mean matching. Substantive model results will be discussed in terms of bias in point estimates and standard errors regarding the respective substantive model of interest, regarding ease of use and, and in terms of computing time.