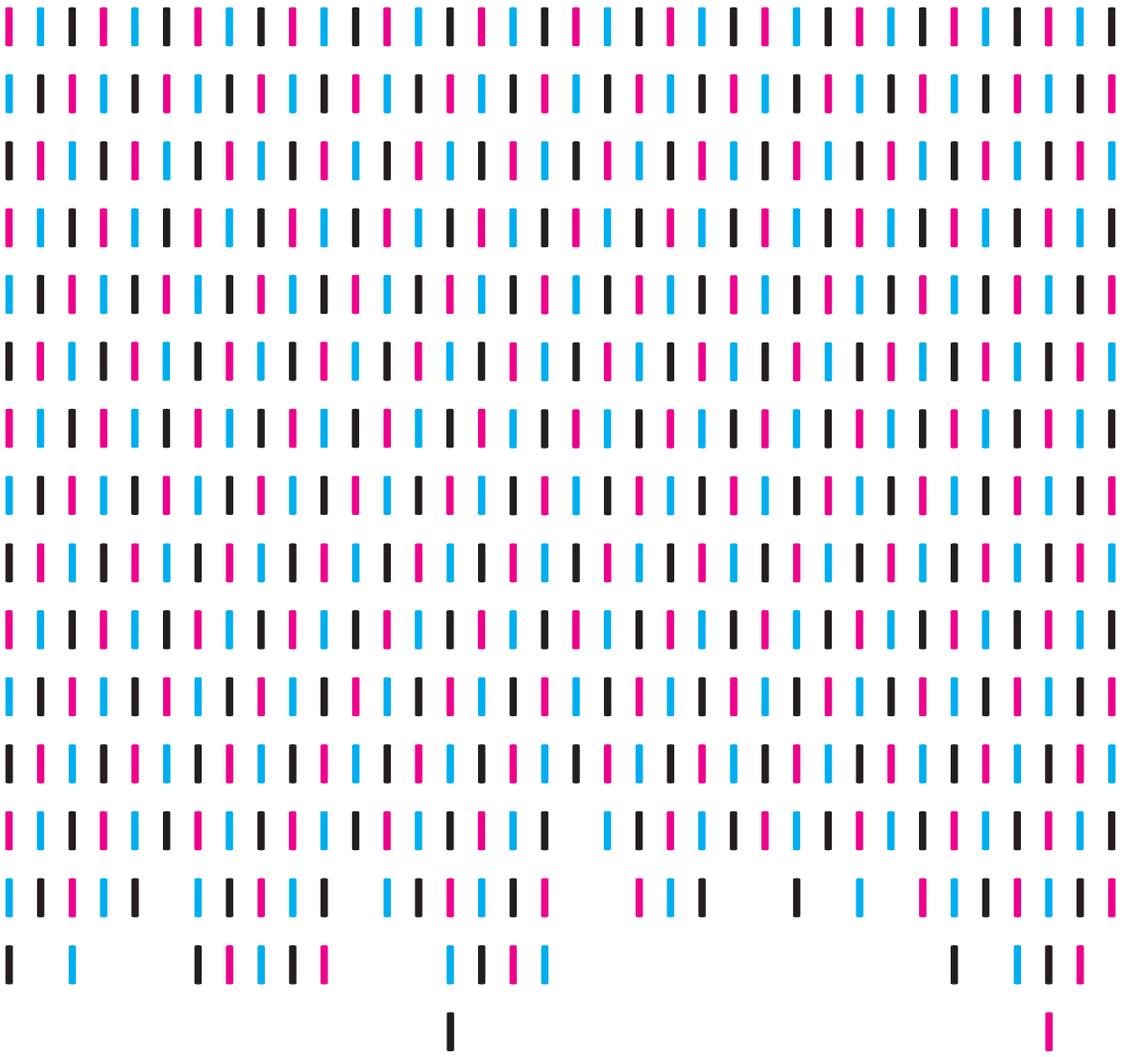

NEPS WORKING PAPERS

Christian Aßmann, Claus H. Carstensen,

Christoph Gaasch, Steffi Pohl

ESTIMATION OF PLAUSIBLE VALUES USING BACKGROUND VARIABLES WITH MISSING VALUES: A DATA AUGMENTED MCMC APPROACH

NEPS Working Paper No. 38
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

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

Estimation of Plausible Values using Background Variables with Missing Values: A Data Augmented MCMC Approach

Christian Aßmann, Leibniz-Institute for Educational Trajectories Bamberg
Claus H. Carstensen, Leibniz-Institute for Educational Trajectories Bamberg
Christoph Gaasch, Leibniz-Institute for Educational Trajectories Bamberg
Steffi Pohl, Freie Universität Berlin

March 2014

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

Bibliographische Angaben:

Aßmann, C., Carstensen, C. H., Gaasch, C., & Pohl, S. (2014). *Estimation of Plausible Values using Background Variables with Missing Values: A Data Augmented MCMC Approach* (NEPS Working Paper No. 38). Bamberg: Leibniz-Institute for Educational Trajectories Bamberg, National Educational Panel Study.

Estimation of Plausible Values using Background Variables with Missing Values: A Data Augmented MCMC Approach

Abstract

The National Educational Panel Study (NEPS) provides data on the development of competencies across the whole life span to educational researchers and politicians. Plausible values as a measure of individual competence are estimated by explicitly including background variables capturing individual characteristics into the corresponding Item Response Theory (IRT) models. Despite tremendous efforts in field work, missing values in the background variables can occur. Adequate estimation routines are needed to reflect the uncertainty stemming from missing values in the background variables in the estimation of plausible values. To achieve this, we propose to adapt an estimation strategy based on Markov Chain Monte Carlo (MCMC) techniques that simultaneously addresses missing values in background variables in the estimation of plausible values for the competence scores. The resulting hybrid sampling scheme establishes a one-step approach for the estimation of plausible values using IRT models that incorporate background variables with missing values. In a simulation study allowing to control the mechanism causing missing values, we evaluate the validity of our approach with respect to statistical accuracy. The results show that the proposed approach is capable to recover the true regression parameters describing the relationship between latent competence scores and background variables. The approach is illustrated on an example using data from the NEPS on mathematical competencies of fifth grade students.

Keywords

Item response theory, bayesian estimation, missing-data imputation, mathematical competence

1. Introduction

In large scale studies, such as the National Educational Panel Study (NEPS), an aim is to provide educational researchers with data that support the investigation of various educational research questions. Typical research questions concern, for example, the explanation of competencies and competence development based on individual characteristics, e.g., gender, socio-economic status, migration background, and context variables, e.g., school characteristics. Competencies in NEPS are assessed via tests on different domains, such as mathematics, reading, science, and information and communication technology, see Weinert et al. (2011). While the questionnaire data are scored based on classical test theory, usually a sum score is provided for scale scores, or not scored at all when considering for single items, most of the competence data are analyzed via Item Response Theory (IRT) models. In IRT models the response on test items are described by a function of the ability of a person as well as characteristics of the item, such as the item difficulty. IRT models allow for aggregation of individual responses towards latent competence scores that are purified from measurement error. Typically, scale scores are provided in form of plausible values, see Mislevy (1991), which usually provide unbiased population estimates of competence distributions and allow for the investigation of latent relationships between competence scores and background variables capturing individual characteristics of the participants as well as context variables. For the estimation of plausible values, the background variables are included in the measurement model. As such, plausible values may be used to investigate the relationship between latent competence scores and these background variables. However, despite tremendous efforts in field work, missing values in these background variables occur. Missing values in background variables pose a great challenge on the estimation of plausible values. We propose to use a Gibbs sampling approach based on the device of data augmentation suggested by Tanner & Wong (1987) to deal with this challenge. In the following paragraphs we first introduce the IRT-model used for scaling the competence data and draw a special focus on the estimation of plausible values in these models. We then focus on the problem of missing responses in the background data and introduce the Markov Chain Monte Carlo (MCMC) method. In the following sections we combine the MCMC method with the estimation of plausible values in IRT models and develop an approach that simultaneously estimates plausible values and imputes missing responses in background variables. The approach is evaluated within a simulation study and demonstrated on a small empirical example measuring one competence dimension and having missing responses in two background variables.

2. IRT model for scaling of competence tests

In the National Educational Panel Study different competence domains are assessed, for example, reading and mathematical competence. The competence domains are assessed by tests that contain a number of items that may be dichotomously scored as correct or incorrect. Some items in the test consist of a couple of dichotomously scored tasks, i.e., complex multiple choice items. For the IRT-analysis they are aggregated to a single polytomous item, see also Andrich (1985). These items indicate the number of correct answers given for a complex MC item. The competence data in NEPS are scaled using the multidimensional random coefficients multinomial logit model, see among others Adams, Wilson, & Wang (1997) and for a description of the scaling model for the competence data in NEPS, see Carstensen & Pohl (2012). A special case of the multidimensional random coefficient multinomial logit model is the Partial Credit Model as introduced in the literature by Masters (1982). In the Partial Credit Model, the Rasch model of Rasch (1960) for dichotomous data is extended to ordered polytomous data. The multidimensional random coefficients multinomial logit model is a general model, which encompasses both the simple Rasch model and the partial credit case. The formulation of these

models as the mixed coefficients multinomial logit was shown by Adams & Wilson (1996). Within that framework, in the unidimensional case the probability of a response being in category m of item j for individual i is given by

$$P(Y_{ijm}=1|\theta_i)=\frac{\exp(b_{jm}\theta_i+\mathbf{a}_{\mathbf{jm}}^{'}\xi)}{\sum_{m=1}^{M_j}\exp(b_{jm}\theta_i+\mathbf{a}_{\mathbf{jm}}^{'}\xi)},\quad i=1,\dots n,\quad j=1,\dots,J,\quad m=1,\dots,M_j, \tag{1}$$

where θ_i is the scalar ability parameter of person i,

$$\mathbf{b}=(b_{11},b_{12},\dots,b_{1M_1},b_{21},\dots,b_{2M_2},\dots,b_{J1},\dots,b_{JM_J}) \tag{2}$$

is a vector of scoring functions with b_{jm} reflecting the performance level (scoring) of each possible item category, $\xi=(\xi_1,\xi_2,\dots\xi_p)$ is a vector of p item difficulty parameters, and

$$\mathbf{A}=(\mathbf{a}_{11},\mathbf{a}_{12},\dots,\mathbf{a}_{1M_1},\mathbf{a}_{21},\dots,\mathbf{a}_{2M_2},\dots,\mathbf{a}_{J1},\dots,\mathbf{a}_{JM_J}) \tag{3}$$

is a design matrix of design vectors $\mathbf{a}_{jm}$, each of length p, giving the empirical characteristics of the item categories. In this formulation, θ_i is regarded as a random parameter with density function $g(\theta_i|\alpha)$ for all i. Mostly, the population distribution $g(\cdot)$ is assumed normal with mean μ and variance σ^2. Adams, Wilson, & Wu (1997) formulated the mixed coefficients multinomial logit model as a multilevel model with persons as level-2 units and responses as level-1 observations. This model allows to simultaneously model item responses and structural relations by allowing the inclusion of explaining variables for the latent competence variable. If such explaining variables (background variables) are included in the model the residual distribution of $g(\cdot)$ is assumed normal with mean $Z_i\gamma$, where Z_i denotes a vector of individual characteristics (background variables) influencing individual ability. This corresponds with the multivariate regression equation

$$\theta_i=Z_i\gamma+\varepsilon_i,\qquad \varepsilon_i\sim\mathcal{N}(0,\sigma^2). \tag{4}$$

The model is easily extended to a multidimensional model and estimation of this model is routinely performed via the maximum likelihood principle, see Adams, Wilson, & Wu (1997) for details. The ability of persons may be estimated using different approaches. One approach is maximum likelihood estimation, see e.g. Warm (1989). The second approach is based on Bayesian statistics. Ability estimates are obtained either as expected values (EAP), mode (MAP) or random draws (plausible values) of the posterior distribution of θ given the item responses, the item parameters, and also background variables. The concept of plausible values was introduced by Mislevy (1991) and is based on the work of Rubin (1987) on multiple imputation. Plausible values are nowadays state of the art, e.g. OECD (2009), since they estimate ability as a latent variable with random effect and, thus, allow to estimate latent relationships of competence scores and background variables. In order to estimate plausible values, the relevant background variables need to be included in the measurement model. Note, that the inclusion of certain background variables for estimating plausible values is essential when one aims at using the plausible values for investigating relationships of competence with these background variables. Adams, Wilson, & Wu (1997) include background variables as level-2 predictors. The estimation of plausible values becomes non-trivial when missing values occur in the background variables. Missing values in questionnaire items are routinely treated via multiple imputation, see Rubin (1987). Since the released data will be used for a variety of research questions - which are not known at the time of data release - providing appropriate data for all these analyses is a great challenge. This is especially prevalent for the estimation of plausible values and for

imputing missing responses. For both estimations, an appropriate imputation model is needed that includes all relevant background variables needed in later analyzes. Specifically questionnaire variables (i.e., background variables) are needed for the estimation of competence scores (plausible values) and questionnaire variables as well as competence scores are needed for the imputation of missing responses in questionnaire items. Other large scale studies, such as the Programme for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP) deal with this problem by aggregating the questionnaire variables to orthogonal factors and using the set of factors as background variables in the IRT measurement model of the competence data, see Allen et al. (2001). Thereby, as many factors as needed to explain 90 percent of the variance of the questionnaire items are considered. However, this approach is a two-step approach that does not incorporate all questionnaire variables and does not depict the uncertainty stemming from missing values in questionnaire items. In the following we will describe a data analysis strategy that applies the multilevel random coefficients multinomial logit model to univariate competence measurement settings in the German National Educational Panel Study. The proposed estimation routine is designed to cope with missing information on individual level variables influencing person abilities. Ensuring the validity of empirical competence measures given the uncertainty stemming from missing values in background variables we adopt a bayesian estimation scheme that allows a conceptually stringent treatment of missing values in observed individual characteristics via the device of data augmentation, see Tanner & Wong (1987). Bayesian estimation is implemented using Markov Chain Monte Carlo (MCMC) techniques, namely Gibbs sampling, which are ideally suited to deal with the hierarchical structure of the model and the incorporation of a missing data imputation step. In addition, the usage of MCMC simulation methods proved straightforward for complex IRT models relative to marginal maximum likelihood as discussed in Patz & Junker (1999). To illustrate the approach, we restrict the distribution of missing values to the normal distribution, where nonparametric distributions provide a valid and highly flexible alternative.

3. Bayesian inference using Markov Chain Monte Carlo techniques

Bayesian inference is concerned about the posterior distribution $p(\psi|S)$ and corresponding moments thereof. A general introduction on the basic principles employed in the following is provided by Geweke (1999) and Koop (2003). Gibbs sampling is a device to produce a sample from the joint posterior distribution of the parameter vector ψ, which can be used to estimate posterior moments and density estimates. Posterior draws of ψ partitioned into convenient blocks $\psi = \{\psi_1, \ldots, \psi_T\}$ are obtained via Gibbs sampling, when direct sampling from the posterior distribution is difficult, but sampling from the full conditional distributions is directly accessible. The functional forms of the full conditional distributions can be deduced from the joint posterior distribution of parameters ψ and the sample data S

$$p(\psi, S) = L(S|\psi)\pi(\psi), \tag{5}$$

where $L(S|\psi)$ denotes the model likelihood and $\pi(\psi)$ denotes the a priori distribution, via isolating the kernel of a single parameter block ψ_t conditional on all other blocks $\psi_1, \ldots, \psi_{t-1}$, $\psi_{t+1}, \ldots, \psi_T$ and the data S, i.e.

$$p(\psi_t|\psi_1, \ldots, \psi_{t-1}, \psi_{t+1}, \ldots, \psi_T, S). \tag{6}$$

Given an initialization $\psi^{(0)}$, the Gibbs sampling algorithm simulates iteratively for $r = 1, \ldots, R$ from the full conditional distributions

$$
\begin{aligned}
&p(\psi_1|\psi_2, \ldots, \psi_T, S),\\
&\vdots\\
&p(\psi_T|\psi_1, \ldots, \psi_{T-1}, S).
\end{aligned}
$$

The iterative sampling constitutes a Markov Chain, which ensures under general regularity conditions given in Chib (2001) convergence to the joint posterior distribution.[1] Since these are fulfilled within the considered class of Rasch models, the convergence of the joint distribution of the sample $\{\psi^{(r)}\}_{r=1}^{R}$ for $R \to \infty$ towards the posterior distribution $p(\psi|S)$ is ensured. Since the functional forms of the full conditional distributions depend on the assumed prior distributions, these are in general conveniently chosen to facilitate sampling from closed form full conditional distributions.

4. Estimation algorithm for Binary Rasch model with missing information in background variables

To illustrate the proposed treatment of missing values, we refer to a simplified version of the model outlined in Equation (1). This simplified version, which allows for closed form sampling from the full conditional distributions employed within the Gibbs sampler, is derived as follows, see also Aßmann & Boysen-Hogrefe (2011) for a general treatment of Bayesian estimation for binary panel probit models. Consider the likelihood of the model stated in Equation (1) given as

$$
\mathcal{L}(S|\{\theta_i\}_{i=1}^{n}) = \prod_{i=1}^{n}\prod_{j=1}^{J}\prod_{m=1}^{M_j}\left(\frac{\exp(b_{jm}\theta_i + \mathbf{a}_{\mathbf{jm}}^{'}\xi)}{\sum_{m=1}^{M_j}\exp(b_{jm}\theta_i + \mathbf{a}_{\mathbf{jm}}^{'}\xi)}\right)^{y_{ijm}}. \tag{9}
$$

[1] Following Chib (2001), the transition from $\psi_t^{(r)}$ to $\psi_t^{(t+1)}$ is accomplished by sampling from the $p(\psi_t|\psi_1, \ldots, \psi_{t-1}, \psi_{t+1}, \ldots, \psi_T, S)$. The transition of the Markov Chain constituting out of T blocks is then described for all continuous distributions as

$$
\Theta(\psi(r), \psi^{(r+1)}) = \prod_{t=1}^{T} p(\psi_t|\psi_1, \ldots, \psi_{t-1}, \psi_{t+1}, \ldots, \psi_T, S). \tag{7}
$$

Sufficient conditions for convergence can then be stated as follows. Let $\Theta(\psi, \psi')$ denotes the transition density of the Gibbs sampler and let $\Theta^R(\psi_0, \psi')$ be the density of ψ' after R iterations of the Gibbs sampler given the initialization ψ_0. Then

$$
||\Theta^R(\psi_0, \psi') - p(\psi|S)|| \to 0 \text{ as } R \to \infty, \tag{8}
$$

where $||\cdot||$ denotes the total variance distance. As it is shown by Roberts & Smith (1994), convergence is ensured under the following conditions

1. $p(\psi|S) > 0$ implies there exists an open neighborhood N_ψ containing ψ and $\xi > 0$ such that, for all $\psi' \in N_\psi$, $p(\psi') \geq \xi > 0$;
2. $\int f(\psi) d\psi_k$ is locally bounded for all k, where ψ_k is the kth block of parameters;
3. the support of ψ is arc connected.

Note that these conditions are not met only for pathological cases.

Setting $M_j = 2$, i.e. considering only dichotomous items, restricting $a_{jm}\xi = \xi_j$ for all m and normalizing $b_{j1} = \frac{\xi_j}{\theta_i}$, $b_{j2} = 1$ and change of notation $y_{ij1} = 1 - y_{ij2} = y_{ij}$ results in the likelihood

$$\mathcal{L}(S|\{\theta_i\}_{i=1}^n) = \prod_{i=1}^{n}\prod_{j=1}^{J}\left(\frac{(\exp\{\theta_i - \xi_j\})^{y_{ij}}}{1+\exp\{\theta_i - \xi_j\}}\right). \tag{10}$$

To solve non-identifiability of the parameters, the sum of the item difficulties equals zero, i. e. $\sum_{j=1}^{J}\xi_j = 0$. In conjunction with a mixing distribution $g(\theta_i)$ given as

$$g(\theta_i|Z_i) = (2\pi)^{-.5}(\sigma^2)^{-.5}\exp\{-\frac{1}{2\sigma^2}(\theta_i - Z_i\gamma)^2\} \tag{11}$$

allows for derivation of the likelihood

$$\mathcal{L}(S) = \prod_{i=1}^{n}\int\prod_{j=1}^{J}\left(\frac{(\exp\{\theta_i - \xi_j\})^{y_{ij}}}{1+\exp\{\theta_i - \xi_j\}}\right)g(\theta_i|Z_i)d\theta_i. \tag{12}$$

As for this likelihood no conjugate priors for parameters exist, facilitating either direct sampling or closed form sampling from the corresponding full conditional distributions, we further change from logit to probit. This allows for Bayesian estimation via Gibbs sampling along the lines suggested by Albert (1992). The likelihood is then given as

$$\mathcal{L}(S) = \prod_{i=1}^{n}\int\prod_{j=1}^{J}\Phi((2y_{ij}-1)(\theta_i - \xi_j))g(\theta_i|Z_i)d\theta_i, \tag{13}$$

with corresponding full conditional distributions given as follows. Additionally, the missing values in background variables are augmented by a parametric method at each iteration that allows the researcher to account for the uncertainty created by a single imputation step. As the model does not involve information concerning the full conditional distribution of the missing values, a hybrid sampling scheme is adopted, where draws for the missing values are obtained from a normal model. After initializing parameters, this leads to the following iterative scheme taking steps $r : 1 \to R$ for repetition r:

Step 1) Sampling the underlying latent variable y_{ij}^* from a truncated normal distribution with corresponding parameters

$$\mu_{y_{ij}^*} = \theta_i - \xi_j, \quad \text{and} \quad \sigma_{y_{ij}^*} = 1,$$

where truncation sphere is $(-\infty, 0)$ for $y_{ij} = 0$ and $(0, \infty)$ for $y_{ij} = 1$.

Step 2) The individual abilities θ_i are sampled from a normal distribution with moments defined as follows

$$\mu_{\theta_i} = \left(\sum_{i=1}^{N} y_{ij}^* + \sum_{i=1}^{N}\xi_i + Z_j\gamma/\sigma^2\right)\left(n + 1/\sigma^2\right)^{-1}, \quad \text{and} \quad \sigma_{\theta_j}^2 = \left(n + 1/\sigma^2\right)^{-1}.$$

Step 3) Let the independent conjugate prior for γ be multivariate normal with moments mean vector ν_γ and covariance matrix Ω_γ. Then draws from full conditional distribution for γ are obtained from a multivariate normal distribution with corresponding moments given as

$$\mu_\gamma = (Z'\theta/\sigma^2 + \Omega_\gamma^{-1}\nu_\gamma)(Z'Z/\sigma^2 + \Omega_\gamma^{-1})^{-1}, \quad \text{and} \quad \Sigma_\gamma = (Z'Z/\sigma^2 + \Omega_\gamma^{-1})^{-1}.$$

Step 4) Choosing the independent conjugate prior for σ^2 inverse gamma with parameters α_0 and β_0, the σ^2 is also distributed inverse gamma with corresponding parameter

$$\alpha = n/2 + \alpha_0, \quad \text{and} \quad \beta = \left(0.5 \sum_{i=1}^{n} (\theta_i - Z_i\gamma)^2 + \beta_0 \right)^{-1}.$$

Step 5) Impute item nonresponse in the $n \times K$ matrix of background variables Z via specifying a univariate normal full conditional distribution for each of the K variables contained in Z. Within the intercourse of the Gibbs sampler, imputed and hence complete variables are at hand for each iteration m, resulting in the following K regression equations given as

$$Z_k = W_k\varphi_k + \epsilon_k, \quad k = 1, \ldots, K,$$

where $W_k = (\iota, Z_{-k}, \theta, SC)$, where SC denotes the vector of score sums for each individual. Imputations are then generated as follows. Each missing values in Z_k is replaced via a draw from a univariate normal distribution with moments $\mu = W'_{mis}\widehat{\varphi}$ and $\sigma^2 = \widehat{\sigma^2_\epsilon}$. Note that instead the least squares estimators $\widehat{\varphi}$ and $\widehat{\sigma^2_\epsilon}$ often draws from the corresponding asymptotic distributions are used for generating draws for the missing values in Z. However, as imputation is performed within each iteration of the Gibbs sampler, the corresponding uncertainty is accounted for. Further, it should be explicitely noted that the estimation scheme introduces the updated draws of the individual abilities θ into the imputation model for each iteration.

Note that the sampler given here assumes knowledge of the item difficulty parameters. Often simultaneous estimation is a straightforward extension of the outlined approach. Given an sample of all model parameters obtained via iterative sequential cycling through the set of full conditional distributions, the plausible values for each individual can be directly taken from the provided Gibbs output. Each sweep $\{\theta_i\}_{r=1}^{R}$, $i = 1, \ldots, n$ from the posterior distribution could be taken as a vector of plausible values.

5. Simulation study

To assess the validity of our approach suggested above, we set up a simulation design comparing the data augmented Gibbs sampler, when missing values occur, with the full sample estimates before deletion. Given this benchmark situation, the relative performance to recover a set of given parameters in the presence of missing data in the background variables can be evaluated. Replication analysis is a method commonly used for this purpose. The data augmented estimation procedure is conducted for $C = 200$ replications of a single data generating process and a missing generating process. Then, the root mean squared error and the proportion of 95% highest posterior density regions that contain the true parameter values (coverages) are computed as the main criteria for comparison. The detailed conditions of the data generating and missing values generating processes are as follows.

For each replication $c = 1, \ldots, C$, the binary response pattern is simulated using the model in (13) with a sample setup of $n = 2000$ individuals facing $J = 10$ items, where the item difficulties are specified as draws from a normal distribution, $\xi_j \sim N(0, 0.5)$. Three background variables X explaining differences in individual abilities θ_i are generated from a standard normal distribution and are having a correlation of 0.5. Then observations in X_2 and X_3 are deleted via a missing process according to two different scenarios I and II. In scenario I, on average 5% and 10% missing values result completely at random for the variables, in scenario II these

rates of missingness increase to 10% and 20% and depend on X_1. The regression weights of the background variables including an intercept take on the values $\gamma = (1, -0.5, 0.5, -0.5)$, while the individual abilities are distributed with variance parameter $\sigma^2 = 1.44$.

Table 1 shows the means of the posterior expected values and their standard deviations over $C = 200$ replications. For both missingness scenarios, our approach reveals an unbiased estimation of all parameters. Also with respect to the error and the coverage rate, the findings support that there is no notable difference to the full sample estimates before deletion reported in the first block columns of the table. Approximately, the observed number of intervals covering the particular parameter corresponds to the theoretical values. Thus, our proposed sampler is a suitable solution for the use of partially observed background variables in the context of IRT models, even with a relatively large amount of missing values present.

Table 1: Mean of the posterior means and standard deviations, root mean squared error and coverage of $C = 200$ replications for a data set before deletion (BD), missing scenario I and missing scenario II.

		$\overline{\text{mean}}$			$\overline{\text{sd}}$			RMSE			coverage		
Parameter	true	BD	I	II	BD	I	II	BD	I	II	BD	I	II
γ_1	1.000	1.003	1.002	1.002	0.035	0.036	0.036	0.037	0.038	0.037	0.935	0.930	0.930
γ_2	−0.500	−0.505	−0.505	−0.504	0.039	0.039	0.040	0.037	0.038	0.038	0.955	0.940	0.940
γ_3	0.500	0.506	0.506	0.506	0.039	0.040	0.041	0.042	0.043	0.043	0.930	0.920	0.940
γ_4	−0.500	−0.506	−0.506	−0.507	0.039	0.040	0.042	0.040	0.042	0.044	0.935	0.930	0.940
σ^2	1.440	1.459	1.457	1.457	0.077	0.078	0.078	0.082	0.082	0.081	0.925	0.925	0.945

6. Empirical Application

To illustrate the usefullness of our approach, we apply the augmented random coefficient IRT probit Gibbs sampler to an exemplary research question. We use data from the National Educational Panel Study (NEPS): Starting Cohort 3 - 5th grade (From Lower to Upper Secondary School), doi:10.5157/NEPS:SC3:1.0.0 (Blossfeld et al., 2011) assessing mathematical competence of students in fifth grade (see Neumann et al., 2013, for the description of the assessment of mathematical competence in NEPS; Duchhardt & Gerdes (2012) for the description of the respective competence data; and Skopek et al. (2013) for the data manual). The data used in this analysis contains information on $n = 5130$ students who have a valid response to at least one of $J = 23$ binary mathematics test items. Missing values in the test item set were ignored, see Pohl et al. (2013) for a comparison of different approaches for treating missing responses in competence tests. In addition to the test results, we consider gender, self-concept beliefs in mathematical skills and satisfaction with school as explanatory variables for the analysis. Descriptive statistics for the data considered in the application are displayed in Table 2. With 6% and 3% missing values for mathematical self-concept and school satisfaction, the amount of missing data is relatively small.

Table 2: Descriptive statistics background variables.

Variable	min	max	mean	sd	missing
female	0	1	0.48	-	0.00
self-concept	1	4	2.94	0.85	0.06
schoolsat	0	10	7.72	2.52	0.03

Notes: $n = 5130$

We applied the proposed data augmented Gibbs sampling approach to the data for estimating

the regression coefficients of the latent mathematics score on gender, mathematical self-concept and school satisfaction. The data augmented Gibbs sampling approach is able to deal with the missing values in two background variables while simultaneously estimating plausible values for the mathematical competence. The algorithm showed a good convergence behavior. The trace plots show no indication of convergence problems (Figure 1), also the autocorrelations become very low (Figure 2) and the cumulative means converge (Figure 3). Taken a burn-in period of 2000 draws, the regression coefficients of gender, school satisfaction and mathematical self-concept were based on $R = 8000$ simulated draws. Table 3 depicts the estimated posterior means and standard deviations, as well as the 95% Highest Density Intervals (HDI). While the results indicate a lower level of competence for females, the other two variables have a positive effect on students mathematical abilities. Note that the regression coefficients reflect the relationship of questionnaire variables with latent mathematics scores that are purified from measurement error. The estimated standard errors of the regression coefficients incorporate not only the uncertainty due to person sampling, but also uncertainty due to missing values in the predictors.

Table 3: Parameter estimates of the random coefficient IRT probit

Parameter	mean	sd	95% HDI
γ_1 (constant)	−0.509	0.053	[-0.611; -0.408]
γ_2 (female)	−0.121	0.019	[-0.158; -0.085]
γ_3 (self-concept)	0.221	0.011	[0.198; 0.243]
γ_4 (schoolsat)	0.026	0.004	[0.019; 0.034]
σ^2	0.330	0.009	[0.314; 0.348]

Notes: $n = 5130$

7. Conclusion

In large scale assessments researchers are usually interested in explaining competence scores by individual characteristics and context variables. Simultaneously accounting for measurement error in competence scores and missing values in background variables capturing individual characteristics and context variables is challenging. We proposed a data augmented MCMC approach that simultaneously estimates plausible values and accounts for missing values in background variables. With this approach latent relationships between competence scores and background variables may be estimated which efficiently incorporate the uncertainty stemming from only partially observed background variables. In a simulation study the proposed approach proved to adequately recover the model parameters to be estimated, even when higher rates of missingness occur in the data. The applicability to educational research data could be illustrated on an empirical example. Especially the iterative use of updated parameter values from posterior sampling for the imputation model showed an appealing feature of our approach. Future research should focus on considerations of an alternative imputation step coping with the often categorical character of background variables in research questions involving a larger set of variables.

References

Adams, R. J., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1-23.

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, *22*, 47-76.

Adams, R. J., & Wilson, M. R. (1996). Formulating the rasch model as a mixed coefficients multinomial logit. In G. Engelhardt & M. Wilson (Eds.), *Objective measurement: Theory into practice* (Vol. 3, p. 143-166). Ablex.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using gibbs sampling. *Journal of Educational Statistics*, *17*, 251-269.

Allen, N. L., Carlson, J. E., Johnson, E. G., & Mislevy, R. J. (2001). Scaling procedures. In N. L. Allen, J. R. Donoghue, & T. L. Schoeps (Eds.), *The naep 1998 technical report.* U. S. Department of Education.

Andrich, D. (1985). A latent trait model for items with response dependencies: Implications for test construction and analysis. In S. E. Embetson (Ed.), *Test design - developments in psychology and psychometrics* (p. 245-275). Academic Press.

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

Blossfeld, H.-P., Roßbach, H.-G., & Maurice, J. von (Eds.). (2011). *Education as a lifelong process. the german national educational panel study (neps).* VS Verlag für Sozialwissenschaften.

Carstensen, C. H., & Pohl, S. (2012). *Neps technical report: Scaling the data of the competence tests (neps working paper no. 14).* (University of Bamberg, Leibniz Institute for Educational Trajectories, National Educational Panel Study)

Chib, S. (2001). Markoc chain monte carlo methods: Computation and inference. In J. J. Heckmann & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, p. 3569-3649). Elsevier.

Duchhardt, C., & Gerdes, A. (2012). *Neps technical report for mathematics: Scaling results of starting cohort 3 in fifth grade (neps working paper no. 19).* (University of Bamberg, Leibniz Institute for Educational Trajectories, National Educational Panel Study)

Geweke, J. (1999). Using simulation methods for bayesian econometric models: Inference, development and communication. *Econometric Reviews*, *18*, 1-73.

Koop, G. (2003). *Bayesian econometrics.* Wiley.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177-196.

Neumann, I., Duchhardt, C., Ehmke, T., Grüßing, M., Heinze, A., & Knopp, E. (2013). Modeling and assessing of mathematical competence over the lifespan. *Journal for Educational Research Online*, *5*, 80-109.

OECD. (2009). *Pisa 2006 technical report.* OECD Publishing.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to markov chain monte carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146-178.

Pohl, S., Gräfe, L., & Rose, N. (2013). Dealing with omitted and not-reached items in competence tests: Evaluating approaches accounting for missing responses in item response theory models. *Educational and Psychological Measurement*, *73*, 1-30.

Rasch, G. W. (1960). *Probabilistic models for some intelligence and attainment tests.* Danish Institute for Educational Research.

Roberts, G. O., & Smith, A. F. M. (1994). Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms. *Stochastic Processes and their Applications*, *49*, 207-216.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* J. Wiley & Sons.

Skopek, J., Pink, S., & Bela, D. (2013). *Starting cohort 3: Grade 5 (sc3). suf version 1.0.0. data manual (neps research data paper).* (University of Bamberg, Leibniz Institute for Educational Trajectories, National Educational Panel Study)

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528-549.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427-450.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). Development of competencies across the life span. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. the german national educational panel study (neps)* (p. 67-86). VS Verlag für Sozialwissenschaften.

Figures

Figure 1: Trace plots for the regression constant (γ_1), the regression coefficients for sex (γ_2), mathematical self-concept (γ_3) and school satisfaction (γ_4), as well as the residual variance (σ^2).

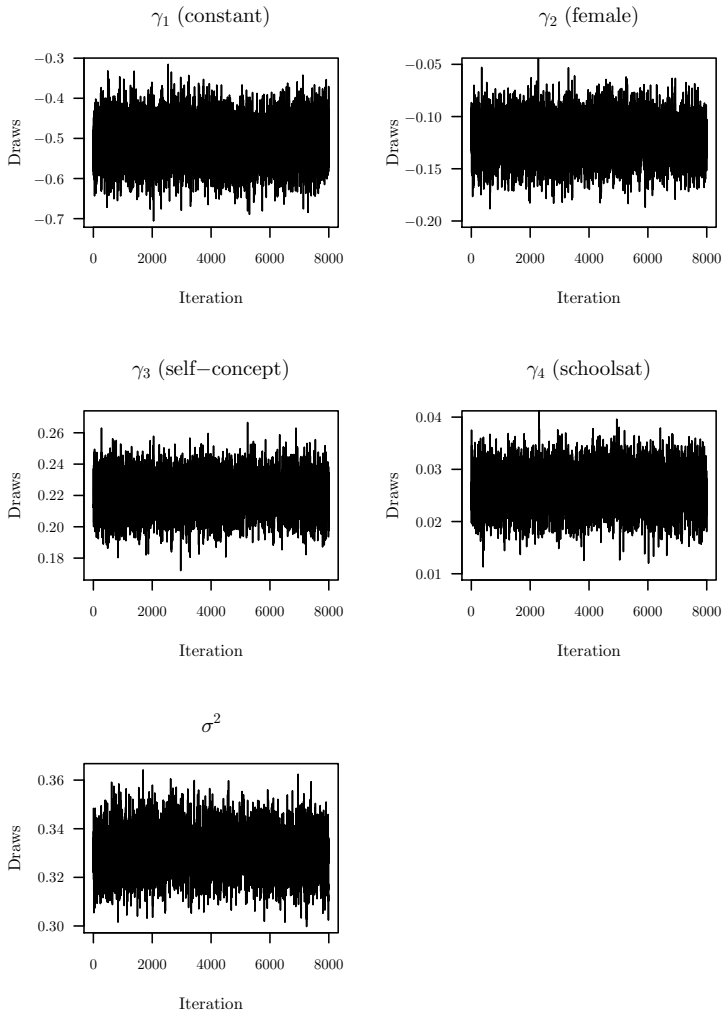

Figure 2: Lag-1 autocorrelation functions for the regression constant (γ_1), the regression coefficients for sex (γ_2), mathematical self-concept (γ_3) and school satisfaction (γ_4), as well as the residual variance (σ^2).

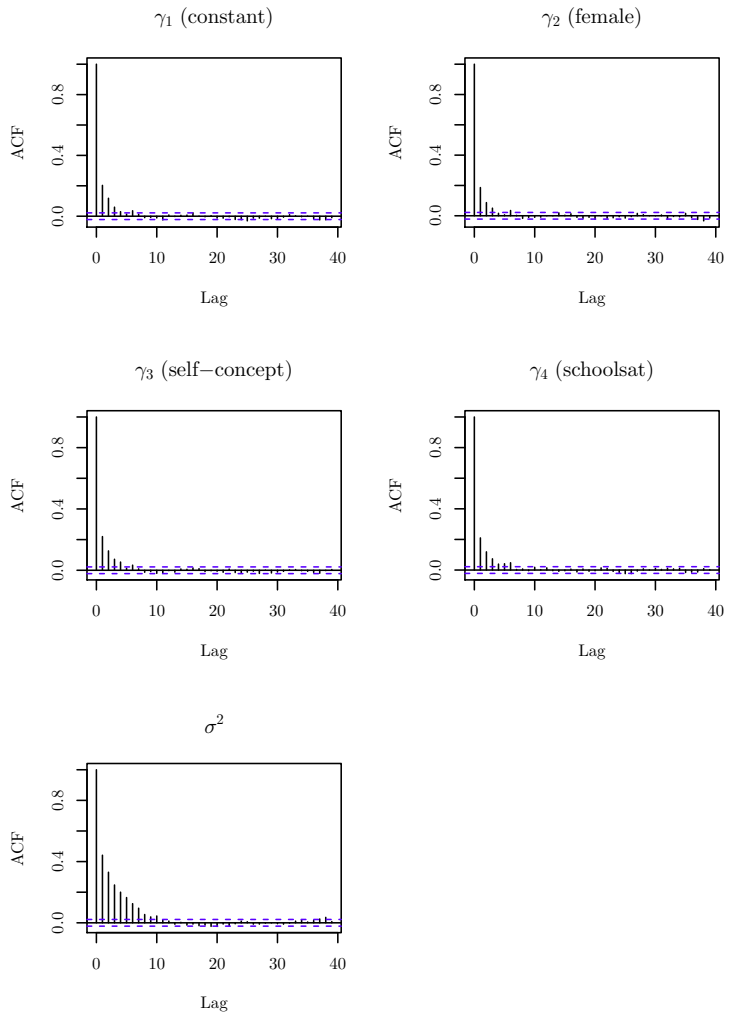

Figure 3: Cumulative mean functions for the regression constant (γ_1), the regression coefficients for sex (γ_2), mathematical self-concept (γ_3) and school satisfaction (γ_4), as well as the residual variance (σ^2).