Choice of Small Area Models Based on Sample Designs and Availability of Auxiliary Data in PIAAC Study

Jianzhu Li[1], Tom Krenzke[1], Leyla Mohadjer[1]
Wendy VanDeKerckhove[1], Lin Li[1]
Weijia Ren[1], Henok Adbaru[1]
[1]Westat, 1600 Research Blvd., Rockville, MD 20850

Abstract
The Programme for the International Assessment of Adult Competencies (PIAAC) is an international survey conducted in about 40 countries to measure adult proficiency in key information-processing skills. Although the survey was designed to produce nationally representative estimates with adequate precision, most countries are also interested in estimates for local areas where the sample size is very small. The application of small area estimation (SAE) approaches provides an affordable option. SAE methods are a set of model-dependent approaches that employ a statistical model using auxiliary information and the survey data to produce indirect estimates when survey data alone are inadequate for direct estimation. The PIAAC participating countries adopted various sample designs from single-stage stratified sample from a registry to multiple-stage area sample involving stratification and clustering. This paper discusses the impact of different design features and sources of auxiliary data on the choice of appropriate small area models to estimate the proportion of adults lacking basic literacy skills in local areas. We used a few countries to demonstrate how the decisions are made.

Key Words: small area estimation, indirect estimate, sample design

1. Introduction

The Programme for the International Assessment of Adult Competencies (PIAAC) is a multicycle survey of adult skills and competencies sponsored by the Organisation for Economic Cooperation and Development (OECD, 2016). The PIAAC sample is designed to produce internationally comparable and nationally representative direct estimates (based solely on survey data) with adequate levels of precision for the nations as a whole and for major population subgroups. However, the OECD and several of the participating countries in Cycle 1 of PIAAC have expressed interest in using PIAAC data to create proficiency estimates for local areas where the PIAAC sample size is too small (or equal to zero) to produce any direct estimates. Small area estimation (SAE) methods facilitate the estimation of the proficiency distribution in subpopulations not initially targeted in large-scale surveys. A considerable amount of research and development in SAE methods has taken place since the text by Rao (2003), which presents a comprehensive overview of the methods, history, and applications of SAE methods. The book has since been updated (Rao and Molina, 2015), and much research and development activity has been ongoing on this topic in recent years. The development of SAE approaches has made it possible to meet the growing demands for more information at lower levels of geography. It is no different for PIAAC. The application of SAE approaches to PIAAC data may provide an affordable option for countries to produce indirect estimates for their small areas of interest.

This paper summarizes the research results from applying SAE methods using PIAAC data from five countries that participated in Cycle 1, with various core national sample designs. We refer to this work as Phase 1 research—using a group of countries to explore what models are suitable for different countries. It is the cornerstone of the Phase 2 (production phase) work in the future for which we will focus more on each country individually, and which will involve more work in selecting covariates and performing model diagnostics to help evaluate various sets of SAEs toward publishable estimates.

Section 2 of the paper provides some background on PIAAC, SAE techniques, and goals of this research. Section 3 contains a description of the sample designs, sample sizes and small areas, and the auxiliary variables (covariates) for each country. In Section 4, the direct estimation process is discussed, which includes the use of the survey regression estimator and variance smoothing. Section 5 introduces the models. The work includes an evaluation of methodology and approaches, including both unit-level and area-level SAE modeling. Section 6 summarizes the overall outcomes and a conclusion on critical factors to be considered.

2. Background

2.1 PIAAC Study

PIAAC examines a range of basic skills in the information age and assesses these adult skills consistently across participating countries. The first cycle of PIAAC includes three rounds: 24 countries participated in 2011–12 (round 1); 9 additional countries participated in 2014–15 (round 2); and 5 additional countries are participating in 2017–18 (round 3). In general, the sampling goal was to achieve 5,000 completed assessments in three domains: Literacy, Numeracy, and Problem-solving in technology-rich environments. The test design for PIAAC is based on a variant matrix sampling (OECD, 2016) whereby each respondent was administered a subset of items from the total item pool. Therefore, item response theory (IRT) scaling was used to derive scores for each domain. To increase the accuracy of the cognitive measurement, PIAAC uses plausible values (multiple imputations) drawn from a posterior distribution by combining the IRT scaling of the cognitive items with a latent regression model using information from the background questionnaire (BQ) in a population model.

2.2 Small Area Estimation

The essence of SAE is to use covariates at the small-area level in combination with survey data to model the small area parameters of interest. As the demand for reliable small area estimates has greatly increased in the past decades, the SAE literature and research findings also has grown rapidly. Section 4 describes the various approaches developed under SAE methodology and used in this research. In general, there are two major types of models: area level and unit level models. The area-level approach models the small area parameter of interest in terms of covariates at the area-level, whereas the unit-level approach models the underlying variable of interest in terms of unit-level covariates, and then aggregating the individual predictions for each small area.

2.3 Goals of Research

The main purpose for the Phase 1 research is to evaluate various SAE approaches across countries of different sizes and with different PIAAC sample designs toward developing an understanding, and guidance, on how SAE can be implemented for PIAAC. Both types of models, area-level and unit-level, were fit to data from each of the five participating countries: Germany, Italy, New Zealand, Slovakia, and Sweden. In this effort, we were interested in producing SAEs of adults at the lower literacy levels, specifically, the proportion in Level 1 or below.[1]

3. Country Data

The SAE procedure can vary depending on the country's sample design, the definitions and sample sizes of the small areas, and the available covariates. To process the small area estimation models, each country provided two files: a PIAAC data file and a population file. The PIAAC data file was to include the following variables for each PIAAC respondent: person identifier, small area (SA) identifier, variance cluster identifier, final full sample and replicate weights, literacy scores (10 plausible values), and covariates. The population file was to include the covariates for the universe of persons for each SA. If the country was not able to include population totals for a full crosstab of the covariates by SA, arrangements were made for countries to provide frequencies or partial cross-tabulations (e.g., involving 1 to 3 variables). For fitting a unit-level model, the covariates on the PIAAC data file should have the same coverage, definitions, and categories as those on the population file.

Germany and Sweden faced confidentiality restrictions in providing microdata. Germany could not provide the SA identifier for each respondent and thus supplied PIAAC data summarized to the area level. Without the respondent-level data, it is not possible to fit a unit-level model, such as a survey regression estimator (SRE) (described in Section 4.1) or a traditional small area unit-level model. Sweden had some interest in producing small area estimates for the 21 counties but could not provide the microdata at this level. They opted to use the eight broader areas identified on the PIAAC public use file. Given the small number of SAs, a model-assisted direct estimation approach was conducted, as well as a unit-level EBLUP model.

3.1 Sample Design

The sample designs varied across countries. Because of the need to conduct the assessment in-person, most countries chose to cluster their sample into primary sampling units (PSUs) to reduce costs of interviewing within households. Table 1 summarizes the sample designs and sample sizes for the five participating countries. All countries but Sweden had clustered samples, with between 260 and 1,000 units at the first stage. The final sample sizes ranged from 4,469 to 6,177.

Table 1: Sample Designs

Country	*Sample design*	*Number of sampled PSUs*	*Number of completes*
Germany	2-stage cluster sample	277	5,465
Italy	3-stage cluster sample	260	4,621
New Zealand	4-stage cluster sample	1,000	6,177
Slovakia	2-stage cluster sample	562	5,723
Sweden	1-stage sample	Not applicable	4,469

[1] In addition, we included statistics on average literacy scores (mean values) in our research to fully examine and evaluate various methods and models. Details can be found in http://www.oecd.org/skills/piaac/PIAACSAEInitialResearchReport10Sept2018.pdf.

Inherent in PIAAC is both informative sampling (clustering and differential base weights) and informative nonresponse (non-ignorable proficiency-related nonresponse), as evident by the steps included in the weighting process which accounts for differential probabilities of selection, nonresponse adjustments, and calibration of the weights. Both informative sample design and nonresponse should be taken into account when generating SAEs.

3.2 Defining Small Areas

The SA definitions for each country are given in Table 2. In all five countries, the SAs are larger areas than the PSUs, meaning that the sample is clustered within an SA. The number of SAs varies from eight for Sweden to 110 for Italy. Germany, Slovakia, and Sweden have PIAAC sample in all areas. Italy and New Zealand have sample in over 80% of areas. In addition, the sample size within an SA varies. For Germany, Italy, and New Zealand, the majority of SAs have between 31 and 100 completed cases. For Slovakia, over 50% have over 100 completed cases, and for Sweden, all SAs have a sample size over 100.

Table 2: Small Area Definitions, Population Counts, and Sample Sizes

				Number of SAs with n =		
Country	*Small area (SA) description*	*Number of SAs*	*Number of SAs with sample*	*1-30*	*31-100*	*101+*
Germany	Collapsed spatial planning regions	85	85	12	60	13
Italy	Provinces	110	91	35	50	6
New Zealand	Territorial Authorities/ Community Boards	87	84	21	47	16
Slovakia	Districts/counties	79	79	15	19	45
Sweden	NUTS2 statistical regions	8	8	0	0	8

3.3 Covariates

The covariates in the SAE models should be highly predictive of the SA estimates of interest. The population data were recommended to include information about age, gender, race/ethnicity, education attainment, employment status, poverty status, and foreign-born status. In addition, the population totals should come from a population census, administrative data, or a large national survey. For fitting a unit-level model, the covariates on the PIAAC data file should have the same coverage, definitions, and categories as those on the population file.

Table 3 shows the covariates provided by each country. It also shows the number of levels for each covariate. Some covariates were available only on the population file and did not have an equivalent variable in the PIAAC data. In addition, education and employment status were often found to match poorly between the two files. Such covariates can be used in area-level models only. The covariates with consistent definitions between the PIAAC and population files are indicated in bold.

Table 3: Covariates on Country Population Files

Country	X_1	X_2	X_3	X_4	X_5	X_6	X_7	*Source of population totals*
Germany	Gender*	Age* (4 levels)	Nationality* (3 levels)	Educational attainment* (5 levels)	Employment Status* (3 levels)			Micro Census (2011)
Italy	**Gender**	**Age** (Exact age)	Citizenship* (2 levels)	**Educational attainment** (6 levels)	**Employment Status** (7 levels)	**Number of people in household** (5 levels)	Marital status* (6 levels)	Census (2011)
New Zealand	**Gender**	**Age** (6 levels)	**Birthplace** (2 levels)	Highest qualification (4 levels)	Work and Labor force status (2 levels)	Ethnic Group (3 levels)		Census of Population and Dwellings (2013)
Slovakia	**Gender**	**Age** (21 levels)	Nationality* (16 levels)	Highest education (9 levels)	Economic activity (13 levels)	Language spoken at home* (14 levels)		Population Census (2011)
Sweden	**Gender**	**Age** (5 levels)	**Birthplace** (2 levels)	Highest education (4 levels)				Swedish register (2012)

* On population file only; not available on PIAAC data file.
NOTE: Bold font indicates consistent definitions between the PIAAC and population files.

4. Direct Estimation

The PIAAC IRT modeling resulted in 10 plausible values (PV) for each respondent, reflecting the uncertainty in the respondents' proficiency estimate. More information can be found in OECD (2016). To handle the plausible values properly, a multiple imputation (MI) approach, as shown in Rubin (1987), was used for calculating direct estimates and the associated variances.

4.1 Survey Regression Estimator (SRE)

The survey regression estimator (SRE) is a model-assisted approach that is used to bring SA population estimates in line with external SA totals and improve the stability of the survey estimates. The SRE process also helps to reduce variances that are used in the SAE modeling process. Rao and Molina (2015, pp. 21-23) describe the use of these estimates in small area estimation, their derivation, and the usual approach to estimating their variance. In addition to the x covariates available for the respondents, the values of the population totals $\mathbf{X}_j$ in SA j must be available for this estimator. For Germany, we did not have the covariates for the respondents, and thus no SRE was produced. For the other countries, the SRE was derived for each plausible value as follows:

$$\hat{y}_{jl}^{surv} = \hat{y}_{jl} + (\bar{X}_j - \bar{x}_j)'\beta \tag{1}$$

where $\hat{y}_{jl}$ = the survey estimate based on the l-th plausible value for SA j;
$\bar{X}_j$ = the vector of population means of the covariates;
$\bar{x}_j$ = the vector of sample means of the covariates; and
β = the vector of regression coefficients from the regression model of the relationship between y and x.

The covariates were limited to variables that were defined consistently for the respondents and the population. The list of covariates used in the SRE model for each country is given in Table 4. Italy has a larger number of covariates and some of them have high correlations with the outcomes at the small area level. On the other end, Slovakia's SRE was limited to using age and gender.

Table 4: List of Covariates for the SRE and Strength of Covariates

Country	*Covariates (correlation with direct estimate in parentheses)*					
Italy	Gender, 1 level (0.02)	Age, Mean (-0.32)		Education attainment, 2 levels (0.41, -0.31)	Employment status, 1 level (-0.36)	Number of people in the household, 2 levels (-0.42, 0.29)
New Zealand	Gender, 1 level (0.05)	Age, 4 levels (0.27, 0.06, -0.26, -0.17)	Birth-place, 1 level (0.02)			
Slovakia	Gender, 1 level (-0.32)	Age, 4 levels (0.46, -0.08, -0.14, -0.15)				
Sweden	Gender, 1 level, (0.43)	Age, 4 levels (-0.14, 0.23, 0.12, -0.28)	Birth-place, 1 level (-0.24)			

We then applied the MI formulae to produce the overall SRE estimate as:

$$\hat{y}_j^{surv} = \frac{1}{10} \sum_{l=1}^{10} \hat{y}_{jl}^{surv}, \tag{2}$$

and the variance as:

$$\hat{\sigma}_{j(SRE)}^2 = \hat{\sigma}_{Wj(SRE)}^2 + \left(\frac{11}{10}\right) \hat{\sigma}_{Bj(SRE)}^2, \tag{3}$$

where $\hat{\sigma}_{Wj(SRE)}^2$ is the within-imputation variance and $\hat{\sigma}_{Bj(SRE)}^2$ is the between-imputation variance for the mean residuals from the SRE model. Variances were calculated using the final replicate weights and appropriate replication method for the country.

4.2 Smoothed Variances

Since the direct or SRE estimates of the variances are subject to substantial sampling error, the true variances (or relative variances $\varphi_j^2 = \sigma_j^2 / p_j^2$) were predicted using a modeling approach. Since the relative variance of an SA estimate depends on the value of the SA's proportion at or below Level 1 in literacy, a two-step approach was implemented to produce model-dependent estimates of the relative variances. The approach followed the one implemented in the 2003 National Assessment of Adult Literacy (NAAL) SAE program (Mohadjer, Kalton, Krenzke, Liu, Van de Kerckhove, Li, Sherman, Dillman, Rao, and White, 2009; Mohadjer, Rao, Liu, Krenzke, and Van de Kerckhove, 2011). In step 1, the proportions at or below Level 1 in literacy were predicted from a simple regression model relating the SRE estimates $\hat{p}_j^{surv}$ to predictor variables.

$$logit(\hat{p}_j^{surv}) = \gamma_0 + \gamma Z_j + e_j, \quad (4)$$

where $\hat{p}_j^{surv}$ = the proportion of adults at or below Level 1 in literacy from the SRE model;
$Z_{\boldsymbol{j}}$ = the predictor variables (given in Table 3-3); and
e_j = the error term.

In step 2, the resulting predicted proportions from step 1 were used in a generalized variance function (GVF) model to smooth the SRE relative variance estimates.

$$\log(\varphi_{j(SRE)}^2) = \eta_0 + \eta_1 \log(\tilde{p}_j) + \eta_2 \log(1 - \tilde{p}_j) + \eta_3 \log(n_j) + e_j, \quad (5)$$

where $\varphi_{j(SRE)}^2$ = the SRE relative variance of the proportion at or below Level 1 in literacy;
$\tilde{p}_j$ = the predicted proportion from step 1;
n_j = the sample size; and
e_j = the error term.

The predicted values of the relative variances for the SA proportions of adults at or below Level 1 in literacy were then computed based on the above GVF regression model, and these predicted values were treated as known relative variances in the small area models. The variance smoothing step was not performed for Sweden because the sample sizes for SAs were adequate. With the exception of Germany, the SRE estimates served as input to the variance smoothing process. For Germany, there were no SRE estimates, so the direct estimates served as inputs.

4.3 Results

Figure 1 compares direct estimates and SRE estimates for estimates of the proportion at or below Level 1 in literacy. The results for each country are shown in a shrinkage plot, with the arrow starting from the direct estimate and ending at SRE estimate. The x-axis is the square root of the sample size. Estimates that changed by more than 0.02 (or 2 percentage points) for proportions are highlighted as red. The results indicate that the SRE had the largest impact on the point estimates for Italy, and it had the least effect on the point estimates for Slovakia. This could be related to the number and strength of available covariates (see Table 4).

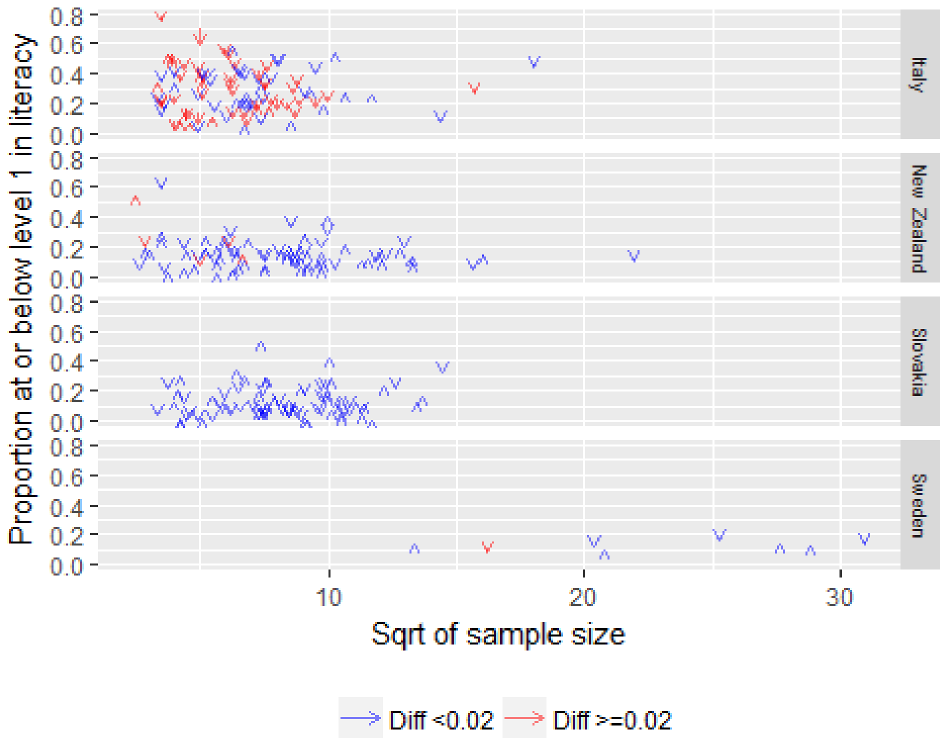

Figure 1: Shrinkage plots comparing the direct and SRE estimates of the proportion at or below Level 1 in literacy.

Figure 2 shows the shrinkage plots comparing the direct and smoothed standard error estimates for the proportion at or below Level 1 in literacy. The smoothing process had a larger impact in the SAs with smaller sample sizes. This is expected, as the direct variance estimates are less stable in such SAs.

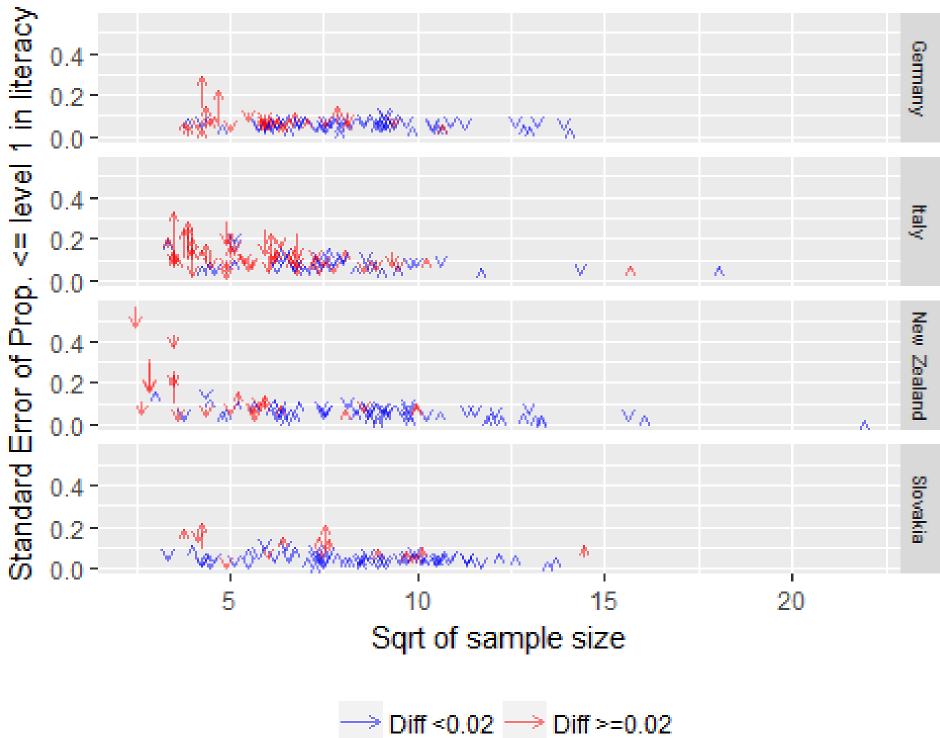

Figure 2: Shrinkage plots comparing the direct and smoothed standard error estimates for the proportion at or below Level 1 in literacy.

5. Small Area Estimation Models

The SAE models need to account for the variance impact from complex samples, which includes differential weighting in direct estimates, and clustering. If the sample of the small areas is not selected as a simple random sample, the sample design is informative. Also, weighting adjustments for nonresponse can reduce bias to the extent that the weighting variables are related to the proficiency scores.

In an area-level model, direct estimates produced at the local area-level are the prime elements in the modeling process. One part of an area-level model is a "sampling model," where survey-weighted estimates are produced for the small areas with sample-design based variance estimates. The other part is the "linking model" (or regression model), which is developed using predictors at the small-area-level and could include variables at higher levels also. One can also distinguish between "matched" and "unmatched" models, where the former has the survey weighted estimate directly as the dependent variable in the model regression, and in the latter case, a functional transformation (e.g., the logit function) provides the link to the predictors; that is, the regression model and sampling model do not blend together directly.

Unlike the area-level approach, the unit-level model is built at a much lower level such as individual persons or households. That is, a unit-level model uses covariates available at the person level to generate person-level values, which are aggregated to compute statistics at the area level. There is potential for smaller MSE and for producing estimates for a wide range of other subgroups of interest. There is no effort to generate sample-design unbiased estimates. The basic unit-level models ignore sample-design based variance estimates at this very low level.

If the model is linear, either the area-level or the unit-level approach could be used for a PIAAC small-area program. In the nonlinear case (e.g., in estimating a small proportion), a full cross-tabulation of the covariates is needed at the small area level. The area-level approach is more design-based, since the basic building blocks are the sample (design-based) estimates at the targeted local level, as well as the sample-design based variance estimates at this level. The unit-level approach is more dependent on the validity of the model, as it disaggregates down to the lowest levels. Sampling weights can be used to estimate the parameters of the model, which can make this portion of the estimation process sample-design consistent. Variance estimates are entirely model dependent. Extensions have included a random-effect term as an attempt to capture the between area variation (see Rao and Molina, 2015).

Operationally, the area-level approach works with a much simpler data set, with one record for each local area rather than one record for each household or person, and in that sense is easier to work with in practice. This is especially useful as the Bayesian methods require numerous iterations with the data set as an input in each iteration.

5.1 Models

The models evaluated in this research were:

- Fay-Herriot (F-H) area-level model

$$\hat{p}_j = x_j'\beta + u_j + e_j \tag{6}$$

where the area-level random effects u_j and the sampling error e_j were assumed to be normally distributed with mean zero and variances σ_u^2, and σ_e^2, respectively.

- Hierarchical Bayes (HB) area-level matched (linear) model

The HB approach has the same model assumptions as the Fay-Herriot (F-H) area-level model and uses a flat prior distributions for β, and gamma priors for σ_e^2, σ_u^2.

- HB area-level unmatched (nonlinear) model

$$\begin{aligned} \hat{p}_j &= p_j + e_j \\ z_j &= x_j'\beta + u_j \end{aligned} \tag{7}$$

where $z_j = \ln(p_j/(1-p_j))$. The other assumptions are similar to the Hierarchical Bayes (HB) area-level matched model.

- Unit-level empirical best linear unbiased predictor (EBLUP)

$$\hat{y}_{jk} = \beta_0 + x'_{jk}\beta + u_j + e_{jk}, \tag{8}$$

where $\hat{y}_{jk}$ is the indicator whether a respondent is at or below Level 1, and x'_{jk} is the covariates for the respondent k in small area j. The random effects u_j and e_{jk} were assumed to be normally distributed with mean zero and variances σ_u^2, and σ_e^2, respectively.

5.2 Results

The results from evaluating the SAE model in Section 4.1 are summarized in graphs. Figure 3 shows the scatterplots of SAE estimates versus direct estimates for each country. In Figure 3, most of the bubbles are located around the 45-degree lines, indicating that the direct estimates and the model estimates are close to each other. Some of the small bubbles, with the sizes of bubbles being proportional to the sample sizes in the small areas, are farther away from the 45-degree lines. This is as expected because the direct estimates contribute less to the model estimates when derived from samples of smaller sizes and associated with higher sampling errors (i.e., less reliable). The bubbles in the plot show that the model estimates are usually smaller than the direct estimates when the estimated proportions are larger than 20 percent, with the Fay-Herriot results being more extreme than the other models. Some investigation into Germany's results has shown that there is at least one very small smoothed standard error that is influential to the Fay-Herriot results. Removal of the influential case provides results very close to the matched HB model results. In Phase 2, the smoothing model would be further investigated to determine the way to address the influential outlier.

The area level covariates used in Germany's models are from the 2011 Micro Census. Weak associations are observed between the area level covariates and the direct estimates. As a result, the models have low predicting power and may not work well for improving the quality of the direct estimates.

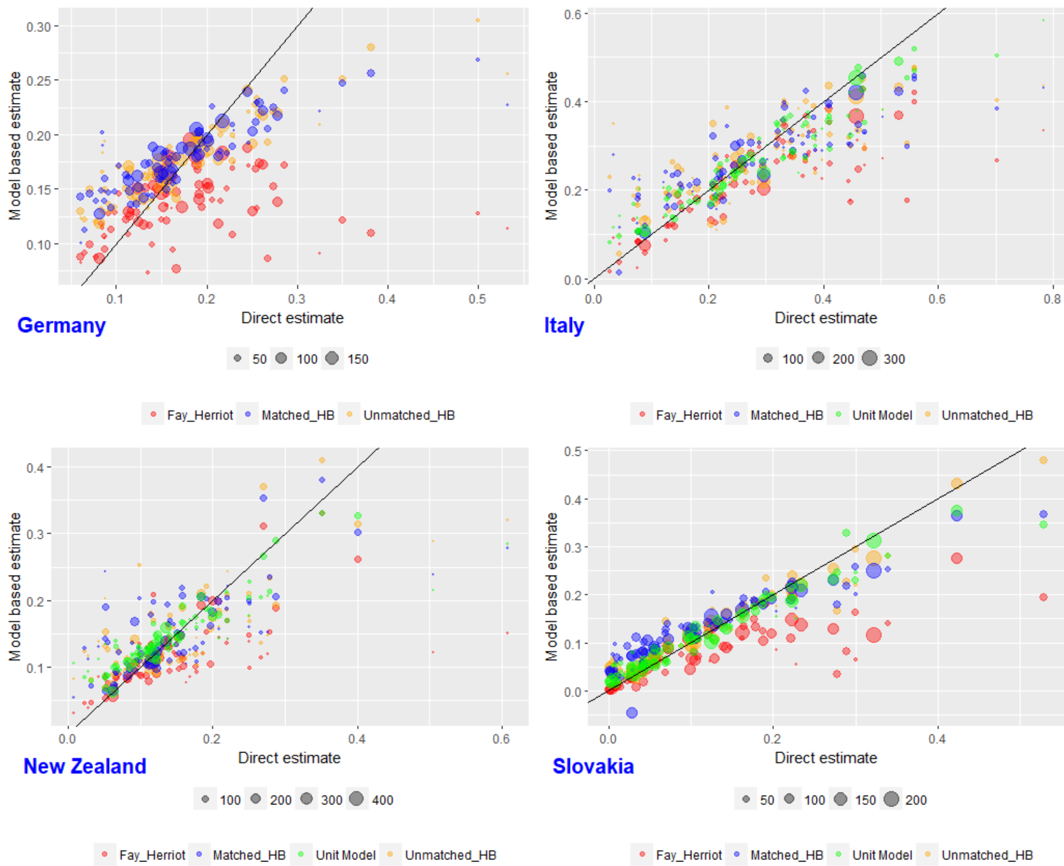

Figure 3: Proportion at or below Level 1: Scatterplot of SAE and direct estimates, with sample size as bubbles.

Figure 4 shows the standard errors of the direct estimates and the MSEs of all model estimates in one plot for each country. For these plots, keep in mind that the MSEs depend on the size of the estimated proportion. Therefore, if the model proportion is different from the direct proportion, the variance will in theory be different; thus, the resulting MSE is not necessarily an improvement to the estimates due to the model. The MSE plot shows that almost all models produce smaller MSEs than the direct estimates, especially for areas of very small sample sizes. For Slovakia there appears to be a less positive impact on the precision from the models. However, some investigation revealed that several direct estimates are close to zero and that SAEs have slightly higher values (likely due to shrinkage). As seen in the formula for the standard error of a proportion, the standard errors for proportions are associated with the magnitude of the proportion, and therefore it is hard to make a conclusion as to the impact on standard errors, especially in the case of Slovakia's proportions.

In general, the unit-level EBLUP does not account for sample weight and design features. Therefore, the MSEs generated from the unit-level EBLUP models show strong correlation with sample sizes.

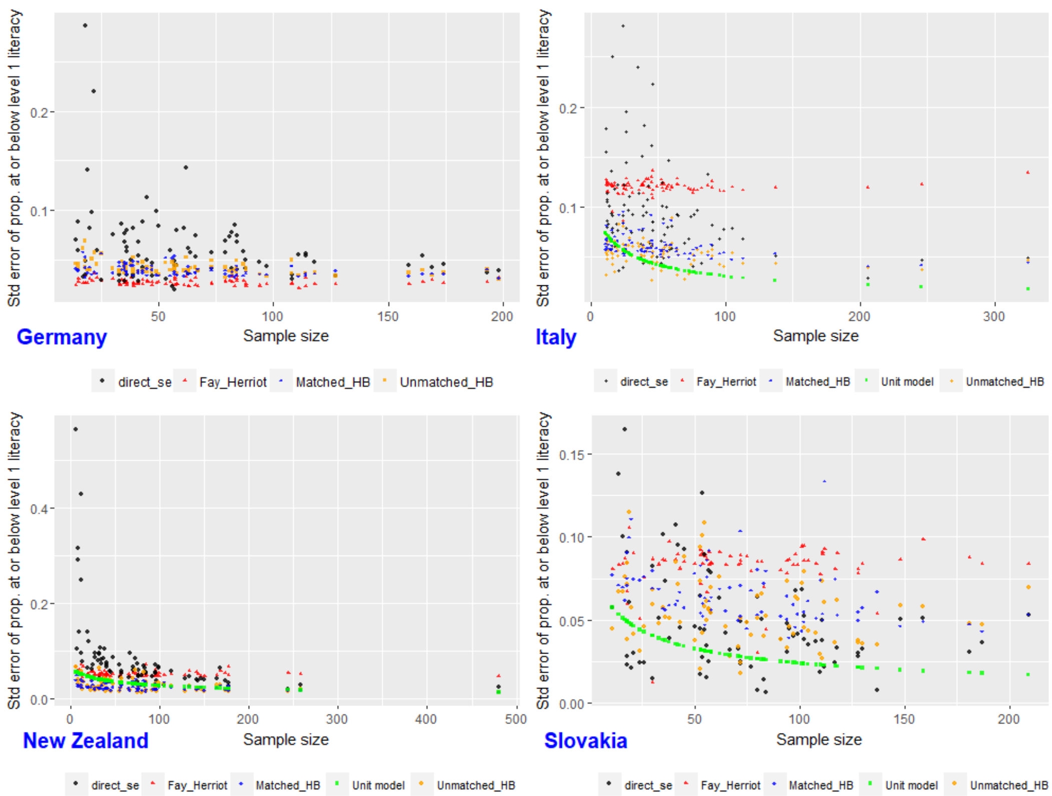

Figure 4: Proportion at or below Level 1: Comparison of standard errors between direct and SAE approaches.

6. General Findings

In general, the covariates used in this research tend to have a good association with the direct estimates. However, the covariate pool was limited, and the predictor search was deterministic due to the limited choices available. The research concluded that area-level models have a good potential for providing reliable PIAAC SAEs for all countries. This is evident by the results, showing that the SAEs are close to direct estimates for SAs with the largest sample sizes, and the SAE models showed impact on SAs with the smallest sample sizes. Also, in general, the MSE plots show that almost all the models produce smaller MSEs than the direct estimates, especially for areas with very small sample sizes. From the evaluation plots, the unmatched and matched HB models have the most potential. Area-level models can rely on area-level data, which may be the only data available. The unit-level model has potential for better estimates for countries with a wealth of registry data and without clustering within their SAEs. One can see the direct association of the standard errors to the sample size for unit-level models in Figure 4, for example, illustrating the effect of ignoring the design effect.

One outcome from the initial research was to identify scenarios for each country that factor into the decision about the SAE model framework for Phase 2 (creating publishable SAEs). These factors are 1) whether external covariate information is available for which the variables match the survey item definitions, 2) whether design contains informative sampling both in terms of clustering within the SAs and variation in the weights, and 3) if the estimated proportion is on average less than 0.20 or not. Table 5 provides the various scenarios and a recommended model to use for SAE. As seen in the initial research, the recommended model is not necessarily as clear-cut as Table 5 shows, and therefore, some investigation of the model-type choice is typically needed in practice.

Table 5: SAE scenarios and recommended model type

Scenario	*Covariates match survey item definitions?*	*Informative sampling*		*Estimated proportion < 0.2?*	*Recommended SAE model for proportions*
		Clustering within small areas?	*Differential weights?*		
1	Y	Y	Y	Y	ALU
2	Y	Y	Y	N	ALM
3	Y	Y	N	Y	ALU
4	Y	Y	N	N	ALM
5	Y	N	Y	Y	PE*
6	Y	N	Y	N	PE*
7	Y	N	N	Y	UL*
8	Y	N	N	N	UL*
9	N	Y	Y	Y	ALU
10	N	Y	Y	N	ALM
11	N	Y	N	Y	ALU
12	N	Y	N	N	ALM
13	N	N	Y	Y	ALU
14	N	N	Y	N	ALM
15	N	N	N	Y	ALU
16	N	N	N	N	ALM

*Note: If non-linear model, then a full cross-tab of covariates is needed at small area level.

ALU = Area-level unmatched model

ALM = Area-level matched model

UL = Unit-level model

PE = Pseudo-EBLUP

Acknowledgements

The authors wish to acknowledge the support from the U.S. National Center for Education Statistics and the Organisation for Economic Cooperation and Development. The authors are grateful to the countries for the help and participation in this research. In addition, the authors acknowledge the invaluable insights and guidance provided by Bob Fay and Jon Rao throughout the small area research. Their vast experience in small area estimation was instrumental in resolving the various challenges faced during the research and development process.

References

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J.N.K., and White, S. (2009). *National assessment of adult literacy indirect county and state estimates of the percentage of adults at the lowest literacy level for 1992 and 2003* (NCES 2009-482). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van de Kerckhove, W. (2011). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics*, 66(1), 1-9.

Organisation for Economic Cooperation and Development (OECD). (2016). *Technical report of the survey of adult skills (PIAAC), 2nd edition*. Paris, France: OECD. Retrieved from https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

Rao, J.N.K. (2003). *Small area estimation* (Wiley Series in Survey Methodology). New York: Wiley.

Rao, J.N.K., and Molina, I. (2015). *Small area estimation* (second edition). (Wiley Series in Survey Methodology). Hoboken, NJ: Wiley.