LIfBi WORKING PAPERS

Hans Walter Steinhauer and Christian Aßmann

MODELLING NONRESPONSE IN EDUCATONAL MULT-INFORMANT STUDIES: A MULTILEVEL APPROACH USING BIVARIATE PROBIT MODELS

NEPS Working Paper No. 74
Bamberg, June 2018

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

Working Papers of the Leibniz Institute for Educational Trajectories (LIfBi)
at the University of Bamberg

The LIfBi Working Papers publish articles, expertises, and findings related to data collected and studies conducted at the Leibniz Institute for Educational Trajectories, first and foremost to the German National Educational Panel Study (NEPS).

The LIfBi Working Papers are edited by a board of researchers representing the wide range of disciplines covered by NEPS and other LIfBi studies. The series started in 2011 under the name “NEPS working papers” and was renamed in 2017 to broaden the range of studies which may be published there.

Papers appear in this series as work in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

Any opinions expressed in this series are those of the author(s) and not those of the LIfBi management or the NEPS Consortium.

The LIfBi Working Papers are available at www.lifbi.de (see section “Institute > Publications”). LIfBi Working Papers based on NEPS data are also available at www.neps-data.de (see section “Data Center > Publications”).

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg − Germany − contact@lifbi.de

Modelling Nonresponse in Educational Multi-informant Studies: A Multilevel Approach Using Bivariate Probit Models

Steinhauer, H. W.
Leibniz Institute for Educational Trajectories

Aßmann, C.
Chair of Statistics and Econometrics, University of Bamberg
Leibniz Institute for Educational Trajectories

E-mail address of lead author:

methoden@lifbi.de

Bibliographic data:

Steinhauer, H. W. & Aßmann, C. (2018). *Modelling nonresponse in educational multi-informant studies: A multilevel approach using bivariate probit models* (LIfBi Working Paper No. 74). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Modelling Nonresponse in Educational Multi-informant Studies: A Multilevel Approach Using Bivariate Probit Models

Abstract

Large-scale educational surveys often use a multi-informant survey approach. Adapting such an approach, the National Educational Panel Study enriches students' tests and survey data with information obtained within a parent telephone interview. Both, students and parents may refuse to participate. We adapt a bivariate probit model with multilevel structure allowing for clustering at the school level to model the participation process. Using simulated maximum likelihood estimation, the empirical results point at significance of explaining factors like school type and family background. Specification tests highlight the importance of considering correlation as well as clustering structures when modelling participation processes.

Keywords
bivariate binary probit, multi-informant survey, nonresponse, multilevel modelling, simulated maximum likelihood, weighting adjustments

1. Introduction

As nonmandatory large-scale educational surveys such as the National Educational Panel Study (NEPS) are facing nonresponse, weighting adjustment procedures are commonly used to correct for unit nonresponse, see Kalton and Kasprzyk (1986), Kalton and Flores-Cervantes (2003) and Brick (2013) among others. Consideration of multiple participation processes is common for large-scale surveys. In household surveys it is also common to address the difference between noncontact and refusal, see Groves (1998), Durrant and Steele (2009), and Steele and Durrant (2011) among others. Distinguishing noncontactability from refusal makes it possible to consider differences in characteristics determining the two components of nonresponse. To analyse the determinants of these response processes, typically sequential univariate models, bivariate sample selection models, multinomial models, and their extensions to multilevel models are used, see O'Muircheartaigh and Campanelli (1999), Durrant and Steele (2009), or Steele and Durrant (2011). Besides individual characteristics, household composition, social environment, or survey design features, these models are enriched with paradata, see for example Couper (1998) and Groves and Heeringa (2006). O'Muircheartaigh and Campanelli (1999) model response processes by incorporating interviewer characteristics, whereas Wood, White, and Hotopf (2006) consider the number of (failed) contact attempts. Micklewright, Schnepf, and Skinner (2010), Pike (2008), and Porter and Whitcomb (2005) analyse nonresponse within the educational context of (higher) education institutions and find gender, school, or academic performance, as well as financial support to be factors affecting the participation decisions of students.

In contrast to household surveys, where respondents can directly provide consent to participate, the participation process in the NEPS, as in other educational surveys such as PISA, involves several actors at multiple stages. Apart from institutional heads giving consent to participate, also consent has to be provided by students and, if underage, by their parents. When employing a multi-informant perspective to enrich data on students via a computer assisted telephone interview (CATI) with one parent, decision processes of students and parents have to be considered. As discussed by Lynn and Kaminska (2010), weights for subgroups of special interest in analysis have to be made available, for example, for students and parents with joint participation in all waves. Given the time structure of these decision processes, where students and parents are asked for participation conditional on the consent of institutional heads, sequential modelling can be used. Further, because an educational survey takes place within schools, clustering at the school level needs to be taken into account. As pointed out by Skinner and D'Arrigo (2011), nonresponse is commonly correlated within clusters because the access to the sampled targets depends on authorities at the cluster level. This is the case for educational surveys typically established via multi-stage sampling designs. Skinner and D'Arrigo (2011) hence suggest using multilevel models to account for clustered nonresponse by adapting random effects. In addition, the decisions of parents and students to participate are likely to be correlated. Therefore, we extend the bivariate probit model respecting the multi-informant perspective by a multilevel structure that accounts for clustering at the school level.

When adjusting weights for the subgroup of students and parents who participate in the panel together, the bivariate setting takes into account the correlation in the participation process of students and parents. Next, the cluster structure of students nested in schools needs to be considered. The cluster structure of students in schools is also present for parents. Within the NEPS, this joint decision process of students and their parents can be found in three out of the six established cohort samples. First, children and parents of the Kindergarten cohort participate together in the survey. Second, students and their parents can participate together in the cohort of secondary school students in grade 5 and in grade 9. In all three cohorts children and their parents are grouped at institutional level. Therefore, the extension of the bivariate probit model respecting this cluster structure is essential for modelling response propensities used in weighting adjustments. This approach can be adopted as long as children and students are grouped in their institutions over time.

In order to provide nonresponse adjusted weights for the subgroup of participating students and parents, a bivariate probit with random intercept allowing for clustering at the institutional (school) level is estimated using a simulated maximum likelihood approach built upon the importance sampler (GHK simulator) documented in Geweke and Keane (2001). To model the response processes, we employ three types of variables. The first type of variables includes sampling characteristics such as stratification variables. The second set of variables consists of characteristics describing sociodemographics, for example, gender or migration background. Finally, the third set of variables involves paradata from call records.

We illustrate the suggested modelling approach for the cohort sample of students in grade 5. The empirical results indicate the importance of jointly considering participation decisions of students and parents. Moreover, it is also important to take into account clustering at the school level in order to correctly gauge the significance of considered determinants such as stratification variables. Variables influencing participation decisions throughout the waves are the students' native language, the number of calls in the parent interview before the first contact was made, as well as the missing indicator for personal characteristics. Besides these, the participation status in wave 1 is a strong predictor in modelling participation propensity for wave 2.

The paper proceeds as follows. Section 2 provides the conceptual frameworks for modelling nonresponse, refusal, and participation. Further, the sequential decision processes involved are described. Section 3 discusses model specifications that were considered in decision modelling and estimation thereof. Section 4 provides an empirical application, describes the data set, and discusses empirical results. Section 5 concludes.

2. Determinants of Multi-stage Participation Processes in Educational Surveys

The sequential decision processes has to reflect the multiple stages used to establish the corresponding cohort sample along the time line. Because the NEPS, see Blossfeld, Roßbach, and von Maurice (2011), is interested in providing data on various aspects of competence development, educational decisions, migration background, returns to education, and especially on learning environments, educational institutions are, whenever possible, used to access the target population. With regard to grade 5 students forming one of the six starting cohorts of the NEPS, access to this particular target population is gained via primary or secondary schools in Germany in the school year of 2010/2011, see Aßmann et al. (2011). Within this cohort survey, a focus is set on the development of students' competencies as well as conditions and prerequisites of educational processes. Surveyed information encompasses information on classes, class composition, school equipment, and information obtained from school teachers and principals, see Frahm et al. (2011) for further details. Additionally, the survey adapts a multi-informant perspective and therefore conducts telephone interviews with the students' parents. This interview is used to validate information provided by the student, as well as to receive information on the student's home learning environment, see Bäumer, Preis, Roßbach, Stecher, and Klieme (2011).

The participation processes resulting in the final panel cohort are embedded in the sampling and recruitment process as follows. The panel cohort sample has been established using a stratified two-stage cluster sampling approach. Stratification reflects the different school systems in Germany via seven explicit strata, see Aßmann et al. (2011) for details. Access to the target cohort, that is, students in grade 5, is gained via schools, thus ensuring the contactability of students in sampled classes. In NEPS, participation is not mandatory and, therefore, unit nonresponse can occur at each level, that is, schools, students, and parents. The process of school recruitment and, subsequently, the student recruitment process is consecutive by nature and thus reflected by sequential modelling. That is, first, school nonresponse is modelled. Second, the panel cohort sample is established on the basis of the active consent to participate in the panel provided by parents, because a fifth-grade student is not of legal age. After correcting for unit nonresponse at the school and student level, each student of the panel cohort is assigned an adjusted design weight. Thereby, nonresponse adjustments on the institutional level take into account sampling information as well as information from the recruitment process. On the individual level, we consider clustering at the school level by specifying random intercepts to account for correlation within schools. Because the students need their parents' permission granted by one parent's signature to participate in the NEPS, a first contact with the parents is already established in the run-up to the survey. The provided consent to participate in the panel survey establishes the panel cohort sample of students. Third, in view of the panel consent being granted by parents on behalf of their children, actual participation in the panel surveys, including testing of students in schools and the telephone interview of parents, also needs to be analysed. The decision processes leading to actual participation within the panel are hence modelled subsequently. For a graphical illustration of the decision processes described see Figure 1 in Appendix A.

However, data availability on students of the cohort depends on actual first-wave participation. Further availability of data provided by parents on the student depends on the participation decision of parents. As a result the decoupled participation decisions–that is, parents may give their children permission to participate but may refuse to participate themselves–parents' participation decision is realised either when they provide consent for their children or during the contact procedure of the telephone interview. Hence, the participation decision of parents and students is modelled jointly.

From the described survey design arises the need for consideration of multilevel structures, as has been well recognised by the literature. For example, O'Muircheartaigh and Campanelli (1999) apply multilevel logistic regression and multilevel multinomial regression to investigate the influence of the interviewer over that of a geographic region on household nonresponse in the British Household Panel Study. Their findings indicate that good interviewers reduce refusal as well as noncontacts, because variability in refusal as well as in noncontact rates is induced by differences between interviewers rather than between geographic regions. Multilevel multinomial regressions are also used by Durrant and Steele (2009) for the 2001 UK Census Link Study, incorporating response outcomes for six major household surveys. Here, the multilevel structure allows for correlation in response probabilities for households allocated to the same interviewer. Their findings, according to the interviewer effects, are in line with those of O'Muircheartaigh and Campanelli (1999). The results indicate that noncontact is related to household and lifestyle characteristics, that is, variables related to the propensity of being at home, whereas refusal is found to be explained by individual characteristics, see for example Durrant and Steele (2009).

An overview concerning alternative modelling strategies for nonresponse is provided by Steele and Durrant (2011). They review sequential models, sample selection models, and their extensions with a random effect and multinomial models. The authors find the sequential model (modelling contact first and refusal second) to be the most commonly used–although sometimes only one of the two is estimated. The sequential modelling approach is also appealing, because it separates the processes of contact and participation, assuming independence of noncontact and nonparticipation. Besides the fact that coefficients are easier to interpret than in the multinomial model, Steele and Durrant (2011) find very similar results. Furthermore, they apply sample selection models allowing for residual correlation between the equations for noncontact and refusal. Mostly however, Steele and Durrant (2011) choose a probit link function in their analysis. They find little difference in estimates using the multinomial and the sequential model. This fact is a result of different sets of variables significantly effecting noncontact and refusal.

To provide weights for the subgroup of students and parents participating jointly, the weights for the panel cohort need to be adjusted. Adjustments are based on response propensity reweighting, harking back to Rosenbaum and Rubin (1983), using models introduced in Section 3. In nonresponse adjustments using auxiliary information, the set of variables is often small, because information on nonrespondents is sparse. When modelling nonresponse the available variables should be good predictors for nonresponse in order to adjust the weights so that the nonresponse bias of the estimate is reduced. Furthermore, weighting adjustments become most effective, thus reducing nonresponse bias without increasing variance, when the variables used in adjusting the weights are also predictive for the variable of interest, as demonstrated by Little and Vartivarian (2003, 2005). The selection of variables in nonresponse adjustments faces the problem of sparse information regarding nonrespondents and, on the other hand, there are only few (if any) variables that are related to response propensity and the key outcome variables, see Kreuter and Olson (2011).

Also, Nicoletti and Peracchi (2005) use a bivariate probit model to account for possible correlations between the ease of contact and the willingness to participate. After controlling for not only personal and household characteristics but also data collection characteristics they find no residual correlation. Skinner and D'Arrigo (2011) point out that nonresponse is commonly correlated within clusters. This is due to access being dependent on authorities at the cluster level. In educational surveys within schools this is commonly the case. Skinner and D'Arrigo (2011) hence suggest using multilevel models to account for clustered nonresponse by adapting random effects. Applications of estimation are given in Yuan and Little (2007). Given the model frameworks employed in the literature, the next section proposes using a bivariate framework accounting for clustering and correlation to model the joint and clustered decision process.

3. Model Framework for Nonresponse Modelling in Institutional Contexts for Multi-Informant Surveys

3.1. Statistical framework

To model participation decisions in social research and related fields, binary regression models using a logit link function seem to be dominant, see Laaksonen (2005). Due to specific parameters of interest and extensions of models estimating participation propensities, we dedicate ourselves to the probit link function. The extensions of the univariate binary probit model will include a random intercept, a bivariate binary probit model, and a combination of these two extensions. The bivariate model setting allows us to consider possible correlations in the decision processes.

To model two decisions, one concerning the participation consent of the student s and one concerning the participation consent of one parent p, a bivariate probit model specification is suitable. We suppose that the two decisions are correlated, mapped by a correlation parameter ρ. Using a binary probit specification, we can then write the joint decision process of individual i, that is student or parent, in cluster j as

$$y_{ij}^s = \begin{cases} 1 & \text{if} \quad \tilde{y}_{ij}^s > 0, \\ 0 & \text{else,} \end{cases} \qquad \text{and} \quad y_{ij}^p = \begin{cases} 1 & \text{if} \quad \tilde{y}_{ij}^p > 0, \\ 0 & \text{else,} \end{cases} \tag{1}$$

with $y_{ij} = (y_{ij}^s, y_{ij}^p)$ and

$$\tilde{y}_{ij}^s = X_{ij}^s \beta^s + \varepsilon_{ij}^s \quad \text{and} \quad \tilde{y}_{ij}^p = X_{ij}^p \beta^p + \varepsilon_{ij}^p. \tag{2}$$

The correlation parameter ρ enters via the correlation matrix Σ of the residuals and

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{ij}^p \\ \varepsilon_{ij}^s \end{pmatrix} \sim N(0, \Sigma) \quad \text{with} \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{3}$$

The bivariate probit model characterises joint participation probabilities given the characteristics $X_{ij} = (X_{ij}^s, X_{ij}^p)$. Note that the set of regressors does not necessarily have to be identical. The bivariate binary probit can also be extended to a bivariate binary probit model with random intercepts. This alters Equation (2) into

$$\tilde{y}_{ij}^s = X_{ij}^s \beta^s + \alpha_j^s + \varepsilon_{ij}^s \quad \text{and} \quad \tilde{y}_{ij}^p = X_{ij}^p \beta^p + \alpha_j^p + \varepsilon_{ij}^p. \tag{4}$$

Here $\alpha_j = (\alpha_j^s, \alpha_j^p)$, with $\alpha_j \sim N(0, \Omega = \text{diag}(\sigma_s^2, \sigma_p^2))$ denotes the random intercept for students and parents grouped in clusters $j = 1, \ldots, m$ of size n_j. The random intercept probit framework occurs as a special case for $\rho = 0$. In this case, the bivariate probit decomposes into two separate univariate probit models, see Greene (2012). The seminal paper of Butler and Moffitt (1982) describes estimation routines for this univariate model framework. Further, the standard probit framework without clustering occurs if no random intercepts are taken into account. Note that a common strategy to model sequential participation decisions taking place over time is to condition current participation decisions on past decisions, thus augmenting the conditioning factors X_{ij} to incorporate lagged participation decisions of students and parents.

Summarising all model parameters as $\vartheta = (\beta^s, \beta^p, \rho, \sigma_s^2, \sigma_p^2)$, the model therefore characterises the joint probability of all individuals i within cluster j, that is,

$$\mathrm{P}(Y_{.j} \quad = y_{.j}|X_{.j}, \vartheta) \tag{5}$$

where $Y_{.j}$ and $X_{.j}$ denote stacked vectors containing all information of cluster j. In order to provide individual participation probabilities serving as the basis for the derivation of adjustment factors, one has to sum over the corresponding joint probabilities, that is,

$$\mathrm{P}(Y_{ij} = y_{ij}|X_{.j}, \vartheta) \quad = \textstyle\sum_{\Delta_j} \mathrm{P}(Y_{.j} = y_{.j}|X_{.j}, \vartheta), \tag{6}$$

where $j = 1, \ldots, m$ and Δ_j denotes the set of combinations of participation decisions for all students and parents within a cluster, that is, the power set of the four individual possibilities $y = (y^p, y^s) = \{(0,0); (1,0); (0,1); (1,1)\}$ for $n_j - 1$ individuals, required in cluster j for marginalisation of the considered probability. As $|\Delta_j|$ consists out of 4^{n_j-1} combinations, computation becomes prohibitively burdensome for $n_j > 20$. To ensure computational feasibility, the probabilities conditional on expected random intercepts $E[\alpha_j|Y_{.j}, X_{.j}, \vartheta] = \hat{\alpha}_j$ are considered, that is,

$$\mathrm{P}(Y_{iJ} = y_{iJ}|X_{.j}, \vartheta, \hat{\alpha}_j), \tag{7}$$

where $\hat{\alpha}_j$ is provided as a byproduct of the estimation routine described below.

This approach allows for straightforward extension toward dynamic participation of students and parents in consecutive waves. The corresponding joint probability of participation beginning in the first wave can be stated as

$$\mathrm{P}(Y_{.j,t} = y_{.j,t}, \ldots, Y_{.j,1} = y_{.j,1}|X_{.j,(t)}, \vartheta_{(t)}), \tag{8}$$

where $X_{.j,(t)} = (X_{.j,t}, \ldots, X_{.j,1})$ denotes the stacked time-specific conditioning factors including lagged participation decision and $\vartheta_{(t)} = (\vartheta_t, \ldots, \vartheta_1)$ all stacked parameters over time. Again, to ensure computational feasibility, we condition on expected cluster-specific effects, thus yielding

$$\mathrm{P}(Y_{.j,t} = y_{.j,t}, \ldots, Y_{.j,1} = y_{.j,1}, |X_{.j,(t)}, \vartheta_{(t)}, \hat{\alpha}_{j,(t)}) = \prod_{z=1}^{t} \mathrm{P}(Y_{ij,z} = y_{ij,z}|X_{.j,(z)}, \vartheta_{(z)}, \hat{\alpha}_{j,(z)}) \tag{9}$$

with $\hat{\alpha}_{j,(t)} = (\hat{\alpha}_{j,t}, \ldots, \hat{\alpha}_{j,1})$ denoting the stacked expected cluster-specific effects over time. Conditioning on the expected cluster-specific effects is also one possibility to circumvent the high computational burden involved in computing marginal effects to gauge the magnitude of change in probability induced by variation in conditioning variables.

3.2. Parameter estimation

Summarising all parameters of the bivariate binary probit model with random intercept (given in Equation (4)) as $\vartheta = \{\beta_p, \beta_s, \rho, \text{diag}(\Omega)\}$, the corresponding likelihood is

$$\mathcal{L}(\vartheta) = \prod_{j=1}^{m} \int \left[\prod_{i=1}^{n_j} \left(\int_{D_{ij}} \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \varepsilon_{ij}' \Sigma^{-1} \varepsilon_{ij} \right\} d\varepsilon_{ij} \right) \right] \frac{1}{2\pi} |\Omega|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \alpha_j' \Omega^{-1} \alpha_j \right\} d\alpha_j \quad (10)$$

where n_j is the number of individuals i in cluster j. The integration regions for the inner integral over the bivariate normal distribution in Equation (10) are limited by

$$D_{ij} = \begin{cases} (-\infty, -\mu_{ij}^p) \times (-\infty, -\mu_{ij}^s), & \text{if } y_{ij}^p = 0, y_{ij}^s = 0 \\ (-\mu_{ij}^p, +\infty) \times (-\mu_{ij}^s, +\infty), & \text{if } y_{ij}^p = 1, y_{ij}^s = 1 \\ (-\infty, -\mu_{ij}^p) \times (-\mu_{ij}^s, +\infty), & \text{if } y_{ij}^p = 0, y_{ij}^s = 1 \\ (-\mu_{ij}^p, +\infty) \times (-\infty, -\mu_{ij}^s), & \text{if } y_{ij}^p = 1, y_{ij}^s = 0 \end{cases} \quad (11)$$

according to the different combinations of participation decisions that are possible for a student and his/her parent with

$$\mu_{ij}^p = X_{ij}^p \beta^p + \alpha_j^p \quad \text{and} \quad \mu_{ij}^s = X_{ij}^s \beta^s + \alpha_j^s.$$

The likelihood of the model can be calculated by the means of simulation. Using a numerical solution to solve the integrals involved is based on the following property, see Geweke and Keane (2001) and Greene (2012). We use Monte Carlo integration and arrange the integral to take the form

$$I = \int_x g(x) f(x) dx,$$

where $f(x)$ denotes a regular density of a random variable x (e.g., a normal density) and $g(x)$ is a smooth function. The Monte Carlo approximation for the integral is expressed as a conditional mean

$$I = \int_x g(x) f(x) dx = E_f[g(x)] \approx \frac{1}{Q} \sum_{q=1}^{Q} g(x_q).$$

The $x_q, q = 1, \ldots, Q$ denote random draws from the density $f(x)$, see Jones, Maillardet, and Robinson (2009). Based on the law of large numbers, this approximation converges in probability to the expectation, see Greene (2012).

Because the computation of the likelihood given in Equation (10) involves evaluations of the distribution function of the bivariate normal distribution, the approach developed by Geweke (1991), Hajivassiliou (1990), and Keane (1994) (GHK-simulator) can be adapted. In general, the GHK-simulator provides an approximation to the K-variate normal distribution

$$I = \int_D (2\pi)^{-\frac{K}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \varepsilon' \Sigma^{-1} \varepsilon \right\} d\varepsilon.$$

The approximation is based on the property that a multivariate distribution can be factored into a set of corresponding conditional distributions also given as normal distributions. The factorisation into conditional distributions can be written as

$$l \approx \hat{l} = \frac{1}{Q}\sum_{q=1}^{Q}\tilde{l}(q) \quad = \tfrac{1}{Q}\sum_{q=1}^{Q}\prod_{l=1}^{K}\Phi(\gamma_{\mathcal{U}}^{(q)}) - \Phi(\gamma_{\mathcal{L}}^{(q)}), \tag{12}$$

with

$$\gamma_{\mathcal{U}}^{(q)} = \left(\frac{D_{l,\mathcal{U}} - L_{1:l-1,l}\cdot\tilde{v}_{1:l-1,q}}{L_{l,l}}\right) \quad \text{and} \quad \gamma_{\mathcal{L}}^{(q)} = \left(\frac{D_{l,\mathcal{L}} - L_{1:l-1,l}\cdot\tilde{v}_{1:l-1,q}}{L_{l,l}}\right), \tag{13}$$

where L denotes the Cholesky decomposition of Σ, that is, $\Sigma = LL'$. Further, $L_{1:l-1,l}, l = 1,\ldots,K$ denotes the vector of elements in columns l up to row $l-1$ and $\{\tilde{v}_{l,q}\}_{q=1}^{Q}$ denotes Q draws from a normal distribution truncated in the region $(D_{l,\mathcal{L}}, D_{l,\mathcal{U}})$ corresponding to D_{ij} in Equation (11) and $\tilde{v}_{1:l-1,q}$ denotes a vector stacking of draws. Conceptually, the simulated likelihood is then given as

$$\tilde{\mathcal{L}}(\vartheta) \quad = \prod_{j=1}^{m}\tfrac{1}{R}\sum_{r=1}^{R}\left[\prod_{i=1}^{n_j}\tfrac{1}{Q}\sum_{q=1}^{Q}\tilde{l}(q|\alpha_j^{(r)})\right] \tag{14}$$

where the conditioning on $\alpha_j^{(r)}, r = 1,\ldots,R$ enters via the lower and upper bounds, that is, $D_l = (D_{l,\mathcal{L}}, D_{l,\mathcal{U}})$ corresponding to D_{ij}. R here denotes the number of random draws. The cluster-specific expected random intercept is then given as

$$\hat{\alpha}_j \quad = E[\alpha_j|Y_{\cdot j}, X_{\cdot j}, \vartheta] = \frac{\frac{1}{R}\sum_{r=1}^{R}\alpha_j^{(r)}\left[\prod_{i=1}^{n_j}\frac{1}{Q}\sum_{q=1}^{Q}\tilde{l}(q|\alpha_j^{(r)})\right]}{\frac{1}{R}\sum_{r=1}^{R}\left[\prod_{i=1}^{n_j}\frac{1}{Q}\sum_{q=1}^{Q}\tilde{l}(q|\alpha_j^{(r)})\right]}. \tag{15}$$

Implementation for the estimation routine is performed in R, see Appendix C for details.

3.3. Simulation-based evaluation

We check the statistical and numerical precision of the estimation routine within a small simulation study. The simulation is based on $D = 100$ different data sets, which include $m = 50$ clusters each of size $n_j = 30$, so that the total number of cases is $N = 1500$. For each of these data sets the bivariate binary probit with random intercept is estimated.

Table 1 in Appendix B gives the results for statistical precision. The true parameters ϑ of the data-generating process are given in the first column ϑ. Further columns report the average estimated parameter $\bar{\hat{\vartheta}}$, the average standard deviation of the parameter estimates $ASE(\hat{\vartheta})$, average bias $ABias(\hat{\vartheta})$, as well as the average mean squared error $AMSE(\hat{\vartheta})$. The next two columns give standard deviations σ for the estimated parameters $\hat{\vartheta}$ and their standard errors $SE(\hat{\vartheta})$. The last column gives the coverage rates $\frac{\mathcal{I}(\vartheta\in CI)}{R}$. Standard errors are computed by inversion of the Hessian. The averages are computed over $D = 100$ results of the different data sets. The results show a low bias for all parameters. Coverage rates indicate good statistical precision. The largest bias is found for $ABias(\sigma_s) = -0.12377$ at a coverage rate of $\frac{\mathcal{I}(\sigma_s\in CI)}{R} = 0.92$.

The lowest coverage rate is found for β_1^s with $\frac{\mathcal{I}(\beta_1^s \in CI)}{R} = 0.89$. Table 1 shows that the average standard errors $ASE(\hat{\vartheta})$ are close to the standard deviations of the estimates $\sigma_{\hat{\vartheta}}$.

To check for numerical precision, the bivariate binary probit with random intercept was estimated using one data set, but this time using $C = 20$ different sets of random numbers $\{\hat{v}_{l,q}\}_{q=1}^{Q}$, see Equation 13. Again ϑ gives the parameters for the data-generating process in Table 2. The average estimate is given in column $\bar{\hat{\vartheta}}$ and column $ASE(\hat{\vartheta})$ reports the average standard error for the estimated $\hat{\vartheta}$s. Columns $\sigma_{\hat{\vartheta}}$ and $\sigma_{SE(\hat{\vartheta})}$ give the corresponding standard deviations for the estimates and their standard errors. The variation of estimates $\sigma_{\hat{\vartheta}}$ and their standard errors $\sigma_{SE(\hat{\vartheta})}$ induced by different sets of random numbers is within reasonable bounds. Comparing the average standard error $ASE(\hat{\vartheta})$ and the standard deviation of the estimate $\sigma_{\hat{\vartheta}}$, we find that they do not overlay although for β_1^p the ratio is $\frac{\sigma_{\hat{\vartheta}}}{ASE(\hat{\vartheta})} = 0.357$.

4. Empirical Results

We illustrate the suggested approach for modelling participation processes within large-scale educational surveys using the NEPS cohort sample of grade 5 students. As in all other cohort samples focusing on the institutional context within the NEPS, surveying and testing of students, is accompanied by a telephone interview with one of the students' parents. In this parent interview some information given by the student is validated and additional background information on the student's environment is collected. The decision processes described above result in the joint participation statuses shown in Table 3. The table gives the participation statuses for students and parents by wave. The panel cohort consists of $N = 6112$ students, of whom 5774 participated in wave 1 (participation rate: 94.47%) and 338 were classified as temporary dropouts due to illness, bad weather conditions, etc. Students' participation rates in SC3 by institution range from 30.77% up to 100% (with median of 96.67%). The students' parents were less likely to participate in the CATI. Altogether, 4151 parents participated in wave 1. The other 1961 parents did not participate in the first wave due to temporary dropout or refusal. For the subgroup of 3974 of wave 1 participants an additional interview with one parent is available. In wave 2 there are fewer students and parents participating together. The subgroup consists of 3727 students and parents in wave 2. In modelling the response propensities for students and parents, the dependent variable is the binary participation status in the corresponding wave. Regressors included in the model comprise variables related to sampling characteristics, such as stratification variables, variables on characteristics describing sociodemographics, for example, gender or migration background, and variables involving paradata from call records.

When establishing the cohort sample, the variety of Federal-State-specific school systems as well as different transitions between primary and secondary school institutions in Germany are respected via seven explicit strata. The first stratum *GY* comprises all Gymnasien (type of school leading to upper secondary education and university entrance qualification), the second stratum *HS* consists of all Hauptschulen (school for basic secondary education), the third stratum *RS* refers to all Realschulen (intermediate secondary school), the fourth to comprehensive schools (stratum *IG*: Integrierte Gesamtschulen, Freie Waldorfschulen), the fifth includes schools with several courses of education (stratum *MB*: Schulen mit mehreren Bildungsgängen). The sixth explicit stratum comprises schools offering schooling to students with special educational needs in the area of learning (stratum *FS*: Förderschule). The seventh explicit stratum comprises all schools providing schooling to grade 5 students, but not to grade 9 students (stratum *N*5). An additional supplement of schools providing access to students with a Turkish migration background or migration background related to the former Soviet Union (ethnic German repatriates) is included as well and is the reference category.

Besides that, information related to the sociodemographic and family background include the age group of the student, gender, native language, and migration background. According to year and month of birth, the students are split by the median into a younger and an older half of the age group (reference category). Gender includes female and male, with male being the reference category. Native language consists of German and other (reference category) and migration background is either Turkish or related to the former Soviet Union (reference category). Information on migration background was available from the school records and provided by teachers in the preliminary stages of the survey. A missing indicator is included for information on missing values for gender or age.

Also paradata, see Couper (1998) and Groves and Heeringa (2006), is available arising from test and telephone interview protocols during fieldwork. Paradata ara available for those parents that were contacted. For parents, the number of calls before first contact is recorded as paradata. It is included in the model as a dummy variable if the number of calls is less than four, with three calls being the median. Models of the second wave are conditioned on the first-waves participation status of students and parents. Besides that, a dummy is included indicating whether a student has left the institutional context of a school and is followed up and surveyed individually. Separation problems occur when using variables together that are missing for nonparticipants because there is no information on this group. This problem is eased by information on nonparticipants of wave 1 participating in wave 2 of the panel. Furthermore, information not available yet (especially for parents, the students' environment, and competencies) may become available later for weighting adjustments in future waves. For all variables used in modelling the participation propensities, Table 4 displays the number of cases (n) and their corresponding proportions ($\frac{n}{N}$) for each category of the variable.

For each wave we show the following settings. First, separate univariate models without random intercepts are estimated for students (I) and parents (II). Second, a joint bivariate model without random intercept (III) is estimated. We then proceed with separate random intercept models for students (IV) and parents (V). Finally, we model joint participation of parents and students using the bivariate probit with random intercepts (VI) as stated in Equation (4). Each model has an additional suffix corresponding to wave 1 (a) and 2 (b), respectively. The values for the log-likelihood $\ln \mathcal{L}$, AIC, and BIC, as well as the test statitistic of the likelihood ratio (LR) test for model comparison can be found in Table 5. To test the significance of the correlation parameter of the bivariate model, we use the likelihood ratio test statistic given as $2 \cdot [\ln \mathcal{L}_{III(a/b)} - (\ln \mathcal{L}_{I(a/b)} + \ln \mathcal{L}_{II(a/b)})]$ or $2 \cdot [\ln \mathcal{L}_{VI(a/b)} - (\ln \mathcal{L}_{IV(a/b)} + \ln \mathcal{L}_{V(a/b)})]$, respectively. In wave 1 and 2 the model settings without clustering show that the consideration of correlation in error terms are fitting the data better at a significance level of 5% in wave 1 and at 0.1% in wave 2. All likelihood ratio tests clearly stress the importance of considering correlation when modelling the decision processes of participation.

With regard to clustering, note that testing for random coefficients is nonstandard, because the variances of the random coefficients lie on the boundary of the parameter space. This violation of the standard regularity conditions causes the invalidity of the asymptotic χ^2-distribution of the LR statistic. Gouriéroux, Holly, and Monfort (1982) derive the correct asymptotic distribution as a mixture of χ^2-distributions. The asymptotic distribution for testing the significance of p random coefficients via a LR-test has the form $\sum_{i=0}^{p} w(p,i)\chi^2(i)$, where $w(p,i) = \frac{\binom{i}{p}}{2^p}$, $\chi^2(i)$ denotes a χ^2-distribution with i degrees of freedom and $\chi^2(0)$ the unit mass at the origin. However, the critical values are lower than those of a corresponding standard χ^2-distribution, thus providing conservative lower significance levels. Bearing this in mind, assessing the significance of random coefficients via a standard χ^2-distribution provides a significance level reaching at most the announced nominal level, see Harvey (1989). To test the significance of the model specification regarding random intercepts, we use the likelihood ratio test given by $2 \cdot (\ln \mathcal{L}_{VI(a/b)} - \ln \mathcal{L}_{III(a/b)})$. In wave 1, as well as in wave 2, considering the cluster structure is strictly preferred. The model settings that respect the cluster structure fit the data significantly better than those that do not.

Focusing on the subgroup of students and parents, Table 6 and 7 provide the estimated models for the participation propensities of students and parents in wave 1. Table 8 and 9 provide the models for wave 2. All model specifications consider the same set of covariates. As discussed above, we use the probit link function because it allows for straightforward extension with regard to both aspects considered, that is, random intercepts to allow for clustering and via bivariate normal setup for correlation between students and parents. The random intercept for both (parents and students) is specified at the school level. This implies a correlation structure possibly capturing unobservable interaction between the participants at the school level, for example, communication with teachers and among parents.

Tables 6 and 7 show the estimated coefficients for different model specifications in wave 1. The main effects remain stable throughout all specifications, that is, they do not change in sign and magnitude. This is also true for the variance parameter of the random intercepts as well as for the correlation coefficient. Comparing the bivariate probit model without and with a random intercept, the correlation increases slightly when considering the clustered structure of the data using random intercepts. The same findings apply to wave 2. The models for wave 2 include 14 observations that are classified as final dropouts. These students withdraw their panel consent between wave 1 and wave 2. A more detailed analysis of these 14 cases is not possible due to the small number. An estimation of the model with and without the final dropouts induced no substantial differences but only small changes in the estimated coefficients. A small difference occurs in the variance parameters of the random intercept model and in the parental equation of the bivariate probit with random intercept.

For wave 1 the bivariate probit with random intercept shows significantly negative effects for all secondary school types except those referring to Stratum *FS*. Negative effects are found for students being educated in a comprehensive school (Stratum *IG*) and schools offering several tracks of education (Stratum *MB*). The missing indicator is also significant and has a large influence on the participation decision. This is mostly due to the fact that information is missing for nonparticipants. A positive effect is found for students speaking German as a native language in both participation decisions. Within the parental participation decision, school types have a positive effect for primary schools educating students in grade 5 (Stratum *N*5) and Gymnasien

(Stratum *GY*). Parents with a child of Turkish migration background influence their participation decision positively. Finally, the dummy for the low number of contact attempts before first contact indicates that parents who are easy to reach (that is, have a higher propensity of being at home and contacted) also have a higher propensity to participate. This has to do with what Durrant and Steele (2009) call lifestyle characteristics, that is, these characteristics influence the propensity of being at home. For both, parents and students, a significantly and large variation in the level of participation propensity across schools was found. Eventually, there is little, albeitsignificant, correlation in the error terms of the model. For wave 2, Tables 8 and 9 show the corresponding models describing the participation propensities. Students' propensity is strongly reduced if they are educated in special schools (Stratum *FS*). The impact of the missing indicator reduces (compared to wave 1) in wave 2. A negative effect on the students' participation decision is found for students in the field of individual retracking. Students are handed over to this field if they cannot be surveyed and tested within the institutional context of their school. The participation status of the parents also positively influences students' propensity to participate, whereas the students' own participation status has a negative sign. Regarding parents, their own participation status and that of their child have a positive effect on wave 2 participation. The number of calls before first contact being less than four is again a strong predictor for parents' participation. The impact of the different school types remains stable and increases for comprehensive schools (Stratum *IG*).

Based on these models, relevant adjustments were derived for the subgroup of students and parents jointly participating in wave 1 and wave 2 of SC3 in the NEPS. Given the design weight d_{ij} for student *i* in school *j*, the adjustment yields an adjusted weight $w_{ij} = d_{ij} \cdot \delta_{ij}^{-1}$, where δ_{ij} is the participation propensity for student *i* in school *j*. The propensity δ_{ij} can be specified according to the subgroup that is to be adjusted. The corresponding probability for students and parents is implied by Equation (9). Further subgroups of interest for weighting adjustments can, for example, be the group of students participating in one particular wave $\tilde{t}$ or the subgroup of students participating in all waves up to wave T. These two probabilities can be derived from the joint bivariate model specification allowing for clustering at the school level, which emerges as the preferred model specification, as follows. For the subgroup of students participating in one particular wave $\tilde{t}$, we have

$$\begin{aligned} \delta_{ij} &= \mathrm{P}(y^s_{ij,\tilde{t}} = 1 | X_{ij,(\tilde{t})}, \vartheta_{(t)}, \hat{\alpha}_{j,(t)}) \\ &= \sum_{\Delta_j^{\tilde{t},s}} \prod_{z=1}^{\tilde{t}} \mathrm{P}(Y_{ij,z} = y_{ij,z} | X_{\cdot j,(z)}, \vartheta_{(z)}, \hat{\alpha}_{j,(z)}), \end{aligned}$$

with $\Delta_j^{\tilde{t},s}$ comprising all participation patterns for parents up to wave $\tilde{t}$ and for students up to wave $\tilde{t}-1$. If $\tilde{t} = 3$ then $|\Delta_j^{\tilde{t},s}| = 32$. For the subgroup of students participating in all waves up to wave T, we have

$$\begin{aligned} \delta_{ij} &= \mathrm{P}(y^s_{ij,(T)} = 1 | X_{ij,(T)}, \vartheta_{(t)}, \hat{\alpha}_{j,(t)}) \\ &= \sum_{\Delta_j^{(T),s}} \prod_{z=1}^{(T)} \mathrm{P}(Y_{ij,z} = y_{ij,z} | X_{\cdot j,(z)}, \vartheta_{(z)}, \hat{\alpha}_{j,(z)}), \end{aligned}$$

with $\Delta_j^{(T),s}$ comprising all participation patterns for a parent up to wave T. If $T = 3$ then $|\Delta_j^{(T),s}| = 8$. We are well aware that the implied summation involves a prohibitively large number of participation constellations. However, preliminary analysis of participation decisions in other starting cohorts of the NEPS provides evidence that conditioning on two previous waves may be sufficient to capture the intertemporal dependencies among the participation decisions of students and parents. The suggested empirical framework allows us to test this assumption.

5. Conclusions

For modelling we suggest considering possibly correlated participation processes arising within large-scale educational surveys a bivariate probit framework with random intercepts, where parameter estimation is accessible to simulated maximum likelihood estimation using the GHK simulator. The proposed model framework serves two important aspects that need to be considered in nonresponse adjustments. First, it was necessary to account for the correlation in participation decisions regarding weighting adjustments in multi-informant surveys, which led us to the bivariate probit framework. Second, extending the bivariate probit with random intercepts was necessary to respect the hierarchical structure of the data provided by students nested in schools. Apart from sociodemographic factors such as language spoken at home, the empirical application emphasizes considering clustering at the institutional level and correlation between decision processes of students and parents.

Furthermore, this approach allows us to extend in a flexible way the introduced model for successive waves of data collection within the NEPS. The model specification facilitates modelling the joint decisions of students and parents in the panel context over time. Using the panel extension will conveniently provide adjustments for specific combinations of subgroups (e.g., parents and students) and time (e.g., wave 1 and wave 3). Having information on temporary dropouts in future waves will make a reestimation necessary; thereby improving the adjustments by the time new information arises. So far, subgroups of interest were primarily given by the multi-informant perspective established in the NEPS.

Future research may focus on integrating the insights derived from the modelling of participation processes into substantial analysis of competence development and determinants of educational decisions.

Acknowledgements:
This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 5, DOI:10.5157/NEPS:SC3:2.0.0. 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.

References

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., … Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: challenges and solutions. In H.-P. Blossfeld, H. G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process: Zeitschrift für Erziehungswissenschaft* (Vol. 14, pp. 51–65). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0181-8

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes.* Retrieved 03.06.2014, from http://CRAN.R-project.org/package=lme4

Bäumer, T., Preis, N., Roßbach, H.-G., Stecher, L., & Klieme, E. (2011). Education processes in life-course-specific learning environments. In H.-P. Blossfeld, H. G. Rossbach, & J. v. Maurice (Eds.), *Education as a lifelong process: Zeitschrift für Erziehungswissenschaft* (Vol. 14, pp. 87–101). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0183-6

Blossfeld, H.-P., Roßbach, H. G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (Vol. 14). Wiesbaden: VS Verlag für Sozialwissenschaften.

Brick, J. M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics*, *29*(3), 329–353.

Butler, J. S., & Moffitt, R. (1982). A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model. *Econometrica*, *50*(3), 761–764.

Couper, M. P. (1998). Measuring Survey Quality in a CASIC Environment. In American Statistical Association (Ed.), *Proceedings of the Survey Research Methods Section* (Vol. 48, pp. 41–49). Retrieved 27.05.2013, from http://www.amstat.org/sections/srms/Proceedings/papers/1998_006.pdf

Durrant, G. B., & Steele, F. (2009). Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK Government surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(2), 361–381. doi: 10.1111/j.1467-985X.2008.00565.x

Frahm, S., Goy, M., Kowalski, K., Sixt, M., Strietholt, R., Blatt, I., … Kanders, M. (2011). Transition and development from lower secondary to upper secondary school. In H.-P. Blossfeld, H. G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process: Zeitschrift für Erziehungswissenschaft* (Vol. 14, pp. 217–232). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0188-1

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. In American Statistical Association (Ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (pp. 571–578). Alexandria.

Geweke, J., & Keane, M. (2001). Chapter 56: Computationally intensive methods for integration in econometrics. In J.J. Heckman & E.E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3463–3568). Elsevier. doi: 10.1016/S1573-4412(01)05009-7

Gouriéroux, C., Holly, A., & Monfort, A. (1982). Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica*, *50*(1), 63–80.

Greene, W. (2012). *Econometric analysis* (7th ed.). Boston: Pearson.

Groves, R. M. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 439–457. doi: 10.1111/j.1467-985X.2006.00423.x

Hajivassiliou, V. (1990). *Smooth Simulation Estimation of Panel Data LDV Models* (Unpublished doctoral dissertation). Department of Economics, Yale University.

Harvey, A. C. (1989). *Forecasting, structural time series models, and the Kalman filter*. Cambridge and New York: Cambridge University Press.

Honaker, J., Owen, M., Imai, K., Lau, O., & King, G. (2013). *bprobit: Bivariate Probit Regression for Two Dichotomous Dependent Variables.* Retrieved 21.01.2014, from http://cran.r-project.org/web/packages/ZeligChoice/vignettes/ZeligChoice-manual.pdf

Jones, O., Maillardet, R., & Robinson, A. (2009). *Introduction to scientific programming and simulation using R*. Boca Raton: CRC Press.

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, *19*(2), 81–97.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, *12*, 1–16.

Keane, M. P. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, *62*(1), 95–116.

Kreuter, F., & Olson, K. (2011). Multiple Auxiliary Variables in Nonresponse Adjustment. *Sociological Methods & Research*, *40*(2), 311–332. doi: 10.1177/0049124111400042

Laaksonen, S. (2005). Does the choice of link function matter in response propensity modelling? *Model Assisted Statistics and Applications*, *1*(2), 95–100.

Little, R., & Vartivarian, S. (2003). On weighting the rates in non-response weights. *Statistics in Medicine*, *22*(9), 1589–1599. doi: 10.1002/sim.1513

Little, R., & Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, *31*(2), 161–168.

Lynn, P., & Kaminska, O. (2010). *Weighting Strategy for Understanding Society: Understanding Society Working Paper No. 2010-05.* Colchester. Retrieved 05.06.2012, from http://research.understandingsociety.org.uk/publications/working-paper/2010-05.pdf

Micklewright, J., Schnepf, S. V., & Skinner, C. (2010). *Non-Response Biases in Surveys of School Children: The Case of the English PISA Samples* (No. 4789). Retrieved 03.06.2013, from http://ftp.iza.org/dp4789.pdf

Nicoletti, C., & Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(4), 763–781. doi: 10.1111/j.1467-985X.2005.00369.x

O'Muircheartaigh, C., & Campanelli, P. (1999). A Multilevel Exploration of the Role of Interviewers in Survey Non-Response. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *162*(3), 437–446.

Owen, M., Imai, K., Lau, O., & King, G. (2012). *ZeligChoice: Zelig Choice Models.* Retrieved 03.06.2014, from http://CRAN.R-project.org/package=ZeligChoice

Pike, G. R. (2008). Using Weighting Adjustments to Compensate for Survey Nonresponse. *Research in Higher Education*, *49*(2), 153–171. doi: 10.1007/s11162-007-9069-0

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The Role of Demographics, Engagement and Personality. *Research in Higher Education*, *46*(2), 127–152. doi: 10.1007/s11162-004-1597-2

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. doi: 10.1093/biomet/70.1.41

Skinner, C. J., & D'Arrigo, J. (2011). Inverse probability weighting for clustered nonresponse. *Biometrika*, *98*(4), 953–966. doi: 10.1093/biomet/asr058

Steele, F., & Durrant, G. B. (2011). Alternative Approaches to Multilevel Modelling of Survey Non-Contact and Refusal. *International Statistical Review*, *79*(1), 70–91. doi: 10.1111/j.1751-5823.2011.00133.x

Wood, A. M., White, I. R., & Hotopf, M. (2006). Using number of failed contact attempts to adjust for non-ignorable non-response. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 525–542. doi: 10.1111/j.1467-985X.2006.00405.x

Yuan, Y., & Little, R. (2007). Model-based estimates of the finite population mean for two-stage cluster samples with unit non-response. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *56*(1), 79–97. doi: 10.1111/j.1467-9876.2007.00566.x

A. Figures

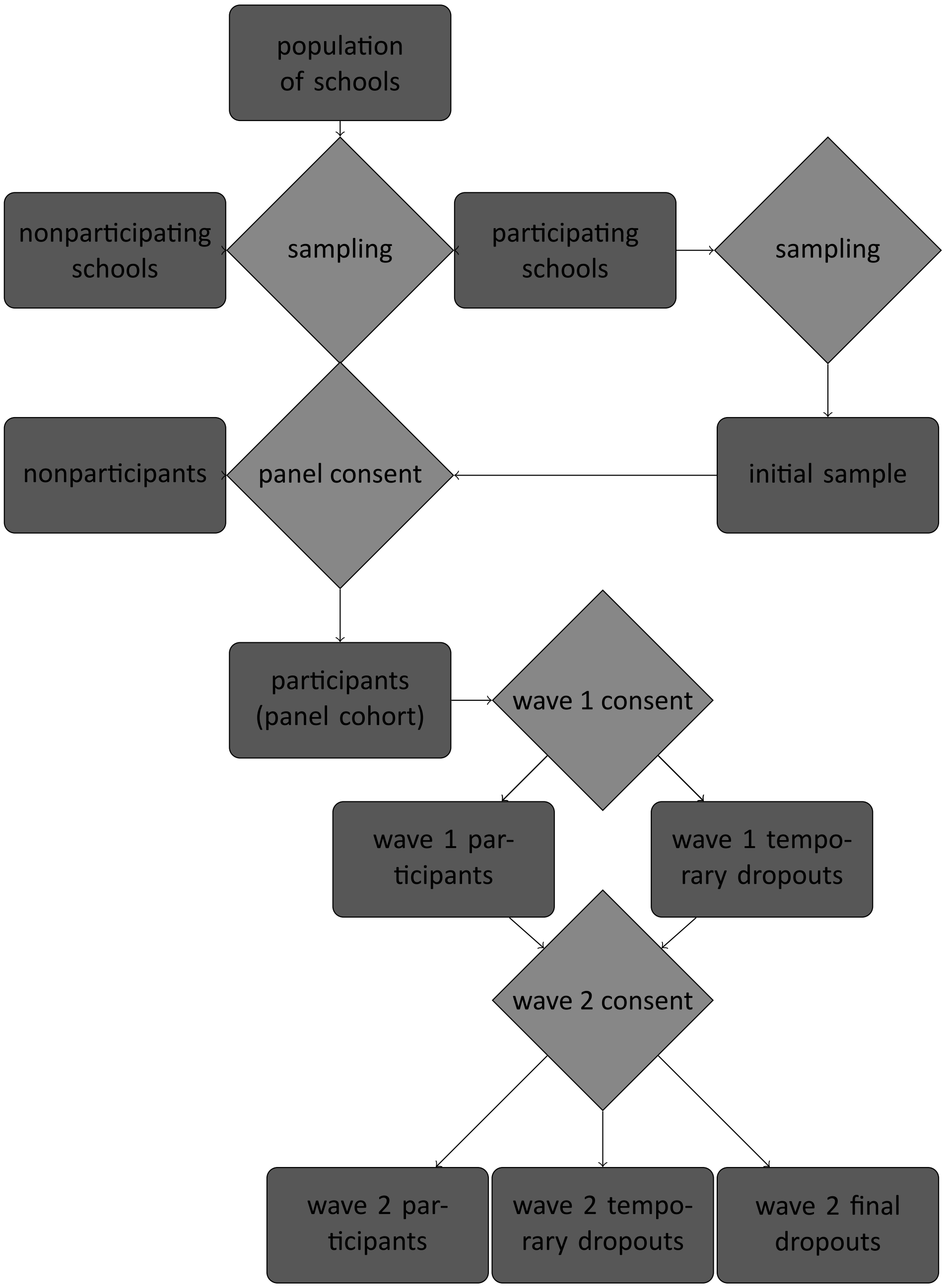

Figure 1: Decision processes of students and parents.

B. Tables

Table 1: Statistical precision over 100 replications.

	$\boldsymbol{\theta}$	$\bar{\boldsymbol{\theta}}$	ASE$(\hat{\boldsymbol{\theta}})$	ABias$(\hat{\boldsymbol{\theta}})$	AMSE$(\hat{\boldsymbol{\theta}})$	$\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}$	$\boldsymbol{\sigma}_{SE(\hat{\boldsymbol{\theta}})}$	$\frac{\mathcal{I}(\boldsymbol{\theta}\in CI)}{R}$
β_1^p	1.0	1.006	0.18373	0.00577	0.03785	0.19544	0.02299	0.92
β_2^p	0.4	0.404	0.02694	0.00370	0.00080	0.02822	0.00182	0.94
β_3^p	0.6	0.607	0.03511	0.00672	0.00121	0.03428	0.00237	0.97
β_1^s	-1.0	-0.955	0.14548	0.04542	0.02203	0.14203	0.01694	0.89
β_2^s	0.5	0.482	0.04406	-0.01822	0.00227	0.04420	0.00498	0.90
β_3^s	-1.5	-1.445	0.11514	0.05482	0.01606	0.11482	0.01443	0.90
$\rho^{p,s}$	0.4	0.376	0.11870	-0.02412	0.01372	0.11521	0.01860	0.95
σ_p	1.2	1.206	0.14635	0.00571	0.02363	0.15439	0.02106	0.94
σ_s	0.8	0.676	0.15604	-0.12377	0.03368	0.13620	0.02271	0.92

Note: Simulation based on $D = 100$ data sets and simulation sample size $Q = 1000$.

Table 2: Numerical precision over $C = 20$ sets or random numbers.

	$\boldsymbol{\theta}$	$\overline{\hat{\boldsymbol{\theta}}}$	$\mathbf{ASE}(\hat{\boldsymbol{\theta}})$	$\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}$	$\boldsymbol{\sigma}_{\mathbf{SE}(\hat{\boldsymbol{\theta}})}$
β_1^p	1.0	1.228	0.18150	0.06476	0.01065
β_2^p	0.4	0.392	0.02659	0.00061	0.00007
β_3^p	0.6	0.605	0.03469	0.00109	0.00011
β_1^s	-1.0	-0.973	0.18414	0.04998	0.00967
β_2^s	0.5	0.513	0.05600	0.00928	0.00329
β_3^s	-1.5	-1.541	0.15388	0.02762	0.01106
$\rho^{p,s}$	0.4	0.471	0.13496	0.03130	0.01573
σ_p	1.2	1.187	0.13834	0.04463	0.01286
σ_s	0.8	0.895	0.21563	0.04875	0.03116

Note: Simulation based on $D = 1$ data set, $C = 20$ different sets of random numbers used in the estimation and simulation sample size $Q = 1000$.

Table 3: Participation statuses for students in SC3 and their parents by wave.

	Parents	
Students	**Participant**	**Nonparticipant**
	Wave 1	
Participant	3974	1800
Nonparticipant	177	161
	Wave 2	
Participant	3727	2063
Nonparticipant	93	229

Table 4: Number of cases (n) and proportion ($\frac{n}{N}$) for variables in models by wave.

	Wave 1		Wave 2	
	n	$\frac{n}{N}$	*n*	$\frac{n}{N}$
Gender				
female	2895	0.4737	2895	0.4737
male	3146	0.5147	3146	0.5147
missing	71	0.0116	71	0.0116
Nationality				
German	5112	0.8364	5112	0.8364
other	344	0.0563	344	0.0563
missing	656	0.1073	656	0.1073
Native language				
German	5225	0.8549	5225	0.8549
other	717	0.1173	717	0.1173
missing	170	0.0278	170	0.0278
Number of calls				
four or more	2318	0.3793	2231	0.3650
less than four	2435	0.3984	2400	0.3927
no calls	1359	0.2223	1481	0.2423
Tracking status				
individual retracking	0	0.0000	444	0.0726
in school	6112	1.0000	5668	0.9274
Age group				
older half	3253	0.5322	3253	0.5322
younger half	2670	0.4368	2670	0.4368
missing	189	0.0309	189	0.0309
Sampling stratum				
MIG	242	0.0396	242	0.0396
N5	458	0.0749	458	0.0749
FS	587	0.0960	587	0.0960
GY	2372	0.3881	2372	0.3881
HS	677	0.1108	677	0.1108
IG	284	0.0465	284	0.0465
MB	352	0.0576	352	0.0576
RS	1140	0.1865	1140	0.1865
Migration background				
Repatriated ethnic Germans	68	0.0111	68	0.0111
Turkish	174	0.0285	174	0.0285
missing	5870	0.9604	5870	0.9604
Missing indicator for personal characteristics	194	0.0317	226	0.0370

Table 5: ln $\mathcal{L}$, *AIC and BIC for considered model specifications.*

Model specifications			**Information criteria** ln $\mathcal{L}$	AIC	BIC
			Wave 1		
No clustering					
Separate	Students	Ia	−1049.599	2125.197	2212.531
	Parents	IIa	−3149.145	6320.290	6394.188
		Ia+IIa	−4198.744		
Joint		IIIa	−4195.952	8441.904	8609.854
LR test	Ia+IIa vs. IIIa		5.584*		
Clustering					
Separate	Students	IVa	−1039.797	2107.595	2201.647
	Parents	Va	−3125.542	6275.084	6355.700
		IVa+Va	−4165.339		
Joint		VIa	−4161.778	8377.555	8558.942
LR test	IVa+Va vs. VIa		7.122**		
LR test	IIIa vs. VIa		68.348***		
			Wave 2		
No clustering					
Separate	Students	Ib	−526.426	1082.853	1183.623
	Parents	IIb	−1985.893	3997.786	4085.120
		Ib+IIb	−2411.555		
Joint		IIIb	−2483.226	4845.998	5047.538
LR test	Ib+IIb vs. IIIb		143.342***		
Clustering					
Separate	Students	IVb	−510.032	886.052	993.503
	Parents	Vb	−1983.092	3961.509	4055.529
		IVb+Vb	−2493.124		
Joint		VIb	−2465.790	4816.750	5031.653
LR test	IVb+Vb vs. VIb		54.668***		
LR test	IIIb vs. VIb		−34.872***		

Notes: ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively.

Table 6: Alternative models estimating individual participation propensity of students and parents for SC3 in Wave 1.

	No clustering		Clustering	
	Students (Ia)	**Parents (IIa)**	**Students (IVa)**	**Parents (Va)**
Intercept	1.195***	−0.708***	1.234***	−0.753***
	(0.243)	(0.167)	(0.270)	(0.184)
Stratum *N5*	−0.361	0.584**	−0.367	0.643**
	(0.264)	(0.179)	(0.299)	(0.204)
Stratum *FS*	0.098	−0.021	0.116	0.001
	(0.271)	(0.175)	(0.303)	(0.196)
Stratum *GY*	−0.297	0.594***	−0.299	0.654***
	(0.248)	(0.169)	(0.278)	(0.189)
Stratum *HS*	−0.339	0.238	−0.331	0.266
	(0.256)	(0.174)	(0.288)	(0.196)
Stratum *IG*	−0.772**	0.472*	−0.750*	0.484*
	(0.265)	(0.185)	(0.310)	(0.218)
Stratum *MB*	−0.602*	0.169	−0.635*	0.218
	(0.266)	(0.181)	(0.302)	(0.208)
Stratum *RS*	−0.207	0.389*	−0.192	0.447*
	(0.253)	(0.171)	(0.284)	(0.193)
Migration background	−0.179	0.407*	−0.170	0.446*
Turkish	(0.281)	(0.194)	(0.313)	(0.214)
Native language	1.099***	0.440***	1.140***	0.454***
German	(0.064)	(0.050)	(0.068)	(0.052)
Age group	−0.072		−0.055	
younger half	(0.063)		(0.067)	
Gender	0.057		0.061	
female	(0.060)		(0.063)	
Missing indicator for	−1.098***		−1.147***	
personal characteristics	(0.111)		(0.116)	
Number of calls		1.299***		1.324***
less than 4		(0.043)		(0.044)
Random intercept				
σ school level			0.311	0.261
ln $\mathcal{L}$	−1049.599	−3149.145	−1039.797	−3125.542
AIC	2125.197	6320.290	2107.595	6275.084
BIC	2212.531	6394.188	2201.647	6355.700
Sample size	6112	6112	6112	6112

Notes: ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parenthesis. To model individual participation, the `glmer` and `glm` functions with a probit link provided by `lme4` (Bates, Maechler, & Bolker, 2012) and `stats` package in R (R Core Team, 2017) was used.

Table 7: Results for the bivariate probit models without and with random intercept estimating the individual participation propensities for students and parents for SC3 in Wave 1.

	Bivariate probit – no clustering		**Bivariate probit – clustering**	
	Parents	**Students**	**Parents**	**Students**
	(IIIa)		**(VIa)**	
Intercept	−0.709***	1.188***	−0.762***	1.197***
	(0.167)	(0.242)	(0.186)	(0.268)
Stratum *N5*	0.586**	−0.352	0.657**	−0.340
	(0.179)	(0.263)	(0.207)	(0.297)
Stratum *FS*	−0.020	0.105	0.010	0.139
	(0.175)	(0.270)	(0.197)	(0.299)
Stratum *GY*	0.596***	−0.287	0.667***	−0.271
	(0.169)	(0.247)	(0.191)	(0.276)
Stratum *HS*	0.239	−0.329	0.274	−0.297
	(0.174)	(0.255)	(0.198)	(0.286)
Stratum *IG*	0.473*	−0.768**	0.502*	−0.747*
	(0.185)	(0.264)	(0.221)	(0.305)
Stratum *MB*	0.170	−0.596*	0.228	−0.613*
	(0.181)	(0.265)	(0.210)	(0.301)
Stratum *RS*	0.390*	−0.201	0.456*	−0.164
	(0.171)	(0.252)	(0.195)	(0.283)
Migration background	0.409*	−0.175	0.458*	−0.157
Turkish	(0.194)	(0.280)	(0.216)	(0.311)
Age group		−0.077		−0.062
younger half		(0.063)		(0.066)
Native language	0.440***	1.099***	0.452***	1.132***
German	(0.050)	(0.064)	(0.052)	(0.069)
Gender		0.061		0.064
female		(0.060)		(0.062)
Missing indicator for		−1.080***		−1.120***
personal characteristics		(0.111)		(0.119)
Number of calls	1.297***		1.315***	
less than 4	(0.043)		(0.044)	
Correlation	0.097*		0.122**	
ρ students parents	(0.049)		(0.044)	
Random intercept				
σ school level			0.261	0.302
ln $\mathcal{L}$	-4195.952		-4161.778	
AIC	8441.904		8377.555	
BIC	8609.854		8558.942	
Sample size	6112		6112	

Notes: ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parenthesis. To model individual participation decisions, the `zelig` function with `bprobit` link provided by `ZeligChoice` package (Owen, Imai, Lau, & King, 2012) in R (R Core Team, 2017) was used. Correlation parameter from the bivariate probit model without random intercept is transformed according to Honaker, Owen, Imai, Lau, and King (2013).

Table 8: Alternative models estimating the individual participation propensity of students and parents for SC3 in Wave 2.

	No clustering		Clustering	
	Students (Ib)	**Parents (IIb)**	**Students (IVb)**	**Parents (Vb)**
Intercept	3.214^{***}	-2.205^{***}	3.769^{***}	-2.251^{***}
	(0.40S2)	(0.236)	(0.489)	(0.242)
Stratum *N5*	−0.201	0.565^{*}	−0.345	0.587^{*}
	(0.377)	(0.231)	(0.517)	(0.240)
Stratum *FS*	-1.591^{***}	0.030	-1.861^{***}	0.047
	(0.352)	(0.228)	(0.438)	(0.235)
Stratum *GY*	−0.453	0.702^{**}	−0.624	0.731^{**}
	(0.354)	(0.220)	(0.437)	(0.227)
Stratum *HS*	−0.492	0.225	−0.533	0.230
	(0.348)	(0.225)	(0.436)	(0.234)
Stratum *IG*	−0.398	0.724^{**}	−0.530	0.755^{**}
	(0.446)	(0.239)	(0.590)	(0.252)
Stratum *MB*	−0.391	0.417	−0.489	0.443
	(0.405)	(0.235)	(0.517)	(0.245)
Stratum *RS*	−0.450	0.535^{*}	−0.632	0.557^{*}
	(0.351)	(0.222)	(0.441)	(0.230)
Migration background	−0.418	−0.049	−0.490	−0.060
Turkish	(0.400)	(0.247)	(0.486)	(0.255)
Native language	0.246^{*}	0.121	0.335^{*}	0.117
German	(0.116)	(0.066)	(0.136)	(0.068)
Student participating in	−0.386	0.216^{*}	−0.483	0.229^{*}
Wave 1	(0.214)	(0.102)	(0.252)	(0.103)
Age group	0.100		0.127	
younger half	(0.097)		(0.115)	
Gender	0.118		0.138	
female	(0.087)		(0.101)	
Missing indicator for	-0.999^{***}		-1.135^{***}	
personal characteristics	(0.155)		(0.181)	
Individual retracking in	-2.640^{***}		-3.197^{***}	
Wave 2	(0.100)		(0.140)	
Number of calls		0.503^{***}		0.513^{***}
less than 4		(0.048)		(0.049)
Parent participating in		2.337^{***}		2.364^{***}
Wave 1		(0.051)		(0.052)
Random intercept				
σ school level			0.533	0.173
ln $\mathcal{L}$	−526.426	−1985.893	−510.032	−1983.092
AIC	1082.853	3997.786	1052.064	3994.184
BIC	1183.623	4085.120	1159.553	4088.236
Sample size	6112	6112	6112	6112

Notes: ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parenthesis. To model individual participation, the `glmer` and `glm` functions with a probit link provided by `lme4` (Bates et al., 2012) and `stats` package in R (R Core Team, 2017) was used.

Table 9: Results for the bivariate probit models without and with random intercept estimating the individual participation propensities for students and parents for SC3 in Wave 2.

	Bivariate probit – no clustering		**Bivariate probit – clustering**	
	Parents	**Students**	**Parents**	**Students**
	(IIIb)		**(VIb)**	
Intercept	-2.163^{***}	3.018^{***}	-2.212^{***}	3.368^{***}
	(0.235)	(0.395)	(0.239)	(0.469)
Stratum *N5*	0.525^{*}	-0.262^{*}	0.554^{*}	-0.479
	(0.230)	(0.370)	(0.237)	(0.466)
Stratum *FS*	0.001	-1.590^{***}	0.023	-1.836^{***}
	(0.226)	(0.347)	(0.230)	(0.412)
Stratum *GY*	0.661^{**}	-0.439^{*}	0.696^{**}	-0.635
	(0.219)	(0.349)	(0.223)	(0.406)
Stratum *HS*	0.186^{*}	-0.519^{*}	0.198	-0.609
	(0.224)	(0.343)	(0.229)	(0.405)
Stratum *IG*	0.685^{**}	-0.332^{*}	0.717^{**}	-0.540
	(0.238)	(0.452)	(0.247)	(0.525)
Stratum *MB*	0.376^{*}	-0.403^{*}	0.411	-0.567
	(0.234)	(0.398)	(0.242)	(0.471)
Stratum *RS*	0.499^{*}	-0.461^{*}	0.524^{*}	-0.709
	(0.221)	(0.346)	(0.226)	(0.408)
Migration background	-0.081	-0.376^{*}	-0.088	-0.517
Turkish	(0.246)	(0.396)	(0.250)	(0.452)
Age group		0.083^{*}		0.092
younger half		(0.096)		(0.101)
Native language	0.122^{*}	0.243^{*}	0.119	0.296^{*}
German	(0.066)	(0.115)	(0.067)	(0.124)
Gender		0.163^{*}		0.177
female		(0.087)		(0.091)
Missing indicator for		-0.933^{***}		-1.006^{***}
personal characteristics		(0.153)		(0.169)
Student participating in	0.217^{*}	-0.379^{*}	0.230^{*}	-0.419
Wave 1	(0.101)	(0.209)	(0.104)	(0.227)
Individual retracking in		-2.589^{***}		-2.899^{***}
Wave 2		(0.099)		(0.162)
Number of calls	0.493^{***}		0.501^{***}	
less than 4	(0.048)		(0.048)	
Parent participating in	2.337^{***}	0.308^{***}	2.361^{***}	0.327^{***}
Wave 1	(0.051)	(0.087)	(0.052)	(0.093)
Correlation	0.415^{**}		0.434^{***}	
ρ students parents	(0.158)		(0.052)	
Random intercept				
σ school level			0.181	0.347
ln $\mathcal{L}$	-2483.226		-2465.790	
AIC	5026.452		4995.580	
BIC	5227.993		4995.580	
Sample size	6112		6112	

Notes: ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively. Standard errors are given in parenthesis. To model individual participation decisions, the `zelig` function with `bprobit` link provided by `ZeligChoice` package (Owen et al., 2012) in R (R Core Team, 2017) was used. Correlation parameter from the bivariate probit model without random intercept is transformed according to Honaker et al. (2013).

C. R source code

C.1. Likelihood

```
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Likelihood for bivariate probit with random effects ---------------------
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
logLikBPRE <- function(param, yy1, yy2, xx1, xx2, k1, k2, m, n_j, S, crn1, crn2, unicrn){
  ### S: number of replications for GHK
  ### crn: martrix mit common random numbers der dim m x S
  ## likelihood evaluation for given starting values of the parameter values
  ## parameters to be defined
  beta1     <- param[1:k1]            # includes intercept
  beta2     <- param[(k1+1):(k1+k2)] # includes intercept
  rho       <- param[k1+k2+1]        # correlation parameter
  sig1      <- param[k1+k2+2]        # var for random intercept
  sig2      <- param[k1+k2+3]        # var for random intercept
  sig       <- matrix(1, 2, 2)       # corvariance matrix
  sig[1,1]  <- 1
  sig[2,2]  <- 1
  sig[1,2]  <- rho                   # correlation parameter
  sig[2,1]  <- rho                   # exp(rho)/(1+exp(rho))
  L         <- t(chol(sig))          # Lower triangular Cholesky root of covariance matrix
  alpha1    <- sig1 * crn1           # random effect 1
  alpha2    <- sig2 * crn2           # random effect 2 = random effect 1

  ### computation of the log-likelihood
  likeli <- rep(NA, m)
  for(j in 1:m){ # looping through the m schools
    gammaLower <- matrix(NA,S,2*n_j[j])
    gammaUpper <- matrix(NA,S,2*n_j[j])
    uniInd <- ((j-1)*S+1):(j*S)
    for(i in 1:n_j[j]){ # looping through the n_j individuals in school j
      ## mu = mu1 and mu2
      mu_ij <- -cbind(xx1[i, ,j] %*% beta1 + alpha1[j, ],
                      xx2[i, ,j] %*% beta2 + alpha2[j, ])
      ## upper truncation
      DUpper <- cbind((yy1[j, i] *  1000 + mu_ij[,1]),
                      (yy2[j, i] *  1000 + mu_ij[,2])) # A_rk bei Greene
      ## lower truncation
      DLower <- cbind(((1-yy1[j, i]) * (-1000) + mu_ij[,1]),
                      ((1-yy2[j, i]) * (-1000) + mu_ij[,2])) # B_rk bei Greene

      ## random numbers form truncated normal
      vhat  <- matrix(NA,S,2)
      ## quantiles for normal distribution
      ind <- (i-1)*2+1
      gammaLower[,ind] <- DLower[,1]/L[1,1]
      gammaUpper[,ind] <- DUpper[,1]/L[1,1]
      ## random numbers form truncated normal
      vhat[,1] <- qnorm(   unicrn[uniInd,1]  * pnorm(gammaUpper[,ind]) +
                        (1-unicrn[uniInd,1]) * pnorm(gammaLower[,ind])
      )

      gammaLower[,ind+1] <- (DLower[,2]-L[2,1]*vhat[,1])/L[2,2]
      gammaUpper[,ind+1] <- (DUpper[,2]-L[2,1]*vhat[,1])/L[2,2]
    }
    likeli[j] <- mean(apply(pnorm(gammaUpper)-pnorm(gammaLower),1,prod))
  }
  logLikelihood <- -sum(log(likeli))
  return(logLikelihood)
}
```

logLikBPRE.R

C.2. Computation

```
library(numDeriv)
library(randtoolbox)
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Estimation routine for bivariate probit with random effects ------------
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
BPREoptim <- function(DataRaw, y1, x1, su1, y2, x2, seed=NULL){
  ### DataRaw is a dataframe
  ### the rest of the arguments are the variable names from the dataframe
  nObs <- nrow(DataRaw)
  Y1  <- DataRaw[ ,y1]    # participation variable eq 1
  X1  <- DataRaw[ ,x1]    # includes a vector of 1 in first column
  SU1 <- DataRaw[ ,su1]   # random variable
  Y2  <- DataRaw[ ,y2]    # participation variable eq 2
  X2  <- DataRaw[ ,x2]    # includes a vector of 1 in first column

  ### 1: prepare data
  ####################
  m <- length(unique(SU1))      # number of schools
  n_j <- as.vector(table(SU1))  # number of children per school

  ## matrices and arrays containing (in-)dependant variables
  yy1 <- matrix(NA, m, max(n_j))
  yy2 <- matrix(NA, m, max(n_j))
  xx1 <- array(NA, dim=c(max(n_j), length(x1), m))
  xx2 <- array(NA, dim=c(max(n_j), length(x2), m))

  for(j in 1:m){
    ## matrix with y1 and y2 for participants per school
    yy1[j, 1:n_j[j]] <- as.vector(Y1[which(SU1 == unique(SU1)[j])])
    yy2[j, 1:n_j[j]] <- as.vector(Y2[which(SU1 == unique(SU1)[j])])
    ## array with X1 and X2 for participatns (nrow) per school (3D)
    xx1[1:n_j[j], 1:length(x1), j] <- as.matrix(DataRaw[which(SU1 == unique(SU1)[j]), x1])
    xx2[1:n_j[j], 1:length(x2), j] <- as.matrix(DataRaw[which(SU1 == unique(SU1)[j]), x2])
  }

  S      <- 1000          # number of draws for GHK
  k1     <- length(x1)   # number of independent variables + constant eq1
  k2     <- length(x2)   # number of independent variables + constant eq2
  rnd1   <- randtoolbox:::halton(S*m) # runif(S*m)
  rnd2   <- rev(rnd1) # halton(S*m) # runif(S*m)
  unicrn <- cbind(rnd1, rnd2) # common random numbers
  crn1   <- matrix(rnorm(S*m), m, S) # common random numbers
  crn2   <- matrix(rnorm(S*m), m, S) # common random numbers

  ## initial parameters
  theta <- c(rep(  0, (k1+k2)), 0,      0.1,     0.1) # initial parameters
  LB    <- c(rep(-5, (k1+k2)), -0.99,   0.001,   0.001)  # lower bounds
  UB    <- c(rep( 5, (k1+k2)),  0.99,   5,       5)      # upper bounds

  ## optimization
  ergMin <- nlminb(start=theta,
                   objective = logLikBPRE,
                   gradient = NULL, hessian = NULL,
                   yy1, yy2, xx1, xx2, k1, k2, m, n_j, S, crn1, crn2, unicrn,
                   control = list(trace=1, iter.max=500, rel.tol=10^-4),
                   lower = LB, upper = UB)

  ## Hessian matrix
  Hesse  <- numDeriv:::hessian(func = logLikBPRE,
                               x = ergMin$par,
                               method = 'Richardson',
                               method.args = list(),
                               yy1, yy2, xx1, xx2, k1, k2, m, n_j, S, crn1, crn2, unicrn)
  ## parameters, standard errors, t- and p-values
  beta   <- ergMin$par
  nPar   <- length(beta)
  seBeta <- sqrt(diag(solve(Hesse)))
  tValue <- beta/seBeta
```

```
  pValue <- 2*(1-pt(abs(tValue), nObs-length(tValue)))
  sig <- rep(NA, nPar)
  sig[pValue >= 0.1]  <- ''
  sig[pValue < 0.1]   <- '.  '
  sig[pValue < 0.05]  <- '*  '
  sig[pValue < 0.01]  <- '** '
  sig[pValue < 0.001] <- '***'

  ## output for coefs
  Final <- data.frame(beta, seBeta, tValue, pValue, sig,
                      row.names=c(paste(x1, ':1', sep=''),
                                  paste(x2, ':2', sep=''),
                                  'rho', 'sigma1', 'sigma2'))
  colnames(Final) <- c('Estimate', 'StdError', 'tValue', 'pValue', '')

  logLik <- -ergMin$objective
  AIC    <- 2*nPar-2*logLik
  BIC    <- nPar*log(nObs)-2*logLik
  ## output list
  OutList <- list('Coefficients' = Final,
                  'logLik'       = logLik,
                  'AIC'          = AIC,
                  'BIC'          = BIC,
                  'N'            = nObs,
                  'm'            = m,
                  'Hessian'      = Hesse,
                  'Optimierung'  = ergMin)

  return(OutList)
}
```

estimateBPRE.R