

RESEARCH PAPER

A new robust Bayesian small area estimation via α -stable model for estimating the proportion of athletic students in California

Shaho Zarei¹ | Serena Arima²  | Giovanna Jona Lasinio³

¹ Department of Statistics, Faculty of Science, University of Kurdistan, Kurdistan, Iran

² Department of History, Social Science and Human Studies, University of Salento, Lecce, Italy

³ Department of Statistical Sciences, University of Rome “La Sapienza”, Roma, Italy

Correspondence

Serena Arima, Department of History, Social Science and Human Studies, University of Salento, Studium 2000, Edificio 5, Via di Valesio, 73100, Lecce, Italy.
Email: serena.arima@unisalento.it

Funding information

Sapienza University, Grant/Award Number: C26W19REHC



This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article are partially reproducible due to their computational complexity.

Abstract

In the last few years, diabetes mellitus and obesity revealed to be one of the fastest-growing chronic diseases in youth in the United States. The number of new diabetes cases is dramatically increasing, and, for the moment, effective therapy does not exist. Experts believe that one of the causes of this increase is the decline in exercise behavior. The California Education Code requires local educational agencies (LEAs) to administer the FITNESSGRAM, the Physical Fitness Test (PFT), to Californian students of public schools. This test evaluates six fitness areas, and experts defined that a passing result on all six areas of the test represents a fitness level that offers some protection against the diseases associated with physical inactivity. We consider 2015–2016 data provided by the California Department of Education (CDE): for each Californian county ($m = 57$), we aim at estimating the county-level proportion of students with a score equal to six. To account for the heterogeneity of the phenomenon and the presence of outlying counties, we extend the standard area-level model by specifying the random effects as a symmetric α -stable (S α S) distribution that can accommodate different types of outlying observations. The model can accurately estimate the county-level proportion of students with a score equal to six. Results highlight some interesting relationships with social and economic situations in each county. The performance of the proposed model is also investigated through an extensive simulation study.

KEYWORDS

area-level model, California FITNESSGRAM, hierarchical Bayesian model, small area estimation, stable distribution

1 | INTRODUCTION

Childhood obesity continues to be one of the most significant health threats to kids and teens across developed countries. In the United States, approximately one in five youth ages 6–19 are obese, according to a recent study (Ogden et al., 2016).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Biometrical Journal* published by Wiley-VCH GmbH.

Childhood overweight and obesity are associated with a plethora of metabolic and clinical constraints, which result in a higher risk for the development of cardiovascular complications and metabolic disease, particularly insulin resistance and type 2 diabetes. Thus, prevention and control for childhood overweight and obesity are urgently needed. Obesity is due to a long-term imbalance between energy intake and energy expenditure: physical activity accounts for 5–35% of total energy expenditure in children and has been regarded as a critical component in obesity prevention (Galuska et al., 2018; Metcalf et al., 2011). However, with social development and changes in people's lifestyles, children and adolescents do not spend enough time on physical activities. The national Youth Risk Behavior Surveillance (YRBS) alerted that in 2011 only 28.7% of nationwide students in the United States were physically active for a total of at least 60 min per day on each of the 7 days before the survey (Eaton et al., 2012). Therefore, the American government deems it necessary to explore effective and feasible strategies to promote childhood physical activity as the best tool for counteracting the diffusion of obesity in the USA. Schools and all educational institutes play a vital role in the dissemination of the culture of physical activities. California Education Code Section 60800 requires local educational agencies (LEAs) to administer the Physical Fitness Test (PFT) annually to students in grades five, seven, and nine. The designated PFT is the FITNESSGRAM. The test evaluates six fitness areas: Aerobic Capacity, Body Composition, Muscle Strength, Endurance and Flexibility, Trunk Extensor Strength and Flexibility, Upper Body Strength and Endurance, and Flexibility. The test is scored based on criterion-referenced standards to evaluate fitness. A passing result in all six areas (6/6) of the test represents a fitness level that offers some protection against the diseases associated with physical inactivity (California Department of Education, 2015). In this paper, we consider data for 2015–2016 provided by the California Department of Education (CDE), and we propose a small area model for estimating the county-level proportion of students with a score equal to six. In order to accommodate counties variability, we extend the standard area-level model by specifying for the random effects a symmetric α -stable ($S\alpha S$) distribution that can capture different types of outlying observations.

The paper is organized as follows. In Section 2, we review the primary literature contribution about the choice of the random effect in small area models when outlying observations are involved. Methodological preliminaries about the $S\alpha S$ distribution are given in Section 3. The Fay–Herriot (FH) model based on $S\alpha S$ for the random effects is introduced in Section 4, and the computational issues related to the model estimation are outlined in Section 5. A simulation study, described in Section 6, is designed to compare the different prior distributions for the random effects in the FH model, highlighting the competitiveness of the proposed model. In Section 7, data about the proportion of athletic students in California are analyzed. The paper concludes with some discussion in Section 8.

2 | RANDOM EFFECTS IN AREA-LEVEL MODEL

Due to the increasing demand for small area statistics, small area estimation (SAE) is becoming one of the most interesting topics for survey statisticians. Direct design-based estimates for small areas usually suffer from significant standard errors due to insufficient or no survey data. Consequently, model-based estimation of small area means is receiving much attention in statistical literature where the goal is to extract information from sources other than the survey (see Molina and Rao, 2015, for a complete review).

The SAE models are mixed-effects models involving random area-specific effects to carry strength to the direct design-based estimates through auxiliary information. The random effects play a crucial role in these models: the bias of model-based small area predictors reduces by adding random effects, but, at the same time, such random effects increase the variability of the predictions, leading, in some cases, to very large confidence intervals (Fay & Graubard, 2001). Therefore, the presence or not of random effects and their specification in the model are issues in the current debate. Because of its simplicity and interpretability, the most well-known area-level model is the FH model (Fay & Herriot, 1979), where the random terms are assumed to be normally distributed with common variance. This assumption is very restrictive, mainly when outliers are detected in the data; hence several alternative models dealing with more flexible specifications of the random effects distribution have been proposed in the literature. In this paper, we discuss in detail the choice of random effect distribution in a Bayesian framework. We contribute to the open debate about the choice of the “optimal” prior distribution for the random effect. In Sinharay and Stern (2003), the authors explained that the distribution of the random effect significantly affects the small area estimates; however, they also showed that violations of the assumptions about the random effects are challenging to be detected using posterior-predictive checks unless the sampling variance is small compared to the population variance. Several extensions of the FH model have been proposed in the literature to make it more robust and flexible. In Bell and Huang (2006), the authors modeled the random effects through a t distribution to deal with outliers in the data. They explained that using a t distribution with few degrees of freedom can decrease

the impact of outliers on the estimation process. In Fabrizi and Trivisano (2010), the authors examine two extensions of the FH model in which the random effects are assumed to be distributed according to either an exponential power distribution (Box & Tiao, 1973) or a skewed exponential power distribution. The authors concluded that the two proposed distributions are two robust alternatives and perform adequately even when normality holds, provided that the number of areas is not too small. Although the skewed exponential power distribution is more flexible with respect to the exponential power distribution, the authors concluded that the two distributions' predictive performance is not dramatically different. They suggested that practitioners use the latter distribution since the parameters have a direct interpretation and its theoretical properties have been extensively explored. A generalized semiparametric approach has been recently proposed in Polettini (2017): the normality assumption for the random effect is replaced by a Dirichlet process that allows accommodating outlying observations. Moreover, design-variances are hierarchically modeled according to a chi-square distribution, embedding in the model the uncertainty on the variances that are often estimated from the same survey data.

Datta and Lahiri (1995) observed that the random effect distribution could not be the same for all areas, especially in the presence of areas where the phenomenon of interest behaves very differently from the other areas. Indeed, they propose to specify different prior distributions for different areas: for the outlying areas, they suggest using a Cauchy distribution while, for the remaining areas, they suggest an appropriate scale mixture of the normal distribution, whose tails are lighter than the Cauchy tails. However, they assume to know which areas are outlying and concluded that their method is appropriate when one or more outliers are found in the data. In Datta et al. (2011), whether or not the presence of the random effects in SAE is necessary is discussed. As already mentioned, random effects reduce the bias of the estimates but increase their variability. When the model is misspecified, and the random effects are not necessary, the predictions' variability is artificially increased with severe consequences on all inferential procedures. To check whether the random effects have to be included in the model, they introduced a test where the null hypothesis is that the common random effect variance is zero. As explained by Torkashvand et al. (2017), "Datta et al. (2011) concluded that including random effects in the model decreases the rate of convergence to the true values of area parameters." The decrease is significant, especially when the sample size is large. They also pointed out that "dropping the random effects can lead to a more accurate point and/or interval estimators, although the flexibility and adaptivity of the area-level (also called Fay-Herriot) model might be lost." Hence, this approach's main drawback is that the random effect will be omitted from all areas, whereas it might be necessary to keep it for some areas. Then, Datta and Mandal (2015) moved to a Bayesian approach and started approaching random effect inclusion as a model selection problem. They proposed to adopt a spike and slab prior distribution, defined as a mixture of a point mass at zero and a zero-mean normal distribution. The spike component describes the probability of a particular random effect in the model to be zero, and then a regression estimate is adequate for that area; the slab component describes nonnegligible random effects when the regression model as such is not adequate. Using a similar rationale, Tang et al. (2018), introduce global-local shrinkage priors for the random effects. These priors allow area-specific variance components for random effects. In areas where random effects are not needed, the variance component is shrunk toward zero; when random effects are needed, the variance component is larger than 0, and it increases with the role that the random effects play in each area. The proposed model is especially advantageous when the number of small areas is large, and the areas are very inhomogeneous for the phenomenon of interest. The global-local shrinkage priors employ two levels of parameters for the variances of normally distributed random effects. The first level consists of the local shrinkage parameters, which are area-specific, while the second level involves the global shrinkage parameter specification, which is the same for all random effects. The global shrinkage parameter captures the overall/global random effect, while the local parameter captures the area-specific random effect. The capability to model the global and local aspects of the random effects is mainly due to the prior distribution's tail. Indeed, the authors state, "If it is appropriately heavy-tailed, both small and large random effects can be well-captured." However, all the aforementioned prior distributions do not account for different types of outlying observations. Atypical observations (outliers) may roughly be divided into two types: mild and gross (Ritter, 2015). A commonly used rule says that a data point is an outlier if it is more than k times the interquartile range above the third quartile or below the first quartile. An outlier is defined as mild for $k \in [1.5; 3)$ while it is gross when $k \geq 3$. Grossly atypical observations are more difficult to predict than mildly atypical observations, and random effects models are often chosen to manage outlying observations, while the goodness of the predictions strongly depends on the distribution of the random effects. Indeed, the standard FH model fails in predicting gross outliers. In this work, we propose to expand FH models by replacing the Gaussian distribution there adopted with the S α S distribution that, being more flexible than the Gaussian, makes the model more robust to outliers effects.

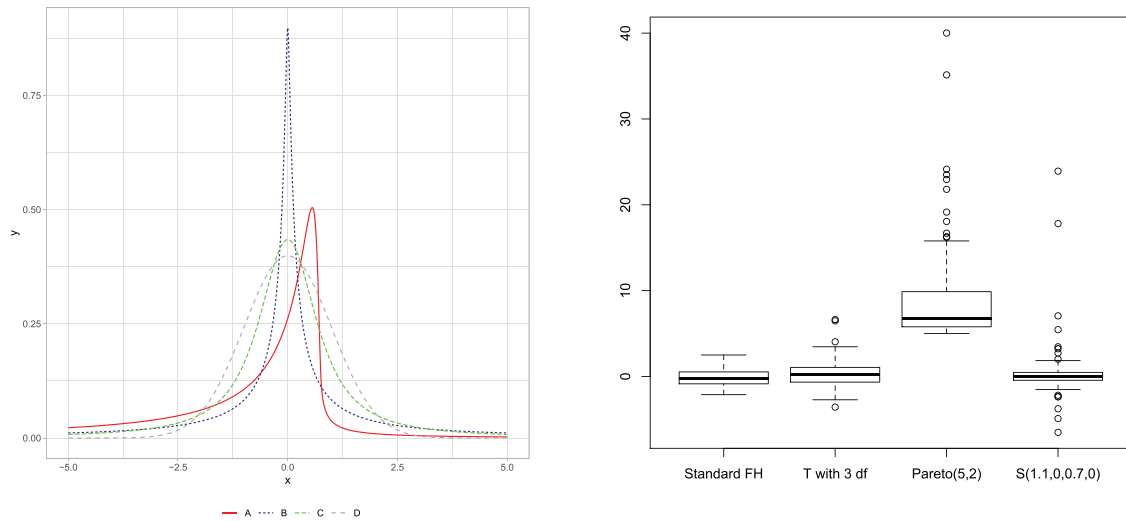


FIGURE 1 Left panel: examples of α -stable distributions: A. $\alpha = 0.5, \beta = -0.8, \gamma = 1, \delta = 0$, B. $\alpha = 0.5, \beta = 0, \gamma = 1/\sqrt{2}, \delta = 0$, C. $\alpha = 1.1, \beta = 0, \gamma = 1/\sqrt{2}, \delta = 0$, D. $\alpha = 2, \beta = 0, \gamma = 1/\sqrt{2}, \delta = 0$. Right panel: boxplots of the data generated from different simulation scenarios

3 | THE α -STABLE DISTRIBUTION

Stable (or α -stable) distributions are a rich class of probability distributions that allow skewness and heavy tails. This class of distributions enjoys many interesting mathematical properties. Initially introduced by Lévy (1925) in his study of sums of independent identically distributed terms, they received little practical attention until the results in Mandelbrot (1961, 1963a), who introduced a simple algorithm for estimating the parameters. He referred to such distributions as “stable Paretian distributions”: in particular, he used those maximally skewed in the positive direction with $1 < \alpha < 2$ as “Pareto-Lévy distributions,” which he regarded as better descriptions of stock and commodity prices than normal distributions (Mandelbrot, 1963b). The univariate α -stable distribution is a four-parameter family, with an index of stability (also called *tail index* or *characteristic exponent*) $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma > 0$, and location $\delta \in \mathbb{R}$. In the following, we denote a stable random variable with aforementioned parameters as $X \sim S(\alpha, \beta, \gamma, \delta)$. The positive stable distributions have $\beta = 1$ and $\alpha < 1$ while the S α S distributions around δ have $\beta = 0$. The S α S distribution is a generalization of the Gaussian distribution having one additional parameter, $\alpha \in (0, 2]$, governing the tails weight. The Gaussian distribution is obtained, as a special case, when $\alpha = 2$, while other symmetric distributions with heavier tails are obtained as α departs from 2. As an example, the Cauchy distribution is obtained for $\alpha = 1$. The left panel of Figure 1 shows the density of the α -stable distribution for different parameters’ different choices. For a more comprehensive discussion about the α -stable distribution, the interested reader can refer to Nolan (2016) and Samorodnitsky and Taqqu (1994).

In general, the variance of the S α S distribution diverges to infinity when $\alpha < 2$ (Nolan, 2016), and this allows the models based on the S α S distribution to be more robust when gross outliers are found in the data. Moreover, S α S satisfies the generalized central limit theorem which states that the only possible nontrivial limit of normalized sums of independent identically distributed terms is stable. Given that the random effects are likely to be the sum of many small terms (maybe with a large variance), a stable model might appropriately describe such effects. The S α S distribution has scale mixtures of normals (SMiN) property. SMiN represents the symmetric stable distribution in conditionally Gaussian form. This property allows standard procedures based on the Gaussian distribution to be reused directly in statistical inferences about models with S α S terms. On the other hand, SMiN induces variances of the random effects to be expressed as the product of two levels of parameters, leading to a new type of global–local prior as introduced by Tang et al. (2018). However, since these distributions do not have a closed form for their densities functions, numerical methods must be used to employ them in modeling, and their properties are studied according to their characteristic functions. There exists many parameterizations of the stable laws characteristic functions (Nolan, 2016), the basic one being

$$\varphi(w; \alpha, \beta, \gamma, \delta) = \exp(iw\delta - |\gamma w|^\alpha (1 - i\beta \operatorname{sign}(w)\Phi)), \quad (1)$$

where $\text{sign}(\cdot)$ is the sign function and

$$\Phi = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\frac{2}{\pi} \log|w| & \alpha = 1. \end{cases}$$

However, (1) shows discontinuities in the parameters when $\alpha = 1$. In Nolan (2016), a simple form for the characteristic function of the $S\alpha S$ distribution is derived as well as some useful algebraic properties. In particular, they introduced the 1-parameterization defined as

$$\phi(w) = \begin{cases} e^{-|w\gamma|^\alpha \left[1 - i \text{sign}(w)\beta \tan\left(\frac{\pi\alpha}{2}\right)\right] + i\mu w} & \alpha \neq 1, \\ e^{-|w\gamma| \left[1 + i \frac{2}{\pi} \text{sign}(w)\beta \log(|w|)\right] + i\mu w} & \alpha = 1. \end{cases} \quad (2)$$

From Equation (2), it can be easily shown that $S(2, 0, \frac{\gamma}{\sqrt{2}}, \mu)$ is the Gaussian distribution with mean μ and variance γ^2 .

4 | BASIC AREA-LEVEL MODEL WITH $S\alpha S$ RANDOM EFFECTS

The FH model (Fay & Herriot, 1979) is defined as follows:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, m, \quad (3)$$

where y_i is the direct estimate for small area i , \mathbf{x}_i is the p -dimensional vector of covariates, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ is a p -dimensional vector of regression coefficients, and m is the number of small areas. The error vector $\mathbf{e} = (e_1, \dots, e_m)^T$ and the random effect vector $\mathbf{v} = (v_1, \dots, v_m)^T$ are assumed to be independent. The elements of \mathbf{e} are independent with $e_i \sim N(0, \psi_i)$, where ψ_i s are typically assumed to be known. When only estimates of the sampling variance are available, a model shrinking both means and variances should be considered. In particular, Maiti et al. (2014) noticed that when the sampling variances are estimated quantities, these are subject to substantial errors because they are often based on equivalent sample sizes as the direct estimates are being calculated. Indeed, they proposed to extend the FH model in order to account the uncertainty of (sampling variance) estimation into the overall SAE strategy. They estimate the model parameters according to an empirical Bayes approach. This idea has been then extended by Sugawara et al. (2017), who propose a Bayesian hierarchical model for the sample variance and discuss different prior choices.

With respect to the area effects, the FH model assumes that v_i s are independent and identically distributed (i.i.d) random variables with $N(0, \sigma_v^2)$. Such an assumption makes the FH model not suitable in the presence of outliers. We proposed a robustification of the FH model for the presence of atypical observations by modeling the random effects v_i with the $S\alpha S$ distribution. In the following, the proposed model will be denoted as Fay-Herriot α -stable (FH α S) model. The $S\alpha S$ distribution allows capturing outlying observation since over the commonly used robustified small area models. The proposed FH α S model is as follows:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, m, \quad (4)$$

where $v_i \stackrel{i.i.d}{\sim} S(\alpha, 0, \gamma_1, 0)$ and $e_i \stackrel{ind}{\sim} N(0, \psi_i)$. The ultimate goal of the proposed model is that it predicts the small area parameter (mean) $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$ with a higher precision than competing approaches.

Following Molina and Rao (2015), the proposed model can be written in the following Bayesian hierarchical way:

- $y_i | v_i, \boldsymbol{\beta}, \gamma_1, \alpha \sim N(\mathbf{x}_i^T \boldsymbol{\beta} + v_i, \psi_i)$,
- $v_i | \boldsymbol{\beta}, \gamma_1, \alpha \sim S(\alpha, 0, \gamma_1, 0)$,
- $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta); \pi(\alpha) \propto \frac{1}{2}; \gamma_1 \sim \pi_1(\cdot)$,

where $\pi_1(\cdot)$ is a density function with positive support adopted as prior distribution for γ_1 .

For the regression parameters β , we assume “flat” prior by specifying a multivariate normal distribution with large variances. Without loosing of generality, we assume $\mu_\beta = 0$ and $\Sigma_\beta = \text{diag}(10^5)$. Adopting the same approach as Salas-Gonzalez et al. (2010), Lombardi (2007), and Tsonas (1999), we choose a Uniform distribution in the interval $(0, 2]$ for the tail index α . Since the closed form of $S(\alpha, 0, \gamma_1, 0)$ is not available, the model as it stands, is hard to be estimated: the joint posterior distribution cannot be computed analytically but also the derivation of full conditional distributions is a complex task. From a computational point of view, it is then convenient to reparameterize the model using the SMIn property of $S\alpha S$ distribution (Samorodnitsky & Taqqu, 1994), more precisely:

Definition 1. Suppose Z is a zero-mean Gaussian random variable with variance σ^2 and $P \sim S(\frac{\alpha}{2}, 1, (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}}, 0)$ is a positive stable random variable, independent of Z then the random variable $X = \mu + \sqrt{P}Z$ is $S(\alpha, 0, \frac{\sigma}{\sqrt{2}}, \mu)$, where μ is a constant.

According to Definition 1, the model can be reparameterized as follows:

- $y_i | v_i, \beta, \gamma_1, \alpha \sim N(\mathbf{x}_i^T \beta + v_i, \psi_i)$,
- $v_i | \lambda_i, \beta, \gamma_1, \alpha \sim N(0, \lambda_i \gamma)$; $\beta \sim N(\mu_\beta, \Sigma_\beta)$,
- $\lambda_i | \alpha \sim S(\frac{\alpha}{2}, 1, (\cos(\frac{\alpha\pi}{2}))^{\frac{2}{\alpha}}, 0)$; $\pi(\alpha) \propto \frac{1}{2}$; $\gamma \sim IG(a, b)$,

where $\gamma = 2\gamma_1^2$ and $IG(a, b)$ denote inverse Gamma distribution with known shape and scale parameters a and b , respectively. Notice that this structure of the random effects is a particular type of the global–local shrinkage prior for the random effects (Tang et al., 2018). The variance parameters of the random effects are area-specific. For each small area, they are expressed as the product of a global parameter γ and a local parameter λ_i : the former captures the overall random effects while the latter brings extra variability to the random effects compared to the FH model. According to the Bayes’ theorem, the posterior distribution is given by

$$\pi(\mathbf{v}, \beta, \gamma, \alpha, \lambda | \mathbf{y}) = \pi(\mathbf{y} | \mathbf{v}, \beta, \gamma, \alpha, \lambda) \pi(\mathbf{v} | \beta, \gamma, \alpha, \lambda) \pi(\lambda | \alpha) \pi(\alpha) \pi(\beta) \pi(\gamma), \quad (5)$$

where $\mathbf{y} = (y_1, \dots, y_m)^T$ is the observations vector and $\lambda = (\lambda_1, \dots, \lambda_m)^T$ is the vector of latent local random effects. Given the complexity of the posterior distribution (5), numerical methods, and Monte Carlo Markov Chain (MCMC) algorithm should be applied in order to sample from it. In particular, we use a combination of the Gibbs sampling, rejection sampling, and Metropolis–Hastings algorithm. Our proposal is developed under a fully Bayesian approach.

5 | COMPUTATIONAL DETAILS

The estimation of the proposed model is tricky and cannot be rephrased in the standard MCMC framework. It requires the application of different sampling algorithms to sample efficiently from the joint posterior distribution. In particular, Gibbs sampling can be involved for β , γ , and v_i ($i = 1, \dots, m$), since full conditional distributions can be derived in closed form. For the other parameters, λ and α , rejection sampling and Metropolis–Hastings algorithm will be used.

5.1 | Updating β , \mathbf{v} , and γ

Parameters β , \mathbf{v} , and γ are updated through Gibbs sampling algorithm according to the following full conditional distributions:

1. $\pi(\beta | \mathbf{v}, \gamma, \alpha, \lambda, \mathbf{y}) \sim N((\mathbf{X}\Sigma^{-1}\mathbf{X}^T + \Sigma_\beta^{-1})^{-1}(\mathbf{X}\Sigma^{-1}(\mathbf{y} - \mathbf{v})), (\mathbf{X}\Sigma^{-1}\mathbf{X}^T + \Sigma_\beta^{-1})^{-1})$
2. $\pi(v_i | \beta, \gamma, \alpha, \lambda, \mathbf{y}) \sim N(\delta_i(y_i - \mathbf{x}_i^T \beta), \delta_i \psi_i)$
3. $\pi(\gamma^{-1} | \mathbf{v}, \alpha, \lambda, \beta, \mathbf{y}) \sim G(\frac{m}{2} + a, \sum_i \frac{v_i^2}{2\lambda_i} + b)$,

where $\Sigma = \text{diag}(\psi_1, \dots, \psi_m)$ and $\delta_i = \frac{\lambda_i \gamma}{\lambda_i \gamma + \psi_i}$ for $i = 1, \dots, m$.

5.2 | Updating the vector of local parameters λ

The posterior distribution of λ is defined as

$$\pi(\lambda|\mathbf{v}, \gamma, \alpha, \boldsymbol{\beta}, \mathbf{y}) \propto \pi(\mathbf{v}|\boldsymbol{\beta}, \gamma, \alpha, \lambda)\pi(\lambda|\alpha).$$

Since $\pi(\lambda|\alpha)$ does not have a closed form, we use rejection sampling for updating λ_i ($i = 1, \dots, m$). Following Godsill and Kuruoglu (1999) and Zarei and Mohammdpour (2020), the posterior distribution can be expressed as

$$\pi(\lambda|\mathbf{v}, \gamma, \alpha, \boldsymbol{\beta}, \mathbf{y}) = \prod_{i=1}^m \pi(\lambda_i|v_i, \gamma, \alpha, \boldsymbol{\beta}, y_i) \propto \prod_{i=1}^m N(v_i|0, \lambda_i\gamma) f_{\frac{\alpha}{2}}(\lambda_i),$$

where $N(v_i|0, \lambda_i\gamma)$ denotes the probability density function of a normal distribution with zero mean and variance equals to $\lambda_i\gamma$ and $f_{\frac{\alpha}{2}}(\cdot)$ is the density function of the positive stable random variable. It is readily seen that the likelihood forms a valid rejection function as it is bounded from above:

$$N(v_i|0, \lambda_i\gamma) \leq \frac{1}{\sqrt{2\pi v_i^2}} \exp\left(-\frac{1}{2}\right).$$

A suitable rejection sampler is:

1. Draw a sample, λ_i^* , from the positive stable distribution with the tail index $\alpha/2$.
2. Draw a sample, u , from the $U(0, \frac{1}{\sqrt{2\pi v_i^2}} \exp(-\frac{1}{2}))$.
3. If $u < N(v_i|0, \lambda_i\gamma)$ accept λ_i^* , otherwise go to 1.

5.3 | Updating the tail index α

Since $\pi(\alpha|\mathbf{v}, \gamma, \lambda, \boldsymbol{\beta}, \mathbf{y}) \propto \pi(\lambda|\alpha)\pi(\alpha)$ and $\pi(\lambda|\alpha) = \prod_{i=1}^m \pi(\lambda_i|\alpha)$ does not have a closed form, the parameter α is estimated using the Metropolis–Hastings algorithm. For each λ_i , we choose the Uniform distribution centered on the current state of the chain, denoted by *cand*, as proposal distribution, that is, $q(\lambda_i|\alpha) \sim U(\text{cand} - 0.15, \text{cand} + 0.15)$. Since the Uniform distribution is symmetric, the acceptance probability in the each iteration is

$$\min \left\{ 1, \frac{\prod_{i=1}^m \pi(\lambda_i|\alpha^{new})}{\prod_{i=1}^m \pi(\lambda_i|\alpha^{(t)})} \right\}. \quad (6)$$

6 | SIMULATION

In this section, we compare the proposed model's performance with the direct estimates, the standard FH model, and the following competing models:

1. The model proposed in Datta and Mandal (2015) (hereafter denoted as DM model). In such a model, the random effects are modeled with spike-and-slab distributions, that is, a mixture of a point mass at zero and a zero-mean normal distribution;
2. The models proposed in Tang et al. (2018), where the random effects follow a global–local prior distribution. We consider two possible specifications of global and local parameters: (A) The global parameter is modeled as Inverse Gamma, and prior distributions for the local parameters set to be normal-Gamma (hereafter denoted as GL-NG). (B) The global parameter is modeled as Inverse Gamma and Laplace distribution (hereafter denoted as GL-LA) as prior distributions for the local parameters.

3. The model in which the random effect follows a Student's t distribution with k degrees of freedom, $v_i \sim t_k$ and k is modeled as a discrete uniform distribution ranging in the interval 0–50 (hereafter denoted as t model);
4. The model proposed in Fabrizi and Trivisano (2010), where the random effects follow an exponential power distribution (hereafter denoted EP).

The data generation settings are adopted from Chakraborty et al. (2016) and Tang et al. (2018). For comparability with the real data applications, the number of small areas, m , is set to 20, 50, and 100. For each choice of m , we generated data from the model (3) with different specifications of the random effects allowing for different outlying behavior of the observed data. Scenario 1 is the standard FH model with normally distributed random effects, $v_i \sim N(0, 1)$. More weight to values on the tails is given in Scenario 2, where the random effects are distributed as a Student t distribution with 3 degrees of freedom, $v_i \sim t_3$ (here after t model). In this setting, Tang et al. (2018) showed that the best fitting model is those with global–local prior for the random effects (GL-LA or GL-NG). In Scenario 3, the model generating the data is the proposed model, $v_i \sim S(1.1, 0, \frac{1}{\sqrt{2}}, 0)$. To investigate the performance of the models when data are generated from a probability distribution that cannot be reconnected to one of the distributions mentioned above, in Scenario 4, we generate data from an FH model with random effects distributed as a Pareto distribution with scale and shape parameters, respectively, equal to 5 and 2, $v_i \sim \text{Pareto}(5, 2)$. Figure 1, right panel, and Figure 1 in the Supplementary Materials present the data's effect when random effects are distributed according to the scenarios above. They show that the data distribution is strongly affected by the distribution of the random effects. When random effects are generated from t_3 , the outliers are mild, while when the random effects are generated according to Scenario 3 and 4, the resulting data are skewed with the gross outlying observations. With the Pareto distribution in Scenario 4, we allow the generation of very large values for the random effects, occasionally inducing very large outliers. The design matrix X includes a column of ones and one explanatory variable sampled from $N(10, 2)$. The coefficient vector β is fixed at $(20, 1)^T$. Errors' variances, ψ_i , are chosen from the set $\{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$, and each value in the set is allocated to the same number of small areas. The hyperparameters of all prior distributions are set to be noninformative. In Scenario 3, we fix $\alpha = 1.1$.

Performance in estimating small area means $\theta_i = \mathbf{x}_i^T \beta + v_i$ are studied according to the following deviance measures: average absolute deviation (AAD), average squared deviation (ASD), average absolute relative deviation (ARB), average squared relative deviation (ASRB), defined as follows:

$$AAD = \frac{1}{m} \sum_{i=1}^m |\hat{\theta}_i - \theta_i|, \quad ASD = \frac{1}{m} \sum_{i=1}^m (\hat{\theta}_i - \theta_i)^2,$$

$$ARB = \frac{1}{m} \sum_{i=1}^m |(\hat{\theta}_i - \theta_i)/\theta_i|, \quad ASRB = \frac{1}{m} \sum_{i=1}^m (\hat{\theta}_i - \theta_i)^2 / \theta_i^2.$$

We also consider the empirical coverage rate (hereafter CR) of 95% credible interval of θ_i . In practice, it is not possible to know the underlying true model, and the true value of θ_i is unknown, so how to choose an appropriate prior for the random effects based on the data is an important task. We use deviance information criterion (DIC; Spiegelhalter et al., 2002) for evaluation of goodness of fit. The following tables show the results of simulations in different scenarios, averaged over the 50 data sets.

Table 1 shows the deviance measures when the random effects are generated under Scenario 1, 2, and 3. As expected, the FH model has the best overall performance; moreover, DM and FH α S perform very similarly, and the deviance measures are not substantially different from those of the FH model. That is not a surprising result since DM model essentially is based on the normal distribution, and FH α S covers the normal distribution as a special case ($\alpha = 2$). Notice that, especially when the number of areas is small ($m = 20$), the Student t model performs similarly and, in some cases, even better than the FH model. That is in agreement with the conclusions in Fabrizi and Trivisano (2010).

In t -student setting, the deviance measures show that the best fitting model is, as expected, the t model. However, when the number of small area increases, the GL-NG model and the FH α S perform very similarly to the model generating the data. When the data are generated from a model with random effects distributed according to the S α S distribution, results confirm that the proposed model outperforms the competing models. Results also highlight that when m is small, the DM model performs better than the other models, while the GL-NG model performs better for large m .

TABLE 1 Models performances according to average absolute deviation (AAD), average squared deviation (ASD), average absolute relative deviation (ARB), average squared relative deviation (ASRB), deviance information criterion (DIC) and coverage rate (CR)

Criteria	Estimate	<i>m</i>			Criteria	Estimate	<i>m</i>				
		20	50	100			20	50	100		
Scenario 1											
ADD	FH α S	0.799	0.708	0.705	ASD	FH α S	1.032	0.818	0.802		
	FH	0.780	0.690	0.701		FH	0.972	0.768	0.783		
	DM	0.764	0.702	0.714		DM	0.925	0.795	0.818		
	GL-NG	0.791	0.720	0.723		GL-NG	1.002	0.838	0.848		
	GL-LA	0.795	0.714	0.717		GL-LA	1.001	0.826	0.829		
	T	0.741	0.691	0.781		T	0.883	0.799	0.731		
	EP	1.07	1.039	0.990		EP	1.864	1.743	1.624		
ARB	Direct	1.270	1.275	1.270	ASRB	Direct	2.759	2.768	2.747		
	FH α S	0.0268	0.0239	0.0237		FH α S	0.0012	0.0009	0.0009		
	FH	0.0262	0.0231	0.0234		FH	0.0011	0.0009	0.0009		
	DM	0.0255	0.0235	0.0239		DM	0.0010	0.0009	0.0009		
	GL-NG	0.0266	0.0242	0.0243		GL-NG	0.0011	0.0009	0.0010		
	GL-LA	0.0266	0.0240	0.0239		GL-LA	0.0011	0.0009	0.0009		
	T	0.0256	0.0232	0.0288		T	0.0010	0.0009	0.0008		
DIC	EP	0.0362	0.0349	0.0331	CR	EP	0.0002	0.0006	0.0003		
	Direct	0.0428	0.0427	0.0426		Direct	0.0032	0.0031	0.0031		
	FH α S	81.67	201.54	400.54		FH α S	0.97	0.96	0.94		
	FH	81.68	201.82	400.17		FH	0.97	0.95	0.94		
	DM	82.17	202.62	403.45		DM	0.88	0.87	0.90		
	GL-NG	82.97	204.05	403.89		GL-NG	0.82	0.81	0.84		
	GL-LA	82.72	203.38	402.85		GL-LA	0.82	0.83	0.88		
Scenario 2	T	81.74	201.83	402.30	CR	T	0.95	0.96	0.98		
	EP	87.684	216.827	433.32		EP	0.90	0.98	0.96		
	Criteria	Estimate	20	50		100	Criteria	Estimate	20	50	100
	ADD	FH α S	0.923	0.876		0.826	ASD	FH α S	1.473	1.289	1.165
		FH	0.929	0.911		0.875		FH	1.517	1.431	1.359
		DM	0.945	0.875		0.839		DM	1.597	1.354	1.245
		GL-NG	0.978	0.870		0.825		GL-NG	1.687	1.289	1.182
GL-LA		0.985	0.874	0.827	GL-LA	1.715		1.304	1.188		
T		0.891	0.841	0.801	T	1.416		1.233	1.127		
EP		1.075	1.074	1.025	EP	1.969		1.934	1.729		
ARB	Direct	1.254	1.329	1.267	ASRB	Direct	2.684	2.958	2.753		
	FH α S	0.0312	0.0296	0.0278		FH α S	0.0017	0.0015	0.0014		
	FH	0.0313	0.0307	0.0300		FH	0.0017	0.0018	0.0030		
	DM	0.0318	0.0295	0.0284		DM	0.0018	0.0017	0.0015		
	GL-NG	0.0329	0.0293	0.0278		GL-NG	0.0019	0.0015	0.0014		
	GL-LA	0.0332	0.0295	0.0280		GL-LA	0.0020	0.0016	0.0014		
	T	0.0401	0.0280	0.0270		T	0.0010	0.0014	0.0013		
DIC	EP	0.0363	0.0363	0.0346	CR	EP	0.0010	0.0080	0.0030		
	Direct	0.0421	0.0446	0.0425		Direct	0.0030	0.0034	0.0031		
	FH α S	83.93	209.78	414.89		FH α S	0.95	0.95	0.94		
	FH	84.46	212.35	421.29		FH	0.94	0.95	0.93		

(Continues)

TABLE 1 (Continued)

Criteria	Estimate	m			Criteria	Estimate	m		
		20	50	100			20	50	100
DIC	DM	85.10	211.44	419.54	CR	DM	0.84	0.88	0.89
	GL-NG	85.61	209.96	415.85		GL-NG	0.80	0.91	0.93
	GL-LA	85.54	210.47	415.50		GL-LA	0.80	0.92	0.94
	T	83.14	209.87	415.32		T	0.99	0.96	0.97
	EP	88.415	219.64	437.77		EP	0.99	0.94	0.94
Scenario 3									
Criteria	Estimate	20	50	100	Criteria	Estimate	20	50	100
	FH1.1S	0.956	0.887	0.843		FH1.1S	1.663	1.451	1.346
	FH α S	0.933	0.888	0.846		FH α S	1.507	1.424	1.367
	FH	1.091	1.082	1.190		FH	2.094	2.056	2.488
ADD	DM	0.962	0.918	0.889	ASD	DM	1.726	1.653	1.587
	GL-NG	0.976	0.921	0.927		GL-NG	1.745	1.510	1.571
	GL-LA	1.012	0.965	0.997		GL-LA	1.866	1.643	1.764
	T	0.965	0.910	0.876		T	1.582	1.524	1.433
	EP	1.159	1.146	1.102		EP	2.272	2.220	2.061
	Direct	1.268	1.248	1.281		Direct	2.785	2.640	2.840
	FH1.1S	0.0330	0.0301	0.0324		FH1.1S	0.0023	0.0018	0.0023
	FH α S	0.0326	0.0310	0.0294		FH α S	0.0026	0.0030	0.0024
	FH	0.0382	0.0370	0.0446		FH	0.0037	0.0029	0.1403
ARB	DM	0.0336	0.0314	0.0330	ASRB	DM	0.0032	0.0025	0.0075
	GL-NG	0.0341	0.0315	0.0326		GL-NG	0.0029	0.0023	0.0039
	GL-LA	0.0354	0.0329	0.0352		GL-LA	0.0032	0.0024	0.0065
	T	0.0332	0.0320	0.0285		T	0.0017	0.0332	0.0098
	EP	0.0388	0.0414	0.0373		EP	0.0015	0.0002	0.0002
	Direct	0.0434	0.0425	0.0455		Direct	0.0036	0.0042	0.0206
	FH1.1S	85.50	209.17	419.80		FH1.1S	0.95	0.95	0.95
	FH α S	85.19	212.32	422.77		FH α S	0.96	0.96	0.95
	FH	88.93	223.17	456.87		FH	0.95	0.95	0.94
DIC	DM	86.94	214.88	438.31	CR	DM	0.85	0.78	0.60
	GL-NG	86.62	212.91	427.44		GL-NG	0.93	0.96	0.97
	GL-LA	87.70	215.92	433.65		GL-LA	0.92	0.96	0.96
	T	85.39	210.97	442.15		T	0.90	0.96	0.96
	EP	89.45	223.046	447.00		EP	0.96	0.96	0.96

Note: All performance diagnostics are reported under Scenario 1, 2, and 3 as described in Section 6. Bold numbers highlight the best performance for each setting.

Also, in this case, the t model performs well when the number of small areas is small. To further investigate the performance of FH α S in Scenario 3, we add a comparison with an S α S model in which α is known (FH1.1S): whether the tail index is known or unknown, the proposed model behaves better than the others, as confirmed by the results in Table 1. That means that the tail index is accurately estimated, and its knowledge leads only to a slight improvement of the estimates.

When the data are generated from a model that cannot be reconnected to one of the models under comparison, Scenario 4, the proposed model's robustness to the competing ones emerges. Indeed, Table 2 shows that for all values of m , FH α S has the best overall performance according to deviance measures and DIC, followed by GL models. This result is quite expected since when data are generated from a Pareto distribution, we faced gross and mild outliers that can be flexibly accommodated by the S α S distribution.

TABLE 2 Comparison of different models performances when data are simulated under Scenario 4, $v_i \sim \text{Pareto}(5, 2)$

Criteria	Estimate	<i>m</i>			Criteria	Estimate	<i>m</i>		
		20	50	100			20	50	100
	FH α S	1.114	1.071	1.064		FH α S	2.196	1.998	1.958
	FH	1.202	1.211	1.232		FH	2.574	2.482	2.542
ADD	DM	1.202	1.118	1.135	ASD	DM	2.523	2.176	2.242
	GL-NG	1.124	1.091	1.088		GL-NG	2.243	2.033	2.016
	GL-LA	1.150	1.125	1.127		GL-LA	2.348	2.152	2.132
	T	1.197	1.095	1.113		T	2.278	2.053	2.135
	ExpPow	1.198	1.209	1.205		ExpPow	2.426	2.509	2.454
	Direct	1.242	1.262	1.271		Direct	2.726	2.716	2.697
	FH α S	0.0284	0.0272	0.0271		FH α S	0.0014	0.0012	0.0012
	FH	0.0307	0.0310	0.0317		FH	0.0017	0.0016	0.0016
ARB	DM	0.0304	0.0283	0.0289	ASRB	DM	0.0016	0.0014	0.0014
	GL-NG	0.0286	0.0278	0.0279		GL-NG	0.0014	0.0013	0.0013
	GL-LA	0.0293	0.0287	0.0289		GL-LA	0.0015	0.0014	0.0014
	T	0.0290	0.0285	0.0280		T	0.0012	0.0012	0.0013
	ExpPow	0.0300	0.0276	0.0310		ExpPow	0.0016	0.0013	0.0013
	Direct	0.0321	0.0326	0.0331		Direct	0.0018	0.0018	0.0018
	FH α S	89.66	220.77	437.32		FH α S	0.95	0.95	0.94
	FH	91.48	229.73	460.02		FH	0.95	0.94	0.95
DIC	DM	92.43	227.37	451.97	CR	DM	0.87	0.78	0.74
	GL-NG	90.40	222.34	441.12		GL-NG	0.95	0.95	0.95
	GL-LA	90.56	224.18	445.22		GL-LA	0.96	0.95	0.95
	T	90.62	225.82	444.778		T	0.94	0.95	0.92
	ExpPow	91.12	227.51	454.270		ExpPow	0.90	0.94	0.95

Note: Model performances are measured according to average absolute deviation (AAD), average squared deviation (ASD), average absolute relative deviation (ARB), average squared relative deviation (ASRB), deviance information criterion (DIC), and coverage rate (CR).

7 | REAL DATA ANALYSIS: ESTIMATING THE PROPORTION OF ATHLETIC STUDENTS IN CALIFORNIA

In the last decades, diabetes mellitus and obesity revealed to be among the most common chronic diseases in youth in the United States. The number of new diabetes cases among youth increased from 3% type 2 cases in 1990 to 45% type 2 cases in 2005. Experts ascribed this significant change to the increase in obesity and a decline in exercise behaviors (Price et al., 2013). This phenomenon dramatically affected the United States, but it is particularly severe in California, where 38.8% of low-income children and adolescents aged 2–19 years, in 2010, were overweight or obese. To study in more detail the diffusion of this phenomenon and determine more specialized political actions, California Education Code Section 60800 requires LEAs to administer the FITNESSGRAM, a PFT, to Californian students. The FITNESSGRAM is a comprehensive, health-related physical fitness battery developed by The Cooper Institute, whose primary goal is to help students establish lifetime habits of regular physical activity. Public school students in grades five, seven, and nine are required to take the PFT, whether or not they are enrolled in a physical education class or participate in a block schedule. These students include those enrolled in LEAs such as elementary, high, and unified school districts, county offices of education, and charter schools. LEAs must also test all students in alternate programs, including, but not limited to, continuation schools, independent study, community day schools, county community schools, and nonpublic schools. The test evaluates six fitness areas: Aerobic Capacity, Body Composition, Muscle Strength, Endurance and Flexibility, Trunk Extensor Strength and Flexibility, Upper Body Strength and Endurance, and Flexibility. The test is scored based on criterion-referenced standards to evaluate fitness. Experts defined that a passing result on all six areas of the test represents a fitness level that offers some protection against the diseases associated with physical inactivity (California Department of Education, 2015). We consider 2015–2016 data provided by the CDE (<http://www.cde.ca.gov/ta/tg/pf/pftresearch.asp>). The

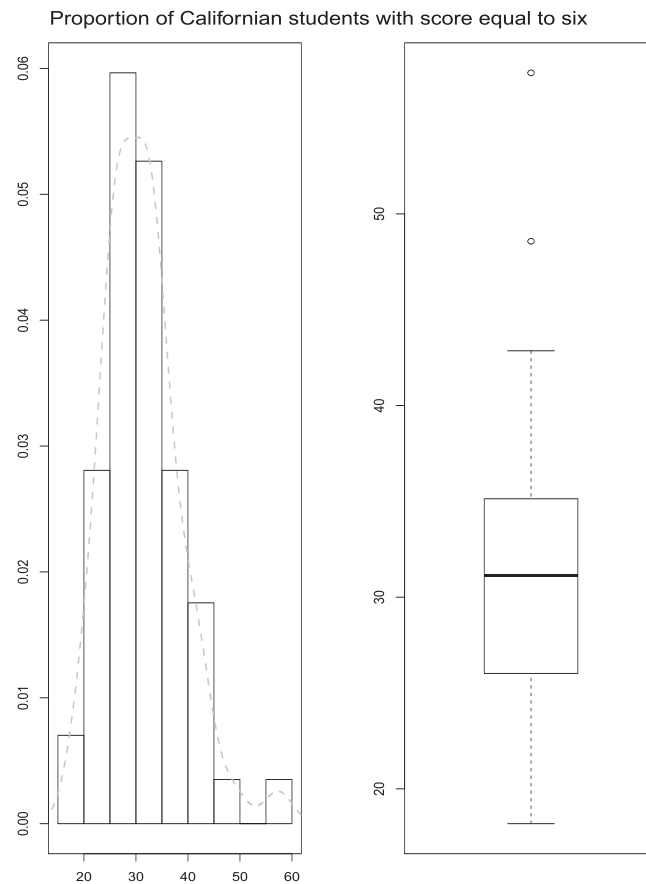


FIGURE 2 PFT data: histogram on left panel and boxplot on the right panel of the proportion of Californian students with a score equal to six

survey is based on a stratified sample where strata correspond to three grades (5th, 7th, 9th), seven race/ethnicity groups (AIAN, AfricanAm, Asian, Latino, Multiple, NHOPI, White), three school types (all schools, noncharter, charter). The direct estimates y_i , that is, the proportion of children that obtain a result 6/6 on the FITNESSGRAM and the design-based variance estimate ψ_i are directly available on the CDE website. The direct estimators are Horvitz–Thompson means, and the design-based variances have been computed with Taylor linearization approximations. We refer to the CDE website for more details about the sampling procedure.

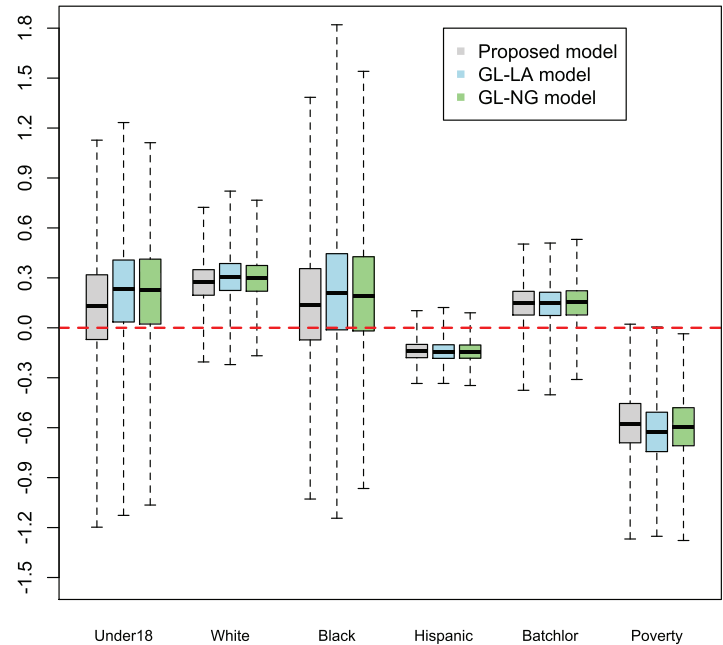
The sample size ($m = 57$) and the proportion of athletic students allow us to use a Gaussian model, although y_i is the mean of binary variables indicating whether a sampled student shows a score equal to six. The y_i s range from 18.18 to 57.36 with a median of 31.14, while standard errors range from 2.24 to 9.6 with a median of 5.374.

As a second data source, we consider the 2015 ACS data available on the Census web page. In particular, for each county in the California state, we collect the following information: percent of under 18 population, percent of the Hispanic population, percent of the White population, percent of the Black population, percent of the population with Bachelor or higher school degree, the median income, and the percent of persons in poverty. All these variables have been used as covariates in the small area model. We also considered many other variables related to the county's urban/rural status and various characteristics of the population, such as percent of the population with a disability and the mean time spent for reaching the workplace. They did not contribute much when adjusting for the other seven covariates included in the model.

Figure 2 shows the histogram and the direct estimator's boxplot, that is, the proportion of Californian students with a score equal to six. The distribution seems to be quite asymmetric, with few potential outlying observations suggesting a more flexible model like the one proposed in the previous sections.

The proposed model has been compared to the aforementioned competing model: the DIC of the proposed model resulted, respectively, equal to 303.35, while for the FH, DM, GL-NG, and GL-LA resulted, respectively, equal to 305.00, 306.69, 303.78, and 303.84. The proposed model fits the data better than other models, followed by the Normal Gamma

FIGURE 3 Posterior distribution of the model parameters: proposed model and competing models



(GL-NG) model and the GL-LA. This result agrees with the simulation study according to which, when data show largely outlying observations, like those generated (see Table 2), the proposed model, GL-NG, and GL-LA models perform very similarly, especially when the number of areas is moderate. In Figure 3, we report the estimates of the model parameters for the proposed model and the two most competing models. According to all models, students' proportion with a score equal to six is positively influenced by the percentage of the White population and the percentage of persons with a high cultural level in the county. On the other hand, such a proportion is negatively influenced by the Hispanic population's percentage and the poverty level. These results agree with the current literature asserting that youth of color and those of low socioeconomic status are disproportionately impacted by diabetes and obesity (Price et al., 2013). Figure 4 shows the small area mean's posterior distribution for the proposed model and the GL-NG model. A very similar plot can be done considering the GL-LA model. The figure highlights the presence of two counties, Marin and Tuolumne show an outlying proportion of students with a score equal to six. Marin's county is well known to be one of the richest counties in the United States, showing the highest pro-capita mean income. Tuolumne county has one of the most recognized and well-administered scholastic systems in the USA. Indeed, schools in that county have an average ranking in the 7 over 10, having one of the highest concentrations of top-ranked public schools in California. The proposed model correctly identifies these two outlying counties while the competing model underestimates them.

In order to investigate the robustness of the model with respect to the presence of the outlying counties, following Arima et al. (2014), we performed a leave-one-out cross-validation: we computed the cross-validation error according to the following formula: $CV = \frac{1}{m} \sum_{i=1}^m (\hat{\theta}_i - \hat{\theta}_i^{(-1)})^2$, where $\hat{\theta}_i$ is the posterior small area mean of the i th area and $\hat{\theta}_i^{(-1)}$ is the posterior small area mean when the i th observation is removed. Figure 3 in the Supplementary Materials shows the small area's posterior distribution means for the proposed model (light blue boxes) and for the proposed model fitted removing the two outlying counties (gray boxes). No significant differences might be grasped. Indeed, when the two counties are removed, the cross-validation error is equal to 0.02, confirming the robustness of the proposed model.

8 | CONCLUSIONS

In this paper, we present an extension of the FH model based on the $S\alpha S$ distribution (FH αS). We act in a fully Bayesian framework to handle large outliers inducing high variability in the random effects. Our proposal can be stated as a hierarchical Bayesian model, and it can be seen as a particular type of the global–local shrinkage prior proposed by Tang et al. (2018). In our model, the random effects' variance parameters are area-specific, and a global parameter γ takes care of the global effects. Our approach's flexibility gives his best performances when large outliers are present: we prove this numerically through several simulated examples. We also apply the proposed model to a real data set regarding the physical

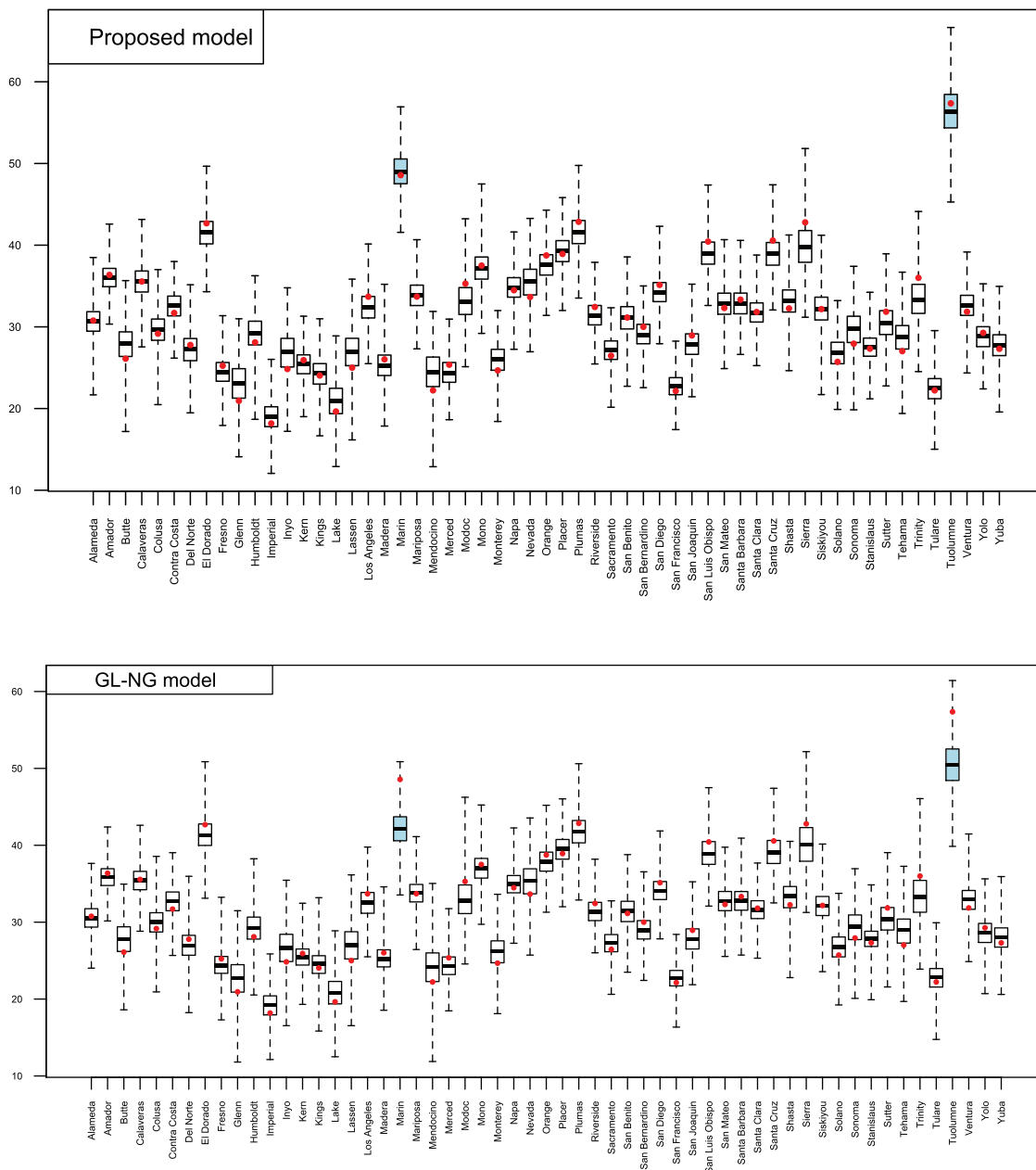


FIGURE 4 Posterior distribution of the small area means of the proposed model (upper panel) and the GL-NG model (lower panel). Red points correspond to the direct estimates

activity of the Californian students. We estimate the county-level proportion of Californian students who obtain a score equal to six to the PFT). Results show that the level of physical activity is significantly related to the county's cultural and economic level, confirming the disparity and heterogeneity among different Californian counties. In real data analysis, our model outperforms its competitors and returns precise and reliable estimates. Here we develop all the necessary detail for implementing the model (the R code is available upon request to the first author).

The FH α S model is more realistic than the Gaussian FH model and has better performance than shrinkage exponential and polynomial-tailed priors used by Tang et al. (2018) in modeling extreme outliers in the random effects. However, the parameter α governing each random effect's tail-weight needs to be evaluated by numerical methods and thus intensify the computational effort needed. However, the computational time remains very reasonable, even with hundreds of areas (a few minutes).

The proposed approach can be extended in several directions. The model ignores the spatial correlation among areas that can be included by modifying the random effects' distribution. We are currently working on a multivariate version

of the proposed model. Another interesting aspect we do not account for is the measurement error in the covariates: we are studying whether specifying a flexible distribution for the random effect could improve small area predictions even in the presence of covariates measured with error.


ACKNOWLEDGMENTS

The authors would like to thank Prof. John Nolan, for allowing the free use of his STABLE R package. Prof. Shaho Zarei was partially supported by the Visting Professor's Grant 2019 from Sapienza University (Anno: 2019 prot. C26W19REHC).

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

 This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article are partially reproducible due to their computational complexity.

ORCID

Serena Arima  <https://orcid.org/0000-0001-6366-1592>

REFERENCES

- Arima, S., Datta, G. S., & Liseo, B. (2014). Bayesian estimators for small area models when auxiliary information is measured with error. *Scandinavian Journal of Statistics*, 42(2), 518–529.
- Bell, W. R., & Huang, E. T. (2006). Using the *t*-distribution to deal with outliers in small area estimation. *Proceedings of Statistics Canada Symposium*.
- Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MS: Addison-Wesley.
- California Department of Education. (2015). *Physical Fitness Testing (PFT)* (<http://www.cde.ca.gov/ta/tg/pf/>). Sacramento, CA: California Department of Education, Physical Fitness Testing.
- Chakraborty, A., Datta, G. S., & Mandal, A. (2016). Robust hierarchical Bayes small area estimation for the nested error linear regression model. *International Statistical Review*, 87, 158–176.
- Datta, G. S., Hall, P., & Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association*, 106(493), 362–374.
- Datta, G. S., & Lahiri, P. (1995). Robust hierarchical Bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis*, 54(2), 310–328.
- Datta, G. S., & Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, 110(512), 1735–1744.
- Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., McManus, T., Chyen, D., Whittle, L., Lim, C., & Wechsler, H. (2012). Youth risk behavior surveillance—United States, 2011. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 61(4), 1–162.
- Fabrizi, E., & Trivisano, C. (2010). Robust linear mixed models for small area estimation. *Journal of Statistical Planning and Inference*, 140, 433–443.
- Fay, M. P., & Graubard, B. I. (2001). Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics*, 57(4), 1198–1206.
- Fay, M. P., & 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(366a), 269–277.
- Galuska, D. A., Gunn, J. P., O'Connor, A. E., & Petersen, R. (2018). Addressing childhood obesity for type 2 diabetes prevention: Challenges and opportunities. *Diabetes Spectrum*, 31(4), 330–335. <https://doi.org/10.2337/ds18-0017>
- Godsill, S., & Kuruoglu, E. E. (1999). Bayesian inference for time series with heavy-tailed symmetric α -stable noise processes. *Proc. Applications of Heavy tailed Distributions in Economics, Engineering and Statistics*.
- Lévy, P. (1925). *Calcul des Probabilités*. Paris: Gauthier-Villars.
- Lombardi, M. J. (2007). Bayesian inference for α -stable distributions: A random walk MCMC approach. *Computational Statistics & Data Analysis*, 51(5), 2688–2700.
- Maiti, T., Ren, H., & Sinha, S. (2014). Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics*, 41, 775–790.
- Mandelbrot, B. (1961). Stable Paretian random functions and the multiplicative variation of income. *Econometrica*, 29(4), 517–543.
- Mandelbrot, B. (1963a). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.

- Mandelbrot, B. (1963b). New methods in statistical economics. *Journal of Political Economy*, 71(5), 421–440.
- Metcalfe, B. S., Hosking, J., Jeffery, A. N., Voss, L. D., Henley, W., & Wilkin, T. J. (2011). Fatness leads to inactivity, but inactivity does not lead to fatness: A longitudinal study in children (EarlyBird 45). *Archives of Disease in Childhood*, 96(10), 942–947.
- Molina, I., & Rao, J. N. K. (2015). *Small area estimation*. Wiley.
- Nolan, J. P. (2016). *Stable distributions: Models for heavy-tailed data*. Boston: Birkhauser. Unfinished manuscript, Chapter 1 online at academic2.american.edu/jpnolan
- Ogden, C. L., Carroll, M. D., Lawman, H. G., Fryar, C. D., Kruszon-Moran, D., Kit, B. K., & Flegal, K. M. (2016). Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *JAMA*, 315(21), 2292–2299. <https://doi.org/10.1001/jama.2016.6361>
- Polettini, S. (2017). A generalised semiparametric Bayesian Fay-Herriot model for small area estimation shrinking both means and variances. *Bayesian Analysis*, 12(3), 729–752.
- Price, J. H., Khubchandani, J., McKinney, M., & Braun, R. (2013). Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States. *BioMed Research International*, 2013, 787616.
- Ritter, G. (2015). *Robust Cluster Analysis and Variable Selection*. Vol. 137 of *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. CRC Press.
- Salas-Gonzalez, D., Kuruoglu, E. E., & Ruiz, D. P. (2010). Modeling with mixture of symmetric stable distributions using Gibbs sampling. *Signal Processing*, 90(3), 774–783.
- Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. New York: Chapman and Hall.
- Sinharay, S., & Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference*, 111, 209–211.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.
- Sugasawa, S., Tamae, H., & Kubokawa, T. (2017). Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*, 44(1), 150–167.
- Tang, X., Ghosh, M., Ha, N. S., & Sedransk, J. (2018). Modeling random effects using global-local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 113(524), 1476–1489.
- Torkashvand, E., Jafari Jozani, M., & Torabi, M. (2017). Clustering in small area estimation with area-level linear mixed models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4), 1253–1279.
- Tsionas, E. G. (1999). Monte Carlo inference in econometric models with symmetric stable disturbances. *Journal of Econometrics*, 88(2), 365–401.
- Zarei, S., & Mohammadpour, A. (2020). Pseudo-stochastic EM for sub-Gaussian α -stable mixture models. *Digital Signal Processing*, 99, 102671.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Zarei, S., Arima, S., Jona Lasinio, G. (2021) A new robust Bayesian small area estimation via α -stable model for estimating the proportion of athletic students in California. *Biometrical Journal*. 63:1309–1324. <https://doi.org/10.1002/bimj.202000235>