NEPS
National Educational Panel Study

Research Data

Additional Study Baden-Wuertemberg (BW) SUF-Version 2.0.0 Data Manual [Supplement]: Weighting

Benno Schönberger and Christian Aßmann

A STUDY BY

LIfBi

LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

Copyrighted Material
Leibniz Institute for Educational Trajectories (LIfBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Hans-Günther Roßbach
Executive Director of Research: Dr. Jutta von Maurice
Executive Director of Administration: Meike Uhde
Bamberg, 2014

Weighting the Additional Study in Baden-Wuerttemberg of the National Educational Panel Study

Technical Report

Benno Schönberger & Christian Aßmann *

National Educational Panel Study (NEPS)
University of Bamberg

version: January 30, 2014

*methods.neps@uni-bamberg.de

1 Structure of samples

Based on a short review of the survey and sampling design, this report provides the results of the weighting procedures concerning the surveys of the additional reform study in Baden-Wuerttemberg conducted within the National Educational Panel Study (NEPS) in 2011, 2012 and 2013, respectively. This study has been implemented to assess possible effects regarding the reform of the organization of the Gymnasiale Oberstufe (final years in upper secondary education qualifying for university education) in Baden-Wuerttemberg. In many Länder of the German Federal Republic, the time spent in Gymnasium (secondary school qualifying for university education) has been cut from 9 years (G9) down to 8 years (G8). In 2004, also Baden-Württemberg decided for a reduction from 9 to 8 years combined with a reorientation of the study schedule. The survey consists of three independent cross-sections scheduled around the year of reform implementation. That means, it is not a panel survey. This report covers weighting schemes for all three waves of the study. Students graduating in 2011 (last cohort before reform implementation with 9 years of schooling) and 2012 (year in which G8 as well as G9 students graduate (Doppeljahrgang)) as well as in 2013 (year after reform) in Baden-Wuerttemberg have been surveyed via competence development tests and questionnaires provided that consent was given.
Access to the population of all students attending upper secondary schools (Gymnasium) in Baden-Wuerttemberg who are affected by the G8 reform was gained via proportional-to-size (pps) sampling of 50 upper secondary schools. The measure of size is the mean number of students in grades 10 to 12 during the 2010/11 school year as best estimate of the total number of students in the four cohorts surveyed within the study over the years 2011, 2012, and 2013. Through the selected upper secondary schools, all four student cohorts that are affected by the reform were reached: I - grade 13 students during the school year of 2010/11, II - grade 12 students, and III - grade 13 students during the school year of 2011/12 (Doppeljahrgang), as well as IV - grade 12 students during the school year of 2012. For each of the four cohorts, a sample of 40 students per selected school is established by simple random sampling. The gross and net sample sizes of each of the cohorts that have been tested are $n^{I}_{gross} = 1858$ (gross sample) and $n^{I}_{net} = 1341$ (net sample) students, $n^{II}_{gross} = 1918$ (gross sample) and $n^{II}_{net} = 1284$ (net sample) students, and $n^{III}_{gross} = 1908$ (gross sample) and $n^{III}_{net} = 1293$ (net sample) students, respectively[1].

We describe the methods used to calculate the weights for the participating students in Section 2. Section 3 addresses corrections for potential systematic nonresponse and section 4 describes the procedures applied for trimming the weights.

2 Weights

60 schools of the eligible 445 Gymnasia in Baden-Württemberg have been excluded from the sampling frame, since 42 of these schools had no G9 structure at all and 18 had only an incomplete G9 structure, resulting in 385 schools meeting all requirements for surveying all four cohorts and thus sampling. Fifty schools have been sampled via a pps-

[1] Detailed information on sample sizes and weighting procedures for cohort IV will be released together with the next scientific use file (SUF Version 3-0-0) via an updated version of this report.

sampling approach, where the inclusion probability is proportional to a certain measure of size (MOS). As each school j has a different number of students within each of the four cohorts *I-IV* to be surveyed, the MOS used for the sampling has been the mean of the number of students in grades 10, 11 and 12 in school year 2010/2011 as best proxy for the number of students to be surveyed with corresponding design weight at the school level given as

$$w_j^{school} = \left(50 \frac{\overline{m}_{j;10^{th},11^{th},12^{th}}}{\sum_{j=1}^{385} \overline{m}_{j;10^{th},11^{th},12^{th}}} \right)^{-1}.$$

For the case, that schools refused to participate for each school two additional schools have been drawn with similar characteristics for replacement. The realized sample comprises 48 schools, where one school refused participation too late in the field process to allow for replacement and two sampled schools have been fused to one school (frame error). Further, the competence development tests could not be conducted in two schools for the first cohort for scheduling reasons. Both participated in the following waves of the study. Therefore cohort I comprises only 46 schools, whereas cohorts II and III consist of 48 schools. School weights for cohort I where hence adjusted by the ratio 50/46 and for cohorts II and III by the ratio 50/48. Within each school, 40 students should have been sampled per cohort via simple random sampling for each of the four surveyed cohorts. There are only slight exceptions for the case that a school had less than 40 students per cohort and for one school where the school itself wanted all students to be sampled. Thus the inclusion probabilities for each student within the same cohort only depend on the total number of students K_{ij} in the corresponding cohort $i \in \{I, II, III\}$ in school j and the actual number of sampled students s_{ij}. The individual inclusion weight w_{ij}^{ind} for all students in cohort i in school j is then the reciprocal of the corresponding inclusion probability p_{ij}

$$w_{ij}^{ind} = \frac{1}{p_{ij}} \quad \text{with} \quad p_{ij} = \frac{s_{ij}}{K_{ij}}.$$

The resulting design weight for each student is then the product of the school level design weight w_j^{school} and the individual inclusion weight w_{ij}^{ind}

$$w_{ij}^{des} = w_j^{school} \cdot w_{ij}^{ind}.$$

3 Weight adjustment for participation

Given the inclusion probabilities and corresponding weights, the willingness to participate in the survey has to be taken into account. For all samples the individual participation and response probability π_{ijk} for students k of each of the four cohorts attending one of the participating schools has been analysed via bayesian probit regression models (see Aßmann et al. (2014), forthcoming) based on the information on all marks of all students. The dependent variable for the probit regressions is the participation status of students. As covariates, we included the main subject marks for every student ("German", "Mathematics" and "English") as well as the mean marks of mathematical-natural-scientific-technical subjects ("NAT") (e.g., physics, biology, chemistry). As further covariates, we consider the individual final grade ("ABI"), as well as a school-specific variable that is the average school mark ("MSM") of all individual final grades. In addition, a school-specific random effect u_j is considered with corresponding variance σ_u^2 in order to adequately reflect the multilevel structure of students being nested within schools.

The probit estimation results and corresponding marginal effects (see Aßmann et al. (2014), forthcoming) for all four analyses are given in Table 1. These analyses showed no significant results regarding selection effects on the individual level for cohorts I and II and IV. Therefore the weight adjustment for participation for these cohorts is done via a nonresponse correction factor calculated as the ratio of all students within the cohort and the participating students. Obviously this yields equal correction factors for all students within a cohort, i.e.

$$w_{ijk}^{nr} = \frac{K_i}{\sum_{k=1}^{K_i} D_{ik}} \qquad \text{for } i \in \{I, II\}, \qquad \text{with } K_i = \sum_{j=1}^{S_i} K_{ij}$$

with S_i as the total number of schools per cohort $i \in \{I, II\}$ and D_{ik} being a participation indicator with $D_{ik} = 1$ if the student k is participating in the survey and $D_{ik} = 0$ otherwise. In comparison to individual selection effects, random effects on the school level can be found for all four cohorts. Participation in the study seems to depend to some extend on unmeasured school characteristics. To adequately compensate for this, a school level correction factor w_{ij}^{corr} has been defined as the average school level effect over the last $m_b = 5000$[2] of all $M = 10000$ draws from the gibbs sampler, that has been used to estimate the bayesian probit models:

$$w_{ij}^{corr} = \frac{1}{\frac{1}{m_b}\sum_{m=1}^{m_b} \Phi(\beta_0^m + u_j^m)} \qquad \text{for } i \in \{I, II\},$$

with β_0 as the coefficient for the intercept and u_j as the school level random effect for school j.

For cohort III - grade 13 students surveyed during the school year of 2011/12 - the probit regression analyses showed some significant results regarding selection effects with respect to the mean final mark within the school and a significant random effect. Thus, participation depends to some extent on (1) the school competence level in terms of the marks of all students and (2) on school characteristics mediated through the random effect u_j. Therefore the nonresponse weights for all students of cohort III have been calculated as

$$w_{ijk}^{nr} = \frac{1}{\pi_{ijk}} \qquad \text{for } i \in \{III\},$$

reflecting the probit response propensity for the whole model. Similar to cohorts I and II, the dependency on school characteristics also has been taken into account in this case. The overall weight corrected for nonresponse is then calculated as the product of the single weight components as follows:

$$cw_{ijk} = w_j^{school} \cdot w_{ij}^{corr} \cdot w_{ij}^{ind} \cdot w_{ijk}^{nr}.$$

[2] The first 5000 draws of the gibbs sampling routine have been discarded for burn-in.

The overall efficiency of the weights for the three cohorts surveyed in years between 2011 and 2012 can be illustrated according to the measure

$$E^i = \frac{(\sum_{k=1}^{n_i} cw_{ijk}^i)^2}{n_i \sum_{k=1}^{n_i}(cw_{ijk}^i)^2}, \quad i \in \{I, II, III\}$$

taking the values 0.903 for cohort I, 0.945 for cohort II and 0.844 for cohort III, respectively.

Table 1: Results of the bayesian probit regression models

	Probit coefficients				Marginal effects			
	par. est.	Std. Error	95% HDR		par. est.	Std. Error	95% HDR	
Cohort I (gross sample size N= 1858)								
Intercept	3.6751	2.2014	-0.7853	7.9241	0.6019	0.2289	-0.1720	0.7379
DEU	0.0048	0.0266	-0.0466	0.0569	0.0013	0.0073	-0.0128	0.0157
ENG	-0.0049	0.0223	-0.0488	0.0383	-0.0013	0.0061	-0.0134	0.0105
MAT	0.0246	0.0177	-0.0097	0.0593	0.0068	0.0049	-0.0027	0.0163
NAT	-0.0025	0.0269	-0.0550	0.0510	-0.0007	0.0074	-0.0150	0.0141
ABI	-0.1728	0.1385	-0.4354	0.1047	-0.0475	0.0381	-0.1203	0.0287
MSM	-1.1204	0.8661	-2.7740	0.5737	-0.3089	0.2375	-0.7562	0.1552
σ_u^2	0.5665	0.1470	0.3484	0.9065				
Cohort II (gross sample size N= 1918)								
Intercept	2.9582	2.8184	-2.4673	8.6752	0.4584	0.3237	-0.3222	0.6856
DEU	0.0436	0.0257	-0.0061	0.0933	0.0132	0.0078	-0.0019	0.0283
ENG	0.0043	0.0218	-0.0388	0.0467	0.0013	0.0066	-0.0117	0.0142
MAT	0.0017	0.0177	-0.0325	0.0364	0.0005	0.0054	-0.0099	0.0109
NAT	0.0269	0.0254	-0.0232	0.0771	0.0082	0.0077	-0.0071	0.0235
abinote	-0.0008	0.1497	-0.2876	0.3010	-0.0003	0.0454	-0.0870	0.0910
msm	-1.3138	1.1267	-3.5801	0.8757	-0.3985	0.3391	-1.0756	0.2630
σ_u^2	0.4615	0.1139	0.2842	0.7253				
Cohort III (gross sample size N= 1908)								
Intercept	7.5063	3.0193	1.5778	13.2993	0.6650	0.0869	0.4715	0.6997
DEU	0.0100	0.0263	-0.0425	0.0618	0.0029	0.0077	-0.0124	0.0180
ENG	0.0080	0.0218	-0.0348	0.0508	0.0023	0.0063	-0.0102	0.0149
MAT	0.0204	0.0184	-0.0155	0.0566	0.0060	0.0054	-0.0046	0.0165
NAT	0.0066	0.0281	-0.0479	0.0633	0.0019	0.0082	-0.0140	0.0185
abinote	-0.1651	0.1833	-0.5220	0.2008	-0.0483	0.0536	-0.1526	0.0585
msm	-2.9089	1.2009	-5.1916	-0.5724	-0.8497	0.3490	-1.5078	-0.1664
σ_u^2	0.5576	0.1364	0.3431	0.8764				

Note: 10000 Gibbs iterations performed, initial 5000 draws were discarded as burn-in. HDR denotes the highest density region. σ_u^2 denotes the variance of the school specific random effect.

4 Weight Trimming

To possibly increase the statistical efficiency of weighted analysis, the adjusted weights were trimmed. The general goal of weight trimming is to reduce sampling variance and, at the same time, to compensate for potential increase in bias. Trimming was performed using the so-called "Weight Distribution" approach (Potter 1990). Here, design weights are assumed to follow an inverse beta distribution with a cumulative distribution function F_w. Parameters of the sampling weight distribution are estimated using the sampling weights, and a trimming level τ is computed whose occurrence probability is 1%, that is, $1 - F_w(\tau) = 0.01$. Sampling weights in excess of τ are trimmed to this level and the excess is distributed among the untrimmed weights. The parameters for the sampling weight distribution are then again estimated using the trimmed adjusted weights, and a revised trimming level $\tilde{\tau}$ is computed. The trimmed adjusted weights are compared to the revised level $\tilde{\tau}$. If any weights are in excess of $\tilde{\tau}$, they are trimmed to this level, and the excess is distributed among the untrimmed weights. This procedure is iteratively repeated until no weights are in excess of a newly revised trimming level. To ease statistical analysis, the trimmed sampling weights are standardized with mean 1.

References

[1] Aßmann, C., Koch, S. and Schönberger, B. (2014, upcoming), Bayesian Analysis Of Binary Probit Models: The Case Of Measurement Error And Sequential Regression Modeling For Missing Explaining Factors, NEPS Working Paper Series.

[2] Rohwer, G. (2011), Use of Probabilistic Models of Unit Nonresponse. *http://www.uni-bamberg.de/fileadmin/inbil/Publikationen/dsw.pdf*

[3] Potter, F. (1990), A Study of Procedures to Identify and Trim Extreme Sampling Weights, in Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 225-230.