NEPS
National Educational Panel Study

Research Data

Curricular Reform Study in Thuringia (TH)
SUF Version 1.0.0
Data Manual [Supplement]: Weighting

Christian Aßmann,
Benno Schönberger

SPONSORED BY THE

Federal Ministry
of Education
and Research

Weighting the Thuringia Samples of the National Educational Panel Study

Technical Report

Christian Aßmann & Benno Schönberger*

National Educational Panel Study (NEPS)
University of Bamberg

version: October 10, 2012

*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 response propensity analyses for the curricular reform study in Thuringia conducted within the National Educational Panel Study (NEPS) in 2010 and 2011, 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 Thuringia. Students graduating in 2010 (last cohort before reform implementation) and 2011 (first cohort after reform implementation) in Thuringia have been surveyed via competence development tests and questionnaires provided that consent was given. Hence, the structural reform study is not a panel survey because each pupil was only tested once.

Access to the population of students attending grade twelve (final grade) in Thuringia was gained via a sample of 32 Gymnasia (grammar schools). All students in grade twelve of these schools were invited to participate in the survey. For the random selection of Gymnasia two explicit strata have been defined. The first stratum comprised all Gymnasia focusing on natural sciences, the second stratum contained all remaining Gymnasia. Within these two strata simple random sampling was performed if appropriate. The procedure was applied to sample schools providing access to pupils surveyed in 2010, where the same schools were contacted again one year later in 2011. The gross and net sample sizes of both samples are 1857 (gross sample) and 1374 (net sample) children in 2010 and 1365 (gross sample) and 886 (net sample) children in 2011, respectively.

We describe the methods used to calculate the design weights for the participating students in Section 2. In Section 3, we briefly address corrections for potential systematic nonresponse. Section 4 concludes with some general remarks on the use of weights.

2 Design weights

Design weights are calculated as inverse sampling probabilities allowing to adjust the sampling design for stratification. That is, assuming an inclusion probability π, the corresponding design weight is $1/\pi$. Of the eligible 105 Gymnasia in Thuringia, 20 schools are excluded from the sampling frame, since 12 of these schools have a specific curricular profile or private sponsorship (e.g., Waldorf-Schulen, Jenaplan-Schulen, reformpädagogische Schulen and Internationale Schulen) and 8 schools already participate in other NEPS studies and double burden for these schools is to be avoided, see also Sibberns et al. (2011). Given this setup, a twofold stratification is implemented. The first stratum contains all Gymnasia with a focus on natural sciences. As this stratum contains only two schools, all schools are included within the sample. Within the second stratum, comprising the remaining 83 schools, a simple random sample of size 30 has been drawn with replacement. Four of the sampled schools refused participation and have thus been replaced by a priori defined schools with similar characteristics. That is, for the two considered strata $l = 1$, 2 the resulting weights w_l^{2010} of schools surveyed in 2010 are

$$w_l^{2010} = \begin{cases} 1, & \text{if } l = 1, \\ \frac{83}{30} = 2.7667 & \text{else.} \end{cases} \tag{1}$$

All schools asked in 2010 were also asked to participate and provide access to pupils in 2011. However, one school refused participation. Hence, a correction factor is included in the weights of schools surveyed in 2011 then given as

$$w_l^{2011} = \begin{cases} 1, & \text{if } l = 1, \\ \frac{83}{30} \cdot \frac{30}{29} = 2.8621 & \text{else.} \end{cases} \tag{2}$$

As all students within the sampled schools were asked to participate in the study, the individual design weight of each student is equal to the design weight of the corresponding school.

3 Weight adjustment for participation

Given the inclusion probabilities and corresponding design weights, we analyze the willingness to participate in the survey. For the samples corresponding to the years 2010 and 2011, the individual participation and response probability of child i attending one of the participating schools π_i^y, $y \in \{2010, 2011\}$, is derived via an extended probit regression model, see Aßmann et al. (2012). The standard probit framework has been extended to cope with a measurement error observed within the students' participation indicator and missing values in the covariates.

Estimation of this extended model framework is based on Markov Chain Monte Carlo (MCMC) techniques, where the iterative sampling scheme is enriched by sequential regression trees, an imputation approach originally described by Breiman et. al (1984) and recently used by Burgette & Reiter (2010). The dependent variable for the probit regressions is the participation status of students. As covariates, we included mean marks for every student according to three fields of subjects, from which they have to choose their subjects for the graduation years. These fields of subjects are linguistic-literary-artistic subjects (fs1) (e.g., German, English, arts, music), social subjects (fs2) (e.g., geography, history, religion) and mathematical-natural-scientific-technical subjects (fs3) (e.g., maths, physics, biology, computer sciences). As further covariates, we regard the gender of students (sex), 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 is considered with corresponding variance denoted as σ_u^2. The bayesian probit estimation results and marginal effects and are given in Table 1.

Table 1: Results of the probit regression models measuring participation probabilities

	2010				2011			
	Probit coefficients							
	par. est.	Std. Error	95% HDR		par. est.	Std. Error	95% HDR	
Intercept	1.1453	1.3044	-1.3631	3.7738	-0.9679	2.7319	-6.3757	4.4437
sex	-0.1555	0.0763	-0.3037	-0.0051	-0.2048	0.0978	-0.3989	-0.0153
fb1	-0.0608	0.0361	-0.1308	0.0101	0.0299	0.0419	-0.0517	0.1119
fb2	0.0117	0.0326	-0.0514	0.0754	-0.0255	0.0381	-0.1011	0.0479
fb3	0.0529	0.0230	0.0078	0.0983	0.0246	0.0273	-0.0289	0.0776
msm	-0.1441	0.5552	-1.2549	0.9185	0.6253	1.2559	-1.8728	3.1242
σ_u^2	0.3817	0.1151	0.2147	0.6545	1.1025	0.3476	0.6040	1.9511
	Marginal effects							
	par. est.	Std. Error	95% HDR		par. est.	Std. Error	95% HDR	
sex	-0.0433	0.0213	-0.0850	-0.0014	-0.0514	0.0246	-0.1002	-0.0039
fb1	-0.0161	0.0095	-0.0342	0.0028	0.0073	0.0101	-0.0123	0.0274
fb2	0.0031	0.0087	-0.0138	0.0201	-0.0061	0.0091	-0.0245	0.0116
fb3	0.0140	0.0061	0.0021	0.0259	0.0059	0.0065	-0.0070	0.0186
msm	-0.0382	0.1469	-0.3291	0.2442	0.1481	0.2984	-0.4494	0.7313
gross sample size	1857				1365			

Note: 20000 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.

In conclusion, we find only slight effects influencing the decision to participate in the study regarding the students' gender with all other variables showing no influence on the participation propensity. This holds true for both the 2010 and 2011 sample. The nonresponse adjusted weight for student i is given as

$$aw_i^y = \frac{1}{\pi_i^y},$$

whereas the combined weights for a student i is

$$cw_i^y = w_{i|l}^y \cdot aw_i^y,$$

where $w_{i|l}^y$, $y = \{2010, 2011\}$, denotes the design weight of student i attending a school in stratum l. Considering the homogenous target population of twelfth graders in Thuringia, no additional calibration according to known population totals, e.g. the number of twelfth graders attending Gymnasia in the particular two years, has been performed. The overall efficiency of the weights can be illustrated according to the measure

$$E^y = (\sum_{i=1}^{n} cw_i^y)^2 / n \sum_{i=1}^{n} (cw_i^y)^2$$

taking the values 0,91 for 2010 and 0,65 for 2011 respectively.

4 Use of weights

No general recommendation is at hand concerning the use of design and adjusted weights, because the choice of which weight to use depends on the analysis and the question the researcher wants to answer. In general, the use of design weights is always recommended when analysis is concerned with sample data comprising data of different schools – presumably the case in most analysis. No general results are available how the use of design or adjusted weights render any possible analysis (see Rohwer (2011) for a general discussion). Nevertheless, the use of weights may help to highlight important features of the analysis under consideration, not at least serving as a robustness check for the performed analysis. The response propensity analysis points at the usual caveat to control for gender effects when analyzing the Thuringia data. Furthermore, as highlighted by the considered school specific random effect, correlation among pupils within a school may prevail.

Provided design weights of participating children are labeled as `weight_design` and participation adjusted weights are labeled as `weight_adj`. The combined weight is given as `weight_total`. Note that also standardized weights with mean one are provided, which are often used in statistical analysis. These are labeled as `weight_design_std`, `weight_adj_std`, and `weight_total_std`. The subsequent table summarizes all type of weights provided:

Type of weight	not standardized	standardized with mean one
design weight	`weight_design`	`weight_design_std`
participation adjusted weight	`weight_adj`	`weight_adj_std`
combined weight	`weight_total`	`weight_total_std`

For further information on weighting please contact `methods.neps@uni-bamberg.de`

References

[1] Sibberns, H., Bundt, S., Gomolka, J., Martin, G. and Knoll, S. (2011), Feld- und Methodenbericht A70, Reform der gymnasialen Oberstufe in Thüringen, Haupterhebung K12, Data Processing and Research Center (DPC).

[2] Aßmann, C., Koch, S. and Schönberger, B. (2012, upcoming), Bayesian Analysis Of Binary Probit Models: The Case Of Measurement Error And Sequential Regression Modeling For Missing Explaining Factors, NEPS Working Paper Series.

[3] Breiman, L. and Friedman, J.H. and Olshen, R.A. and Stone, C.J. (1984), Classification and Regression Trees, Chapman and Hall, New York.

[4] Burgette, L.F. and Reiter, J.P. (2010), Multiple Imputation for Missing Data via Sequential Regression Trees, American Journal of Epidemiology (172):9, pp. 1070-1076.

[5] Chib, S. and Greenberg, E. (1995), Understanding the Metropolis-Hastings Algorithm, The American Statistician (49):4, pp. 327-335, American Statistical Association.

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