NEPS WORKING PAPERS

Christian Aßmann, Solange Goßmann, Benno Schönberger

BAYESIAN ANALYSIS OF BINARY PANEL PROBIT MODELS: THE CASE OF MEASUREMENT ERROR AND MISSING VALUES IN EXPLAINING FACTORS

NEPS Working Paper No. 35
Bamberg, March 2014

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

Working Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Working Papers publish articles, expertises, and findings related to the German National Educational Panel Study (NEPS).
The NEPS Working Papers are edited by a board of researchers representing the wide range of disciplines covered by NEPS. The series started in 2011.

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 NEPS Consortium.

The NEPS Working Papers are available at
https://www.neps-data.de/projektübersicht/publikationen/nepsworkingpapers

Editorial Board:
Jutta Allmendinger, WZB Berlin
Cordula Artelt, University of Bamberg
Jürgen Baumert, MPIB Berlin
Hans-Peter Blossfeld, EUI Florence
Wilfried Bos, University of Dortmund
Edith Braun, DZHW Hannover
Claus H. Carstensen, University of Bamberg
Henriette Engelhardt-Wölfler, University of Bamberg
Frank Kalter, University of Mannheim
Corinna Kleinert, IAB Nürnberg
Eckhard Klieme, DIPF Frankfurt
Cornelia Kristen, University of Bamberg
Wolfgang Ludwig-Mayerhofer, University of Siegen
Thomas Martens, DIPF Frankfurt
Manfred Prenzel, TU Munich
Susanne Rässler, University of Bamberg
Marc Rittberger, DIPF Frankfurt
Hans-Günther Roßbach, LIfBi
Hildegard Schaeper, DZHW Hannover
Thorsten Schneider, University of Leipzig
Heike Solga, WZB Berlin
Petra Stanat, IQB Berlin
Volker Stocké, University of Kassel
Olaf Struck, University of Bamberg
Ulrich Trautwein, University of Tübingen
Jutta von Maurice, LIfBi
Sabine Weinert, University of Bamberg

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

Bayesian analysis of binary panel probit models: The case of measurement error and missing values in explaining factors

Christian Aßmann, Solange Goßmann, Benno Schönberger
Leibniz Institute for Educational Trajectories, Germany
Department for Statistics and Econometrics, University of Bamberg, Germany

March 2014

E-Mail-Adresse des Autors:
christian.assmann@uni-bamberg.de

Bibliographische Angaben:

Aßmann, C., Goßmann, S. & Schönberger, B. (2014). *Bayesian analysis of binary panel probit models: The case of measurement error and missing values in explaining factors* (NEPS Working Paper No. 35). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Bayesian analysis of binary panel probit models: The case of measurement error and missing values in explaining factors

Abstract

Since large panel data sets, e.g. on educational or epidemiological issues, are despite tremendous efforts in field work almost inevitably plagued by missing data and measurement error, the development of appropriate estimation techniques is necessary. Bayesian analysis facilitated via Markov Chain Monte Carlo (MCMC) sampling algorithms allows for conceptually straightforward treatment of measurement error and missing values based on the device of data augmentation. Augmenting the parameter vector by the missing values allows for direct incorporation of the uncertainty stemming from missing values into parameter estimation. Full conditional distributions for missing values are provided on the basis of a nonparametric sequential regression modeling approach. For empirical illustration the proposed methodology is applied for students participating in a survey of the National Educational Panel Study (NEPS) assessing the impact of curricular reforms. The empirical application points at the necessity to cope with missing data and measurement errors in order to avoid biased estimation. Additionally a simulation study is performed documenting the adequacy of the proposed estimation methodology.

Keywords

MCMC ; Bayesian Analysis ; Binary Probit ; Measurement Error ; Panel Data; Nonresponse ; CART ; Metropolis-Hastings ; Missing Values ; Imputation

1. Introduction

Binary panel probit models serve as workhorse analytical tools, especially in the context of large panel surveys. Next to substantial empirical analysis addressing various topics in economics, sociology and demography, see e.g. Hyslop (1999); Fleming & Klera (2008); Holm & Jæger (2011) and Contoyannis et al. (2004), binary probit allow for documentation of panel participation and panel attrition. While analysing complete case data likelihood based estimation routines are available for all kinds of variations of the basic binary framework, e.g. extensions dealing with serial correlation and latent heterogeneity, see e.g. Albert & Chib (1993) or Casella & George (1992), only few approaches are documented within the literature dealing with analysis of incomplete data setups, see Münnich & Rässler (2005). A conceptually straightforward way to deal with missing values is provided within a bayesian framework via the device of data augmentation as suggested by Tanner & Wong (1987); Geman & Geman (1984) and Gelfand & Smith (1990). With the fundamentals of bayesian estimation provided by Albert & Chib (1993), augmenting the parameter vector with the missing values allows for incorporation of the uncertainty of missing values into parameter estimation via iterative sampling from the corresponding full conditional distributions. The main goal of this article is the integration of measurement error and missing values in a binary probit model framework. For demonstration, the participation status is used as binary variable as it opens up the method for inference about (non-) response. Additionally, an institutional context (hierarchical structure of the data) is assumed as it fits to the empirical application. The article is structured as follows. First, the general model formulation and estimation is described. Second, the empirical application data is introduced. Third, the setting and the results of a simulation study are presented. Fourth, the results of the empirical application are described and finally a summary concludes.

2. Model formulation and estimation

We want to handle the uncertainty of measurement error and item nonresponse within parameter estimation. As we use the participation status as dependent variable, a binary probit or logit model is suitable to detect possible effects of covariates on the decision to participate or not in the study. We decided in favor of the probit model, because the inherent distributional assumptions in comparison to the logit assumptions are more suitable for our model, where we want to perform a bayesian model parameter estimation and some imputations. When a standard probit regression is used the question whether selection effects exist is decided based on data neglecting the hierarchical struture of the data. Furthermore, it is obvious, that an estimation solely based on complete cases would completely ignore the additional information that can be contributed by the integration of partially missing information. As we want to incorporate as much information as possible to get valid estimates, ways have to be found how to cope with this challenge. The current literature on missing values, measurement errors or imputation methods (e.g. Rubin (1987); Little (1992); Albert & Chib (1993); Schafer (1997); Xie & Paik (1997); Schafer (1999); Raghunathan et al. (2001); Dunson et al. (2003); Nasrollahzadeh (2007)) suggest the use of a bayesian Gibbs sampling approach, where all kinds of parameters can be estimated and uncertainty due to missing data can be incorporated as well. The different parts of our estimation algorithm will be explained in detail in the following. Our general model framework is based on a bayesian probit regression model with second level random effects and will be described first. As we have to deal with some kinds of missing values, the steps necessary to cope with this problem are discussed in the second part of this chapter. Missing independent values are imputed via an adapted classification and regression tree (CART) approach originally described by Breiman et al. (1984) and recently used for imputation by Burgette & Reiter (2010). Uncertainty arising from missing values in the dependent Y variable of the binary probit model is incorporated via an Metropolis-Hastings based imputation step (Chib & Greenberg, 1995) we call *Probit Forecast Draw*.

Note that the description uses an assumed school as institutional context and students as individuals as it eases the transfer to the empirical application and is more descriptive. First of all, a binary probit model with random second-level-effects will be set up in order to correctly gauge possible school level effects and the hierarchical structure of students being in homogenous contexts and learning environments. This model allows for the integration of several imputation steps needed to deal with the uncertainty in the data due to the mentioned measurement error and missing values as well. Therefore, a Gibbs sampling approach is to be defined in the second part of this section. For the specification of the bayesian probit model, let y_{ij} denote the observed dichotomous variable, with $i = 1, \ldots, N_j$ and $j = 1, \ldots, J$, where N_j denotes the number of students in school j and J denotes the number of schools. Then a link between observed explaining factors and the observed binary variables is provided via the latent variable z_{ij}:

$$y_{ij} = \begin{cases} 1, & \text{if } z_{ij} \geq> 0, \\ 0, & \text{if } z_{ij} < 0, \end{cases} \tag{1}$$

where $z_{ij} = X_{ij}\beta + u_j + e_{ij}$ and e_{ij} is an independent identically normally distributed error term with unit variance and u_j a cluster-specific random error term with $\mathcal{N}(0, \sigma_u^2)$. Pooling hence yields the complete likelihood

$$\mathcal{L}_P(Y|\beta, X, u_j) = \prod_{j=1}^{J} \prod_{i=1}^{N_j} \Phi\left[(2y_{ij} - 1)(X_{ij}\beta + u_j)\right], \tag{2}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of a standard normal distribution. The posterior distribution using data augmentation (Tanner & Wong, 1987) is then:

$$\begin{aligned} \pi(\beta, Z, u_j, \sigma_u^2|Y, X) \propto \quad & \pi(\beta)\pi(\sigma_u^2) \prod_{j=1}^{J} \prod_{i=1}^{N_j} \{I(z_{ij} > 0)I(y_{ij} = 1) \\ & + I(z_{ij} \leq 0)I(y_{ij} = 0)\}\phi(z_{ij}; x'_{ij}\beta + u_j, 1)\phi(u_j; 0, \sigma_u^2) \end{aligned} \tag{3}$$

This joint distribution is complicated in the sense that it cannot be sampled from directly. But via Gibbs sampling and the calculation of the full conditional distributions the marginal posterior distribution of β can be estimated. Following Nasrollahzadeh (2007), the posterior distribution of β given Y, u, σ_u^2, Z and X ist then

$$\pi(\beta|Y, u, \sigma_u^2, Z, X) \approx \pi(\beta) \prod_{j=1}^{J} \prod_{i=1}^{N_j} \phi(z_{ij}; x'_{ij}\beta + u_j, 1). \tag{4}$$

With the proper conjugate $\mathcal{N}(\beta^*, B^*)$ prior it follows $\mathcal{N}(\tilde{\beta}, \tilde{B})$ with parameters

$$\tilde{\beta} = (B^{*-1} + X'X)^{-1}(B^{*-1}\beta^* + X'(Z - u)) \tag{5}$$

and

$$\tilde{B} = (B^{*-1} + X'X)^{-1}. \tag{6}$$

The posterior distribution of $Z|X, \beta, u_j, \sigma_u^2$ also has a simple form truncated normal distribution $\mathcal{N}(x'_{ij}\beta + u_j, 1)$, truncated at the left by 0 if $y_{ij} = 1$ and truncated at the right by 0 if $y_{ij} = 0$. The full conditional of $u_j|Z, X, \beta, \sigma_u^2$ is given as a $\mathcal{N}(\mu_{u_j}, \sigma_{u_j}^2)$ with parameters

$$\mu_{u_j} = \left(N_j + \frac{1}{\sigma_u^2}\right)^{-1} \left(\sum_{i=1}^{N_j} (z_{ij} - x'_{ij}\beta)\right) \tag{7}$$

and

$$\sigma_{u_j}^2 = \left(N_j + \frac{1}{\sigma_u^2} \right)^{-1}. \tag{8}$$

The covariance matrix σ_u^2 of the random coefficients is sampled from independent inverse gamma distributions $\mathcal{IG}(\alpha_{\sigma_u^2}, \beta_{\sigma_u^2})$ with parameters

$$\alpha_{\sigma_u^2} = \frac{J}{2} + \alpha_{\sigma_u^2}^0 \tag{9}$$

and

$$\beta_{\sigma_u^2} = \frac{1}{2} \sum_{j=1}^{J} u_j^2 + \beta_{\sigma_u^2}^0 \tag{10}$$

where the parameters of the conjugate inverse gamma prior distribution $\mathcal{IG}(\alpha_{\sigma_u^2}^0, \beta_{\sigma_u^0}^0)$ are $\alpha_{\sigma_u^2}^0 = 1$ and $\beta_{\sigma_u^2}^0 = 1$.

Estimation

To determine the necessary steps for the estimation and imputation procedures, four different data situations can be distinguished (see figure 1): complete cases (1), cases with missing participation indicator and complete values in explanatory variables (2), cases with complete participation indicator and missing values in explanatory variables (3) and cases with missing values in the participation indicator and in explanatory variables (4). Whereas situation (1) is straightforward to estimate with standard probit regression, missing values for cases (2)-(4) are handled via two different imputation approaches, that will be explained in the following.

CART is a non-parametric algorithm for recursive partition respective to the dependent variable and was invented by Breiman et al. (1984). CART divides the data into disjoint subgroups within each partition step. Binary splits are used for the partition comparable to e.g. QUEST (Loh & Shih, 1997) and GUIDE (Loh & Vanichsetakul, 1988), whereas other procedures, e.g. FIRM (Hawkins, 1990), use multiway splits. The split criterion for each of those partition steps is the maximization of the decrease of heterogenity. As measure of heterogenity the variance for continuous and the entropy for non-continuous variables is used. The usage as imputation procedure has been shown by Burgette & Reiter (2010) and is described in the following.

When imputing and using multivariate imputation by chained equations (MICE), conditional models have to be specified for all variables with missing data, including interactive and non-linear relations between variables when they occur. When the knowledge about the conditional distribution is low Burgette & Reiter (2010) propose a CART-based MICE, defined as follows. When using MICE they replace filling in initial values by draws from the predictive distribution conditional on the matrix of all variables with complete and yet imputed cases and also drawing from the predictive distribution conditional on Y_{-i} (all variables besides the one that has to be imputed) by CART using a bayesian bootstrap. Using the CART approach of Burgette & Reiter (2010) simplifies the users effort to impute the data as it is based on recursive partitioning and not on defined conditional models.
Our approach is an extension and adaption of Burgette & Reiter (2010). When defining initial values, they use CART creating a decision tree and drawing an existing value of a leaf that has the same predictive distribution. The ability to create this tree is limited by the structure of missing values in the data. If there are only a few or no variables without missing values, the

creation of a tree is difficult or impossible. What we suggest therefore, is to sample initial values unconditional referred to the empirical data with replacement. It can be shown that the effect of sampling those initial values has no effect when the CART approach is repeated and updated several thousand times.
An innovation of our approach is, that possible clusters in the data are regarded. CART is using the effect of a cluster, for example a common mean value, to impute by describing the effect by the value of the cluster. But when imputing the missing value for all elements within a cluster, for every element a single draw is taken. So we built up a two-CART-model. The first step is the CART approach as described above, using the information of the initial value for the cluster level, which has to be the same for all elements within this cluster, which means that sampling initial values has to be adapted, too. In this first CART model all values are updated by draws from the predictive distribution conditional besides the cluster variables. Then a second CART model is done using the information of all variables aggregated on cluster level. The values that are drawn can then be added to the data for the single elements.
The most important extension and adaption is, that CART is not used as a single approach to impute the data, but within a Gibbs sampler in combination with a bayesian probit analysis which will be described in the following.

Additional to missing values in explanatory variables X, some values of the dependent binary variable Y are missing as well. In the standard case of listwise or casewise deletion - the default in most statistical software - the whole information of these cases would be neglected in a model with Y being missing. According to Little (1992) „cases with Y missing can provide [only] a minor amount of information for the regression of interest“. Nevertheless, as we want to incorporate as much information as possible in our approach, we impute these missing values and use all the information on X available for our estimation in the next step. As Y is the dependent binary variable of the probit model, it is necessary to draw new 0/1 coded values for the missings. We use an adapted Metropolis-Hastings algorithm for that purpose see e.g. Chib & Greenberg (1995).
Suppose the individuals can be classified according to some known classification criteria such as country, state, region, affiliation to organizations or - as in our empirical application - classes of students. Then the data set can be clustered into B different blocks and the procedure can be applied for each block separately. Whether there is additional information regarding the maximum number (in our application the maximum is known for each class of students, conditional on sex) or upper bound $\overline{c_b}$ of missing Y's being 1 within $b = 1, \ldots, B$ defined blocks of cases with $i = 1, \ldots, N_b$ cases within each block is available, this approach even yields more appropriate and probable values. Then, the candidates for the Metropolis-Hasting-sequence are generated as follows:
As a first step, one possible candidate set $Y_b^{(t)}$ with $\sum_{i=1}^{N_b} y_{bi} \leq \overline{c_b}$ is sampled to replace the missing values in Y_b. The probability of this candidate set given the data X and the parameters β and u_j

$$p_b^{(t)} = \prod_{i=1}^{N_b} Pr(Y_{bi} = 1|X_{bi} = x) = \prod_{i=1}^{N_b} \Phi(x'_{bi}\beta + u_j) \tag{11}$$

compared to the probability of the candidate set $p_b^{(t-1)}$ then defines the probability of move from $Y_b^{(t-1)}$ to $Y_b^{(t)}$:

$$prob(Y_b^{(t-1)}, Y_b^{(t)}) = \begin{cases} \min\left[\frac{p_b^{(t)}}{p_b^{(t-1)}}, 1\right], & \text{if } p_b^{(t-1)} > 0 \\ 1, & \text{otherwise.} \end{cases} \tag{12}$$

This Metropolis-Hastings sequence runs T times with $t = 1, \ldots, T$ (in our case $T = 10$) and the final candidate set $Y_b^{(T)}$ for each block is used to replace missing values in Y. The complete $Y^{(T)} = \{Y_{obs}, Y_b^{(T)}\}$ is then used in the next sequence of the Gibbs sampler as the dependent variable for the probit model. As this imputation approach is used to predict missing values in Y for our bayesian probit model by drawing new values we call this part **Probit Forecast Draw**.

Now the estimation routine is presented on the whole. Based on the data augmentation device proposed by Albert & Chib (1993) for bayesian estimation of binary probit models, we establish a Gibbs sampler dealing the model complications of measurement error within the dependent and missing values within the explaining factor. The estimation routine then consists out of the following steps:

Step (I):	Initialize all missing values and parameters
Step (Ia):	Unconditionally draw new values for X_{mis} from X_{obs}
Step (Ib):	Maximum likelihood estimation results based on complete cases provide starting values for the β coefficients (informative prior for β)
Step (Ic):	Generate one run of the Metropolis-Hastings sequence to draw new values for Y_{mis} (measurement error) based on the complete values from (Ia) and (Ib)
Step (II):	Generate new values for X_{mis} for level 1 and level 2 from full conditional distributions provided by CART analysis
Step (III):	Generate one run of the Metropolis-Hastings sequence to draw values for Y_{mis} (measurement error) based on the complete values from (Ib) and (II)
Step (IV):	Generate new random effects variance-components σ_u^2 and u_j
Step (V):	Calculate new β coefficients based on steps (II), (III) and (IV)
Step (VI):	Return to step (II)

Following Cowles & Carlin (1996) the convergence of a markovian updating scheme or moreover the accuracy of the resulting estimates from the joint posterior distribution of interest strongly depends on the length of the sampling sequence. Some authors already examined the possibility to use multiple parallel chains and to combine the results. For an overview of the literature see Cowles & Carlin (1996). In their summarization they point out the inefficiency of discarding many initial values from multiple chains for the burn-in-phase and emphasize that the last iterations of single long chains are likely to be closer to the true distribution than those reached by any of the shorter chains compare Raftery & Lewis (1992). In order to examine possible dependencies from initial values or the specification of priors we chose to run multiple long chains of length $M = 20000$ $(m = 1, \ldots, l, \ldots, M)$ with various starting values. The results from the Gibbs output then have to be examined to determine the length of the necessary burn-in for values that have to be discarded for more correct estimates (see chapter 5. for details). The values from the remaining iterations after the burn-in-phase then have to be combined. With the burn-in length determined at iteration l the Bayes posterior mean vector of unknown parameters $\widehat{\Theta_i} = \{\widehat{\beta}, \widehat{\sigma_u^2}\}$ is then the mean of the remaining values

$$\overline{\widehat{\Theta}} = \frac{1}{M - l} \sum_{i=l+1}^{M} \widehat{\Theta_i}. \tag{13}$$

To correctly gauge the relevant effects for our probit analysis of participation, the inspection of marginal effects is necessary, too. Following Aßmann & Boysen-Hogrefe (2011), the marginal effects, evaluated for a particular value of the covariates x^*, which are represented by the mean of all sample covariates in the case of continuous variables or the mode for binary variables, with

Θ summarizing all model parameters, is then calculated as

$$\frac{\partial}{\partial x} Pr(y = 1|x = x^*, \Theta) \tag{14}$$

for continuous variables and for binary variables it is given as

$$(-1)^{1-x^*}(Pr(y = 1|x = x^*, \Theta) - Pr(y = 1|x = 1 - x^*, \Theta)). \tag{15}$$

The specified priors for all parameters can be found in table 3.

3. Empirical data

The two data sets we used for the estimations stem from an organisational reform study of the National Educational Panel Study (NEPS) - a large panel with more than 150 different single studies. Among many other topics the NEPS also conducted a study in Thuringia with two waves in 2010 and 2011 to assess possible effects on the competence development of students after an organisational reform of the „Gymnasiale Oberstufe“ in Thuringia. This study was conducted at 32 upper secondary schools for the graduation years 2010 (last year group which is not affected by the reform) and 2011 (first reformed year group), where all students of the 12th grade at the selected schools were tested and questioned one time (provided given consent). To record possible effects of the reform, achievement tests (Fachleistungstests) in the fields of mathematics, physics, biology and English, questions about the students' social background, a test on cognitive abilities as well as questionnaires were applied. In addition, the parents and subject teachers of the 12th grade were integrated in the survey. In total, 1857 students were asked for participation in 2010 and 1374 in 2011.
Because of the voluntary nature of this study some students, parents or teachers chose not to take part. Crucial for the definition of the participation status is only the participation of the students. Although the NEPS made enormous efforts in order to achieve a high data quality and especially to get all the requested data, some minor errors occurred. Some schools refused to follow the detailed instructions regarding the delivery process of the students' marks (complete marks of students were not available at the time of testing because the school year was not finished yet. The marks should be delivered by the school after the final exams for all students). As a consequence, partially missing mark information has been delivered or schools completely refused to pass on information about students that did not participate in the study. Therefore we have to deal with a small amount of missing data for the students' marks. Additionally another type of uncertainty comes into play. The concrete procedure of data privacy protection required all individual identifiers (IDs) to be removed for all nonparticipating students from the mark file by the school. This partially anonymous data set then has been sent to the NEPS. During the matching of these data sets it came clear, that some schools failed to exactly record the actual participation status of their students and therefore in some cases erroneously removed IDs for participants. In these cases we have to deal with some kind of uncertainty regarding the actual participation status. How we model these measurement errors will be in more detail explained in chapter 2..
After these shortcomings of the first waves' data delivery procedures were known, they have been refined for the 2nd wave in order to automate as many processes as possible and to obtain an even higher data quality. The students' marks for the 2nd wave have therefore been automatically extracted from the school mark software (schools that do not use software for mark management had to extract the marks manually from the students' files). The resulting improvement of data quality and the reduction of missing values can be seen in table 1.
As german students in upper secondary schools have to choose their subjects for the final two years from three different fields of subjects, we also used this structure to avoid an overspecification of our models through the integration of many single mark variables. Therefore we

calculated mean marks for every student according to the three fields of subjects. These fields of subjects are (1) linguistic-literary-artistic subjects (e.g. german, english, arts, music), (2) social subjects (e.g. geography, history, religion) and (3) mathematical-natural-scientific-technical subjects (e.g. maths, physics, biology, computer sciences). Although the rules of subject choice were slightly changed from the 1st to the 2nd wave of the study through the reform process, the mean marks for the students of the 2nd wave were calculated on the basis of the 1st wave structure in order to maintain comparability between the estimations. Our assumptions is, that the reform has no causal inference on nonresponse or selectivity and has therefore no effect on our analysis. Further information to be included in the estimation is the sex of the students as well as the multi-level structure of the students' population. Students being in the same school are exposed to the same learning environment or context effects and tend to make similar decisions. This structure of dynamic decision making within the class context especially considering the decision to participate or not in the NEPS study has to be taken into account. Due to little information on contextual variables in the data sets, we chose to integrate a mean school mark variable into our estimations to reflect the hierarchical data structure as good as possible regarding the data situation. The mean school mark represents the mean of all final exam marks of all students within a school.
In summary, for our estimation we have to deal with a hierarchical data structure and shortcomings of the data delivery processes in terms of measurement error and missing values. To estimate possible selection effects we are in the comfortable situation that additional mark information is available for respondents as well as for nonrespondents. The used variables in our models are the actual participation status as the dependent variable and sex, the mean marks in three fields of subjects as well as the mean school mark as covariates.

4. Simulation Study

In order to efficiently assess the accuracy of our estimation routine in different settings a small simulation study has been set up. As the estimator comprises of various components dealing with missing values on the one side and measurement errors on the other side, we decided to check the estimators performance in four different scenarios analogous to the empirical application: (1) complete cases only (2) with missing values (3) with measurement error and (4) missing values and measurement error. Therefore we decided to construct 20 artificial complete data sets of size $N = 4000$ with $N_j = 40$ students being in $J = 100$ schools.

Data generating process

The structure of this artificial data set should be similar to that of our real data sets. Therefore 3 continuous variables $x_{ij} = \{x_{ij}^1; x_{ij}^2; x_{ij}^3\}$ have been sampled from a multivariate truncated normal distribution $\mathcal{MVTN}(M, V)$, where M equals the means of three variables from the real 2010 data set and V is the corresponding variance-covariance matrix, respectively. The truncation points are set to the empirically observed minimum and maximum values of the corresponding variables from the real data set in order to simulate values as similar as possible. Additionally and in contrast to the original data set 2 binary variables $d_{ij} = \{d_{ij}^1; d_{ij}^2\}$ have been simulated, where d_{ij}^1 is unconditional on all other variables with

$$Pr(d_{ij}^1 = 1) = 0{,}45. \tag{16}$$

These values of the first binary variable have been arbitrarily set in order to simulate values similar to the distribution of the sex variable in the real data sets. The second binary variable

d_{ij}^2 has been sampled conditional on the continuous variable x_{ij}^1 according to

$$Pr(d_{ij}^2 = 1) = \phi(0{,}9 \cdot x_{ij}^1). \tag{17}$$

Furthermore, for the school level variable x_j, 100 values were randomly sampled from x_{ij}^3 and were randomly assigned to the schools. The construction of the binary dependent variable z_{ij} of the probit model was done via a linear regression model with arbitrarily chosen parameters $\tilde{\beta} = \{1.5, 0.1, 0.05, 0.20, 0.30, -1.5\}$ providing a latent variable

$$z_{ij}^* = \tilde{\beta}_{(1)} \cdot const + \tilde{\beta}_{(2)} \cdot d_{ij}^1 + \tilde{\beta}_{(3)} \cdot x_{ij}^1 + \tilde{\beta}_{(4)} \cdot x_{ij}^2 + \tilde{\beta}_{(5)} \cdot d_{ij}^2 + \tilde{\beta}_{(6)} \cdot x_j + e_{ij} + u_j \tag{18}$$

with e_{ij} as an individual specific standard normal distributed error term and u_j as a school level specific error term following a $\mathcal{N}(0, 0.8)$. The binary variable z_{ij} for the probit model of the estimation routine is then

$$z_{ij} = \begin{cases} 1, & \text{if } z_{ij}^* \geq 0, \\ 0, & \text{if } z_{ij}^* < 0. \end{cases} \tag{19}$$

As a next step, missing values were randomly generated according to the proportions given for each scenario in table 4. Whereas the missingness patterns for the variables $z_{ij}, d_{ij}^1, d_{ij}^2$ and x_j are independent of each other, we decided to sample missing values for the variables x_{ij}^1 and x_{ij}^2 together. Therefore in 8% of all cases, these two variables are both missing. Note that the missing values applied for scenario 2 are the same as in scenario 4 whereat the measurement error in scenario 4 is identical to that generated for the scenario 3 data set. Due to computational costs we only performed $N_{sim^s} = 20$ independent runs with $M = 20000$ Gibbs iterations for each of the four scenarios $s = 1, \ldots, 4$. To check the robustness of these estimates we additionally tested scenario (1) with $N_{sim^1} = 100$ runs. Tables 5 and 6 provide the respective pooled estimates. The coverage was calculated as

$$coverage = \frac{\sum_{i=1}^{N_{sim^s}} \tilde{\beta} \in KI[\overline{\widehat{\Theta}}]^i}{N_{sim^s}}, \tag{20}$$

and $HDR[\overline{\widehat{\Theta}}]^i$ is the 95% high density region of the Bayes posterior means of all parameters of scenario i. The coverage is therefore the percentage of how often the true parameter values $\tilde{\beta}$ are covered by the estimated values. By construction it should be around 95% if the estimation is correct. Furthermore, the overall mean of the means, the standard deviation of the means and the overall mean of the standard deviation as well as the standard deviation of the standard deviations have been calculated.

Simulation results

The results from this small simulation study point to the efficiency of our new estimation and imputation routine. The coverage for all four scenarios and all single estimates look promising. Although only 20 runs have been performed, the overall means of the bayesian posterior mean coefficients are very close to the true values. With a look at the results from an additional analysis with 100 runs, which was only performed for scenario (1) (see table 6) due to computational costs, this impression is even reinforced. Consequently, coverage rates below 90% or slight differences between the estimated and the true values seem to be directly attributable to the small number of performed simulation runs. Another quality criterion is that the standard deviations of all means and the means of all standard deviations are quite similar.
Therefore, with this small simulation study we have proof of the efficiency of our estimation and

imputation routine with respect to the estimated parameters. Our estimator seems to adequately sample values from the distribution of interest, even if missing values or measurement errors have occurred. With these findings in mind we will now turn to the inspection of the results from the empirical application with NEPS data.

5. Results

One of the main goals of our work was the estimation of possible selection effects based on two data sets provided by the National Educational Panel Study, where the effects of a curricular reform study have been evaluated for students within grammar schools in Thuringia in 2010 and 2011. First results from a standard probit regression model with complete cases indicate the presence of selection effects as can be seen in the upper part of table 2. It seems that participation in the study depends significantly on sex and the marks in mathematical-natural-scientific-technical subjects (e.g. maths, physics, biology, computer sciences). Without further examination of the data the analyst would assume the data to be distorted and the calculation of nonresponse correcting weights would be necessary. Through the incorporation of the hierarchical data structure with students being within homogeneous contexts of schools through a standard bayesian probit model with random effects, these findings turn out to be misleading. The lower part of table 2 provides the respective estimates. The consideration of second-level random effects for schools leads to nonsignificant estimates regarding the covariates. Assuming the probit regression with random second level effects to be the right model the participation in the NEPS study is seemingly unrelated to covariates and can be considered as a random process at the individual level. This conclusion holds for both data sets. The question is whether these results change through the incorporation of the additional uncertainty stemming from item and unit non response. In order to answer this question we constructed a new estimation routine consisting of several steps within a bayesian framework using a Gibbs sampling approach. All missing values in the participation status as dependent variable and in explaining factors of the binary probit model have been imputed within each iteration of the approach in order to approximate the posterior distribution of all parameters and covariates. Plots for the cumulative means can serve as a visual tool to determine the duration until the Gibbs sampler reaches stability and how many iterations have to be deleted as a burn-in. Furthermore, the convergence behavior of the Gibbs sampler can be inspected (see figure 2). For all parameters the stationary distribution is reached after 5000 initial iterations. Therefore, further calculations of posterior means, standard deviations or high density regions (HDR) are based on the last 15000 iterations. The trace plots for the whole Gibbs sequence (see figure 3) show a good mixing behavior of our Gibbs estimator for the whole bandwidth of the distribution. If the sampler would be stuck in certain regions of the distributions for many successive iterations, then a refinement of the estimation routine would be necessary. The iterative nature of Gibbs sampling or markov chains in general also entails the fact, that successive iterations always depend on the preceding iterations. Therefore the autocorrelation functions (ACF) for each parameter have to be inspected, too (see figure 4). For the individual level the ACF plots for covariates *sex* and the three *mark* variables look very good with only moderate dependencies up to the last 10 iterations. Due to the construction of the second level random effect, a higher dependency can be observed for the intercept as well as the school level variable *mean school mark*, as has been expected.

The Bayes posterior means, standard deviations and high density regions of the coefficients are given in table 7. Although in many applications especially with long single chain Gibbs samplers the dependency of the results from specified priors can be neglected Lopes & Tobias (2011), estimation results stemming from different prior specifications (estimates with different precision for the β priors, where the precision is the reciprocal of the prior variance) are presented,

too. In our application, the results are independent of the chosen prior specifications and do not change substantially. Nevertheless, for out future work on this topic, further evaluation of different starting parameters especially regarding the parametrization of the random effect sampling components and different β priors is planned.
The inspection of the bayesian posterior means provides an interesting picture with respect to the initial results from the standard probit regression without and with random effects. Whereas the standard model based on complete cases could lead to the conclusion that significant selection effects can be observed and the participation probability strongly depends on convariates, no significant results could be observed anymore for the model with integration of the random effect components to reflect the hierarchical data structure. For the 2010 data set all of our new bayesian probit estimations provide nearly identical results. The participation in the NEPS study is significantly depending on the students' sex. Girls tend to participate at a higher level than boys, although the marginal effect is quite small with about 4% (table 8) in 2010 and 2011. The fact, that in model (III.2) the effect of sex is not significant is considered as a random variation due the estimation process. This assumption is furthermore supported by the corresponding high density region, which is including the null.
Thus, these results bring us to the following conclusion. Standard probit regression can be heavily distorted and lead to the wrong results because so much information is neglected through the concentration on complete values. Also the structure of the data or the examined population respectively, has to be taken into account to find the „true“ model. Furthermore the integration of uncertainty due to the missing data within bayesian estimation routines is extremely important to get estimates that are close to reality. Of course, the validity of these results and our new estimation routine in general will in detail be tested in our future work.

6. Summary

This paper focuses on bayesian estimation of binary probit models with measurement error in the dependent variable and missing values in model covariates. Our starting point for this kind of analysis was an analysis of possible selection effects in data sets from a curricular reform study of the National Educational Panel Study in grammar schools in Thuringia in 2010 and 2011. These data sets provide detailed information on all marks of the two graduation years for participating students and nonparticipating students as well and are therefore well suitable for a sophisticated analysis of participation rates.
We set up a binary probit model for the estimation of participation in the NEPS study with some covariates. A bayesian approach was chosen in order to explicitly include uncertainty due to missing values in model covariates and measurement error in the dependent variable. For the imputation of missing values in model covariates we adapted the Classification and Regression Tree imputation approach from Burgette & Reiter (2010) for the hierarchical sampling structure of the NEPS data set. A Metropolis-Hastings updating procedure (Chib & Greenberg, 1995) was used to sample possible candidates for missing values in the actual participation status stemming from measurement error in the data collection. A Gibbs sampling approach then allows to simulate draws from the posterior distribution of interest via iterative sampling from the full conditional distributions of all model parameters conditional on the data. A small simulation study was performed in order to check the estimators efficiency in terms of correct parameter estimation with missing information. Therefore we divided the estimation routine into its single components and tested them in four different scenarios with 20 runs each. Furthermore, sensitivity checks have been performed with the empirical data in order to assess estimation stability and the dependence from prior assumptions. In this paper we only varied the precision of the coefficient prior, but further analysis on this topic as well as a more extensive simulation study on model stability and estimation precision is already planned for future work. This new

estimation and imputation routine can naturally be extended from the binary to the multinomial probit regression case and is therefore flexible enough to answer a lot of analytical questions. With a few adaptations it is also suitable to the imputation of even more complex data situations including filter questions or Rasch scaled competence scores.
The empirical illustration is based on the NEPS data sets. It has been shown that the inclusion of additional information from partially missing data, that would completely be ignored in standard probit regression models based on complete cases (cc) leads to reasonable results regarding the model coefficients. Participation in the NEPS study is significantly dependent from the students gender as girls tend to participate at a higher level than boys. These results, especially in contrast to results from complete case analysis, highlight the importance to include all available information into the model to avoid incorrect model specifications and wrong conclusions.

References

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422), p 669–679.

Aßmann, C. & Boysen-Hogrefe, J. (2011). A bayesian approach to model-based clustering for binary panel probit models. *Computational Statistics and Data Analysis*, 55(1), pp. 261–279.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* New York: Chapman and Hall.

Burgette, L. F. & Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology*, 172(9), 1070–1076.

Casella, G. & George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3), 167–174.

Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4), pp. 327–335.

Contoyannis, P., Jones, A. M., & Rice, N. (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, 19(4), 473–503.

Cowles, M. K. & Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434), pp. 883–904.

Dunson, D. B., Watson, M., & Taylor, J. A. (2003). Bayesian latent variable models for median regression on multiple outcomes. *Biometrics*, 59(2), 296–304.

Fleming, C. M. & Klera, P. (2008). I?m too clever for this job: a bivariate probit analysis on overeducation and job satisfaction in australia. *Applied Economics*, 40(9), 1123–1138.

Gelfand, A. & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398–409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions and the baysian restoration of images. *IEEE Trans. Pattern Anal. & Mach. Intell.*, 6, 721–741.

Hawkins, D. (1990). Firm: Formal inference-based recursive modeling. *Technical Report 546, School of Statistics, University of Minnesota.*

Holm, A. & Jæger, M. M. (2011). Dealing with selection bias in educational transition models:the bivariate probit selection model. *Research in Social Stratification and Mobility*, 29(3), 311–322.

Hyslop, D. (1999). State dependence, serial correlation and heterogeneity in intertemporal labour force participation of married women. *Econometrica*, 3(6), 1255–1294.

Little, R. J. A. (1992). Regression with missing x's: A review. *Journal of the American Statistical Association*, 87(420), pp. 1227–1237.

Loh, W.-Y. & Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7, 815–840.

Loh, W.-Y. & Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with discussion). *Journal of the American Statistical Association*, 83(403), 715–728.

Lopes, H. F. & Tobias, J. L. (2011). Confronting prior convictions: On issues of prior sensitivity and likelihood robustness in bayesian analysis. *Annual Review of Economics*, 3, pp. 107–131.

Münnich, R. & Rässler, S. (2005). Prima: A new multiple imputation procedure for binary variables. *Journal of Official Statistics*, 21(2), 325–341.

Nasrollahzadeh, S. (2007). The analysis of bayesian probit regression of binary and polychotomous response data. *IJE Transactions B: Applications*, 20(3), 237–248.

Raftery, A. E. & Lewis, S. (1992). How many iterations in the gibbs sampler? *Bayesian statistics*, 4(1), pp. 763–773.

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27, 85–95.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Boca Raton and London and New York and Washington D.C.: Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8(1), 3–15.

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), pp. 528–540.

Xie, F. & Paik, M. C. (1997). Multiple imputation methods for the missing covariates in generalized estimating equation. *Biometrics*, 53(4), pp. 1538–1546.

Appendix

Table 1: Overview of missing values

nr.	variable	percent of missings	
		2010	2011
1	participation status	1.3 %	0.4 %
2	sex	3.8 %	2.4 %
3	field of subjects 1	11.9 %	5.0 %
4	field of subjects 2	12.0 %	5.0 %
5	field of subjects 3	12.1 %	5.0 %
6	mean school mark	1.0 %	0.0 %
	complete cases	**85.0 %**	**93.6 %**

Table 2: Comparison of a standard probit model without and with random effects

	2010				2011			
	Standard Probit regression (complete cases: N=1578)				Standard Probit regression (complete cases: N=1304)			
	Estimate	Std. Error	95% CI		Estimate	Std. Error	95% CI	
Intercept	2.2685	0.5812	1.1482	3.4018	0.7200	0.6150	-0.4892	1.9320
sex	-0.1349	0.0763	-0.2842	0.0143	-0.2325	0.0832	-0.3963	-0.0689
fs1	-0.0703	0.0318	-0.1324	-0.0084	-0.0127	0.0340	-0.0792	0.0538
fs2	0.0217	0.0285	-0.0340	0.0773	-0.0141	0.0302	-0.0729	0.0447
fs3	0.0496	0.0211	0.0083	0.0909	0.0468	0.0217	0.0057	0.0879
msm	-0.6324	0.2181	-1.0550	-0.2166	-0.1672	0.2336	-0.6293	0.2934
Bayes Probit incl random effects (complete cases)					Bayes Probit incl random effects (complete cases)			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	1.9000	1.7080	-1.4816	5.2895	-0.8209	2.8212	-6.4948	4.6262
sex	-0.1321	0.0856	-0.2997	0.0366	-0.1885	0.1001	-0.3860	0.0070
fs1	-0.0671	0.0372	-0.1393	0.0065	0.0336	0.0428	-0.0507	0.1165
fs2	0.0280	0.0332	-0.0370	0.0922	-0.0109	0.0392	-0.0864	0.0654
fs3	0.0390	0.0234	-0.0073	0.0843	0.0022	0.0257	-0.0484	0.0526
msm	-0.3698	0.7374	-1.8310	1.1034	0.6009	1.2724	-1.8854	3.0999
σ_u^2	0.6037	0.2326	0.2920	1.1836	1.2770	0.4139	0.6874	2.2772

Table 3: Prior distributions

Parameter	Distribution	Mean	Variance
β_i	$\mathcal{N}$	0	100
u_j	$\mathcal{N}$	0	0,5
σ_u^2	$\mathcal{IG}$	1	1

Table 4: Proportions of missing values

variable	scenario (1)	scenario (2)	scenario (3)	scenario (4)
z_{ij}	complete	complete	2 %	2 %
d_{ij}^1	complete	5 %	complete	5 %
x_{ij}^1 and x_{ij}^2	complete	8 %	complete	8 %
x_{ij}^1	complete	2 %	complete	2 %
x_{ij}^2	complete	2 %	complete	2 %
d_{ij}^2	complete	20 %	complete	20 %
x_j	complete	2.5 %	complete	2.5 %

Table 5: Probit regression coefficients for different scenarios of the simulation study

Scenario (1): complete cases ($N_{sim} = 20$)						
Var	true value	coverage	$mean_m$	sd_m	$mean_{sd}$	sd_{sd}
Intercept	1.50	0.95	1.9991	1.0333	1.2242	0.1078
d^1_{ij}	0.10	0.85	0.0945	0.0669	0.0508	0.0009
x^1_{ij}	0.05	0.90	0.0498	0.0290	0.0223	0.0006
x^2_{ij}	0.20	1.00	0.2052	0.0196	0.0200	0.0006
d^2_{ij}	0.30	0.95	0.2975	0.0675	0.0631	0.0011
x_j	-1.50	0.95	-1.7423	0.5056	0.5377	0.0474
σ^2_u	0.64	0.90	0.6953	0.1225	0.1284	0.0241
Scenario (2): missing values ($N_{sim} = 20$)						
Var	true value	coverage	$mean_m$	sd_m	$mean_{sd}$	sd_{sd}
Intercept	1.50	0.95	1.7060	0.9255	1.1589	0.0904
d^1_{ij}	0.10	0.85	0.0912	0.0662	0.0512	0.0010
x^1_{ij}	0.05	0.85	0.0458	0.0290	0.0227	0.0005
x^2_{ij}	0.20	0.95	0.1973	0.0190	0.0204	0.0005
d^2_{ij}	0.30	0.95	0.2541	0.0713	0.0656	0.0013
x_j	-1.50	0.95	-1.5481	0.4438	0.5077	0.0384
σ^2_u	0.64	0.90	0.6783	0.1170	0.1219	0.0182
Scenario (3): measurement error ($N_{sim} = 20$)						
Var	true value	coverage	$mean_m$	sd_m	$mean_{sd}$	sd_{sd}
Intercept	1.50	0.95	1.8916	1.0626	1.1739	0.1071
d^1_{ij}	0.10	0.80	0.0870	0.0701	0.0504	0.0011
x^1_{ij}	0.05	0.80	0.0487	0.0314	0.0226	0.0007
x^2_{ij}	0.20	0.85	0.1878	0.0417	0.0203	0.0007
d^2_{ij}	0.30	0.95	0.2814	0.0734	0.0639	0.0014
x_j	-1.50	0.95	-1.6260	0.5385	0.5148	0.0465
σ^2_u	0.64	1.00	0.6012	0.1000	0.1139	0.0353
Scenario (4): missing values and measurement error ($N_{sim} = 20$)						
Var	true value	coverage	$mean_m$	sd_m	$mean_{sd}$	sd_{sd}
Intercept	1.50	0.9167	1.9084	1.0882	1.1104	0.1002
d^1_{ij}	0.10	0.9167	0.0929	0.0590	0.0513	0.0009
x^1_{ij}	0.05	0.8333	0.0382	0.0316	0.0232	0.0006
x^2_{ij}	0.20	0.9167	0.1905	0.0203	0.0208	0.0005
d^2_{ij}	0.30	0.8333	0.2305	0.0636	0.0668	0.0016
x_j	-1.50	0.9167	-1.5922	0.5287	0.4858	0.0431
σ^2_u	0.64	1.0000	0.6025	0.0829	0.1059	0.0156

Table 6: Robustness check for scenario (1) of the simulation study

Scenario (1): complete cases ($N_{sim} = 100$)						
Var	true value	coverage	$mean_m$	sd_m	$mean_{sd}$	sd_{sd}
Intercept	1.50	0.97	1.5535	1.1984	1.2413	0.1356
d_{ij}^1	0.10	0.94	0.0987	0.0530	0.0505	0.0012
x_{ij}^1	0.05	0.99	0.0497	0.0183	0.0222	0.0006
x_{ij}^2	0.20	1.00	0.2006	0.0176	0.0199	0.0005
d_{ij}^2	0.30	0.93	0.3097	0.0660	0.0627	0.0017
x_j	-1.50	0.97	-1.5236	0.5177	0.5451	0.0595
σ_u^2	0.64	0.94	0.6669	0.1146	0.1222	0.0254

Table 7: Bayesian Probit estimation with different prior precision

	2010				2011			
(I.1) Gibbs Cart MH P = 0.01					(I.2) Gibbs Cart MH P = 0.01			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	1.2011	1.3131	-1.3491	3.7941	-0.8278	2.5411	-5.9724	4.1328
sex	-0.1528	0.0749	-0.2975	-0.0058	-0.1912	0.0962	-0.3791	-0.0017
fs1	-0.0582	0.0339	-0.1247	0.0079	0.0286	0.0410	-0.0504	0.1087
fs2	0.0193	0.0309	-0.0417	0.0801	-0.0098	0.0374	-0.0840	0.0622
fs3	0.0387	0.0218	-0.0046	0.0811	0.0105	0.0253	-0.0393	0.0598
msm	-0.1603	0.5586	-1.2717	0.9051	0.5588	1.1597	-1.7002	2.9225
σ_u^2	0.3867	0.1156	0.2195	0.6673	1.1512	0.3549	0.6373	2.0165
(II.1) Gibbs Cart MH P = 0.02					(II.2) Gibbs Cart MH P = 0.02			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	1.1325	1.2841	-1.3911	3.7243	-0.6130	2.5406	-5.6655	4.1640
sex	-0.1530	0.0754	-0.3025	-0.0070	-0.1925	0.0980	-0.3863	-0.0002
fs1	-0.0568	0.0336	-0.1226	0.0090	0.0287	0.0412	-0.0525	0.1098
fs2	0.0178	0.0303	-0.0411	0.0774	-0.0115	0.0376	-0.0845	0.0624
fs3	0.0394	0.0218	-0.0033	0.0820	0.0105	0.0252	-0.0395	0.0596
msm	-0.1318	0.5503	-1.2348	0.9443	0.4624	1.1531	-1.7344	2.7384
σ_u^2	0.3890	0.4201	0.2153	0.6604	1.1531	0.3555	0.6338	2.0121
(III.1) Gibbs Cart MH P = 0.05					(III.2) Gibbs Cart MH P = 0.05			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	1.2086	1.3278	-1.3657	3.8252	-0.5243	2.4254	-5.2583	4.3361
sex	-0.1509	0.0755	-0.2990	-0.0035	-0.1901	0.0988	-0.3845	0.0028
fs1	-0.0570	0.0338	-0.1233	0.0098	0.0290	0.0409	-0.0514	0.1083
fs2	0.0183	0.0306	-0.0412	0.0785	-0.0107	0.0374	-0.0844	0.0631
fs3	0.0389	0.0216	-0.0034	0.0807	0.0098	0.0252	-0.0401	0.0587
msm	-0.1665	0.5649	-1.2735	0.9383	0.4320	1.0963	-1.7518	2.5616
σ_u^2	0.3867	0.1373	0.2167	0.6627	1.1523	0.3572	0.6388	2.0180

Note: Initial 5000 draws were discarded for burn-in

Table 8: Marginal effects of the Bayesian Probit estimation with different prior precision

	2010				2011			
(I.1) Gibbs Cart MH P = 0.01					(I.2) Gibbs Cart MH P = 0.01			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	0.3180	0.3462	-0.3587	0.9947	-0.1926	0.5988	-1.3899	0.9858
sex	-0.0426	0.0209	-0.0830	-0.0016	-0.0472	0.0238	-0.0936	-0.0004
fb1	-0.0154	0.0089	-0.0329	0.0021	0.0069	0.0098	-0.0118	0.0264
fb2	0.0051	0.0082	-0.0110	0.0212	-0.0024	0.0089	-0.0201	0.0147
fb3	0.0103	0.0058	-0.0012	0.0214	0.0025	0.0060	-0.0094	0.0141
msm	-0.0425	0.1476	-0.3356	0.2396	0.1304	0.2727	-0.4065	0.6803
(II.1) Gibbs Cart MH P = 0.02					(II.2) Gibbs Cart MH P = 0.02			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	0.3005	0.3399	-0.3721	0.9807	-0.1420	0.5981	-1.3144	0.9957
sex	-0.0427	0.0210	-0.0843	-0.0020	-0.0475	0.0242	-0.0955	-0.0001
fb1	-0.0151	0.0088	-0.0322	0.0024	0.0069	0.0098	-0.0122	0.0266
fb2	0.0047	0.0081	-0.0110	0.0206	-0.0028	0.0089	-0.0201	0.0147
fb3	0.0105	0.0058	-0.0009	0.0217	0.0025	0.0060	-0.0093	0.0141
msm	-0.0351	0.1459	-0.3267	0.2510	0.1078	0.2710	-0.4148	0.6345
(III.1) Gibbs Cart MH P = 0.05					(III.2) Gibbs Cart MH P = 0.05			
	Estimate	Std. Error	95% HDR		Estimate	Std. Error	95% HDR	
Intercept	0.3206	0.3513	-0.3640	1.0114	-0.1219	0.5709	-1.2220	1.0239
sex	-0.0421	0.0210	-0.0832	-0.0010	-0.0470	0.0244	-0.0954	0.0007
fb1	-0.0151	0.0089	-0.0324	0.0027	0.0070	0.0098	-0.0120	0.0264
fb2	0.0049	0.0081	-0.0110	0.0210	-0.0026	0.0089	-0.0202	0.0149
fb3	0.0103	0.0057	-0.0009	0.0214	0.0023	0.0060	-0.0097	0.0139
msm	-0.0442	0.1498	-0.3362	0.2473	0.1010	0.2575	-0.4175	0.5945

Note: Initial 5000 draws were discarded for burn-in

Figure 1: Data situation

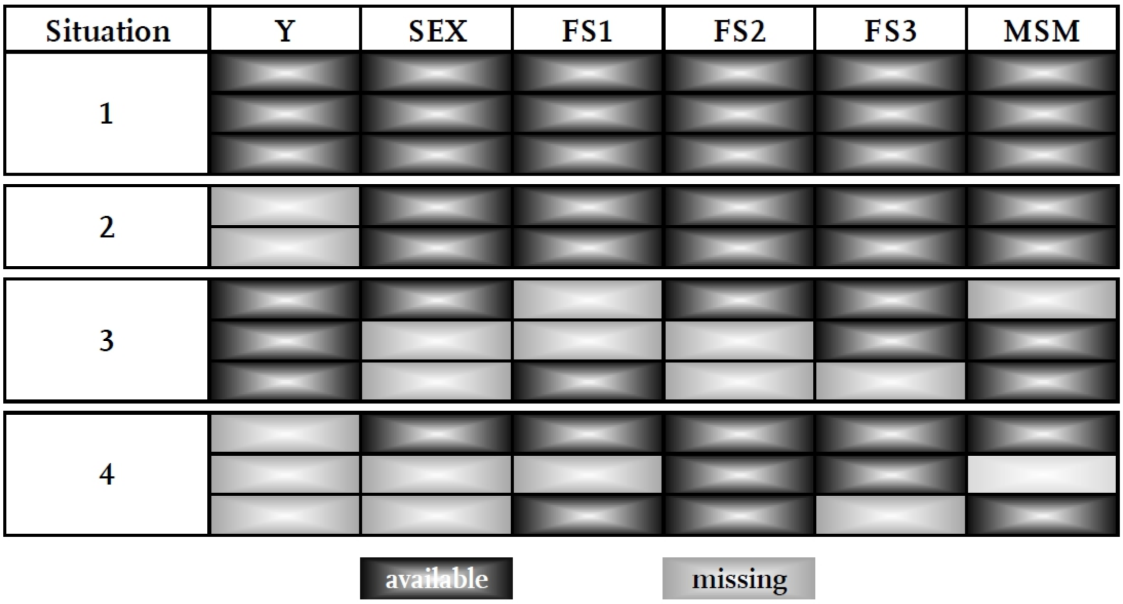
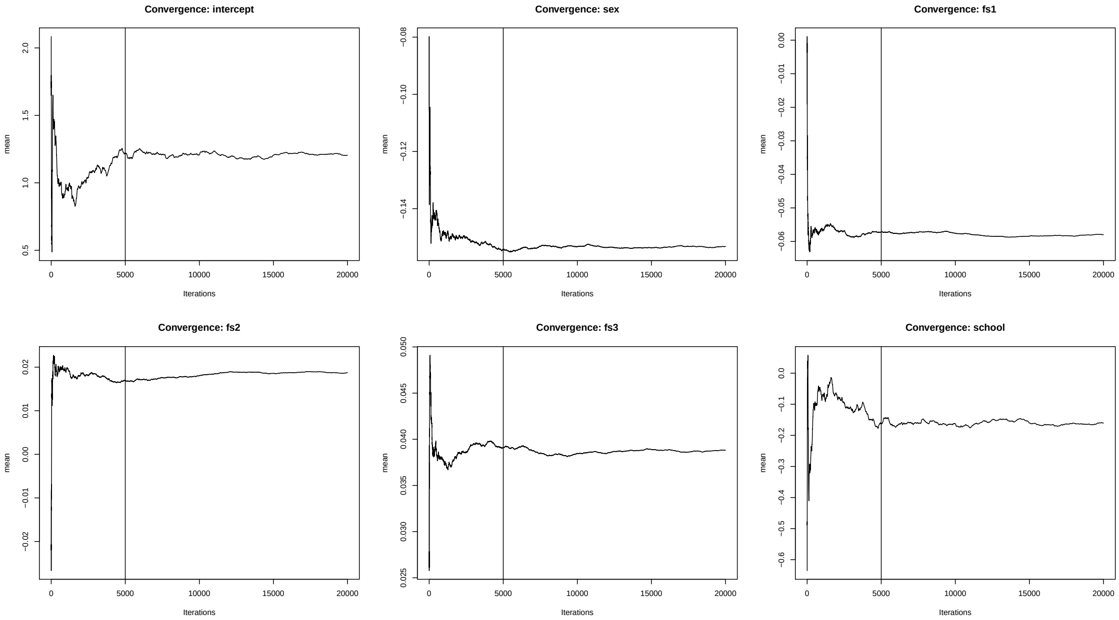
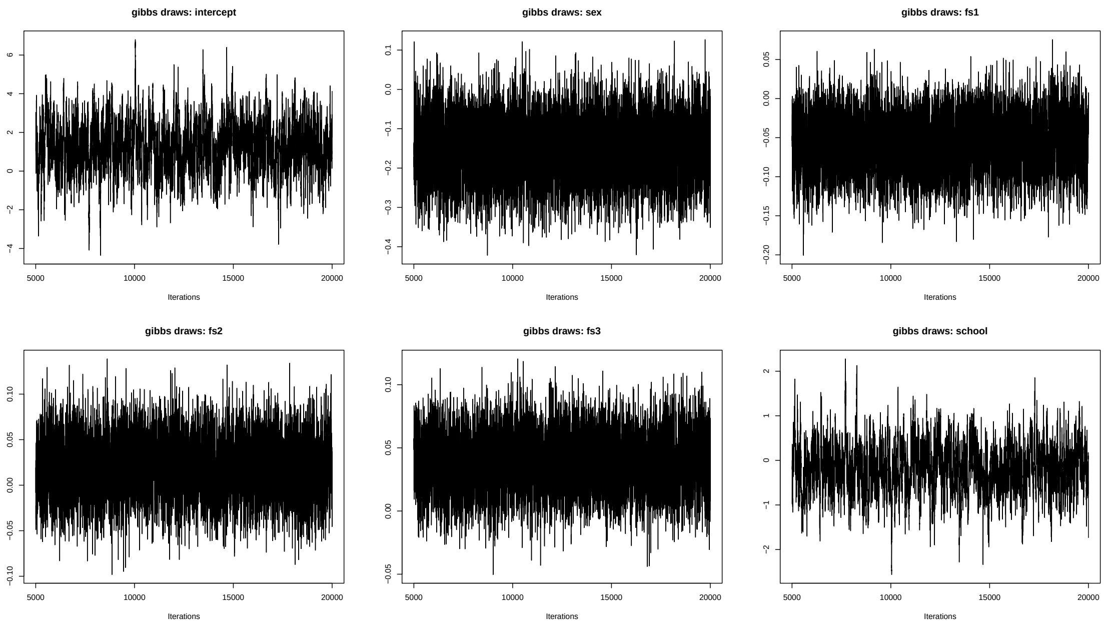


Figure 2: Convergence of the means



Note: Estimations are based on 20000 Gibbs iterations, where initial 5000 draws were discarded for burn-in.

Figure 3: Draws from the Gibbs sampler



Note: Estimations are based on 20000 Gibbs iterations, where initial 5000 draws were discarded for burn-in.

Figure 4: Plots of the autocorrelation functions (ACF)