Date: 24.06.2019 (Version 1.0.0)

This is the online material for the publication (currently under review)

Analyzing Nonresponse in Longitudinal Surveys Using Bayesian Additive Regression Trees:

A Nonparametric Event History Analysis

By Sabine Zinn1\* & Timo Gnambs2,3\*

1 German Institute for Economic Research

2 Leibniz Institute for Educational Trajectories

3 Johannes Kepler University Linz

Abstract

Increasing nonresponse rates are a pressing issue for many longitudinal panel studies. Respondents frequently either refuse participation in single survey waves (temporary dropout) or discontinue participation altogether (permanent dropout). Contemporary statistical methods that are used to elucidate predictors of survey nonresponse are typically limited to small variable sets and ignore complex interaction patterns. The innovative approach of Bayesian additive regression trees (BART) is an elegant way to overcome these limitations. We present a BART event history analyses that allow identifying predictors for different types of nonresponse to anticipate response rates for upcoming survey waves. We apply our novel method to data from a German large-scale assessment including *N* = 4,559 students in grade 5 that observed nonresponse rates of up to 36% across five waves. The results highlight the potential of Bayesian discrete time event modeling for the prediction of participation rates across multiple survey waves. Finally, potential applications of this approach for operational use in survey management are outlined.

We use data from the National Educational Panel Study (NEPS): Starting Cohort Grade 5, doi:10.5157/NEPS:SC3:8.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Data access requires the conclusion of a Data Use Agreement with the Leibniz Institute for Educational Trajectories (LIfBi), see <https://www.neps-data.de/en-us/datacenter/dataaccess/datauseagreements.aspx>.

This GitHub project contains the source code for data preparation and analysis (with BART), as well as for plotting the results.

1. loadData.R (Editing and preparing data using the NEPS SUF files.)
2. modelIndF.R (BART model for passing over to permanent drop out, in our case this means leaving the NEPS school context.)
3. modelParticipation.R (BART model for temporary dropout in the school context.)
4. combineResults.R (Predicting participation probabilities in the distinct survey waves for new samples)