*Chilean Journal of Statistics*
Volume 15, Number 1 (June 2024), Pages 1–26
DOI: 10.32372/chjs.15-01-01

Sampling and Statistical Modeling
Research Article

# Small area estimation using multiple imputation in three-parameter logistic models

Cristian Téllez-Piñérez[1,*], Leonardo Trujillo[2], Juan Sosa[2], and Andrés Gutiérrez-Rojas[3]

[1]Departamento de Estadística, Fundación Universitaria Los Libertadores, Bogotá, Colombia
[2]Departamento de Estadística, Facultad de Ciencias,
Universidad Nacional de Colombia, Sede Bogotá, Bogotá, Colombia
[3]División de Estadística, ECLAC, Santiago, Chile

(Received: 13 March 2024 · Accepted in final form: 20 May 2024)

**Abstract**

This article presents a new method that combines item response theory techniques with small-area estimation approaches to handle missing data. We propose an unbiased estimator for the average skill parameter of three-parameter logistic models using plausible values as the imputation method for missing data. We conduct a thorough simulation study to compare our estimator with the Horvitz-Thompson estimator in complex sampling. Synthetic data experiments demonstrate that our proposal has lower standard errors than its competitor. Additionally, we apply our method to the results in mathematics of the 2015 Program for International Student Assessment and compare our findings with previous studies. These findings indicate that our method is a competitive alternative for generating accurate official statistics.

**Keywords:** Education assessment · Item response theory · Missing data · PISA tests · Statistical modeling · Survey sampling.

**Mathematics Subject Classification:** Primary 62D05 · Secondary 97D60.

## 1. Introduction

Educational assessment can be understood as the process of using collected information about attitudes, beliefs, knowledge, and skills to improve learning in academic programs (Allen, 2004). These data are typically obtained from standardized tests applied to students for assessing planned learning goals (Kuh et al., 2014; ICFES, 2015; OECD, 2016; UNESCO, 2019). However, standardized tests for educational assessment are currently facing a serious issue: A decreasing number of participants per application due to lack of access and financial restrictions (Sevilla et al., 2021).

Then, the problem of missing data is common in response strings both for traditional surveys (San Martín and Alarcón-Bustamante, 2022) and also for standardized tests for educational assessment (Finch, 2008; Rose et al., 2014). Internal expert discussions reveal that if a student answered all the questions designed to measure a particular competency, it would take more than fifteen hours to complete all the questions on the entire exam.

*Corresponding author. Email: cftellezp@libertadores.edu.co (C. Téllez-Piñérez)

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

Exposing each student to a fifteen-hour test is not only a counterproductive strategy for the final scores but also promotes cognitive conflicts from a pedagogical point of view (Álvarez et al., 2007), not to mention anxiety, health, and stress issues. This is also the case with well-known tests such as the Program for International Student Assessment (PISA) (OECD, 2014), Trends in the International Study of Mathematics and Sciences (TIMS; Mullis et al., 2015), and Saber 3, 5, and 9 tests (ICFES, 2015), among many others around the globe. Based on the above, balanced incomplete blocks are implemented to administer a standardized test without forcing students to excessive time constraints. These blocks involve distributing all questions assessing a specific competency to different groups of students within the same school. This approach ensures that all questions are evaluated without requiring students to answer each individually. However, the school does assess all questions related to the competencies (van der Linden and Veldkamp, 2004).

Estimating and analyzing standardized tests mostly employ item response theory (IRT) methods. IRT models intend to explain the relationship between latent traits, that is, unobservable characteristics or attributes and their manifestations, observed outcomes, responses, or performance (Lord, 1980; Martínez, 1995; Muñiz, 1997; van der Linden and Hambleton, 1997; Reckase, 2009; Fox, 2010; Embretson and Reise, 2013; Hambleton and Swaminathan, 2013; Ariza-Hernández and Gutiérrez-Peña, 2021). In particular, three-parameter logistic (3PL) models (Paek and Cole, 2019) are very popular for such an end. According to Sulis and Porcu (2017), these models characterize the probability of responses to any particular question (item) as a function of a location parameter (basal position along the latent trait), a discrimination parameter (item capability to discriminate individuals with different latent trait values), and then, an ability parameter (intensity of the latent trait).

This article aims to estimate the average ability parameter in a three-parameter logistic model in the presence of missing data. To the best of our knowledge, there is a substantial lack of research about non-response methods in the context of standardized tests (Adams and Darwin, 1982; Baker and Kim, 2004; Sulis and Porcu, 2017). The problem of missing data in IRT models was initially addressed by Mislevy (1991), Mislevy et al. (1992), and Khorramdel et al. (2020) using plausible values, and this methodology is currently employed for obtaining aggregated statistics in various international assessments such as TIMSS (Foy et al., 2008) and PISA (OECD, 2016). Since the technique used by standardized tests to generate aggregated results consists of generating $K$ plausible values, based on Rubin (1987) and relying on the 3PL model, this falls short when generating results in unobserved domains. Therefore, our foundational aim relies on extending IRT methodologies accounting for missing data while achieving similar quality indicators at lower costs. To explain the problem addressed in this article, suppose a standardized test has been administered to a probabilistic sample of students (subjects), where each selected student belongs to a particular school or educational institution (domains) and only answers a subset of all items. Suppose further that the goal is to report the student's ability for all domains, yet not all domains are observed. The usual way to estimate an average of such an ability (latent variable) is using IRT models, particularly the 3PL model. However, because students respond to only a subset of items, measuring their abilities is subject to considerable measurement error. In this article, we propose to use small area estimation (SAE) techniques (Avila et al., 2021; Rodríguez et al., 2021), particularly the Fay-Herriot model, to address this issue.

A small area refers to any domain or subpopulation for which direct estimators do not have satisfactory accuracy due to insufficient (or null) sample size. According to Pfeffermann (2002), Rao and Molina (2015) and Morales et al. (2022), to obtain precise and reliable estimates in SAE, it is necessary to use alternative estimators that not only rely on the sampling design but also incorporate indirect estimators or predictors borrowing information from other domains, periods or auxiliary information.

SAE methods gained popularity through the works of Fay and Herriot (1979), Battese et al. (1988), and Ghosh and Rao (1994). Other works in this field have also significantly contributed to the development of SAE, such as Chambers and Tzavidis (2006); Falorsi and Righi (2008); Chambers et al. (2009); Pfeffermann and Correa (2012). Additionally, Pfeffermann (2013) conducted a comprehensive literature review on SAE, including the works of Pfeffermann (2002), Rao (2005),Jiang and Lahiri (2006a), Jiang and Lahiri (2006b), Rao (2008), Datta (2009), and Lehtonen and Veijanen (2009); among others.

Missing data may be encountered in these areas, complicating the use of SAE. One of the earliest approaches to this problem was addressed by Longford (2004), who employed the work by Rubin and Schenker (1991) for multiple imputation using plausible values, obtaining estimates in the small areas of interest. One of the most relevant conclusions of this work is that the inferences have good properties concerning sampling mechanisms and non-response. Another recent article on SAE with missing values is Burgard et al. (2019). Thus, the proposal in this article is to combine the traditional estimation in item response theory methods when data is subject to missing data with SAE techniques. This article was motivated given the importance of this methodology in the educational sector and in generating inputs for educational public policy

In contrast with standard methods, our proposal incorporates estimates of the ability parameter using the approach presented in Fay and Herriot (1979) within the framework of IRT modeling, where multiple imputation tasks are needed. Such a methodology has in itself two significant contributions. It has profound practical implications to deal appropriately with problematic data structures involving missing data, and on the other, it resolves to challenge theoretical issues associated with producing reliable official statistics. We highlight that this is a complex task demanding auxiliary variables correlated with the ability of the students, such as parental socio-economic level, parental education, and school infrastructure, among others, according to Treviño et al. (2010), as well as sophisticated statistical tools for computing the resulting estimator together with its mean square error (MSE).

This rest of this article is structured as follows. Section 2 presents the theoretical development of our estimator along with its MSE under some restrictions. In Section 3, we conduct an exhaustive simulation study in which the proposed estimator is compared with estimators existing in the literature. Also, in this section, we illustrate our proposal with an application with real data to estimate the average ability parameter in the 2015 PISA test. In Section 4, our findings and some relevant aspects for future research are discussed.

## 2. Theoretical development of the estimator and background

### 2.1. Notations

A boldfaced version of a variable denotes a vector with entries consisting of the subscripted variables. For example, $\boldsymbol{x} = (x_1, \ldots, x_n)$ denotes an $n \times 1$ column vector with entries $x_1, \ldots, x_n$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the column vector with all entries equal to 0 and 1, respectively, and $\boldsymbol{I}$ to denote the identity matrix. A subindex in this context refers to the corresponding dimension; for instance, $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. The transpose of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^\top$; analogously for matrices. Moreover, if $\boldsymbol{X}$ is a square matrix, we use $\mathrm{tr}(\boldsymbol{X})$ to denote its trace and $\boldsymbol{X}^{-1}$ to denote its inverse. The norm of $\boldsymbol{x}$, given by $\sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$, is denoted by $\|\boldsymbol{x}\|$.

### 2.2. Three-parameter logistics model

The 3PL model incorporates parameters of difficulty, discrimination, and guessing to model the probability of responding correctly to the item. This model assumes that the characteristic function is given by

$$P_i\left(\xi_{ij}=1 \mid \theta_j,\, a_i,\, b_i,\, c_i\right)=c_i+\frac{(1-c_i)}{1+\exp\left(-a_i\left(\theta_j-b_i\right)\right)},$$

where $b_i$ is the item difficulty, $a_i$ is the discrimination, and $c_i$ is the pseudo-chance parameter, for $i=1,\ldots,l$. $\xi_{ij}$ is a dichotomous variable that is one if the response is correct and zero if it is not. Note that $\theta_j$ is the ability or latent trait of individual $j$ and $P_i(\xi_{ij}=1|\cdot)$ is the probability that individual $j$ responds correctly to item $i$ given an ability $\theta_j$. Observe that $\theta_j$ and $b_i$ are on the same scale and take values in the interval $(-\infty,\, \infty)$. Note that, in $b_i$, the point of inflection in the item characteristic curve, defined as a mathematical function relating the probability of correctly answering an item to the measured level of the latent trait (Paek and Cole, 2019) is present. In addition, $a_i$ is proportional to the slope of the item characteristic curve at the point $b_i, c_i$ indicates the probability that an individual responds correctly to item $i$ when their attribute value tends to be minus infinity. It is worth noting that the two-parameter model is obtained when the pseudo-chance parameter $c_i=0$ in the 3PL model, and the one-parameter model is obtained when the discrimination parameter $a_i$ is a constant in addition to the guessing parameter.

### 2.3. Ability estimation using plausible values and 3PL models

We have a finite universe $U=\bigcup_{d=1}^{D} U_d$ of size $N=\sum_{d=1}^{D} N_d$ distributed in $d$ domains, where $U_d$ is the population in domain $d$ of size $N_d$. Also, let $s=\bigcup_{d=1}^{D} s_d$ be the sample of size $n=\sum_{d=1}^{D} n_d$ under consideration, where $s_d$ is the sample in domain $d$ of size $n_d$ drawn under a particular sampling design $p$. For some domain $d$, $s_d$ may be empty, or $n_d$ is not large enough to enable reliable estimations. Moreover, indicator variables $\xi_{dij}$ registering whether individual $j$, in the domain $d$ answers item $i$ correctly, $\xi_{dij}=1$, or not, $\xi_{dij}=0$, are observed, for $j=1,\ldots,n_d$, $d=1,\ldots,D$, and $i=1,\ldots,I$, where $I$ represents the total number of items assessed.

To explain the problem addressed in this article, suppose a standardized test has been administered to a probabilistic sample of students, where each selected student belongs to a domain $d$ (schools or educational institutions), for $d=1,\ldots,D$, and such student only answers a subset of all items. Suppose further that the goal is to report the statistic $\gamma_d=\mathrm{E}(\theta)$, where $\theta$ is the student's ability for all domains, yet not all domains are observed. The usual way to estimate a student ability, given it is a latent variable, relies in using IRT models, particularly the 3PL model.

IRT modeling assumes that individuals are endowed with the ability to answer items correctly. Let $\boldsymbol{\theta}$ be a vector containing the ability parameters of all the individuals under consideration. The length of this vector corresponds to the number of individuals or students in the sample, that is, the sample is of size $n$. The vector $\boldsymbol{\theta}$ varies for each student since it represents the latent capacity or aptitude that each student possesses to perform a specific task, solve a problem, or understand a particular concept. To find the distribution of $\boldsymbol{\theta}$, there must be available known individual-level auxiliary information, typically stored in a vector of variables $\boldsymbol{x}_j^{\mathrm{I}}$ (the superscript I emphasizes the notion of auxiliary information at the level of individuals or subjects), for each $j=1,\ldots,n_d$. Thus, the probability distribution of $\boldsymbol{\theta}$ in the population is not only conditional on the observed indicator variables $\boldsymbol{\xi}_{\mathrm{obs}}$, but also on the auxiliary information $\boldsymbol{X}_{\mathrm{I}}$, that is,

$$\mathrm{P}\left(\boldsymbol{\theta} \mid \boldsymbol{\xi}_{\mathrm{obs}},\, \boldsymbol{X}_{\mathrm{I}}\right) \propto \mathrm{P}\left(\boldsymbol{\xi}_{\mathrm{obs}} \mid \boldsymbol{\theta}, \boldsymbol{X}_{\mathrm{I}}\right) \mathrm{P}\left(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathrm{I}}\right), \tag{2.1}$$

where $\boldsymbol{X}_{\mathrm{I}}$ is a matrix storing all the individual-level auxiliary information.

Assuming conditional independence between $\boldsymbol{\xi}_{\text{obs}}$ and $\boldsymbol{X}_{\text{I}}$ in Equation (2.1), which is reasonable in practice as in Rubin and Schenker (1991, Section 6.1), we get

$$\mathrm{P}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_{\text{obs}}, \boldsymbol{X}_{\text{I}}) \propto \mathrm{P}\left(\boldsymbol{\xi}_{\text{obs}} \mid \boldsymbol{\theta}\right) \mathrm{P}\left(\boldsymbol{\theta} \mid \boldsymbol{X}_{\text{I}}\right).$$

Under this setting, the main goal relies on finding the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\xi}_{\text{obs}}$ and $\boldsymbol{X}_{\text{I}}$, which in turn depends on two conditional distributions. Firstly, $\mathrm{P}(\boldsymbol{\xi}_{\text{obs}} \mid \boldsymbol{\theta})$, the response chain distribution of students, given their ability, is considered here as a standard 3PL model (Bock and Aitkin, 1981). Secondly, $\mathrm{P}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\text{I}})$, the ability distribution of students given the auxiliary information, regarded here as a multivariate normal distribution with mean $\boldsymbol{X}_{\text{I}}\boldsymbol{\Gamma}$ and covariance matrix $\boldsymbol{\Sigma}$, where both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ need to be estimated. On this point, we are implicitly stating that the parameter space of each component of $\boldsymbol{\theta}$ is the real line, even though most of the ability mass lies between $-3$ and 3 (Andrade et al., 2000).

It is straightforward to see that the conditional distribution $\mathrm{P}(\boldsymbol{\theta} \mid \boldsymbol{\xi}_{\text{obs}}, \boldsymbol{X}_{\text{I}})$ is completely determined by handling the unknown parameters in $\mathrm{p}(\boldsymbol{\xi}_{\text{obs}} \mid \boldsymbol{\theta})$ as well as those in $\mathrm{p}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\text{I}})$. Existing IRT methods typically deal with this setting by first using the expectation maximization (EM) algorithm (Bock and Aitkin, 1981) to estimate the unknown parameters in $\mathrm{P}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\text{I}})$, and then, using the Metropolis-Hastings (MH) algorithm (Fox, 2010) to draw $L$ plausible values (abilities estimates) for each individual $j$ in domain $d$. In this spirit, let $\theta^{\text{pv}}_{dj\ell}$ be the $\ell$-th MH plausible value (hence the upper-index pv) of individual $j$ in domain $d$, for $\ell = 1, \ldots, L$, with $j = 1, \ldots, J$ and $d = 1, \ldots, D$. Also, let $\gamma_d$ the average of the plausible values $\theta^{\text{pv}}_{dj\ell}$ in the $d$-th domain. Thus, an estimate of $\gamma_d$ and its corresponding variance can be found by noticing that

$$\mathrm{P}(\gamma_d \mid \boldsymbol{\xi}_{\text{obs}}) = \int \mathrm{P}\left(\gamma_d \mid \boldsymbol{\xi}\right) \mathrm{P}\left(\boldsymbol{\xi}_{\text{unobs}} \mid \boldsymbol{\xi}_{\text{obs}}\right) \mathrm{d}\boldsymbol{\xi}_{\text{unobs}}, \tag{2.2}$$

where $\boldsymbol{\xi}_{\text{unobs}}$ is composed of all those unobserved indicators variables not given in $\boldsymbol{\xi}_{\text{obs}}$, and the integral is carried out over the space parameter of $\boldsymbol{\xi}_{\text{unobs}}$. As a consequence of Equation (2.2), for the average of the individuals abilities $\gamma_d$, that means the mean of the $\theta^{\text{pv}}_{dj\ell}$, it follows that

$$\begin{aligned}
\mathrm{E}\left(\gamma_d \mid \boldsymbol{\xi}_{\text{obs}}\right) &= \mathrm{E}\left[\mathrm{E}\left(\gamma_d \mid \boldsymbol{\xi}_{\text{unobs}}, \boldsymbol{\xi}_{\text{obs}}\right) \mid \boldsymbol{\xi}_{\text{obs}}\right] \simeq \frac{1}{L}\sum_{\ell=1}^{L} \hat{\theta}^{\text{pv}}_{d\ell} = \hat{\gamma}_d \\
\mathrm{Var}\left(\gamma_d \mid \boldsymbol{\xi}_{\text{obs}}\right) &= \mathrm{E}\left[\mathrm{Var}\left(\gamma_d \mid \boldsymbol{\xi}_{\text{unobs}}, \boldsymbol{\xi}_{\text{obs}}\right) \mid \boldsymbol{\xi}_{\text{obs}}\right] + \mathrm{Var}\left[\mathrm{E}\left(\gamma_d \mid \boldsymbol{\xi}_{\text{unobs}}, \boldsymbol{\xi}_{\text{obs}}\right) \mid \boldsymbol{\xi}_{\text{obs}}\right] \\
&\simeq \frac{1}{L}\sum_{\ell=1}^{L} \mathrm{Var}(\hat{\theta}^{\text{pv}}_{d\ell}) + \left(1 + \frac{1}{L}\right)\frac{1}{L-1}\sum_{\ell=1}^{L}(\hat{\theta}^{\text{pv}}_{d\ell} - \hat{\gamma}_d)^2,
\end{aligned} \tag{2.3}$$

where $\hat{\theta}^{\text{pv}}_{d\ell}$ is the $\ell$-th plausible value in domain $d$, and $\mathrm{Var}(\hat{\theta}^{\text{pv}}_{d\ell})$ depends on the specific formulation of the sampling design and corresponds to the average of the estimated variances for each plausible value.

## 2.4. Proposed predictor

Here, we take a step further and go beyond the existing IRT literature as presented above to carry out our main task: combining IRT methods with SAE strategies to account for missing data. Thus, inspired by SAE methods, we propose estimating $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_D)$, by using a predictor based on an area-level linear mixed model fitted to $\hat{\boldsymbol{\gamma}}$ stated as

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{X}_{\text{A}}\boldsymbol{\beta} + Zu + e, \tag{2.4}$$

where $\boldsymbol{X}_{\mathrm{A}}$ is a fixed-effects matrix storing all the area-level auxiliary information (the sub-index A emphasizes the notion of auxiliary information at the level of areas or domains), $\boldsymbol{\beta}$ is a vector of unknown constants, $\boldsymbol{Z}$ is a random-effects design matrix, and then, $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent random vectors such that $\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0},\, \boldsymbol{V}_u)$ and $\boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0},\, \boldsymbol{V}_e)$, with known $\boldsymbol{V}_u$ and $\boldsymbol{V}_e$. Thus, under the previous specification it follows directly that $\mathrm{Var}\,(\hat{\boldsymbol{\gamma}}) = \boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}^\top + \boldsymbol{V}_e = \boldsymbol{V}$, and also, following standard results about linear mixed-effects models (Morales et al., 2022), it can be shown that the optimal unbiased linear predictor of $\boldsymbol{\tau} = \boldsymbol{L}\boldsymbol{\beta} + \boldsymbol{M}\boldsymbol{u}$ is $\hat{\boldsymbol{\tau}} = \boldsymbol{L}\hat{\boldsymbol{\beta}} + \boldsymbol{M}\hat{\boldsymbol{u}}$, with associated estimators of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$ given by $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\right)^{-1}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{u}} = (\boldsymbol{Z}\boldsymbol{V}_u)^\top\boldsymbol{V}^{-1}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{X}_{\mathrm{A}}\hat{\boldsymbol{\beta}}\right)$.

The mathematical construction of the best linear unbiased predictor of $\boldsymbol{\tau}$ is stated next. Building on ideas given in Harville (1977) and Morales et al. (2022), we consider an unbiased linear predictor of the form $\hat{\boldsymbol{\tau}} = \boldsymbol{Q}_0 + \boldsymbol{Q}_1\hat{\boldsymbol{\gamma}}$, where $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ are two conformable matrices. Since $\hat{\boldsymbol{\tau}}$ is unbiased, we have that $\mathrm{E}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) = \boldsymbol{0}$, but $\mathrm{E}\,(\hat{\boldsymbol{\tau}}) = \boldsymbol{Q}_0\mathrm{E}\,(\hat{\boldsymbol{\gamma}}) + \boldsymbol{Q}_1 = \boldsymbol{Q}_0\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta} + \boldsymbol{Q}_1$ and $\mathrm{E}\,(\boldsymbol{\tau}) = \boldsymbol{L}\boldsymbol{\beta}$. Then, it follows that

$$\boldsymbol{0} = \mathrm{E}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) = \boldsymbol{Q}_0\mathrm{E}\,(\hat{\boldsymbol{\gamma}}) + \boldsymbol{Q}_1 - \boldsymbol{L}\boldsymbol{\beta} = (\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} - \boldsymbol{L})\,\boldsymbol{\beta} + \boldsymbol{Q}_0,$$

which requires that both $\boldsymbol{Q}_0 = \boldsymbol{0}$ and $\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} = \boldsymbol{L}$ hold. Thus, the best predictor $\hat{\tau}$ can be found by minimizing $\mathrm{Var}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ subject to $\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} = \boldsymbol{L}$. Thus,

$$\begin{aligned}
\mathrm{Var}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) &= \mathrm{Var}\,(\boldsymbol{Q}_1\hat{\boldsymbol{\gamma}} - \boldsymbol{L}\boldsymbol{\beta} - \boldsymbol{M}\boldsymbol{u}) \\
&= \mathrm{Var}\,(\boldsymbol{Q}_1\hat{\boldsymbol{\gamma}}) + \mathrm{Var}\,(\boldsymbol{M}\boldsymbol{u}) - 2\mathrm{Cov}\,(\boldsymbol{Q}_1\hat{\boldsymbol{\gamma}}, \boldsymbol{M}\boldsymbol{u}) \\
&= \boldsymbol{Q}_1\boldsymbol{V}\boldsymbol{Q}_1^\top + \boldsymbol{M}\boldsymbol{V}_u\boldsymbol{M}^\top - 2\boldsymbol{Q}_1\boldsymbol{C}\boldsymbol{M}^\top,
\end{aligned} \tag{2.5}$$

with $\boldsymbol{C} = \mathrm{Cov}\,(\hat{\boldsymbol{\gamma}},\, \boldsymbol{u})$ in Equation (2.5). Since $\boldsymbol{M}\boldsymbol{V}_u\boldsymbol{M}^\top$ does not depend on $\boldsymbol{Q}_1$, the minimization problem given above can be restated as minimize $\boldsymbol{Q}_1\boldsymbol{V}\boldsymbol{Q}_1^\top - 2\boldsymbol{Q}_1\boldsymbol{C}\boldsymbol{M}^\top$ subject to $\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} = \boldsymbol{L}$, whose corresponding Lagrangian function is presented as

$$\ell\,(\boldsymbol{Q}_1,\, \boldsymbol{\Lambda}) = \boldsymbol{Q}_1\boldsymbol{V}\boldsymbol{Q}_1^\top - 2\boldsymbol{Q}_1\boldsymbol{C}\boldsymbol{M}^\top + 2\,(\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} - \boldsymbol{L})\,\boldsymbol{\Lambda}.$$

Taking partial derivatives with respect to both $\boldsymbol{Q}_1$ and $\boldsymbol{\Lambda}$ and equalizing to zero, we get the Equations stated as

$$\begin{aligned}
\frac{\partial\ell\,(\boldsymbol{Q}_1,\, \boldsymbol{\Lambda})}{\partial\boldsymbol{Q}_1} &= 2\boldsymbol{V}\boldsymbol{Q}_1^\top - 2\boldsymbol{C}\boldsymbol{M}^\top + 2\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\Lambda} = \boldsymbol{0}, \quad \boldsymbol{V}\boldsymbol{Q}_1^\top + \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\Lambda} = \boldsymbol{C}\boldsymbol{M}^\top, \\
\frac{\partial\ell\,(\boldsymbol{Q}_1, \boldsymbol{\Lambda})}{\partial\boldsymbol{\Lambda}} &= 2\,(\boldsymbol{Q}_1\boldsymbol{X}_{\mathrm{A}} - \boldsymbol{L}) = \boldsymbol{0}, \quad \boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{Q}_1^\top = \boldsymbol{L}^\top,
\end{aligned}$$

that, in turn, can be rewritten in matrix form as

$$\begin{pmatrix} \boldsymbol{V} & \boldsymbol{X}_{\mathrm{A}} \\ \boldsymbol{X}_{\mathrm{A}}^\top & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{Q}_1^\top \\ \boldsymbol{\Lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}\boldsymbol{M}^\top \\ \boldsymbol{L}^\top \end{pmatrix} \Rightarrow \begin{pmatrix} \boldsymbol{Q}_1^\top \\ \boldsymbol{\Lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V} & \boldsymbol{X}_{\mathrm{A}} \\ \boldsymbol{X}_{\mathrm{A}}^\top & \boldsymbol{0} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{C}\boldsymbol{M}^\top \\ \boldsymbol{L}^\top \end{pmatrix}.$$

Thus, using results from matrix algebra and letting $\boldsymbol{G} = \left(\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}\boldsymbol{X}_{\mathrm{A}}\right)^{-1}$, we obtain

$$\begin{pmatrix} \boldsymbol{Q}_1^\top \\ \boldsymbol{\Lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1} & \boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{G} \\ \boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^{a}\top\boldsymbol{V} & -\boldsymbol{G} \end{pmatrix}\begin{pmatrix} \boldsymbol{C}\boldsymbol{M}^\top \\ \boldsymbol{L}^\top \end{pmatrix},$$

obtaining an expression for $\boldsymbol{Q}_1$ as

$$\begin{aligned}\boldsymbol{Q}_1 &= \left[\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{L}^\top + \boldsymbol{V}^{-1}\left(\boldsymbol{I} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\right)\boldsymbol{C}\boldsymbol{M}^\top\right]^\top \\ &= \boldsymbol{L}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1} + \boldsymbol{M}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\left(\boldsymbol{I} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\right),\end{aligned}$$

where $\boldsymbol{I}$ is the identity matrix.

Recall from our earlier discussion that the linear predictor takes the form $\hat{\boldsymbol{\tau}} = \boldsymbol{Q}_1\hat{\boldsymbol{\gamma}}$ because it is assumed to be unbiased from the beginning. Thus, substituting $\boldsymbol{Q}_1$, we get that the best linear unbiased predictor of $\boldsymbol{\tau}$ is given by the Equations stated as

$$\begin{aligned}\hat{\boldsymbol{\tau}} &= \left[\boldsymbol{L}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1} + \boldsymbol{M}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\left(\boldsymbol{I} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\right)\right]\hat{\boldsymbol{\gamma}} \\ &= \boldsymbol{L}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}} + \boldsymbol{M}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}} - \boldsymbol{M}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}} \\ &= \boldsymbol{L}\boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}} + \boldsymbol{M}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{X}_{\mathrm{A}}\hat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{L}\hat{\boldsymbol{\beta}} + \boldsymbol{M}\hat{\boldsymbol{u}},\end{aligned}$$

where $\hat{\boldsymbol{\beta}} = \boldsymbol{G}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{u}} = \boldsymbol{C}^\top\boldsymbol{V}^{-1}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{X}_{\mathrm{A}}\hat{\boldsymbol{\beta}}\right)$. Now, our goal is to see through the previous expression to adapt the Fay-Herriot model stated as

$$\hat{\gamma}_d = \boldsymbol{x}_{\mathrm{A}\,d}^\top\boldsymbol{\beta} + u_d + e_d, \qquad d = 1, \ldots, D, \tag{2.6}$$

where $\boldsymbol{x}_{\mathrm{A}\,d}^\top\boldsymbol{\beta} = \sum_{k=1}^{p} x_{\mathrm{A}\,dk}\,\beta_k$ is a linear predictor of fixed effects, $u_d$ is a domain-specific random effect, and $p$ is the number of parameters in the model. All the random effects are assumed to be independent and identically distributed with zero mean and variance $\sigma_u^2$. Furthermore, $e_d$ is the sampling error associated with the sampling design $p$, which is also assumed to be independent of $u_d$, and also, such that $\mathrm{E}_p[e_d \mid \hat{\gamma}_d] = 0$ and $\mathrm{Var}_p[e_d \mid \hat{\gamma}_d] = \sigma_d^2$. As a final remark, each $\sigma_d^2$ can be calculated in a straightforward fashion using Equation (2.3). Using the results provided in this section and rewriting the model in Equation (2.6) in the matrix form of Equation (2.4), it follows that $\hat{\boldsymbol{\gamma}} = \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$ is equivalent to the expression given by

$$\begin{pmatrix}\hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_D\end{pmatrix} = \begin{pmatrix} x_{\mathrm{A}\,11} & \cdots & x_{\mathrm{A}\,1p} \\ \vdots & \ddots & \vdots \\ x_{\mathrm{A}\,D1} & \cdots & x_{\mathrm{A}\,Dp}\end{pmatrix}\begin{pmatrix}\beta_1 \\ \vdots \\ \beta_p\end{pmatrix} + \begin{pmatrix}u_1 \\ \vdots \\ u_D\end{pmatrix} + \begin{pmatrix}e_1 \\ \vdots \\ e_d\end{pmatrix},$$

which means that $\boldsymbol{Z} = \boldsymbol{I}_D$ and $\boldsymbol{V} = \mathrm{diag}\left(\sigma_u^2 + \sigma_1^2, \ldots, \sigma_u^2 + \sigma_D^2\right)$. As a consequence, we have that the best linear unbiased estimator of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ and $\boldsymbol{u} = (u_1, \ldots, u_D)$ are shown respectively in the formulations presented as

$$\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}^{-1}\boldsymbol{X}_{\mathrm{A}}\right)^{-1}\boldsymbol{X}_{\mathrm{A}}^\top\boldsymbol{V}\hat{\boldsymbol{\gamma}}, \quad \tilde{\boldsymbol{u}} = \boldsymbol{C}^\top\boldsymbol{V}^{-1}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{X}_{\mathrm{A}}\tilde{\boldsymbol{\beta}}\right), \tag{2.7}$$

with $\boldsymbol{V}^{-1} = \mathrm{diag}\left(1/\left(\sigma_u^2 + \sigma_1^2\right), \ldots, 1/\left(\sigma_u^2 + \sigma_D^2\right)\right)$ and $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_D)$.

Therefore, the best linear unbiased predictor proposed in this article is stated as

$$\hat{\gamma}_d^{\mathrm{B}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_d^2}\,\hat{\gamma}_d + \frac{\sigma_d^2}{\sigma_u^2 + \sigma_d^2}\,\boldsymbol{x}_{\mathrm{A}\,d}^\top\tilde{\boldsymbol{\beta}}. \tag{2.8}$$

Also, the empirical best linear unbiased predictor of the mean of the small area $\gamma_d$ under the model in Equation (2.6) can be obtained just by replacing $\sigma_u^2$ by its corresponding estimate $\hat{\sigma}_u^2$ in Equation (2.8) as

$$\hat{\gamma}_d^{\mathrm{P}} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_d^2}\,\hat{\gamma}_d + \frac{\sigma_d^2}{\hat{\sigma}_u^2 + \sigma_d^2}\,\boldsymbol{x}_{\mathrm{A}\,d}^\top \tilde{\boldsymbol{\beta}} = (1 - B_d)\,\hat{\gamma}_d + B_d\,\boldsymbol{x}_{\mathrm{A}\,d}^\top \tilde{\boldsymbol{\beta}}, \tag{2.9}$$

with $B_d = \sigma_d^2/(\hat{\sigma}_u^2 + \sigma_d^2)$. The proposed predictor $\hat{\gamma}_d^{\mathrm{P}}$ in Equation (2.9) is unbiased for $\gamma_d$, for $d = 1, \ldots, D$. Indeed, consider the difference stated in

$$\hat{\gamma}_d^{\mathrm{P}} - \gamma_d = (1 - B_d)\,\hat{\gamma}_d + B_d\,\boldsymbol{x}_{\mathrm{A}\,d}^\top \tilde{\boldsymbol{\beta}} - \left[B_d\,\gamma_d + (1 - B_d)\,\gamma_d\right].$$

Now, by letting $\alpha_d = 1 - B_d$, $\gamma_d = \boldsymbol{x}_{\mathrm{A}\,d}^\top \boldsymbol{\beta} + u_d$, and $\hat{\gamma}_d = \gamma_d + e_d$, we have

$$\begin{aligned}
\hat{\gamma}_d^{\mathrm{P}} - \gamma_d &= \alpha_d\,(\gamma_d + e_d) + (1 - \alpha_d)\,\boldsymbol{x}_{\mathrm{A}\,d}^\top \tilde{\boldsymbol{\beta}} - \alpha_d \gamma_d - (1 - \alpha_d)\,\gamma_d \\
&= \alpha_d e_d + (1 - \alpha_d)\,\boldsymbol{x}_{\mathrm{A}\,d}^\top \tilde{\boldsymbol{\beta}} - (\alpha_d)\left[\boldsymbol{x}_{\mathrm{A}\,d}^\top \boldsymbol{\beta} + u_d\right] \\
&= \alpha_d e_d + (1 - \alpha_d)\left[\boldsymbol{x}_{\mathrm{A}\,d}^\top (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] - (1 - \alpha_d)\,u_d.
\end{aligned}$$

Then, taking expected values presented as

$$\mathrm{E}\left(\hat{\gamma}_d^{\mathrm{P}} - \gamma_d\right) = \mathrm{E}\left(\alpha_d e_d + (1 - \alpha_d)\left[\boldsymbol{x}_{\mathrm{A}\,d}^\top (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] - (1 - \alpha_d)\,u_d\right) = 0. \tag{2.10}$$

The results in Equation (2.10) follow because $\mathrm{E}\,(e_d) = \mathrm{E}\,(u_d) = 0$, and also, $\mathrm{E}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{0}$ since $\tilde{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.

### 2.5. Mean square error of the proposed predictor

Next, the estimation of the MSE for the proposed predictor is presented. We follow very Kackar and Harville (1984), Prasad and Rao (1990), and Ghosh and Rao (1994) to obtain an expression for the MSE of the proposed predictor $\hat{\gamma}_d^{\mathrm{P}}$. Specifically, we consider the variance component estimation method, which requires us to find three quantities explicitly, namely, $g_{1d}(\sigma_u^2)$, $g_{2d}(\sigma_u^2)$, and $g_{3d}(\sigma_u^2)$, for $d = 1, \ldots, D$. We refer the reader to the previous references for details about such a method. However, we outline some fundamental details below.

First, to calculate $g_{1d}\,(\sigma_u^2)$, we need to take into account that $\boldsymbol{V}_u = \sigma_u^2 \boldsymbol{I}_D$ and $\boldsymbol{V}_e = \sigma_d^2 \boldsymbol{W}_N^{-1} \boldsymbol{V}_s^{-1} = \mathrm{diag}\,(\boldsymbol{V}_{s_1}^{-1}, \ldots, \boldsymbol{V}_{s_D}^{-1})$, where $\boldsymbol{W}_N$ is an $N \times N$ diagonal matrix of weights induced by the sampling design $p$, with $N$ the population size, $s = \cup_{d=1}^{D} s_d$ is the full sample under consideration, with $s_d$ the sample in domain $d$, and

$$\boldsymbol{V}_{s_d}^{-1} = \frac{1}{\sigma_d^2}\left(\boldsymbol{W}_{s_d} - \frac{B_d}{w_d}\,\boldsymbol{w}_{n_d} \boldsymbol{w}_{n_d}^\top\right), \tag{2.11}$$

where $w_d$ is the weight of domain $d$, for $d = 1, \ldots, D$ in Equation (2.11).

Thus, it follows that

$$
\begin{aligned}
\boldsymbol{V}_u \boldsymbol{Z}_s^\top \boldsymbol{V}_{e\,s}^{-1} \boldsymbol{Z}_s \boldsymbol{V}_u &= \frac{\sigma_u^4}{\sigma_d^2} \operatorname{diag}\left(\mathbf{1}_{n_1}^\top, \ldots, \mathbf{1}_{n_D}^\top\right) \operatorname{diag}\left(\boldsymbol{W}_{n_1} - \frac{B_1}{n_1}\boldsymbol{w}_{n_1}\mathbf{1}_1^\top, \ldots, \boldsymbol{W}_{n_D} - \frac{B_D}{n_D}\boldsymbol{w}_{n_D}\mathbf{1}_D^\top\right) \\
&\quad \times \operatorname{diag}\left(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_D}\right) \\
&= \sigma_u^2 \operatorname{diag}\left(\frac{B_1}{w_1}\boldsymbol{w}_{n_1}^\top, \ldots, \frac{B_D}{w_D}\boldsymbol{w}_{n_D}^\top\right) \operatorname{diag}\left(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_D}\right) \\
&= \sigma_u^2 \operatorname{diag}\left(B_1, \ldots, B_D\right),
\end{aligned}
$$

and therefore we have that

$$
\boldsymbol{T}_s = \boldsymbol{V}_u(\boldsymbol{I}_D - \boldsymbol{Z}_s^\top \boldsymbol{V}_{e\,s}^{-1} \boldsymbol{Z}_s \boldsymbol{V}_u) = \sigma_u^2 \operatorname{diag}\left(1 - B_1, \ldots, 1 - B_D\right), \tag{2.12}
$$

with $\boldsymbol{Z}_s = \operatorname{diag}\left(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_D}\right)$ in Equation (2.12) and

$$
\boldsymbol{V}_{e\,s} = \operatorname{diag}\left(\boldsymbol{W}_{n_1} - \frac{B_1}{n_1}\boldsymbol{w}_{n_1}\mathbf{1}_{w_1}^\top, \ldots, \boldsymbol{W}_{n_D} - \frac{B_D}{n_D}\boldsymbol{w}_{n_D}\mathbf{1}_D^\top\right).
$$

The previous result is useful because we want to estimate the average of the plausible values for all the individuals within the domain $d$, $\eta = \boldsymbol{a}^\top \boldsymbol{\gamma}_{\mathrm{U}}$, with

$$
\boldsymbol{a}^\top = \tfrac{1}{N_d}\left(\mathbf{0}_{N_1}^\top, \ldots, \mathbf{0}_{N_{d-1}}^\top, \mathbf{1}_{N_d}^\top, \mathbf{0}_{N_{d+1}}^\top, \ldots, \mathbf{0}_{N_D}^\top\right),
$$

and $\boldsymbol{\gamma}_{\mathrm{U}}$ is the population parameter, leading us to consider all those individuals included in the sample and those that did not, denoted by $s$ and $r$, respectively. As a consequence, we get for $g_{1d}\left(\sigma_u^2\right)$ that

$$
\begin{aligned}
g_{1d}\left(\sigma_u^2\right) &= \boldsymbol{a}_r^\top \boldsymbol{Z}_r \boldsymbol{T}_s \boldsymbol{Z}_r^\top \boldsymbol{a}_r \\
&= \left(\mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d-n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top\right) \operatorname{diag}\left(\mathbf{1}_{N_1-n_1}, \ldots, \mathbf{1}_{N_d-n_d}\right) \\
&\quad \times \frac{\sigma_u^2}{N_d^2} \operatorname{diag}\left(1 - B_1, \ldots, 1 - B_D\right) \\
&\quad \times \operatorname{diag}\left(\mathbf{1}_{N_1-n_1}^\top, \ldots, \mathbf{1}_{N_D-n_D}^\top\right)\left(\mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d-n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top\right) \\
&= \frac{\sigma_u^2}{N_d^2}\left(1 - B_d\right)\left(N_d - n_d\right)^2 \\
&\simeq \sigma_u^2\left(1 - B_d\right),
\end{aligned} \tag{2.13}
$$

for $n_d << N_d$.

Now, to compute $g_{2d}\left(\sigma_u^2\right)$, we need to get

$$
\begin{aligned}
\boldsymbol{Z}_r \boldsymbol{T}_s \boldsymbol{Z}_s^\top &= \operatorname{diag}\left(\mathbf{1}_{N_1-n_1}, \ldots, \mathbf{1}_{N_D-n_D}\right) \sigma_u^2 \operatorname{diag}\left(1 - B_1, \ldots, 1 - B_D\right) \operatorname{diag}\left(\mathbf{1}_{n_1}^\top, \ldots, \mathbf{1}_{n_D}^\top\right) \\
&= \sigma_u^2 \operatorname{diag}\left(\left(1 - B_1\right)\, 1_{N_1-n_1}\mathbf{1}_{n_1}^\top, \ldots, \left(1 - B_D\right)\, \mathbf{1}_{N_D-n_D}\mathbf{1}_{n_D}^\top\right),
\end{aligned}
$$

which means that

$$
\begin{aligned}
\boldsymbol{a}_r^\top \boldsymbol{Z}_r \boldsymbol{T}_s \boldsymbol{Z}_s^\top \boldsymbol{V}_{es}^{-1} \boldsymbol{X}_{\mathrm{A}\,s} &= \frac{1}{N_d} \frac{\sigma_u^2}{\sigma_d^2} \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d - n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right) \\
&\quad \times \mathrm{diag}\left( (1 - B_1)\, \mathbf{1}_{N_1 - n_1} \mathbf{1}_{n_1}^\top, \ldots, (1 - B_D)\, \mathbf{1}_{N_D - n_D} \mathbf{1}_{n_D}^\top \right) \boldsymbol{W}_s \boldsymbol{X}_{\mathrm{A}\,s} \\
&= \frac{1}{N_d} \frac{\sigma_u^2}{\sigma_d^2} (1 - B_d)(N_d - n_d) \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \boldsymbol{w}_{n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right) \boldsymbol{X}_{\mathrm{A}\,s} \\
&= (1 - f_d)\, B_d\, \hat{\bar{\boldsymbol{x}}}_d\,, \qquad (2.14)
\end{aligned}
$$

with $f_d = n_d / N_d$ is the sampling fraction and $\hat{\bar{\boldsymbol{x}}}_d = 1/\sum w_d \sum_{k \in s_d} x_{dk} \boldsymbol{w}_{dk}$ in Equation (2.14). Moreover, we have that

$$
\boldsymbol{a}_r^\top \boldsymbol{X}_{\mathrm{A}\,r} = \frac{1}{N_d} \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d - n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right) \boldsymbol{X}_{\mathrm{A}\,r} = (1 - f_d)\, \bar{\boldsymbol{x}}_d, \qquad (2.15)
$$

with $\bar{\boldsymbol{x}}_d = (1/N_d) \sum_{k \in d} x_{dk}$. Using together Equation (2.15), along with $n_d << N_d$, we get the expression given by

$$
g_{2d}\left(\sigma_u^2\right) = \left( \bar{\boldsymbol{x}}_d - B_d\, \hat{\bar{\boldsymbol{x}}}_d \right) \left( \boldsymbol{X}_{\mathrm{A}\,s}^\top \boldsymbol{V}_s^{-1} \boldsymbol{X}_{\mathrm{A}\,s} \right)^{-1} \left( \bar{\boldsymbol{x}}_d - B_d\, \hat{\bar{\boldsymbol{x}}}_d \right)^\top. \qquad (2.16)
$$

This, in order to $g_{3d}\left(\sigma_u^2\right)$, we need to get

$$
\begin{aligned}
\boldsymbol{b}^\top &= \frac{1}{N_d} \frac{\sigma_u^2}{\sigma_d^2} \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d - n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right) \mathrm{diag}\left( \mathbf{1}_{N_1 - n_1}, \ldots, \mathbf{1}_{N_D - n_D} \right) \mathrm{diag}\left( \mathbf{1}_{n_1}^\top, \ldots, \mathbf{1}_{n_D}^\top \right) \\
&\quad \times \mathrm{diag}\left( \boldsymbol{W}_{n_1} - \frac{B_1}{w_1} \boldsymbol{w}_{n_1} \boldsymbol{w}_{n_1}^\top, \ldots, \boldsymbol{W}_{n_D} - \frac{B_D}{w_D} \boldsymbol{w}_{n_D} \boldsymbol{w}_{n_D}^\top \right) \\
&= \frac{1}{N_d} \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \mathbf{1}_{N_d - n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right) \mathrm{diag}\left( \frac{B_1}{w_1} \mathbf{1}_{N_1 - n_1} \boldsymbol{w}_{n_1}^\top, \ldots, \frac{B_D}{w_D} \mathbf{1}_{N_D - n_D} \boldsymbol{w}_{n_D}^\top \right) \\
&= \left( \mathbf{0}^\top, \ldots, \mathbf{0}^\top, \frac{B_d}{w_d} \frac{N_d - n_d}{N_d} \boldsymbol{w}_{n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \right).
\end{aligned}
$$

Then, we obtain

$$
\nabla \boldsymbol{b}^\top = \begin{pmatrix} \mathbf{0}^\top, \ldots, \mathbf{0}^\top, (1 - f_d) \frac{\partial B_d}{\partial \sigma_d^2} \frac{1}{w_d} \boldsymbol{w}_{n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \\ \mathbf{0}^\top, \ldots, \mathbf{0}^\top, (1 - f_d) \frac{\partial B_d}{\partial \sigma_u^2} \frac{1}{w_d} \boldsymbol{w}_{n_d}^\top, \mathbf{0}^\top, \ldots, \mathbf{0}^\top \end{pmatrix}
$$

and therefore we get the expression for $g_{3d}\left(\sigma_u^2\right)$ given by

$$
\begin{aligned}
g_{3d}\left(\sigma_u^2\right) &= (1 - f_d)^2 \left( \sigma_u^2 + \frac{\sigma_d^2}{w_d} \right) \mathrm{tr} \left\{ \begin{pmatrix} \left( \frac{\partial B_d}{\partial \sigma_d^2} \right)^2 & \frac{\partial B_d}{\partial \sigma_d^2} \frac{\partial B_d}{\partial \sigma_u^2} \\ \frac{\partial B_d}{\partial \sigma_d^2} \frac{\partial B_d}{\partial \sigma_u^2} & \left( \frac{\partial B_d}{\partial \sigma_u^2} \right)^2 \end{pmatrix} \begin{pmatrix} \mathrm{Var}\left(\hat{\sigma}_d^2\right) & \mathrm{Cov}\left(\hat{\sigma}_d^2, \hat{\sigma}_u^2\right) \\ \mathrm{Cov}\left(\hat{\sigma}_d^2, \hat{\sigma}_u^2\right) & \mathrm{Var}\left(\hat{\sigma}_u^2\right) \end{pmatrix} \right\} \\
&= \left( \sigma_u^2 + \frac{\sigma_d^2}{w_d} \right)^{-3} \frac{1}{w_d^2} \mathrm{Var}\left( \sigma_u^2 \hat{\sigma}_d^2 - \sigma_d^2 \hat{\sigma}_u^2 \right) \qquad (2.17)
\end{aligned}
$$

for $n_d << N_d$.

The estimation of the MSE for the proposed predictor $\hat{\gamma}_d^{\mathrm{P}}$ can now be obtained by considering our findings given in Equations (2.13), (2.16) and (2.17). Specifically, considering $\sigma_u^2$ instead of $\hat{\sigma}_u^2$, it follows that

$$g_{1d}\left(\hat{\sigma}_u^2\right) = \frac{\hat{\sigma}_u^2 \sigma_d^2}{\hat{\sigma}_u^2 + \sigma_d^2} = \sigma_d^2\,(1 - B_d)\,, \quad g_{2d}\left(\hat{\sigma}_u^2\right) = \left(\frac{\sigma_d^2}{\hat{\sigma}_u^2 + \sigma_d^2}\right)^2 a = B_d^2\, a,$$

where $B_d = \sigma_d^2/(\hat{\sigma}_u^2 + \sigma_d^2)$ and $a = \boldsymbol{x}_{d\,\mathrm{A}} \boldsymbol{F}^{-1} \boldsymbol{x}_{d\,\mathrm{A}}^\top$, with $\boldsymbol{F} = (\hat{\sigma}_u^2 + \sigma_d^2)^{-1} \sum_{d=1}^{D} \boldsymbol{x}_{d\,\mathrm{A}}\, \boldsymbol{x}_{d\,\mathrm{A}}^\top$. The component $g_{3d}\left(\hat{\sigma}_u^2\right)$ depends on the estimation method of the variance components. Either way, $g_{3d}\left(\hat{\sigma}_u^2\right)$ takes the form stated as

$$g_{3d}\left(\hat{\sigma}_u^2\right) = \frac{\sigma_d^4}{(\hat{\sigma}_u^2 + \sigma_d^2)^3}\, \mathrm{Var}\left(\hat{\sigma}_u^2\right),$$

where

$$\mathrm{Var}\left(\hat{\sigma}_u^2\right) = \frac{2}{D}\left[\hat{\sigma}_u^4 + \frac{2\hat{\sigma}_u^2}{D}\sum_{d=1}^{D}\sigma_d^2 + \frac{1}{D}\sum_{d=1}^{D}\sigma_d^4\right] = 2\left[\sum_{d=1}^{D}\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^{-2}\right]^{-1},$$

using the moment estimator proposed by Prasad and Rao (1990) or the expression in the right, either the maximum likelihood (ML) or the restricted maximum likelihood (REML). Thus, $g_{3d}\left(\hat{\sigma}_u^2\right)$ simplifies to

$$g_{3d}\left(\hat{\sigma}_u^2\right) = \left(\frac{1}{\hat{\sigma}_u^2 + \sigma_d^2}\right) \frac{2\,B_d^2}{\sum_{d=1}^{D}\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^2}.$$

Therefore, on the one hand, if either the Prasad-Rao or REML estimator are used, then the estimator for the MSE is given by

$$\widehat{\mathrm{MSE}}(\,\hat{\gamma}_d^{\mathrm{P}}\,) = g_{1\,d}\left(\hat{\sigma}_u^2\right) + g_{2\,d}\left(\hat{\sigma}_u^2\right) + 2g_{3\,d}\left(\hat{\sigma}_u^2\right), \tag{2.18}$$

but on the other hand, if the ML estimator is used, then we get for the estimator of the MSE that

$$\widehat{\mathrm{MSE}}(\,\hat{\gamma}_d^{\mathrm{P}}\,) = g_{1\,d}\left(\hat{\sigma}_u^2\right) + g_{2\,d}\left(\hat{\sigma}_u^2\right) + 2g_{3\,d}\left(\hat{\sigma}_u^2\right) - b\,\nabla g_1,$$

with

$$b = -\frac{1}{\left[\sum_{d=1}^{D}\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^{-2}\right]}\mathrm{tr}\left\{\left(\sum_{d=1}^{D}\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^{-1}\boldsymbol{x}_{d\,\mathrm{A}}\,\boldsymbol{x}_{d\,\mathrm{A}}^\top\right)^{-1}\left(\sum_{d=1}^{D}\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^{-2}\boldsymbol{x}_{d\,\mathrm{A}}\,\boldsymbol{x}_{d\,\mathrm{A}}^\top\right)\right\}$$

and $\nabla g_1 = \sigma_d^4\left(\hat{\sigma}_u^2 + \sigma_d^2\right)^{-2}$.

Taking into account the above results presented and the developments made by Ghosh and Rao (1994) and Prasad and Rao (1990), the estimation of the MSE for $\hat{\gamma}_d^{\mathrm{P}}$ can be written depending on the method of estimation of the variance components.

The expressions for $g_{1\,d}\left(\hat{\sigma}_u^2\right)$ and $g_{2\,d}\left(\hat{\sigma}_u^2\right)$ are presented as

$$
\begin{aligned}
g_{1\,d}\left(\hat{\sigma}_u^2\right) &= \frac{\hat{\sigma}_u^2\sigma_d^2}{\hat{\sigma}_u^2+\sigma_d^2} = \sigma_d^2\,\left(1-B_d\right)\\
g_{2\,d}\left(\hat{\sigma}_u^2\right) &= \left[\frac{\sigma_d^2}{\left(\hat{\sigma}_u^2+\sigma_d^2\right)}\right]^2 \boldsymbol{x}_d^A \left[\frac{\sum_{d=1}^D \boldsymbol{x}_d^A\,\left(\boldsymbol{x}_d^A\right)^\top}{\hat{\sigma}_u^2+\sigma_d^2}\right]^{-1}\left(\boldsymbol{x}_d^A\right)^\top\\
&= B_d^2\,\boldsymbol{x}_d^A\left[\frac{\sum_{d=1}^D \boldsymbol{x}_d^A\,\left(\boldsymbol{x}_d^A\right)^\top}{\hat{\sigma}_u^2+\sigma_d^2}\right]^{-1}\left(\boldsymbol{x}_d^A\right)^\top.
\end{aligned}
$$

It is worth noting that the component $g_{3\,d}\left(\hat{\sigma}_u^2\right)$ will depend on the method of estimation of the variance components used and their possible expressions appear in

$$
g_{3\,d}\left(\hat{\sigma}_u^2\right) = \frac{\sigma_d^4}{\left(\hat{\sigma}_u^2+\sigma_d^2\right)^3}\frac{2}{D}\left[\hat{\sigma}_u^4+\frac{2\hat{\sigma}_u^2}{D}\sum_{d=1}^D\sigma_d^2+\frac{1}{D}\sum_{d=1}^D\sigma_d^4\right],
$$

and rewritten as

$$
g_{3\,d}\left(\hat{\sigma}_u^2\right) = \left(\frac{1}{\hat{\sigma}_u^2+\sigma_d^2}\right)\,\frac{2\,B_d^2}{\sum_{d=1}^D\left(\hat{\sigma}_u^2+\sigma_d^2\right)^2}.
$$

When the ML estimator or REML estimator are used for the variance component estimates, the term $g_3\left(\hat{\sigma}_u^2\right)$ is given by

$$
g_{3\,d}\left(\hat{\sigma}_u^2\right) = \frac{\sigma_d^4}{\left(\hat{\sigma}_u^2+\sigma_d^2\right)^3}\mathrm{Var}\left(\hat{\sigma}_u^2\right),
$$

with

$$
\mathrm{Var}\left(\hat{\sigma}_u^2\right) = 2\left[\sum_{d=1}^D\left(\hat{\sigma}_u^2+\sigma_d^2\right)^{-2}\right]^{-1}.
$$

The MSE of the proposed predictor has two alternative expressions depending on the estimation method. If we use the Prasad and Rao or the REML estimation method, the MSE estimator follows that

$$
\widehat{\mathrm{MSE}}\left(\hat{\gamma}_d^{Ep}\right) = g_{1\,d}\left(\hat{\sigma}_u^2\right)+g_{2\,d}\left(\hat{\sigma}_u^2\right)+2g_{3\,d}\left(\hat{\sigma}_u^2\right).
$$

Meanwhile, if the ML estimation method is used, the MSE estimator is given by

$$
\widehat{\mathrm{MSE}}\left(\hat{\gamma}_d^{\mathrm{Ep}}\right) = g_{1\,d}\left(\hat{\sigma}_u^2\right)+g_{2\,d}\left(\hat{\sigma}_u^2\right)+2g_{3\,d}\left(\hat{\sigma}_u^2\right)-b\,\nabla g_1,
$$

where the last terms are stated as

$$
\begin{aligned}
\nabla g_1 &= \sigma_d^4\left(\hat{\sigma}_u^2+\sigma_d^2\right)^{-2},\\
b &= -\mathrm{tr}\left\{\left(\sum_{d=1}^D\left(\hat{\sigma}_u^2+\sigma_d^2\right)^{-1}\left(x_d^A\right)^\top x_d^A\right)^{-1}\left(\sum_{d=1}^D\left(\hat{\sigma}_u^2+\sigma_d^2\right)^{-2}\left(x_d^A\right)^\top x_d^A\right)\right\}\left[\sum_{d=1}^D\left(\hat{\sigma}_u^2+\sigma_d^2\right)^{-2}\right]^{-1}.
\end{aligned}
$$

## 3. Numerical Applications

### 3.1. Simulation Study

Next, we conduct a simulation study to assess the statistical properties of the proposed predictor. This simulation aims to analyze the relative bias (RB), $\text{RB}_d = (\gamma_d - \hat{\gamma}_d)/\hat{\gamma}_d$, the squared root mean square error $\text{RMSE}\left(\hat{\gamma}_d\right) = \sqrt{\text{MSE}\left(\hat{\gamma}_d\right)}$ and the relative standard error $\text{RSE}_d = (\text{RMSE}(\hat{\gamma}_d)/\gamma_d) \times 100\%$, associated with the proposed predictor versus the Horvitz-Thompson (Narain, 1951; Horvitz and Thompson, 1952), the calibration and the composite estimator. Our simulation study follows the next steps:

(1) Simulate a population of $N = 100,000$ students with 150 items per student and $D = 500$ domains, along with two auxiliary variables associated with the ability $\theta$ for each student.
   a) The first step is to define the number of students that will be in each of the 500 domains. For this, each student is assigned to one of the 500 domains with a probability generated using a beta distribution. Then, the number of students assigned to each domain is counted. This ensures that the domains do not have the same number of students while respecting the fact that the sum of the domain sizes is $N = 100,000$.
   b) Expected abilities are generated for each domain using a sequence of numbers between -3 and 3 as most abilities fall within this range (Andrade et al., 2000).
   c) To generate the students abilities in each domain, a normal distribution is used with a mean equal to the sequence mentioned in step 3 and a random variance (between 0.3 and 0.7 per student).
   d) To generate the items that replicate the ability created in the previous step for each student, a 3PL model and the `simdata` function from the `mirt` library in `R` are used.
   e) To define the auxiliary variables associated with the expected abilities defined in step (c), the variance $\left(\sigma_i^2\right)$ of the random component in a linear regression model of the form $X_{d,j} = \beta_{dj}\theta_{dj} + e_{dj}$ is increased, where $\theta_{dj}$ is the ability of the $j$-th student in domain $d$, $\beta_{dj}$ is a constant obtained from a uniform distribution, and $e_{dj} \sim \text{N}\left(0, \sigma_i^2\right)$, until achieving approximate high ($> 80\%$), medium ($60\%$, $80\%$), and low ($< 60\%$) correlations.

(2) Set 10%, 20%, and 30% of missing responses per student completely at random in the population. Each student is induced with 10%, 20%, or 30% of missing data randomly, depending on the scenario, using the `sample` function to select the responses that will be removed randomly.

(3) Define two auxiliary variables at the domain level correlated with $\theta_d$ on three levels (high ($> 80\%$), medium ($60\%$, $80\%$) and low ($< 60\%$)) using the methodology described in the step (1), item (e).

(4) Estimate five plausible values for each student.
   a) The parameters of the 3PL model are estimated using the EM algorithm. This is done using the `tam.mml.3p` function from the `TAM` library.
   b) Five values are generated using the `tam.pv` function from the `TAM` library.

(5) Consider a different number of domains in the sample using different sampling fractions as $f_d = 30\%$, $50\%$ and $70\%$ on the total number of domains in the population.

(6) For the domains selected in Step 5., select a random sample from the population using simple random sampling with sampling fraction $f_n$=5%, 10% and 20%, considering the Horvitz-Thompson estimator, $\hat{\gamma}_d^{\text{Dir}}$, the calibration estimate, $\hat{\gamma}_d^{\text{Cal}}$, and the composite estimator, $\hat{\gamma}_d^{\text{Comp}}$, the proposed predictor, $\hat{\gamma}_d^{\text{P}}$ using the REML method for each selected domain, along with its standard deviation.

(7) Calculate $\mathrm{RB}_d$ and the relative standard errors, $\mathrm{RSE}_d$, by domain, for each of the estimators $\hat{\gamma}_d^{\mathrm{Dir}}$, $\hat{\gamma}_d^{\mathrm{Cal}}$, $\hat{\gamma}_d^{\mathrm{Comp}}$, and $\hat{\gamma}_d^{\mathrm{P}}$, and compute:
   a) $\mathrm{MRB} = (\sum_{d=1}^{D} \mathrm{SB}_d)/D \times 100\%$ (mean RB).
   b) $\mathrm{MRSE} = (\sum_{d=1}^{D} \mathrm{EERP}_d)/D \times 100\%$ (mean relative standard error)

(8) Repeat steps 5 to 7 for 100,000 random samples and compute:

   a) $$\overline{\mathrm{MRB}} = \left[\left(\sum_{r=1}^{100000} \mathrm{MRB}_r\right) \times 100\%\right]/100000$$
   (mean of the average RB).

   b) $$\overline{\mathrm{MRSE}} = \left[\left(\sum_{r=1}^{100000} \mathrm{MRSE}_r\right) \times 100\%\right]/100000$$
   (mean of the average relative standard error).

The statistical software `R` (Version 4.3.2) was used to conduct this simulation. Some packages used to carry out the simulation include `sae`, `emdi`, `survey`, `TeachingSampling`, `mirt`, and `TAM`. This simulation took approximately four days of computational time due to the number of scenarios in which the estimator was tested and the complexity of the calculations. These calculations were performed on computers with core i7 processors and 32GB of RAM. All script details are in Téllez-Piñérez (2020, Appendix).

Table 1 considers the simulation scenario when the percentage of missing values is 10%, and there is a high correlation between the auxiliary variables and the mean ability. Our findings show that the proposed predictor is unbiased as it was theoretically shown. In all the scenarios, the proposed predictor has a lower $\overline{\mathrm{MRSE}}$ and a lower MSE than the alternative estimators. One of the scenarios where the $\overline{\mathrm{MRSE}}$ is higher for all the considered estimators is when the percentage of domains is 70% (350 domains) and the sample fraction is 5% (5,000 individuals).

Such a configuration is not very convenient in practical terms since, on average, each selected domain will have fifteen observations. Also, in this particular scenario, the Horvitz-Thompson estimator becomes less efficient than the calibration and the composite estimators. However, the proposed predictor has the property of high efficiency with a low $\overline{\mathrm{MRSE}}$ for all the simulation scenarios. Another result in Table 1 is that if the sample percentage increases ($f_n \uparrow$), keeping the number of domains fixed, then the $\overline{\mathrm{MRSE}}$ in all the estimators decreases. Since the sampling fraction inside each domain increases, the estimates variance tends to decrease. In addition, if the number of domains increases ($f_d \uparrow$) keeping the sample percentage fixed, the $\overline{\mathrm{MRSE}}$ for all estimators increases. This is because, as the number of domains increases, the sample size per domain decreases, which makes the estimates less efficient.

Table 1. 10% missing and high correlation.

| $f_d$ (%) | $f_n$(%) | $\overline{\mathrm{MRSE}}\hat{\gamma}_d^{\mathrm{Dir}}$(%) | $\overline{\mathrm{MRSE}}\hat{\gamma}_d^{\mathrm{Cal}}$(%) | $\overline{\mathrm{MRSE}}\hat{\gamma}_d^{\mathrm{Comp}}$ (%) | $\overline{\mathrm{MRSE}}\hat{\gamma}_d^{\mathrm{P}}$(%) | $\overline{\mathrm{MRB}}\hat{\gamma}_d^{\mathrm{P}}$ (%) |
|---|---|---|---|---|---|---|
| 30% | 5% | 1.53 | 1.20 | 1.03 | 0.85 | 0.08 |
| 30% | 10% | 1.23 | 1.07 | 1.00 | 0.85 | 0.02 |
| 30% | 20% | 1.03 | 1.00 | 0.98 | 0.85 | -0.01 |
| 50% | 5% | 1.87 | 1.38 | 1.09 | 0.88 | 0.14 |
| 50% | 10% | 1.45 | 1.16 | 1.02 | 0.86 | 0.05 |
| 50% | 20% | 1.17 | 1.04 | 0.99 | 0.87 | 0.01 |
| 70% | 5% | 2.16 | 1.56 | 1.15 | 0.96 | 0.22 |
| 70% | 10% | 1.63 | 1.24 | 1.05 | 0.87 | 0.09 |
| 70% | 20% | 1.29 | 1.09 | 1.01 | 0.89 | 0.03 |

The results obtained for the remaining scenarios are consistent with the ones in Table 1, which can be found in Tables 2 and 3.

Table 2. 10% and 20% missing and several degrees of correlation.

| $f_d$ (%) | $f_n$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Dir}}$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Cal}}$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Comp}}$ (%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{P}}$(%) | $\overline{\text{MRB}}\hat{\gamma}_d^{\text{P}}$ (%) |
|---|---|---|---|---|---|---|
| 10% missing and average correlation | | | | | | |
| 30% | 5% | 1.48 | 1.47 | 1.01 | 0.92 | 0.06 |
| 30% | 10% | 1.17 | 1.15 | 0.94 | 0.88 | 0.05 |
| 30% | 20% | 0.96 | 0.96 | 0.90 | 0.87 | 0.01 |
| 50% | 5% | 1.84 | 1.84 | 1.11 | 0.96 | 0.09 |
| 50% | 10% | 1.38 | 1.37 | 0.99 | 0.88 | 0.07 |
| 50% | 20% | 1.09 | 1.09 | 0.92 | 0.86 | 0.04 |
| 70% | 5% | 2.11 | 2.17 | 1.22 | 1.11 | 0.11 |
| 70% | 10% | 1.58 | 1.57 | 1.04 | 0.92 | 0.08 |
| 70% | 20% | 1.22 | 1.21 | 0.95 | 0.87 | 0.05 |
| 10% missing and low correlation | | | | | | |
| 30% | 5% | 1.49 | 2.84 | 1.33 | 1.22 | -0.06 |
| 30% | 10% | 1.19 | 1.93 | 1.10 | 1.08 | -0.03 |
| 30% | 20% | 1.00 | 1.24 | 0.97 | 1.01 | 0.01 |
| 50% | 5% | 1.84 | 3.85 | 1.61 | 1.45 | -0.14 |
| 50% | 10% | 1.40 | 2.60 | 1.26 | 1.15 | -0.04 |
| 50% | 20% | 1.13 | 1.72 | 1.06 | 1.01 | 0.01 |
| 70% | 5% | 2.12 | 4.71 | 1.95 | 1.86 | -0.19 |
| 70% | 10% | 1.58 | 3.12 | 1.40 | 1.23 | -0.11 |
| 70% | 20% | 1.24 | 2.11 | 1.14 | 1.03 | -0.02 |
| 20% missing and high correlation | | | | | | |
| 30% | 5% | 1.57 | 1.25 | 1.09 | 0.92 | 0.07 |
| 30% | 10% | 1.27 | 1.12 | 1.05 | 0.91 | 0.02 |
| 30% | 20% | 1.10 | 1.06 | 1.04 | 0.91 | -0.01 |
| 50% | 5% | 1.90 | 1.42 | 1.14 | 0.94 | 0.03 |
| 50% | 10% | 1.48 | 1.20 | 1.08 | 0.93 | 0.06 |
| 50% | 20% | 1.22 | 1.10 | 1.05 | 0.92 | 0.01 |
| 70% | 5% | 2.18 | 1.59 | 1.20 | 1.02 | 0.24 |
| 70% | 10% | 1.67 | 1.29 | 1.10 | 0.94 | 0.1 |
| 70% | 20% | 1.33 | 1.14 | 1.06 | 0.95 | 0.03 |
| 20% missing and average correlation | | | | | | |
| 30% | 5% | 1.53 | 1.52 | 1.09 | 0.98 | 0.07 |
| 30% | 10% | 1.22 | 1.22 | 1.02 | 0.95 | 0.04 |
| 30% | 20% | 1.04 | 1.04 | 0.98 | 0.94 | 0.00 |
| 50% | 5% | 1.85 | 1.89 | 1.18 | 1.04 | 0.09 |
| 50% | 10% | 1.44 | 1.43 | 1.08 | 0.97 | 0.07 |
| 50% | 20% | 1.17 | 1.16 | 1.01 | 0.94 | 0.03 |
| 70% | 5% | 2.13 | 2.24 | 1.29 | 1.19 | 0.09 |
| 70% | 10% | 1.62 | 1.61 | 1.11 | 0.98 | 0.09 |
| 70% | 20% | 1.28 | 1.27 | 1.03 | 0.95 | 0.05 |
| 20% missing and low correlation | | | | | | |
| 30% | 5% | 1.52 | 2.83 | 1.34 | 1.16 | -0.07 |
| 30% | 10% | 1.20 | 1.93 | 1.12 | 1.02 | 0.01 |
| 30% | 20% | 1.02 | 1.26 | 1.00 | 0.96 | 0.02 |
| 50% | 5% | 1.83 | 3.83 | 1.61 | 1.41 | -0.07 |
| 50% | 10% | 1.41 | 2.57 | 1.27 | 1.08 | -0.04 |
| 50% | 20% | 1.15 | 1.73 | 1.08 | 0.96 | 0.01 |
| 70% | 5% | 2.12 | 4.70 | 1.92 | 1.86 | -0.16 |
| 70% | 10% | 1.60 | 3.11 | 1.41 | 1.19 | -0.08 |
| 70% | 20% | 1.26 | 2.11 | 1.16 | 0.99 | -0.01 |

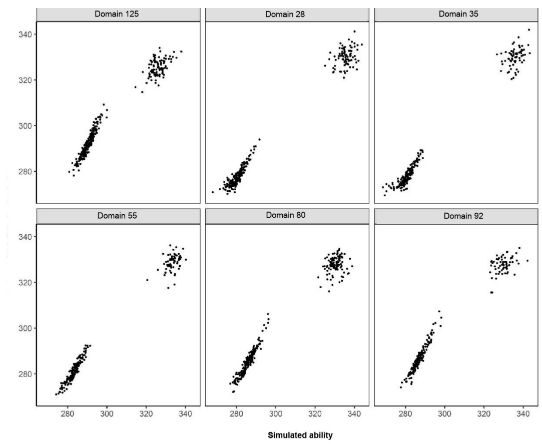
Table 3. 30% missing and several degrees of correlation.

| $f_d$ (%) | $f_n$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Dir}}$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Cal}}$(%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{Comp}}$ (%) | $\overline{\text{MRSE}}\hat{\gamma}_d^{\text{P}}$(%) | $\overline{\text{MRB}}\hat{\gamma}_d^{\text{P}}$ (%) |
|---|---|---|---|---|---|---|
| 30% missing and high correlation | | | | | | |
| 30% | 5% | 1.57 | 1.26 | 1.11 | 0.90 | 0.1 |
| 30% | 10% | 1.29 | 1.13 | 1.07 | 0.89 | 0.04 |
| 30% | 20% | 1.11 | 1.07 | 1.05 | 0.90 | 0.01 |
| 50% | 5% | 1.89 | 1.41 | 1.16 | 0.92 | 0.17 |
| 50% | 10% | 1.48 | 1.21 | 1.10 | 0.90 | 0.07 |
| 50% | 20% | 1.23 | 1.11 | 1.07 | 0.92 | 0.03 |
| 70% | 5% | 2.16 | 1.58 | 1.21 | 1.01 | 0.25 |
| 70% | 10% | 1.67 | 1.29 | 1.12 | 0.92 | 0.11 |
| 70% | 20% | 1.34 | 1.15 | 1.08 | 0.93 | 0.05 |
| 30% missing and average correlation | | | | | | |
| 30% | 5% | 1.64 | 1.65 | 1.24 | 1.04 | 0.09 |
| 30% | 10% | 1.36 | 1.36 | 1.18 | 1.04 | 0.04 |
| 30% | 20% | 1.20 | 1.20 | 1.15 | 1.03 | 0.02 |
| 50% | 5% | 1.94 | 2.00 | 1.33 | 1.11 | 0.12 |
| 50% | 10% | 1.54 | 1.57 | 1.22 | 1.04 | 0.07 |
| 50% | 20% | 1.31 | 1.31 | 1.17 | 1.03 | 0.04 |
| 70% | 5% | 2.22 | 2.34 | 1.42 | 1.25 | 0.13 |
| 70% | 10% | 1.72 | 1.74 | 1.26 | 1.07 | 0.09 |
| 70% | 20% | 1.41 | 1.42 | 1.19 | 1.05 | 0.06 |
| 30% missing and low correlation | | | | | | |
| 30% | 5% | 1.60 | 2.95 | 1.45 | 1.24 | -0.07 |
| 30% | 10% | 1.32 | 2.03 | 1.24 | 1.08 | -0.01 |
| 30% | 20% | 1.16 | 1.39 | 1.14 | 1.03 | 0.02 |
| 50% | 5% | 1.91 | 3.93 | 1.70 | 1.46 | -0.14 |
| 50% | 10% | 1.51 | 2.67 | 1.38 | 1.14 | -0.05 |
| 50% | 20% | 1.26 | 1.84 | 1.21 | 1.04 | 0.00 |
| 70% | 5% | 2.18 | 4.83 | 2.00 | 1.94 | -0.14 |
| 70% | 10% | 1.68 | 3.20 | 1.51 | 1.26 | -0.08 |
| 70% | 20% | 1.37 | 2.20 | 1.28 | 1.07 | -0.01 |

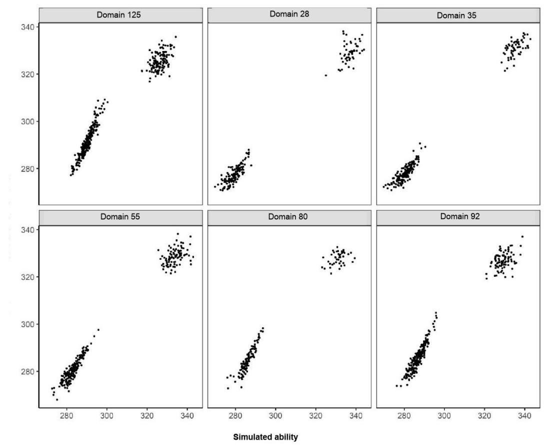
Figures 1(a)-(d) illustrate ability estimations using the methodology proposed in this article compared to the simulated true ability of students in randomly selected domains and specific simulation scenarios. Across the different graphs, the behavior of the ability estimations, in comparison to the true ability, is similar, and no outlier behaviors are observed in the different tested scenarios.

### 3.2. Application with real data

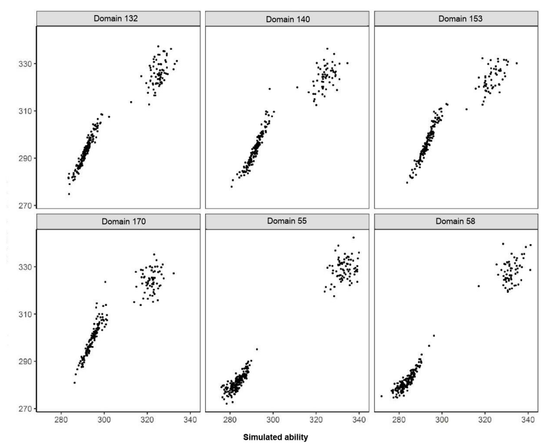
The dataset for this section can be found on the OECD website (https://www.oecd.org/pisa/data/, accessed on 20 May 2024). Plausible PISA 2015 Mathematics Test values are available for 55 countries in this dataset and will be used to estimate their average results. In addition, the auxiliary information that we consider to carry out this application is composed of variables related to the learning context that, according to Treviño et al. (2016), have a direct relationship with academic achievement. In particular, the auxiliary variables considered for each country, in this case, are: (i) gross domestic product —GDP—; (ii) expenditure per student at the secondary level —% of GDP per capita—; (iii) unemployment total —% of total participation in the labour force as a national estimate—; (iv) number of articles in scientific and technical publications; (v) expenditure on research and development —% of GDP—; (vi) public expenditure on education total —% of GDP—, (vii) Gini index; (viii) percentage of schools with access to drinking water service; and (ix) percentage of schools with access to electric service.



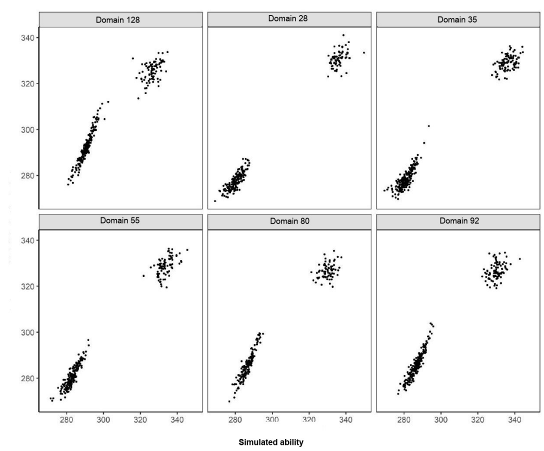

(a) 10% missing data




(b) 20% missing data




(c) 20% missing data




(d) 30% missing data

Figure 1. Observed ability versus estimated ability by domain with high (a), (b) and (d), and median (c) correlation.

First, the sampling design employed in PISA for each country is probabilistic, stratified, and multi-stage. In the first stage, schools are selected, and in the second stage, 15-year-old students are selected. The item response theory model that PISA uses to generate the estimated abilities of students is presented as

$$\mathrm{P}_i\left(\xi_{ik} = 1 \mid \theta_k,\, a_i,\, b_i,\right) = \frac{\exp(1.7a\left(\theta_k - b_i\right))}{1 + \exp(1.7a\left(\theta_k - b_i\right))}. \tag{3.19}$$

Once the students abilities have been estimated using the IRT model stated in Equation (3.19), the Fay-Herriot model is then adjusted with the covariates mentioned in the introduction of this section. The structure of the model is formulated as

$$\hat{\gamma_d} = \beta_0 + \beta_1 X_1 + \cdots + \beta_9 X_9 + \mu_d + \epsilon_d.$$

Once the model was adjusted (the `emdi` library in the `R` software and the `fh` function were used), a stepwise method was performed to obtain the covariates that will be ultimately used in the model (the `emdi` library in the `R` software and the `step` function were used). Table 4 presents the variables and estimates of the selected parameters.

Table 4. Coefficients estimates, $\beta$, and their p-values.

| Covariate | $\hat{\beta}$ |
|---|---|
| Intercept | $-5669.07$ |
| Public spending on education, total (% of GDP) | 14.52 |
| % of schools with access to electricity service | 54.60 |
| % of schools with access to drinking water service | 6.26 |
| Research and development expenditure (% of GDP) | 10.41 |
| Unemployment, total (% of total labor force participation) | $-1.50$ |

We conducted a normality test on the errors and random effects. To compute the estimated ability mean using the proposed predictor, we use Equation (2.9) directly. In doing so, first, we estimate the variance of the random effect by means of Equation (2.7), obtaining that $\hat{\sigma}_u = 986.58$. The variances $\sigma_d^2$ and the direct estimates $\hat{\gamma}_d$ are those reported by PISA 2015. Thus, to calculate the MSE of $\hat{\gamma}_d^{\mathrm{P}}$, we must compute $g_{1d}\left(\hat{\sigma}_u^2\right)$, $g_{2d}\left(\hat{\sigma}_u^2\right)$, and $g_{3d}\left(\hat{\sigma}_u^2\right)$ as shown in Section 2.5.

Table 5 shows all the estimates for $\hat{\gamma}_d^{\mathrm{P}}$ by country and the components for its calculation in Equation (2.8). In these results and for the countries with available PISA 2015 Mathematics Tests, the abilities in Mathematics are quite similar among countries. However, Eastern Asian countries such as Singapore, Hong Kong, Macao, Japan, and South Korea, followed by some European countries like Estonia, Switzerland, Denmark, Finland, Netherlands, and Slovakia (also Canada), have the highest abilities in Mathematics. Countries with lower abilities in Mathematics among the considered countries were Tunisia, Jordan, Brazil, and Colombia in that order.

Also, Table 6 shows the corresponding MSE values of the proposed predictor $\hat{\gamma}_d^{\mathrm{P}}$ with their corresponding three variance components in Equation (2.18). In these results, Vietnam has the highest MSE, followed by Turkey in second place and Bulgaria in third.

Table 7 shows the estimates $\hat{\gamma}_d$ of the mean ability by country obtained in the Mathematics test, along with their estimated coefficient of variation (CVE) taking as inputs the results from Tables 5 and 6. To understand the variability of these countries in relation to their results, the standard errors of the estimates are in Table 7.

Table 7 also displays the corresponding estimates according to the proposed predictor $\hat{\gamma}_d^{\mathrm{P}}$ together with the values of

$$\mathrm{RSE}_d = \frac{\sqrt{\mathrm{MSE}}}{\hat{\gamma}_d^{\mathrm{P}}} \times 100\%,$$

$$\mathrm{Dif}_{\mathrm{rel}} = \frac{\left(\sigma_d^2 - \mathrm{MSE}\left(\hat{\gamma}_d^{\mathrm{P}}\right)\right)}{\sigma_d^2} \times 100\%.$$

The latter expression is the relative difference between the measures of variability of $\hat{\gamma}_d$ and $\hat{\gamma}_d^{\mathrm{P}}$, to compare the reduction in the variability of the estimation of the mean ability in the Mathematics test using the proposed predictor. From Table 7, we see that the $\mathrm{RSE}_d$, in all those countries that participated in the test are lower than the CVE published by PISA. However, it is not surprising that using the proposed predictor, the estimation error decreased, as shown in the $\mathrm{Dif}_{\mathrm{rel}}$ (100%) column, even though they were already very small.

Table 5. Estimate of $\gamma_d^{\mathrm{P}}$ by country.

| Countries | $\sigma_d^2$ | $B_d$ | $1 - B_d$ | $\boldsymbol{x}_{d\mathrm{A}}^\top \hat{\boldsymbol{\beta}}$ | $\hat{\gamma}_d$ | $\hat{\gamma}_d^{\mathrm{P}}$ |
|---|---|---|---|---|---|---|
| Albania | 11.90 | 0.01 | 0.99 | 430.00 | 413.00 | 413.00 |
| Arab Emirates | 5.81 | 0.01 | 0.99 | 461.00 | 427.00 | 427.00 |
| Australia | 2.59 | 0.00 | 1.00 | 476.00 | 494.00 | 494.00 |
| Austria | 8.18 | 0.01 | 0.99 | 519.00 | 497.00 | 497.00 |
| Belgium | 5.52 | 0.01 | 0.99 | 496.00 | 507.00 | 507.00 |
| Brazil | 8.18 | 0.01 | 0.99 | 417.00 | 377.00 | 377.00 |
| Bulgaria | 15.60 | 0.02 | 0.98 | 458.00 | 441.00 | 441.00 |
| Canada | 5.34 | 0.01 | 0.99 | 502.00 | 516.00 | 516.00 |
| Chile | 6.45 | 0.01 | 0.99 | 464.00 | 423.00 | 423.00 |
| Colombia | 5.24 | 0.01 | 0.99 | 359.00 | 390.00 | 390.00 |
| Costa Rica | 6.10 | 0.01 | 0.99 | 416.00 | 400.00 | 400.00 |
| Croatia | 7.67 | 0.01 | 0.99 | 465.00 | 464.00 | 464.00 |
| Czech Republic | 5.76 | 0.01 | 0.99 | 513.00 | 492.00 | 492.00 |
| Denmark | 4.71 | 0.00 | 1.00 | 513.00 | 511.00 | 511.00 |
| Estonia | 4.16 | 0.00 | 1.00 | 497.00 | 520.00 | 520.00 |
| Finland | 5.34 | 0.01 | 0.99 | 507.00 | 511.00 | 511.00 |
| France | 4.41 | 0.00 | 1.00 | 470.00 | 493.00 | 493.00 |
| Germany | 8.35 | 0.01 | 0.99 | 510.00 | 506.00 | 506.00 |
| Greece | 14.06 | 0.01 | 0.99 | 447.00 | 454.00 | 454.00 |
| Hong Kong | 8.88 | 0.01 | 0.99 | 505.00 | 548.00 | 548.00 |
| Hungary | 6.40 | 0.01 | 0.99 | 487.00 | 477.00 | 477.00 |
| Iceland | 3.96 | 0.00 | 1.00 | 542.00 | 488.00 | 488.00 |
| Indonesia | 9.49 | 0.01 | 0.99 | 420.00 | 386.00 | 386.00 |
| Ireland | 4.20 | 0.00 | 1.00 | 453.00 | 504.00 | 504.00 |
| Israel | 13.18 | 0.01 | 0.99 | 512.00 | 470.00 | 471.00 |
| Italy | 8.12 | 0.01 | 0.99 | 469.00 | 490.00 | 490.00 |
| Japan | 9.00 | 0.01 | 0.99 | 519.00 | 532.00 | 532.00 |
| Jordan | 7.02 | 0.01 | 0.99 | 420.00 | 380.00 | 380.00 |
| Latvia | 3.50 | 0.00 | 1.00 | 461.00 | 482.00 | 482.00 |
| Lithuania | 5.43 | 0.01 | 0.99 | 454.00 | 478.00 | 478.00 |
| Luxembourg | 1.61 | 0.00 | 1.00 | 477.00 | 486.00 | 486.00 |
| Macao | 1.23 | 0.00 | 1.00 | 491.00 | 544.00 | 544.00 |
| Mexico | 5.02 | 0.01 | 0.99 | 398.00 | 408.00 | 408.00 |
| Montenegro | 2.13 | 0.00 | 1.00 | 437.00 | 418.00 | 418.00 |
| New Zealand | 5.15 | 0.01 | 0.99 | 470.00 | 495.00 | 495.00 |
| Netherlands | 4.88 | 0.00 | 1.00 | 504.00 | 512.00 | 512.00 |
| Norway | 4.97 | 0.01 | 0.99 | 482.00 | 502.00 | 502.00 |
| Peru | 7.34 | 0.01 | 0.99 | 424.00 | 387.00 | 387.00 |
| Poland | 5.71 | 0.01 | 0.99 | 480.00 | 504.00 | 504.00 |
| Portugal | 6.20 | 0.01 | 0.99 | 482.00 | 492.00 | 492.00 |
| Qatar | 1.61 | 0.00 | 1.00 | 472.00 | 402.00 | 402.00 |
| Romania | 14.36 | 0.01 | 0.99 | 457.00 | 444.00 | 444.00 |
| Russian Federation | 9.67 | 0.01 | 0.99 | 456.00 | 494.00 | 494.00 |
| Singapore | 2.16 | 0.00 | 1.00 | 511.00 | 564.00 | 564.00 |
| Slovakia | 7.08 | 0.01 | 0.99 | 478.00 | 475.00 | 475.00 |
| Slovenia | 1.59 | 0.00 | 1.00 | 495.00 | 510.00 | 510.00 |
| South Korea | 13.76 | 0.01 | 0.99 | 529.00 | 524.00 | 524.00 |
| Spain | 4.62 | 0.00 | 1.00 | 458.00 | 486.00 | 486.00 |
| Sweden | 10.05 | 0.01 | 0.99 | 491.00 | 494.00 | 494.00 |
| Switzerland | 8.53 | 0.01 | 0.99 | 519.00 | 521.00 | 521.00 |
| Thailand | 9.18 | 0.01 | 0.99 | 456.00 | 415.00 | 415.00 |
| Tunisia | 8.70 | 0.01 | 0.99 | 355.00 | 367.00 | 367.00 |
| Turkey | 17.06 | 0.02 | 0.98 | 473.00 | 420.00 | 421.00 |
| United States | 10.05 | 0.01 | 0.99 | 505.00 | 470.00 | 470.00 |
| Vietnam | 19.89 | 0.02 | 0.98 | 439.00 | 495.00 | 494.00 |

Table 6. MSE of $\hat{\gamma}_d^{\mathrm{P}}$ by country.

| Countries | $g_{1d}\left(\hat{\sigma}_u^2\right)$ | $g_{2d}\left(\hat{\sigma}_u^2\right)$ | $g_{3d}\left(\hat{\sigma}_u^2\right)$ | $MSE\left(\hat{\gamma}_d^{\mathrm{P}}\right)$ |
|---|---|---|---|---|
| Albania | 11.7606 | 0.0149 | 0.0051 | 11.7810 |
| Arab Emirates | 5.7741 | 0.0016 | 0.0012 | 5.7770 |
| Australia | 2.5853 | 0.0006 | 0.0002 | 2.5860 |
| Austria | 8.1123 | 0.0048 | 0.0024 | 8.1200 |
| Belgium | 5.4918 | 0.0011 | 0.0011 | 5.4940 |
| Brazil | 8.1123 | 0.0096 | 0.0024 | 8.1240 |
| Bulgaria | 15.3596 | 0.0167 | 0.0087 | 15.3850 |
| Canada | 5.3074 | 0.0013 | 0.0010 | 5.3100 |
| Chile | 6.4097 | 0.0031 | 0.0015 | 6.4140 |
| Colombia | 5.2164 | 0.0145 | 0.0010 | 5.2320 |
| Costa Rica | 6.0634 | 0.0045 | 0.0014 | 6.0690 |
| Croatia | 7.6137 | 0.0050 | 0.0022 | 7.6210 |
| Czech Republic | 5.7266 | 0.0023 | 0.0012 | 5.7300 |
| Denmark | 4.6865 | 0.0013 | 0.0008 | 4.6890 |
| Estonia | 4.1441 | 0.0007 | 0.0006 | 4.1450 |
| Finland | 5.3074 | 0.0018 | 0.0010 | 5.3100 |
| France | 4.3904 | 0.0021 | 0.0007 | 4.3930 |
| Germany | 8.2820 | 0.0038 | 0.0025 | 8.2880 |
| Greece | 13.8649 | 0.0496 | 0.0071 | 13.9220 |
| Hong Kong | 8.8012 | 0.0101 | 0.0029 | 8.8140 |
| Hungary | 6.3596 | 0.0013 | 0.0015 | 6.3620 |
| Iceland | 3.9443 | 0.0045 | 0.0006 | 3.9490 |
| Indonesia | 9.3961 | 0.0079 | 0.0033 | 9.4070 |
| Ireland | 4.1847 | 0.0010 | 0.0007 | 4.1860 |
| Israel | 13.0032 | 0.0311 | 0.0062 | 13.0410 |
| Italy | 8.0562 | 0.0024 | 0.0024 | 8.0610 |
| Japan | 8.9186 | 0.0063 | 0.0029 | 8.9280 |
| Jordan | 6.9729 | 0.0048 | 0.0018 | 6.9790 |
| Latvia | 3.4845 | 0.0005 | 0.0005 | 3.4850 |
| Lithuania | 5.3992 | 0.0023 | 0.0011 | 5.4030 |
| Luxembourg | 1.6103 | 0.0001 | 0.0001 | 1.6100 |
| Macao | 1.2306 | 0.0002 | 0.0001 | 1.2310 |
| Mexico | 4.9922 | 0.0048 | 0.0009 | 4.9980 |
| Montenegro | 2.1270 | 0.0005 | 0.0002 | 2.1280 |
| New Zealand | 5.1261 | 0.0025 | 0.0010 | 5.1300 |
| Netherlands | 4.8600 | 0.0010 | 0.0009 | 4.8620 |
| Norway | 4.9480 | 0.0020 | 0.0009 | 4.9510 |
| Peru | 7.2898 | 0.0083 | 0.0020 | 7.3000 |
| Poland | 5.6792 | 0.0011 | 0.0012 | 5.6820 |
| Portugal | 6.1614 | 0.0021 | 0.0014 | 6.1650 |
| Qatar | 1.6103 | 0.0004 | 0.0001 | 1.6110 |
| Romania | 14.1580 | 0.0259 | 0.0074 | 14.1910 |
| Russian Federation | 9.5782 | 0.0077 | 0.0034 | 9.5890 |
| Singapore | 2.1562 | 0.0002 | 0.0002 | 2.1570 |
| Slovakia | 7.0252 | 0.0020 | 0.0018 | 7.0290 |
| Slovenia | 1.5850 | 0.0001 | 0.0001 | 1.5850 |
| South Korea | 13.5747 | 0.0281 | 0.0068 | 13.6100 |
| Spain | 4.6009 | 0.0040 | 0.0008 | 4.6060 |
| Sweden | 9.9476 | 0.0123 | 0.0037 | 9.9640 |
| Switzerland | 8.4533 | 0.0058 | 0.0026 | 8.4620 |
| Thailand | 9.0963 | 0.0093 | 0.0031 | 9.1090 |
| Tunisia | 8.6264 | 0.0234 | 0.0028 | 8.6530 |
| Turkey | 16.7670 | 0.0109 | 0.0103 | 16.7880 |
| United States | 9.9476 | 0.0050 | 0.0037 | 9.9560 |
| Vietnam | 19.4985 | 0.1504 | 0.0139 | 19.6630 |

Table 7. $\hat{\gamma}_d$ and $\hat{\gamma}_d^{\mathrm{P}}$ along with their quality measures by country.

| Countries | $\hat{\gamma}_d$ | CVE (100%) | $\hat{\gamma}_d^{\mathrm{P}}$ | $\mathrm{RSE}_d$ (100%) | $\mathrm{Dif}_{\mathrm{rel}}$ (100%) |
|---|---|---|---|---|---|
| Albania | 413.0000 | 0.8354 | 413.0000 | 0.8308 | 0.9810 |
| Arab Emirates | 427.0000 | 0.5644 | 427.0000 | 0.5627 | 0.5146 |
| Australia | 494.0000 | 0.3259 | 494.0000 | 0.3256 | 0.2202 |
| Austria | 497.0000 | 0.5755 | 497.0000 | 0.5732 | 0.7043 |
| Belgium | 507.0000 | 0.4635 | 507.0000 | 0.4624 | 0.4953 |
| Brazil | 377.0000 | 0.7586 | 377.0000 | 0.7555 | 0.6457 |
| Bulgaria | 441.0000 | 0.8957 | 441.0000 | 0.8891 | 1.3383 |
| Canada | 516.0000 | 0.4477 | 516.0000 | 0.4467 | 0.4747 |
| Chile | 423.0000 | 0.6005 | 423.0000 | 0.5984 | 0.5547 |
| Colombia | 390.0000 | 0.5872 | 390.0000 | 0.5868 | 0.2132 |
| Costa Rica | 400.0000 | 0.6175 | 400.0000 | 0.6158 | 0.4957 |
| Croatia | 464.0000 | 0.5970 | 464.0000 | 0.5950 | 0.6502 |
| Czech Republic | 492.0000 | 0.4878 | 492.0000 | 0.4865 | 0.4978 |
| Denmark | 511.0000 | 0.4247 | 511.0000 | 0.4238 | 0.4118 |
| Estonia | 520.0000 | 0.3923 | 520.0000 | 0.3916 | 0.3724 |
| Finland | 511.0000 | 0.4521 | 511.0000 | 0.4510 | 0.4652 |
| France | 493.0000 | 0.4260 | 493.0000 | 0.4253 | 0.3645 |
| Germany | 506.0000 | 0.5711 | 506.0000 | 0.5690 | 0.7332 |
| Greece | 454.0000 | 0.8260 | 454.0000 | 0.8222 | 0.9518 |
| Hong Kong | 548.0000 | 0.5438 | 548.0000 | 0.5422 | 0.7138 |
| Hungary | 477.0000 | 0.5304 | 477.0000 | 0.5288 | 0.5770 |
| Iceland | 488.0000 | 0.4078 | 488.0000 | 0.4071 | 0.2573 |
| Indonesia | 386.0000 | 0.7979 | 386.0000 | 0.7941 | 0.7996 |
| Ireland | 504.0000 | 0.4067 | 504.0000 | 0.4062 | 0.3698 |
| Israel | 470.0000 | 0.7723 | 471.0000 | 0.7676 | 0.9874 |
| Italy | 490.0000 | 0.5816 | 490.0000 | 0.5797 | 0.7283 |
| Japan | 532.0000 | 0.5639 | 532.0000 | 0.5619 | 0.7682 |
| Jordan | 380.0000 | 0.6974 | 380.0000 | 0.6948 | 0.5869 |
| Latvia | 482.0000 | 0.3880 | 482.0000 | 0.3874 | 0.3140 |
| Lithuania | 478.0000 | 0.4874 | 478.0000 | 0.4864 | 0.4650 |
| Luxembourg | 486.0000 | 0.2613 | 486.0000 | 0.2611 | 0.1439 |
| Macao | 544.0000 | 0.2040 | 544.0000 | 0.2040 | 0.0966 |
| Mexico | 408.0000 | 0.5490 | 408.0000 | 0.5481 | 0.3736 |
| Montenegro | 418.0000 | 0.3493 | 418.0000 | 0.3489 | 0.1740 |
| Netherlands | 512.0000 | 0.4316 | 512.0000 | 0.4307 | 0.4370 |
| New Zealand | 495.0000 | 0.4586 | 495.0000 | 0.4577 | 0.4341 |
| Norway | 502.0000 | 0.4442 | 502.0000 | 0.4434 | 0.4253 |
| Peru | 387.0000 | 0.7003 | 387.0000 | 0.6978 | 0.5727 |
| Poland | 504.0000 | 0.4742 | 504.0000 | 0.4731 | 0.5140 |
| Portugal | 492.0000 | 0.5061 | 492.0000 | 0.5048 | 0.5459 |
| Qatar | 402.0000 | 0.3159 | 402.0000 | 0.3156 | 0.1256 |
| Romania | 444.0000 | 0.8536 | 444.0000 | 0.8483 | 1.1518 |
| Russian Federation | 494.0000 | 0.6296 | 494.0000 | 0.6274 | 0.8207 |
| Singapore | 564.0000 | 0.2606 | 564.0000 | 0.2604 | 0.1922 |
| Slovakia | 475.0000 | 0.5600 | 475.0000 | 0.5582 | 0.6316 |
| Slovenia | 510.0000 | 0.2471 | 510.0000 | 0.2469 | 0.1436 |
| South Korea | 524.0000 | 0.7080 | 524.0000 | 0.7041 | 1.0728 |
| Spain | 486.0000 | 0.4424 | 486.0000 | 0.4417 | 0.3465 |
| Sweden | 494.0000 | 0.6417 | 494.0000 | 0.6391 | 0.8131 |
| Switzerland | 521.0000 | 0.5605 | 521.0000 | 0.5584 | 0.7269 |
| Thailand | 415.0000 | 0.7301 | 415.0000 | 0.7267 | 0.7536 |
| Tunisia | 367.0000 | 0.8038 | 367.0000 | 0.8019 | 0.5419 |
| Turkey | 420.0000 | 0.9833 | 421.0000 | 0.9738 | 1.5146 |
| United States | 470.0000 | 0.6745 | 470.0000 | 0.6710 | 0.8854 |
| Vietnam | 495.0000 | 0.9010 | 494.0000 | 0.8981 | 1.0803 |

## 4. Concluding remarks

This article reflects the practical and methodological utility of integrating small area estimation, item response theory, and multiple imputation. Specifically, the article proposes an estimator for the average ability parameter of three-parameter logistic models. It was demonstrated through simulation and theory that the proposed predictor for the average of skills is unbiased. Additionally, in all scenarios where the proposed predictor was tested, it exhibited a lower average relative standard error compared to other estimators studied, considering a probabilistic sample. More specifically, by varying the sample fractions ($f_n$) and domain fractions ($f_d$), and having high correlations in auxiliary variables, it was observed that the proposed predictor had a lower average relative standard error compared to other estimators used in the simulation. This behavior persisted for medium and low correlations.

When varying the percentage of missing data, it was observed that although the relative standard errors increased as the percentage of missing data increased, the proposed predictor always obtained the lowest relative standard errors compared to the other estimators studied in the simulation. Furthermore, in the scenario where 30% of missing data and a low sampling fraction were induced, the relative standard errors did not exceed, on average, 5%, which implies that these predictions, according to Särndal (1992), are low.

The new statistical methodology was also applied to the Mathematics results of the 2015 PISA. It is worth noting that the proposed methodology must be handled with caution, and a discussion on the appropriateness of replacing the test administration with estimation generated by statistical models is necessary. Institutions in charge of educational assessments in different countries should always ensure that the resources and the operative work are feasible to avoid missing values. We have shown the advantages from a statistical and theoretical perspective. However, predicted values are neither a way to replace observed values nor to avoid the necessary resources for education assessments and adequate data collection. Therefore, the results generated by this methodology can be used to construct a monitoring system for the performance of educational establishments, which provides a forecast for the establishments reflecting, among other things, the score they would have obtained if they were selected in the sample.

A subsequent comparison between the generated forecast and the results reported by the score will allow the detection of establishments that obtain scores much lower or higher than expected, following the SAE methodology.

It is worth noting that the use of this methodology can be extended to other standardized tests, for example, Program for International Student Assessment for Schools (https://www.oecd.org/pisa/data/, accessed on 20 May 2024) and TIMMS and PRILS (https://www.iea.nl/data-tools/repository, accessed on 20 May 2024). A possible extension of this work is to extend it to the multivariate case, for example, estimating a vector of proportions using the Dirichlet distribution and investigating whether including correlations significantly alters the results already shown in this work.

Author contributions Conceptualization: A.G., C.T., L.T.; investigation: C.T., L.T.; methodology: C.T., L.T.; software: C.T.; writing —original draft: C.T., J.S.; writing —review and editing C.T., J.S., L.T. All the authors have read and agreed to the published version of the article.

Acknowledgements The authors would like to thank the editors and two anonymous reviewers for their valuable comments and suggestions, which helped us significantly improve the quality of this article.

Funding This research received no external funding.

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

## References


Adams, L.M. and Darwin, G., 1982. Solving the quandary between questionnaire length and response rate in educational research. Research in Higher Education, 17, 231–240.

Allen, T.M., 2004. Assessing Academic Programs in Higher Education. Jossey-Bass, Anker Publishing, San Francisco, CA, US.

Álvarez, L., González-Castro, P., Núñez, J.C., González-Pienda, J.A., Álvarez, D., and Bernardo A., 2007. Desarrollo de los procesos atencionales mediante actividades adaptadas. Papeles del Psicólogo, 28, 211–217.

Ariza-Hernández, F. and Gutiérrez-Peña, E., 2021. Bayesian analysis of an item response model with an AEP-based link function. Chilean Journal of Statistics, 12, 103–122.

Ávila, J.L., Huerta, M., Leiva, V., Riquelme, M., and Trujillo, L., 2020. The Fay-Herriot model in small area estimation: EM algorithm and application to official data. REVSTAT Statistical Journal, 18, 613–635.

Baker, F.B. and Kim, S.H., 2004. Item Response Theory: Parameter Estimation Techniques. Dekker, New York, NY, US.

Battese, G.E., Harter, R.M., and Fuller, W.A., 1988. An error-components model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association, 83, 28–36.

Bock, R.D. and Aitkin, M., 1981. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443–459.

Burgard, J., Morales, D., and Wolwer, A., 2019. Area level small area estimation with missing values. Research Papers in Economics, 14/19. Technical Report, Universitat Trier, Germany.

Chambers, R., Chandra, H., and Tzavids, N., 2009. On bias-robust mean squared error estimation for pseudo-linear small area estimators. Survey Methodology, 37, 153–170.

Chambers, R. and Tzavids, N., 2006. M-quantile models for small area estimation. Biometrika, 93, 255–268.

Datta, G., 2009. Model-based approach to small area estimation. In Pfeffermann, D. and Rao, C., editors, Handbook of Statistics. Sample Surveys: Inference and Analysis, volume 29B, pp. 251-288. Elsevier, Oxford, UK.

De Andrade, D.F., Tavares, H.R., and da Cunha Valle, R., 2000. Teoria da Resposta ao Item: Conceitos e Aplicações (in Portuguese). ABE, Sao Paulo, Brazil.

Embretson, S.E. and Reise, S.P., 2008. Item Response Theory. Psychology Press, Routledge, Philadelphia, PA, US.

Falorsi, P.D. and Righi, P., 2008. A balanced sampling approach for multi-way stratification designs for small area estimation. Survey Methodology, 34, 223–234.

Fay, R.E. and Herriot, R.A., 1979. Estimates of income for small places: An application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269–277.

Finch, H., 2008. Estimation of item response theory parameters in the presence of missing data. Journal of Educational Measurement, 45, 225–245.

Fox, J.P., 2010. Bayesian Item Response Modeling: Theory and Applications. Springer, New York, NY, US.

Foy, P., Galia, J., and Li, I., 2008. Scaling the data from the TIMSS 2007 Mathematics and Science assessments. TIMSS 2007 Technical Report, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA, US.

Ghosh, M. and Rao, J.N.K., 1994. Small area estimation: An appraisal. Statistical Science, 9, 55–83.

Hambleton, R.K. and Swaminathan, H., 2013. Item Response Theory: Principles and Applications. Springer, New York, US.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, 320–338.

Horvitz, D.G., and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663–685.

ICFES, 2015. Manual de calificación. Prueba cognitiva Saber 3, 5, 7 y 9. Technical report, Instituto Colombiano para el Fomento de la Educación Superior - ICFES, Bogotá, Colombia.

Jiang, J. and Lahiri, P., 2006a. Estimation of finite population domain means: A model-assisted empirical best prediction approach. Journal of the American Statistical Association, 101, 301–311.

Jiang, J. and Lahiri, P., 2006b. Mixed model prediction and small area estimation. TEST, 15, 1–96.

Kackar, R.N. and Harville, D.A., 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the American Statistical Association, 79, 853–862.

Khorramdel, L., von Davier, M., Gonzalez, E., and Yamamoto, K., 2020. Plausible values. Principles of item response theory. In Moehler, D. and Rammstealt, B., editors, Large-Scale Cognitive Assessment. Analyzing PIAAC Data. Methodology of Educational Measurement and Assessment. pp. 27–47. Springer-Verlag, Cham, Switzerland.

Kuh, G.D., Jankowski, N. and Ikenberry, S.O., 2014. Knowing what students know and can do: The current state of learning outcomes assessment in US colleges and universities. Technical report, University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment, Illinois, US.

Lehtonen, R. and Veijanen, A., 2009. Design-based methods of estimation for domains and small areas. In Pfeffermann, D. and Rao, C., editors, Handbook of Statistics. Sample Surveys: Inference and Analysis, volume 29B, pp. 219–249. Elsevier, Oxford, UK.

Longford, N.T., 2004. Missing data and small area estimation in the UK Labour Force Survey. Journal of the Royal Statistical A, 167, 341–373.

Lord, F.M., 1980. Applications of Item Response Theory to Practical Testing Problems. Routledge, Philadelphia, PA, US.

Martínez, R., 1995. Psicometria: Teoria de los Tests Psicológicos y Educativos (in Spanish). Editorial Sintesis, Madrid, Spain.

Mislevy, R., 1991. Randomization-based inference about latent variables from complex surveys. Psychometrika, 56, 177–196.

Mislevy, R., Beaton, A., Kaplan, B. and Sheehan, K., 1992. Estimating population characteristics from sparse matrix samples of item responses. Journal of Educational Measurement, 29, 133–161.

Morales, D., Esteban, M., Pérez, A., and Hobza, T., 2022. A Course on Small Area Estimation and Mixed Models. Springer, New York, NY, US.

Mullis, I., Martin, M.O., Kennedy, A.M., Trong, K.L., and Sainsbury, M.P., 2015. Assessment Frameworks. TIMMS and PIRLS International Study Center. Boston College, Boston, MA, US.

Muñiz, F.J., 1997. Introducción a la Teoría de Respuesta a los Items (in Spanish). Ediciones Pirámide, Mexico.

Narain, R.D., 1951. On sampling without replacement with varying probabilities. Journal of the Indian Society of Agricultural Statistics, 3, 169–174.

OECD, 2014. PISA 2012. Technical Report. Available at https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf (accessed on 20 May 2024).

OECD, 2016. PISA 2015. Resultados Claves. Technical Report. Available at https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf (accessed on 20 May 2024).

Paek, I. and Cole, K., 2019. Using R for Item Response Theory Model Applications. Routledge, Philadelphia, PA, US.

Pfeffermann, D., 2002. Small area estimation: New developments and directions. International Statistical Review, 70, 125–143.

Pfeffermann, D., 2013. New important developments in small area estimation. Statistical Science, 28, 40–68.

Pfeffermann, D. and Correa, S., 2012. Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation. Biometrika, 99, 457–472.

Prasad, N.G.N. and Rao, J.N.K., 1990. The estimation of the mean squared error of small-area estimators. Journal of the American Statistical Association, 85, 163–171.

Rao, J.N.K., 2008. Some methods for small area estimation. Rivista Internazionale di Scienze Sociali, 116, 387–346.

Rao, J.N.K., 2005. Interplay between survey theory and practice: An appraisal. Survey Methodology, 31, 117–138.

Rao, J.N.K. and Molina, I., 2015. Small Area Estimation. Wiley, New York, NY, US.

Reckase M., 2009. Multidimensional Item Response Theory. Springer, New York, NY, US.

Rodríguez, M., Leiva, V., Huerta, M., Lillo, C., Tapia, A., and Ruggeri, F., 2021. An asymmetric area model-based approach for small area estimation applied to survey data. REVSTAT Statistical Journal, 19, 399–420.

Rose, N., von Davier, M. and Xu, X., 2014. Modeling nonignorable missing data with item response theory. Technical report, ETS Research Report Series, Princeton, NJ, US.

Rubin, D., 1987. Multiple imputation for Nonresponse in Surveys. Wiley, New York, NY, US.

Rubin, D. and Schenker, N., 1991. Multiple imputation in health-care databases: An overview and some applications. Statistics in Medicine, 10, 585–598.

San Martín, E. and Alarcón-Bustamante, E., 2022. Dissecting Chilean surveys: The case of missing outcomes. Chilean Journal of Statistics, 13, 17–46.

Särndal, C. E., Swensson, B., and Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York, NY, US.

Sevilla, A., Cuevas-Ruiz, P. and Sanz, I., 2021. La brecha de género en el rendimiento en matemáticas (TIMSS 2019, in Spanish). Análisis: El Blog de Sociedad y Educación. Fundación Europea Sociedad y Educación, Madrid, Spain, available at https://www.sociedadyeducacion.org/blog/la-brecha-de-genero-en-el-rendimiento-en-matematicas-timss-2019 (accessed on 20 May 2024).

Sulis, I. and Porcu, M., 2017. Handling missing data in item response theory. Assessing the accuracy of a multiple imputation procedure based on latent class analysis. Journal of Classification, 34, 327–359.

Téllez-Piñérez, C., 2020. Estimación de áreas pequeñas utilizando imputación múltiple en modelos logísticos de tres parámetros. Universidad Nacional de Colombia, Doctorado en Ciencias-Estadística, Bogotá, Colombia. Available at https://repositorio.unal.edu.co/handle/unal/78956 (accessed on 20 May 2024).

Treviño, E., Fraser, P., Meyer, A., Morawietz, L., Hinostrosa, P. and Naranjo, E., 2016. Informe de Resultados del Tercer Estudio Regional Comparativo y Explicativo - TERCE. Factores Asociados 2015. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. OREALC/UNESCO. Santiago, Chile.

Treviño, E., Valdés, H., Castro, M., Costilla, R., Pardo, C. and Donoso-Rivas, F., 2010. Factores Asociados al Logro Cognitivo de los Estudiantes de América Latina y el Caribe, Segundo Estudio Regional Comparativo y Explicativo - SERCE (in Spanish). OREAL-C/UNESCO, Santiago, Chile.

UNESCO, 2019. The promise of large-scale learning assessments: Acknowledging limits to unlock opportunities. Technical report, United Nations Educational, Scientific and Cultural Organization, Paris, France.

van der Linden, W. and Hambleton, R.K., 1997. Handbook of Modern Item Response Theory. Springer, New York, NY, US.

van der Linden, W.J., Veldkamp, B.P., and Carlson, J.E., 2004. Optimising balanced incomplete block designs for educational assessments. Applied Psychological Measurement, 23, 317–331.