

Jan De Neve

In several applications the outcome of interest is not measured directly, but instead a proxy is used. Examples include the body mass index as a proxy for the body fat percentage, the fluorescence intensity as a proxy for gene expression or blood pressure as a proxy for stress. The relationship between the outcome of interest and the observed proxy can be complex, but is typically assumed to be monotone. Several methods, however, assume that this relationship is linear (e.g. structural equation models). We illustrate via simulations how slight deviations from linearity can have a substantial impact on the validity of the statistical inference of the methods. We further present a semiparametric estimation strategy to quantify the effect of observed covariates on a summary measure of the unobserved outcome. We consider the probabilistic index as a summary measure, i.e. the probability that the outcome of one subject exceeds the outcome of another subject, conditional on covariates. Since this effect size is invariant under monotone transformations, we do not need to model the relationship between the unobserved outcome and the proxy. By considering this relationship as nuisance, the estimation strategy can be applied to settings where the outcome of interest cannot be observed (e.g. a latent cognitive process) and hence only weak assumptions (such as monotonicity) about the relationship can be imposed.