"Estimating a contextual effects model given error-prone measures of contextual variables and missing data"

Yongyun Shin

In a three-level regression model where occasions are nested within children attending schools, the association between a time-varying (level-1) covariate and the outcome may be different at different levels to yield a contextual effects model. The conventional analysis is to use the covariate and its aggregated sample means as regressors. The effects of the sample means controlling for the covariate are defined as the contextual effects at the individual and school levels (levels 2 and 3). This approach, however, produces biased contextual effects unless sample sizes are large at levels 1 and 2. In addition, the biased contextual effects introduce bias in the variance estimates and the effects of other level-2 and level-3 covariates that are correlated with the covariate. Furthermore, the outcome as well as covariates may be subject to missingness at any of the levels to make efficient estimation challenging. In this paper, the author expresses the desired contextual effects model given the latent population group means of the covariate having a contextual effect instead of the unreliable sample means, shows the bias terms in the conventional approach, and introduces an estimation method for the desired model that corrects the bias as well as efficiently handles missing data at any of the levels under an assumption of ignorable missing data. The key idea is to reexpress the desired model as the joint distribution of variables, including the outcome and the covariates having contextual effects, that are subject to missingness given all of the covariates that are completely observed; efficiently estimate the joint model by maximum likelihood; produce multiple imputation of completed data including the latent population means given the estimated joint model; and analyze the desired model given the completed data. The multiple imputation facilitates exploration of multiple contextual effects models for model selection. The approach will be illustrated with a study of youth outcomes analyzing a national sample of children.