Chilean Journal of Statistics
Vol. 8, No. 1, April 2017, 3-24

Invited Paper

A Hierarchical Mixture Beta Dynamic Model of School Performance in the Brazilian Mathematical Olympiads for Public Schools

Alexandra M. Schmidt[2,1,*], Caroline P. de Moraes[3,1], and Helio S. Migon[1]

[1]Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil,
[2]Dept. of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada,
[3]Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro, Brazil

(Received: March 1, 2017 · Accepted in final form: April 15, 2017)

Abstract

We propose a hierarchical mixture dynamic model to investigate the performance of schools in the Brazilian Mathematical Olympiads for Public Schools (OBMEP) across different educational levels, from 2006 until 2013. As not necessarily all the schools in the sample took part in the second phase of the OBMEP across all these years, we propose a mixture dynamic hierarchical model. More specifically, we assume that the score of a school j, in a particular year t, and educational level i, is a realization from a mixture between a Bernoulli distribution with probability of success and a beta distribution. This probability of success describes the probability of presence of a school j, in educational level i and year t, in the second phase of the OBMEP. Inference procedure follows the Bayesian paradigm, meaning that it is performed under a single framework, and uncertainty about unknowns in the model are naturally accounted for. We fit different versions of the proposed model. Our model is able to provide estimates of the performance of a school in a particular year, even if it has not taken part in the second phase of that year. It also provides the probability of presence in the second phase as a function of covariates. Our study indicates that the performance of schools is mainly affected by the school's administrative level and the human development index of the municipality the schools is located in.

Keywords: Bayesian inference · Dynamic linear models · Educational data · Multilevel models · OBMEP.

1. Introduction

Our aim in this study is to understand the performance of schools, as a function of a set of covariates, in the second phase of the Brazilian Mathematical Olympiads for Public Schools (OBMEP). OBMEP has been promoted in Brazil, yearly, since 2005, for three educational levels. To this end we propose a hierarchical beta mixture dynamic model to investigate the performance of schools across different educational levels, from 2006

[*]Corresponding author. Alexandra M. Schmidt, Department of Epidemiology, Biostatistics and Occupational Health McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, QC, Canada, H3A 1A2. *E-mail*: alexandra.schmidt@mcgill.ca. *Homepage*: http://alex-schmidt.research.mcgill.ca/

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

http://www.soche.cl/chjs

until 2013. The mixing component is present because not necessarily all the schools in the sample took part in the second phase of the OBMEP across all these years. More specifically, we consider that the score of a school j, in a particular year t, and educational level i, is a realization of a mixture between a Bernoulli distribution with probability of success θ_{tij}, and a beta probability density function. In order to borrow strength across different years and educational levels, we assume that the parameters involved in θ_{tij}, in the mean, and scale of the beta distribution, follow hierarchical dynamic models. This allows the coefficients of the covariates to evolve smoothly across years. The proposed model is able to provide an estimate of the probability of presence of a school in the second phase of the OBMEP in a given year and educational level. It also provides an estimate of a school's average score even if it has not taken part in the second phase of OBMEP in a particular year and educational level. The data that motivated this study is described in the following subsection.

1.1 The structure of OBMEP

Some authors claim that competitions can be used by educators to develop the talents of the gifted. Campbell and Walberg (2010), for example, analyze data from the USA that involve 345 adult Olympians from different fields that have assumed positions in universities or research institutions and make important contributions to the productivity of the USA. Losada and Rejali (2015) mention that in several countries of the Latin America and Caribbean, the mathematical olympiads have been a very effective vehicle for promoting mathematics and identifying highly talented students even in remote and low-income areas.

The OBMEP has been promoted since 2005 by the Ministries of Science and Technology, and of Education, and organized by *Instituto Nacional de Matemática Pura e Aplicada* (IMPA). The OBMEP is focused on Brazilian public schools, wherein the Brazilian educational system faces serious challenges. The aims of the OBMEP are to stimulate the study of mathematics by students in Brazilian public schools, develop and improve the training of teachers, influence the improvement of public education, in addition to discovering young talents.

The OBMEP is held every year since 2005, when there were over 31,000 schools registered, comprising over 10.5 million students. In 2013 there were over 47,000 schools registered, involving nearly 19.2 million students, covering approximately 99,5% of the municipalities in Brazil.

The OBMEP is structured as follows. The educational school system in Brazil comprises 12 years of basic education, the first 9 years comprise the primary school and the remaining 3 are the secondary school. Compared to other countries, the first 5 years can be compared to primary school, the next four grades can be compared to a low secondary school, and the last three grades are the secondary school or high school (Biondi et al., 2012).

The Brazilian educational public system has three different types of administration: municipal, state and federal. Any of these schools are allowed to register for the OBMEP. The registration is done by the schools, and each school indicates how many students will take part in the first phase of the OBMEP. The students are divided into three different levels:

- Level 1: students in the 6^{th} and 7^{th} grades of the primary school;
- Level 2: students in the 8^{th} and 9^{th} grades of the primary school;
- Level 3: students in high school.

The OBMEP is performed in two phases: first, students take a multiple choice exam with 20 questions for each educational level. The correction of the exams in the first phase is done locally, that is, they are corrected by the school's own teachers. Approximately 5% of students with the highest scores in each level of each school, are approved for the second phase of the OBMEP. Students who obtained zero are not qualified for the second phase, even if his/her school has not reached the proportion of students expected to be in the second phase. In the second phase, students write a discursive examination comprising 6 questions. Each question is worth 20 points, so that the mark varies in the range $(0, 120)$. The questions have sub-items and the division of the 20 points among the sub-items is decided by the organizing committee. In the second phase, the exams are also divided by the level of education. The exams are corrected regionally by committees formed by the OBMEP organizing committee whose members receive a scoring rubric for the questions. Typically, the members of the committee are mathematical researchers from universities in the region, who have experience with Mathematics Olympiads. For every edition, the various regional committees define a cutoff point that will be considered to give the prizes. The scores are reviewed by a national committee who establishes the prizes that will be awarded in that edition.

We aim at studying the performance of schools across Brazil that have taken part in the OBMEP from 2006 until 2013, the latest year that we have information available. Understanding what covariates influence the performance of schools in the OBMEP is important as it might help defining, or revising, strategies about teaching mathematics and attract more students to the area.

The structure of the paper is as follows: next subsections describe the dataset available and how the sample to be analyzed was obtained. Section 2 provides a brief literature review of beta regression models and hierarchical dynamic models. Section 3 describes the proposed model and the inference procedure. Section 4 presents the data analysis, by describing the different fitted models we consider and the model comparison criteria used to choose the best model among those fitted. Then, the results under the best model are discussed. Finally, Section 5 discusses our findings and describes some possible avenues of future research about the OBMEP.

1.2 Dataset description

We have information available from three different sources. The organizers of OBMEP provided information on all the schools that registered for editions of OBMEP between 2005 and 2013. For each year, we have information on the performance of each student within each school in both phases. We also have available the name and the national code of the schools. These can be linked to the schools' census data, which is collected nationwide, every year, by *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP, `http://portal.inep.gov.br/`). The census data have information about local characteristics of the schools. Previous studies about the OBMEP have suggested that the performance of students is strongly related to the geographical region the schools are located in. As in Brazil the geographical regions are strongly related to the human development index (HDI), we also obtained information about the 2010 HDI of each Brazilian municipality present in the data. This is available from `http://www.pnud.org.br/IDH/DH.aspx`.

This initial study focuses on the average scores of the schools, in the different educational levels, that took part in the second phase of the OBMEP between 2006 and 2013. The population under study is relatively big. We aim at fitting highly structured stochastic models to the data that are able to accommodate important features of the observations. For this reason, we start by selecting a sample from the population under study, such that we *ease the computational burden* when fitting the proposed model without loosing important characteristics of the population.

The locations of the schools that take part in the OBMEP are divided between urban and rural areas. In 2013, 70.1% of the schools that participated in the OBMEP are located in urban areas, among these, 0.6% are federal, 42.8% are state and 26.6% are municipal. The remaining 0.1% are private schools that incorporate some students from the public system and offer a curriculum similar to the public one. These private schools are excluded from this study. Throughout the years the distribution of urban and rural schools taking part in the OBMEP follows similar patterns. As the rural schools involve too many particularities we opted to focus only on schools located in urban areas.

Our aim is to model the average score of the schools in the second phase of the OBMEP. In the Brazilian public educational system, schools do not necessarily have all educational levels. Table 1 shows the number of schools, divided by educational level, registered for the second phase of the OBMEP from 2006 until 2013. Although not shown here, the proportion of schools in both phases tend to be around 90% every year. It is worth mentioning that a single school might be registered in more than one level of the OBMEP. For this reason, in Table 1, the column showing the total of schools registered *is smaller than* the sum of the schools registered in the different educational levels, as schools might register more than once, one for each educational level.

1.3 Reducing the size of the data to be analyzed

In order to minimize the variance of the mean average scores we propose a stratified random sampling scheme with Neyman allocation (Thompson, 2012) based on the population of schools that were registered in the second phase of the 2005 edition. This sample will be considered throughout the years. Note that, not necessarily, schools which were present in the second phase of the 2005 edition, will be present in the following years. Our proposed model in Section 3 has a component that captures this feature of the data.

The strata are defined by the following three auxiliary variables: the educational level (1, 2, and 3), the administrative level of the school (federal, state or municipal), and different levels of the HDI. The behavior of the HDI across Brazil is strongly related to the country geographical regions[1], assuming high values in the south, and smaller values in the north and north-east regions of the country. We expect this variable to capture local characteristics of where the school is located in. We assume $z = HDI \in (0, 1)$ with probability density function $f(z)$. Let z_0 and z_U be the smallest and largest values of z in the population. We obtain stratum boundaries, $z_1, z_2, \cdots, z_{U-1}$, by minimizing $V(\overline{z}) = \frac{1}{n}\sum_{h=1}^{U} W_h S_h^2$ and ignoring the finite population correction factor (Dalenius and Hodges, 1959). In the previous equation, $W_h = N_h/N$ is the stratum weight, with N denoting the population size, N_h the stratum size, and S_h^2 the true variance of the stratum. Following this procedure, HDI was divided into 5 categories. When the ranges of the three auxiliary variables are combined 45 strata result.

The sample was obtained by selecting approximately 5% of the schools that took part in the 2005 edition of the OBMEP. This percentage showed to be adequate to attain good precision for our inference procedure. The final sample size comprises $n = 1,501$ schools, with 33 schools registered in at least two levels. This sample size was decided after investigating the behaviour of the estimates of the mean performance of schools by year and educational level. We noted that this sample size gave reasonable results when compared to those obtained under bigger sample sizes.

As the information about the student's gender only started to be collected from 2006 onwards, we do not include the data from 2005 in our sample. The last row of Table 1 shows

[1] See e.g. https://en.wikipedia.org/wiki/List_of_Brazilian_federative_units_by_Human_Development_Index.

the distribution of the number of schools in the sample, by educational level, whereas the last three columns show the number of schools in the sample that actually took part in the second phase in each year. Note that the sum of the schools in all educational levels is greater than the total number of schools in the sample. This is because there are 33 schools in the sample that are registered in more than one educational level.

	Population information (in the 2nd Phase)				Sample information		
	Distribution of schools by educational level ($\times 10^3$)			No. of schools registered ($\times 10^3$)	Distribution of schools by educational level, registered for the 2nd phase		
Year	Level 1	Level 2	Level 3		Level 1	Level 2	Level 3
2006	26.9	26.2	13.2	29.6	517	421	219
2007	32.3	31.5	15.2	35.4	533	454	225
2008	32.6	32.1	15.5	35.9	548	437	215
2009	35.7	34.9	16.3	39.3	565	453	234
2010	36.0	35.5	16.6	39.9	538	433	233
2011	35.7	35.1	16.5	39.9	539	435	216
2012	36.2	35.8	16.7	40.7	506	430	209
2013	37.3	36.8	17.5	42.4	508	423	231

Distribution of schools by educational level in all years		
Level 1	Level 2	Level 3
674	568	259
Total no. of schools in the sample in all years		1468

Table 1. Distribution of the number of schools in the population, and sample, registered in the second phase of the OBMEP, by educational level and year, from 2006 until 2013. There are 33 schools in the sample that took part in more than one educational level and, for this reason, the sum across the columns of the second last row is greater than 1468.

Obtaining the schools average score Let W_{tij} be the average score of school j within level i in year t, $i = 1, 2, 3$, $j = 1, 2, \cdots, n_i$, $t = 1, 2, \cdots, 8$. Define $Y_{tij} = \frac{W_{tij}}{120} \in (0, 1)$, which is the response variable to be modelled in Section 4. Panels of Figure 1 show the boxplots of Y_{tij} as defined above, based on the sample that took part in the second phase of OBMEP. The plots are divided by year and educational level (columns). From these panels it is clear that the majority of schools have quite low scores, and that the distribution of the average scores differ across years and educational levels. Next, we briefly review beta regression models.

2. Literature review

From the panels of Figure 1 it is clear that our observations lie in the interval $(0, 1)$. Also the different boxplots do not seem to be symmetric. Two other important features of the data are that they are observed across years, and not every school in the sample took part in the second phase of all editions between 2006 and 2013. Based on these characteristics we focus our attention on beta regressions, and hierarchical dynamic models.

The use of the beta distribution to model rates and proportions as a function of covariates is relatively recent in the literature. Paolino (2001) proposes a beta regression model for variables observed in a closed interval, and maximum likelihood estimation is performed

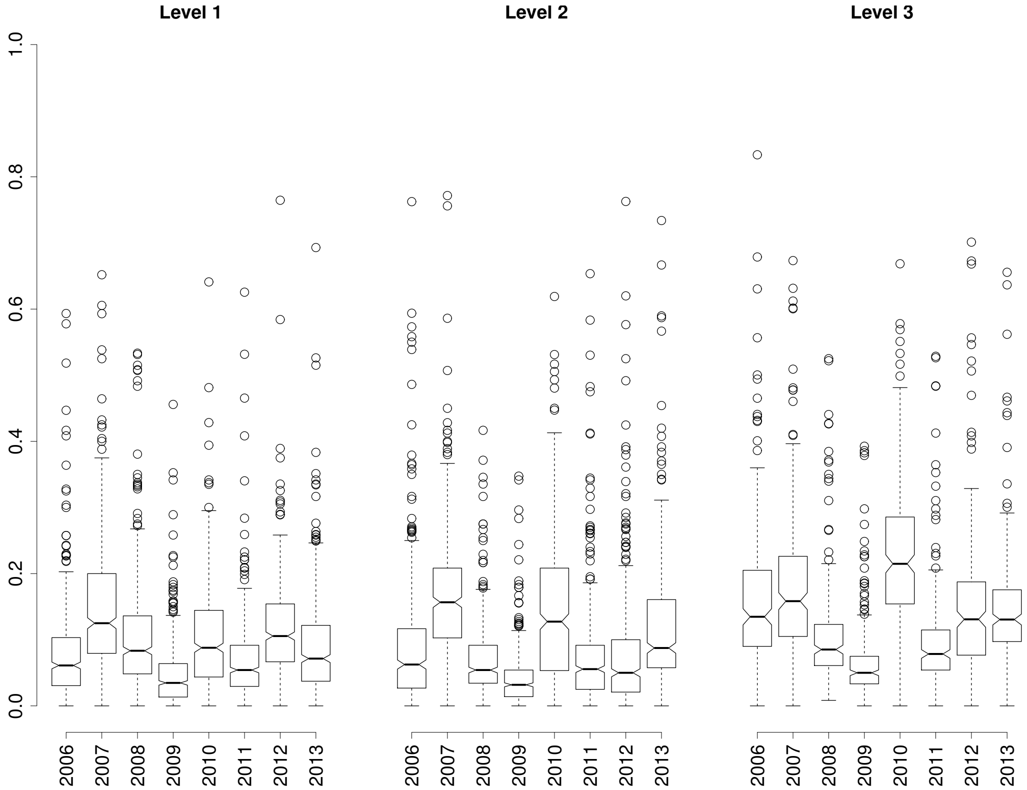

Figure 1. Box plots of the average schools' scores, $Y_{tij} = W_{tij}/120$ in the sample, for each year t, and educational level $i = 1, 2, 3$.

through Monte Carlo simulations. Ferrari and Cribari-Neto (2004) also propose a beta regression model for rates and proportions. They provide closed-form expressions for the score function, for the Fisher's information matrix and perform hypothesis testing of the coefficients using approximations based on the asymptotic normality of the maximum likelihood estimator. In particular, Ferrari and Cribari-Neto (2004), assume that if $Y \sim beta(\mu, \phi)$ then $f(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{((1-\mu)\phi-1)}$, such that $E(Y) = \mu$ and $Var(Y) = \frac{\mu(1-\mu)}{1+\phi}$, for $y \in (0,1)$, $\mu \in (0,1)$, and $\phi > 0$. They focus on the modelling of a transformation of μ as a function of covariates, and assume the precision as a nuisance parameter.

On the other hand, Smithson and Verkuilen (2006) propose a beta regression model wherein the mean and the precision parameters are described as functions of covariates and estimation of the parameters is performed through maximum likelihood. Branscum et al. (2007) discuss beta regression from a Bayesian point of view. In their model, the mean depends on covariates through a logistic link function. They also propose a semiparametric beta regression, and model fitting is performed using `WinBUGS` (Lunn et al., 2000). Zimprich (2010) extend the beta regression model by including random effects in the model of the mean and precision of the beta distribution.

Bayes et al. (2012) propose a beta rectangular regression model which allows more flexibility in the modelling of the tails and of the precision parameter when compared to the beta regression model. The inference procedure follows the Bayesian paradigm and they also use the software `WinBUGS` to obtain samples from the resultant posterior distribution of the model parameters.

Beta regression models have already been used to model behavioral or educational data. Smithson et al. (2011) propose a finite mixture of beta distributions to model response styles, polarization and anchoring in probability judgements. Verkuilen and Smithson (2012) extend the former model to accommodate discrete and continuous mixtures of beta distributions, which enables modeling dependent data. Cepeda-Cuervo and Núñez Antón (2013) propose a spatial double generalized beta regression model to describe the quality of education in Colombia. As observations are obtained in the Departmental level of the country, they propose that the mean and precision parameters of the beta distribution include a spatial lag specification introduced as explanatory variable.

Our proposed model allows the parameters to borrow strength across years and educational levels. To this end we resort to dynamic linear models (West and Harrison, 1997). Fernandes et al. (2009) propose a mixture dynamic linear model to account for zero inflation in spatio-temporal processes. Da-Silva et al. (2011) develop a Bayesian dynamic beta regression model for time series of rates or proportions. They propose to approximate the posterior distribution of the state parameters through Bayesian linear estimation and Gaussian quadrature, avoiding the use of Markov chain Monte Carlo (MCMC) methods.

Because of the hierarchical structure present in the data, our proposed model is related to the class of hierarchical dynamic model proposed by (Gamerman and Migon, 1993) and extended by Da-Silva and Migon (2016) to accommodate beta distributed response variables. Our contribution lies in considering that our observations come from a hierarchical mixture beta dynamic model.

3. Proposed model

As described in Section 1.3 the average scores of the schools were transformed to have Y_{tij} lying in the interval $(0, 1)$. We assume the performance of school $j = 1, 2, \cdots, n_i$ in each level $i = 1, 2, 3$ and year $t = 1, 2, \cdots, T$, is a realization from a mixture distribution. Following the ideas on zero-inflated models, we assume that each observation is generated from a random variable whose probability density function is given by

$$p(y_{tij} \mid \mu_{tij}, \phi_{tij}) = \begin{cases} (1-\theta_{tij}), & y_{tij} = 0, \\ \theta_{tij} p(y_{tij} \mid \mu_{tij}, \phi_{tij}), & 0 < y_{tij} < 1. \end{cases}$$

For ease of notation the equation above assumes that $y_{tij} = 0$ represents that educational level i of school j did not take part in the second phase of OBMEP in year t. Now let Z_{tij} be an indicator variable being equal to 1 if school j, within educational level i took part in the second phase of OBMEP in year t, and 0 otherwise. Conditioned on $Z_{tij} = 1$, let Y_{tij} be the score as computed in the previous section. Then the joint distribution of Y_{tij} and Z_{tij}, conditional on a set of parameters, is written as

$$p(y_{tij}, z_{tij} \mid \theta_{tij}, \mu_{tij}, \phi_{tij}) = [\theta_{tij} p(y_{tij} \mid \mu_{tij}, \phi_{tij})]^{z_{tij}} [1-\theta_{tij}]^{(1-z_{tij})}, \tag{1}$$

where θ_{tij} is the probability that school j, within educational level i, takes part in the second phase of OBMEP in year t, such that $Z_{tij} \mid \theta_{tij} \sim \textit{Bernoulli}(\theta_{tij})$, with $0 < \theta_{tij} < 1$. If a school takes part in the second phase of OBMEP, that is, if $Z_{tij} = 1$, then we assume the average score of school j, within educational level i, and year t, follows a beta distribution with probability density function (pdf) $p(y_{tij} \mid z_{tij} = 1, \mu_{tij}, \phi_{tij})$, where μ_{tij} represents the mean of $Y_{tij} \mid Z_{tij} = 1$, and $\phi_{tij} > 0$ is a scale parameter (Ferrari and Cribari-Neto, 2004). In what follows we describe the proposed model for the components θ_{tij}, μ_{tij} and ϕ_{tij}.

3.1 Dynamic Hierarchical prior specification

Following the model in equation (1) we propose a hierarchical dynamic linear (Gamerman and Migon, 1993) prior specification for each of the parameters that are involved in the joint distribution of Y_{tij} and Z_{tij}, that is, θ_{tij}, the probability of presence in the second phase, μ_{tij} the mean of the average score, and ϕ_{tij} the precision of the distribution of Y_{tij} (conditioned on $Z_{tij} = 1$), of school j, in educational level i and year t.

Let $\boldsymbol{\eta}_{tij} = (\eta_1, \eta_2, \eta_3)'_{tij} = (g_1(\theta_{tij}), g_2(\mu_{tij}), g_3(\phi_{tij}))'$ be a three-dimensional vector, such that $g_k(\cdot)$ represents a transformation of the parameter of interest to the real line. As $\theta_{tij} \in (0, 1)$ one can assume $g_1(\theta_{tij}) = \log \frac{\theta_{tij}}{1-\theta_{tij}}$. As $p(. \mid \mu_{tij}, \phi_{tij})$ is the pdf of a beta distribution, then $\mu_{tij} \in (0, 1)$, and we assume $g_2(\mu_{tij}) = \log \frac{\mu_{tij}}{1-\mu_{tij}}$. As we assume ϕ_{tij} to be the precision parameter of the beta distribution, a natural choice is to assume $g_3(\phi_{tij}) = \log \phi_{tij}$. We assume the $k - th$ component of $\boldsymbol{\eta}_{tij}$ is modelled as

$$(\eta_k)_{tij} = g_k(\cdot) = \boldsymbol{\beta}^k_{ti}\mathbf{X}^k_{tij} + \delta^k_j, \text{ with} \tag{2a}$$

$$\boldsymbol{\beta}^k_{ti} = \mathbf{F}'_k\boldsymbol{\alpha}^k_t + \mathbf{v}^k_{ti}, \quad \mathbf{v}^k_{ti} \sim N_{p_k}(\mathbf{0}, \mathbf{V}^k_i),\ t = 1, \cdots, T, \tag{2b}$$

$$\boldsymbol{\alpha}^k_t = \mathbf{G}_k\boldsymbol{\alpha}^k_{t-1} + \boldsymbol{\omega}^k_t, \quad \boldsymbol{\omega}^k_t \sim N_{p_k}(\mathbf{0}, \mathbf{W}^k) \tag{2c}$$

$$\boldsymbol{\alpha}^k_0 \sim N(\mathbf{m}^k_0, \mathbf{C}^k_0), \tag{2d}$$

where $\mathbf{X}^k_{tij}$ is a p_k-dimensional vector of covariates that affect a known transformation of the $k - th$ component of $\boldsymbol{\eta}_{tij}$, and δ^k_j is a random effect associated with the $j-$th school, that captures unobserved school's characteristics, after adjusting $(\eta_k)_{tij}$ to the covariate vector $\mathbf{X}^k_{tij}$. As we observe a short period of time, we assume this random effect fixed across the years. The coefficients $\boldsymbol{\beta}^k_{ti}$ in equation (2a) vary with the educational level i, and year t. Equation (2b) describes the hierarchical structure of the coefficients $\boldsymbol{\beta}^k_{ti}$, as schools belonging to the same educational level follow the same prior distribution. In particular, *a priori*, the coefficients $\boldsymbol{\beta}^k_{ti}$ follow a dynamic linear model (West and Harrison, 1997). That is, $\mathbf{F}_k$ is a q_k-dimensional vector, and $\mathbf{G}_k$ is a q_k-dimensional matrix. Equations (2b) and (2c) and, in particular, $\mathbf{F}_k$ and $\mathbf{G}_k$ allow the model to accommodate different temporal structures, such as a linear trend and/or a seasonal pattern, to describe the behaviour of the coefficients $\boldsymbol{\beta}^k_{ti}$. The parameter vector $\boldsymbol{\alpha}^k_t = (\alpha^k_{0t}, \cdots, \alpha^k_{(q_k-1)t})'$ is a q_k-dimensional vector, with each component representing the overall effect of the l^{th}_k component of $\mathbf{X}^k_{tij}$ on the transformation of parameter $(\eta_k)_{tij}$. The components $\mathbf{v}^k_{ti}$ and $\boldsymbol{\omega}^k_t$ are assumed mutually and internally independent, for all k, i and t, each following a p_k-dimensional multivariate normal distribution. The covariance matrix of the prior distribution of the coefficients, $\mathbf{V}^k_i$, also varies with the educational level i, such that $\mathbf{V}^k_i$ is a q_k-dimensional diagonal matrix, with elements V^k_{im}, $m = 0, 1, 2, \cdots, q_k - 1$. And $\mathbf{W}^k$ is a q_k-dimensional diagonal matrix, with each element of the diagonal representing the variance of the evolution in time of component α^k_{tm}, $m = 0, \cdots, q_k - 1$. The vector $\boldsymbol{\alpha}^k_0$ represents the initial information of the overall effect vector $\boldsymbol{\alpha}^k_t$, and *a priori*, it follows a multivariate normal distribution with known mean vector $\mathbf{m}^k_0$ and covariance matrix $\mathbf{C}^k_0$. Figure A1 in the Appendix, depicts a directed acyclic graph of the proposed model for $\delta^3_j = 0$.

3.2 Likelihood function, prior specification and inference procedure

We follow the Bayesian paradigm to perform the inference procedure. One of the main advantages is that it is performed under a single framework and uncertainty about parameters' estimates is naturally obtained. To complete model specification and following equations (2) we are left to assign the prior distribution of the hyperparameters δ_j^k, the diagonal elements of $\mathbf{V}_i^k$, and of $\mathbf{W}_i^k$ for $i, k = 1, 2, 3$. We assume prior independence among the hyperparameters. For the school's random effects we assume independent, zero mean, normal distributions with unknown variance σ_k^2. For the variances, we assign independent, inverse gamma prior distributions with infinite variance and prior mean fixed at some reasonable value, e.g. the maximum likelihood estimate based on independent fits for each year, and educational level. In other words, following the DAG in Figure A1, we fix $c = a_W = a_j = 2$, for $j = 1, 2, 3$, and b_j is based on previous estimates obtained from independent fits for each year and educational level, such that we have a prior distribution with infinite variance and location fixed at a reasonable value.

Likelihood function Let $\mathbf{y}$ be the vector comprising the average scores of the schools stacked across the different educational levels and years, $\mathbf{z}$ a vector of 0s and 1s indicating if school j, within educational level i, took part in the second phase of OBMEP in year t. And let $\boldsymbol{\Theta}$ be the parameter vector comprising all the parameters and hyperparameters in equation (2). As we assume that $(Y_{tij} \mid Z_{tij} = 1)$ follows a beta distribution, the likelihood function, $f(\mathbf{y} \mid \boldsymbol{\Theta})$, is given by

$$f(\mathbf{y}, \mathbf{z} \mid \boldsymbol{\Theta}) = \prod_{t=1}^{T} \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left\{ \theta_{tij} \frac{\Gamma(\phi_{tij})}{\Gamma(\mu_{tij}\phi_{tij})\Gamma\left((1-\mu_{tij})\phi_{tij}\right)} y_{tij}^{[\mu_{tij}\phi_{tij}-1]} \left(1 - y_{tij}\right)^{[(1-\mu_{tij})\phi_{tij}-1]} \right\}^{z_{tij}} \\ \times \left[1 - \theta_{tij}\right]^{(1-z_{tij})},$$

where $\Gamma(\cdot)$ is the usual Gamma function.

Following the Bayes' theorem, the posterior distribution of $\boldsymbol{\Theta}$, $p(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{z})$, is proportional to the likelihood function times the prior distribution. As we assume independence among the hyperparameters, it follows that

$$p(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{z}) \propto f(\mathbf{y}, \mathbf{z} \mid \boldsymbol{\Theta}) \prod_{k=1}^{3} \left\{ \prod_{i=1}^{I} \prod_{t=1}^{T} \left[p(\boldsymbol{\beta}_{it}^k \mid \boldsymbol{\alpha}_t^k, V_i^k) p(\boldsymbol{\alpha}_t^k \mid \boldsymbol{\alpha}_{t-1}^k, \mathbf{W}^k) \right] \left[\prod_{m=0}^{p_k-1} p(V_{im}^k)\, p(W_m^k) \right] p(\boldsymbol{\alpha}_0^k \mid \mathbf{m}_0^k, C_0^k) \right\},$$

which does not have a closed analytical form. We make use of MCMC methods to obtain samples from the posterior distribution above. In particular we use a hybrid Gibbs sampler with some steps of the Metropolis-Hastings algorithm. The resultant posterior full conditional distributions of $\boldsymbol{\beta}_{it}$ and $\boldsymbol{\delta}_{it}$ do not have a closed form, and are sampled using the Metropolis-Hastings algorithm. In particular, the MCMC algorithm is implemented using the `JAGS` software (Plummer, 2003).

4. Data Analysis

We fit 3 different versions of the proposed model in Section 3 that assume different structures in the mean of the beta distribution, and in the probability of presence θ_{tij}. As mentioned in Section 1.2, we had available information on different schools' characteristics (covariates) through the census data made available by *INEP*. The covariates that enter in the models fitted below were chosen after performing a thorough exploratory data analysis based on independent fits of beta regression models per year and educational level. All 3 models consider the same set of covariates for the different parameters, such that $\mathbf{X}^1_{tij} = (1, ADM, HDI, LIB, LAB, NEL)'_{tij}$, $\mathbf{X}^2_{tij} = (1, ADM, HDI, LIB, LAB, BOYS, NEL)'_{tij}$, and $\mathbf{X}^3_{tij} = (1, nstudent)'_{tij}$, where 1 represents an intercept, which captures a yearly-varying level for each of the parameters, θ_{tij}, μ_{tij}, and ϕ_{tij}. The symbols for the covariates represent the following:

- ADM is a dummy covariate equals 1 if the school is under the federal administration, and 0 otherwise;
- HDI is the standardized human development index of the municipality the school belongs to;
- LIB is a dummy variable being 1 if the school has a library, and 0 otherwise;
- LAB is another dummy variable indicating the presence of a laboratory in the school;
- NEL is a dummy variable equals 1 if the school has more than one educational level;
- $BOYS$ is the standardized number of boys present in the second phase of the OBMEP in educational level i of school j in year t;
- $nstudent$ is the number of students in school j, in educational level i, and year t, present in the second phase of the OBMEP.

We assume a dynamic regression model such that in equations (2b) and (2c), we have that $\mathbf{F}'_1 = \mathbf{I}_6$, $\mathbf{F}'_2 = \mathbf{I}_7$, $\mathbf{F}'_3 = \mathbf{I}_2$, and $\mathbf{G}_1 = \mathbf{I}_6$, $\mathbf{G}_2 = \mathbf{I}_7$, $\mathbf{G}_3 = \mathbf{I}_2$, where $\mathbf{I}_p$ is the p-dimensional identity matrix. For the initial instant in time, $\boldsymbol{\alpha}^k_0$, the prior distribution follows a zero mean multivariate normal distribution, with $\mathbf{C}^k_0 = 100\,\mathbf{I}_{p_k}$. Note that because of the structure of $\mathbf{G}_k$, the evolution of the coefficients across time is through a random walk, which is the simplest structure of a DLM. This is because we only have 8 observations in time, therefore we do not have temporal information to capture more complex structures from the data.

Table 2 describes the 3 models we fit to the data. For each model we run two parallel chains starting from different starting points. We let each of the chains run for 45,000 iterations, considered the first 5,000 iterations as burn in, and stored every 40th iteration from each chain to avoid possible autocorrelation among the sampled values. The final sample size is 2,000. Convergence was checked using the diagnostic tools in the `R` package `coda` (Plummer et al., 2006). In particular, we used the criteria proposed by Geweke (1992) and Gelman and Rubin (1992).

Model	$g_1(\theta_{tij})$	$g_2(\mu_{tij})$	$g_3(\phi_{tij})$
M1	$logit(\theta_{tij}) = \mathbf{X}^1_{tij}\boldsymbol{\beta}^1_{ti}$	$logit(\mu_{tij}) = \mathbf{X}^2_{tij}\boldsymbol{\beta}^2_{ti}$	$log(\phi_{tij}) = \mathbf{X}^3_{tij}\boldsymbol{\beta}^3_{ti}$
M2	$logit(\theta_{tij}) = \mathbf{X}^1_{tij}\boldsymbol{\beta}^1_{ti}$	$logit(\mu_{tij}) = \mathbf{X}^2_{tij}\boldsymbol{\beta}^2_{ti} + \delta^2_j$	$log(\phi_{tij}) = \mathbf{X}^3_{tij}\boldsymbol{\beta}^3_{ti}$
M3	$logit(\theta_{tij}) = \mathbf{X}^1_{tij}\boldsymbol{\beta}^1_{ti} + \delta^1_j$	$logit(\mu_{tij}) = \mathbf{X}^2_{tij}\boldsymbol{\beta}^2_{ti} + \delta^2_j$	$log(\phi_{tij}) = \mathbf{X}^3_{tij}\boldsymbol{\beta}^3_{ti}$

Table 2. Summary of the fitted models. All of them consider the same set of covariates in the probability of presence, the mean and precision parameters of the beta distribution, and these are, respectively, $\mathbf{X}^1_{tij} = (1, ADM, HDI, LIB, LAB, NEL)'_{tij}$, $\mathbf{X}^2_{tij} = (1, ADM, HDI, LIB, LAB, BOYS, NEL)'_{tij}$, and $\mathbf{X}^3_{tij} = (1, nstudent_{tij})'$. Index t indicates the year, j the school, and i the educational level.

4.1 Model comparison

In this Section we describe the different model comparison criteria used to compare the different fitted models. In particular we use the deviance information criterion proposed by Spiegelhalter et al. (2002), and three other criteria based on proper scoring rules.

Deviance Information Criterion (DIC) The DIC is a generalization of the AIC based on the posterior distribution of the deviance, $D(\boldsymbol{\Theta}) = -2\log p(\mathbf{y}, \mathbf{z} \mid \boldsymbol{\Theta})$ (Spiegelhalter et al., 2002). More formally, the DIC is defined as

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\mathbf{\Theta}}),$$

where $\overline{D}$ defines the posterior expectation of the deviance, $\overline{D} = E_{\mathbf{\Theta}|\mathbf{y},\mathbf{z}}(D)$, p_D is the effective number of parameters, with $p_D = \overline{D} - D(\overline{\mathbf{\Theta}})$, and $\overline{\mathbf{\Theta}}$ represents the posterior mean of the parameters. $\overline{D}$ might be seen as a goodness of fit measurement, whereas p_D indicates the complexity of the model. Smaller values of DIC indicate better fitting models.

Gneiting and Raftery (2007) consider proper scoring rules for assessing the quality of probabilistic forecasts. Following Gschlößl and Czado (2008), we use the same data for estimation and computation of the scores, as our focus is on understanding the relationship between the schools' performance and the covariates other than prediction. We use three different scoring rules:

Continuous ranked probability score ($CRPS$) For each y_{tij}, the CRPS can be expressed as

$$CRPS(y_{tij}) = E|y_{tij}^{rep} - y_{tij}| - \frac{1}{2}E|y_{tij}^{rep} - \tilde{y}_{tij}^{rep}|,$$

where y_{tij} is the observed average score of the j^{th} school within level i in year t, y_{tij}^{rep} and $\tilde{y}_{tij}^{rep}$ are independent replicates from the posterior predictive distribution of the respective model.

Assuming there is a sample of size L from the posterior distribution of the parameters in the model, we can obtain roughly independent replicates, y_{tij}^{rep} and $\tilde{y}_{tij}^{rep}$, from the respective posterior predictive distribution. The components $E|y_{tij}^{rep} - y_{tij}|$ and $E|y_{tij}^{rep} - \tilde{y}_{tij}^{rep}|$ can be approximated using Monte Carlo integration through $\frac{1}{L}\sum_{l=1}^{L}|y_{tij}^{rep^{(l)}} - y_{tij}|$ and $\frac{1}{L}\sum_{l=1}^{L}|y_{tij}^{rep^{(l)}} - \tilde{y}_{tij}^{rep^{(l)}}|$, and

$$CRPS = \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{n_i} RPS(y_{tij}),$$

where n is the total number of schools across all the years and educational levels. Smaller values of $CRPS$ indicate the best model among the fitted ones. In the implementation of $CRPS$ for the mixture model we considered in its computation only sampled values that returned $Z_{tij} = 1$. This avoids the use of the missing values, which occur when schools are not present in the second phase of OBMEP.

Logarithmic score ($LogS$) The logarithmic score is defined as $-\log p(y_{tij}, z_{tij})$, where $p(y_{tij}, z_{tij})$ is the probability density function at the observed average score of school j in the i^{th} level and year t. Considering the observed sample $\mathbf{y}$, $LogS$ is computed as

$$LogS = \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{n_i} -\log\, p(y_{tij}, z_{tij}),$$

where n is the total number of schools across all the years and educational levels. Smaller values of $LogS$ indicate the best model among the fitted ones.

Assuming there is a sample from the posterior distribution of the parameters of size L available, the predictive distribution $p(y_{tij})$ is approximated using Monte Carlo integration, that is,

$$p(y_{tij}, z_{tij}) = \int_{\boldsymbol{\Theta}} p(y_{tij}, z_{tij} \mid \boldsymbol{\Theta}) p(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{z}) d\boldsymbol{\Theta} \approx \frac{1}{L} \sum_{l=1}^{L} p(y_{tij}, z_{tij} \mid \boldsymbol{\Theta}^{(l)}),$$

where $p(y_{tij}, z_{tij} \mid \boldsymbol{\Theta}^{(l)})$ is the probability density function shown in equation (1), conditioned on the respective l^{th} sampled value of the parameter vector $\boldsymbol{\Theta}$, evaluated at y_{tij} and z_{tij}.

Dawid-Sebastiani score (DSS) This scoring rule is defined as

$$DSS(y_{tij}) = \left(\frac{y_{tij} - \mu_{p_{tij}}}{\sigma_{p_{tij}}} \right)^2 + 2 \log \sigma_{p_{tij}},$$

where $\mu_{p_{tij}}$ and $\sigma^2_{p_{tij}}$ denote the mean and variance of the predictive distribution $p(y_{tij}, z_{tij})$ for year t, school j and educational level i.

Assuming there is available a sample of size L from the posterior distribution of the parameter vector $\boldsymbol{\Theta}$, samples from the posterior predictive distribution for each school can be obtained, and these provide an estimate of $\mu_{p_{tij}}$ and $\sigma_{p_{tij}}$, say $\hat{\mu}_{p_{tij}}$ and $\hat{\sigma}_{p_{tij}}$. Then $DSS(y_{tij}) = \left(\frac{y_{tij} - \hat{\mu}_{p_{tij}}}{\hat{\sigma}_{p_{tij}}} \right)^2 + 2 \log \hat{\sigma}_{p_{tij}}$, and

$$DSS = \frac{1}{n} \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{j=1}^{n_i} DSS(y_{tij}),$$

where n is the total number of schools across all years and educational levels.

For all scoring rules, smaller values indicate better fitting models. See Gschlößl and Czado (2007) and Czado et al. (2009) for further details on the properties and implementation of $CRPS$, $LogS$, DSS, and other criteria based on proper scoring rules.

Table 3 shows the values of the different model comparison criteria obtained under each fitted model. Different criteria point to different models, but in general, they suggest that the inclusion of school random effects in the mean μ_{tij}, and/or in the logit of the probability of presence, θ_{tij}, improves model performance. The results shown next are based on model M3.

Fitted Model	Inclusion of School random effect (ν_j^k)	p_D	DIC	$CRPS$	$LogS$	DSS
M1	-	92.55	219692.1	0.025	-7.34	-5.38
M2	$logit(\mu_{tij})$	1957.90	210004.4	0.033	*-8.81*	-4.68
M3	$logit(\theta_{tij})$, $logit(\mu_{tij})$	3625.36	*203785.8*	*0.025*	-7.96	*-5.43*

Table 3. Model comparison criteria, DIC and its component, p_D, RPS, $LogS$, and DSS, under each fitted model. Numbers in italics indicate best model under the respective criterion.

4.2 Posterior predictive distribution of the scores for unobserved schools

Assume that one wants to perform some sort of cross validation, and forecast the performance of a set of schools in the second phase of OBMEP for year 2013 that were not used in the model fitting step. If the covariates of these set of schools are available, following the Bayesian framework, the prediction is performed by obtaining summaries from the predictive posterior distribution. Let y^{new} be the score of a school in the second phase of OBMEP in 2013 that was not present in the sample. We are interested on the distribution of $p(y^{new} \mid \mathbf{y})$, which is given by

$$p(y^{new}, z^{new} \mid \mathbf{y}) = \int_{\Theta} p(y^{new}, z^{new} \mid \mathbf{\Theta}, \mathbf{y})\ p(\mathbf{\Theta} \mid \mathbf{y}, \mathbf{z})\ d\mathbf{\Theta} = \int_{\Theta} p(y^{new} \mid \mathbf{\Theta})\ p(\mathbf{\Theta} \mid \mathbf{y}, \mathbf{z})\ d\mathbf{\Theta}, \quad (3)$$

as given $\mathbf{\Theta}$, $\mathbf{y}$, $\mathbf{z}$ and y^{new}, z^{new} are independent. Note that $p(y^{new}, z^{new} \mid \mathbf{\Theta}, \mathbf{y})$ is the probability density function shown in equation (1). As we have available a sample of size L from the posterior distribution of $\mathbf{\Theta}$, the posterior predictive density above can be approximated through Monte Carlo integration by assuming $p(y^{new}, z^{new} \mid \mathbf{y}, \mathbf{z}) \approx \frac{1}{L}\sum_{l=1}^{L} p(y^{new}, z^{new} \mid \mathbf{\Theta}^{(l)})$. Samples from the predictive posterior distribution can be generated by drawing samples from each realization of the posterior distribution, that is, from $p(y^{new}, z^{new} \mid \mathbf{\Theta}^{(l)})$. With samples from this distribution we can obtain summaries for the distribution in equation (3). As conditional on the parameter vector $\mathbf{\Theta}$, the scores of different schools are independent, this process can be repeated for a set of different schools.

Under model M3, there are also the school's random effects. When predicting for schools that were left out from the inference procedure, as they are not present in the likelihood function, their posterior distribution are equal to their respective prior distribution. Then, samples from these school's random effects, that enter in the equations for $\theta_{j_{new}}$ and $\mu_{j_{new}}$, are obtained by sampling, for each iteration l, from $\delta^{1^{(l)}}_{j_{new}} \sim N(0, \sigma_1^{2^{(l)}})$, and $\delta^{2^{(l)}}_{j_{new}} \sim N(0, \sigma_2^{2^{(l)}})$, respectively; note that $\sigma_1^{2^{(l)}}$ and $\sigma_2^{2^{(l)}}$ are samples from the posterior distribution of σ_1^2 and σ_2^2, respectively.

Panels of Figure 2 show the summary (mean and 95% credible intervals) of the posterior predictive distribution of the scores for 90 schools that were left out from the inference procedure under models M3, and all took part in the second phase of OBMEP in 2013. These schools were randomly chosen among the different administrative and educational levels. We picked 10 schools in each combination of administrative and educational levels. Among the 90 schools, for only 6, the observed value of the score did not fall within the 95% posterior predictive credible intervals.

4.3 Analysis of the results

Panels of Figure 3 show the posterior summary of the coefficients of the covariate effects (rows) present in the equation for the probability of presence in the 2nd phase of OBMEP, $log\frac{\theta_{tij}}{1-\theta_{tij}}$, under the different educational levels (columns) and across the years. The estimated values of the intercept across time for levels 1 and 2 show a smoother behavior when compared to the one for level 3. The 95% posterior credible interval of the coefficient of HDI in level 3 includes 0 across the years, indicating that for level 3, HDI does not affect the logit of the probability of presence of a school in the second phase. On the other hand, for levels 1 and 2, the coefficient of HDI is estimated at around -1, suggesting that one standard deviation increase in HDI, increases the probability of presence of a school in the second phase by approximately 0.27(=exp(-1)/(1 + exp(-1)) if all the other covariates are held fixed. The coefficients of ADM are estimated at high values indicating that federal schools have the probability of presence greatly increased in the second phase

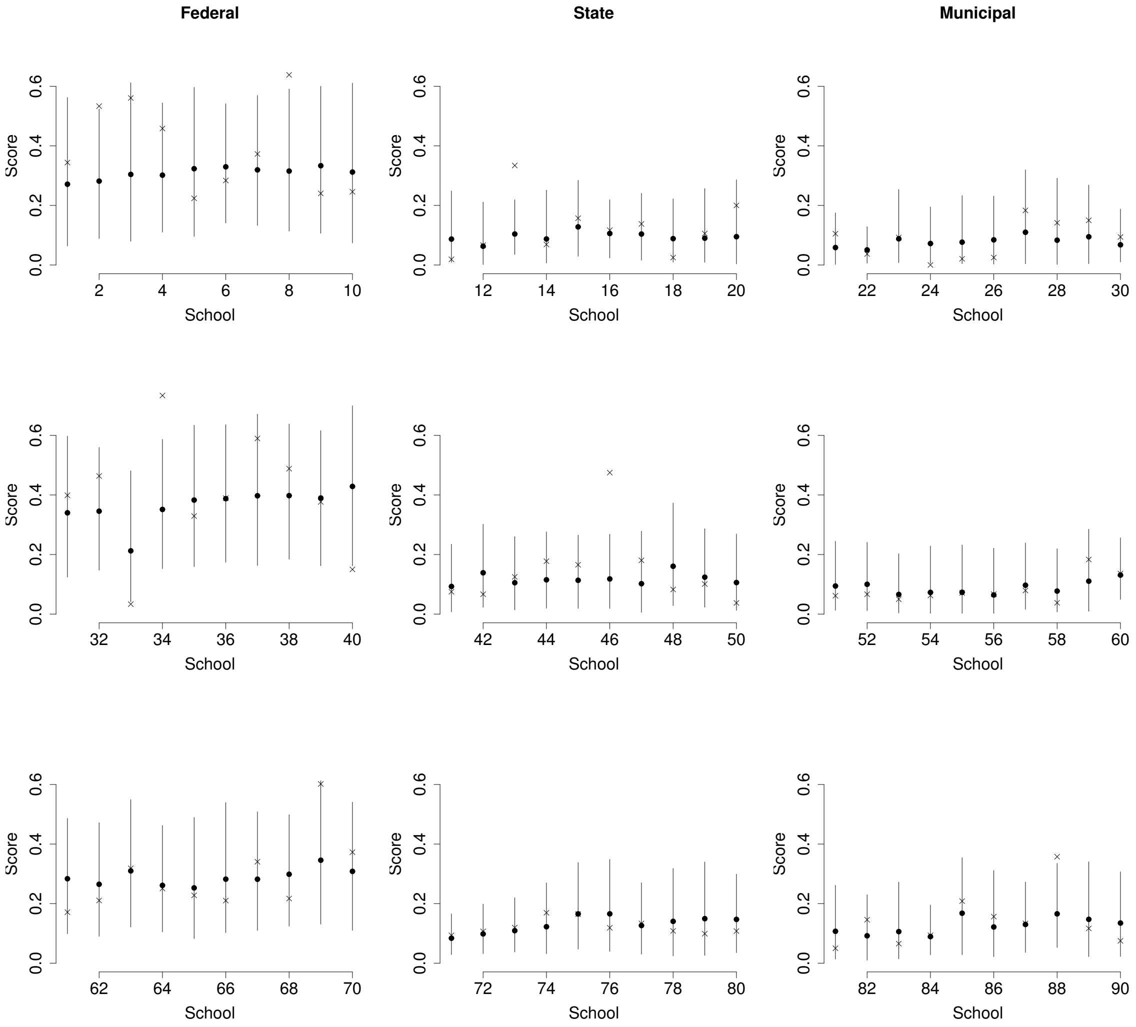

Figure 2. Posterior summary (mean: solid circle, and limits of the 95% credible intervals, under model M3): solid lines of the predictive distribution of the scores of 90 schools that actually took part in the 2nd phase in 2013, and were left out from the inference procedure for predictive purposes. The crosses represent the observed scores. The results are shown by administrative (columns) and educational levels (rows 1, 2 and 3).

of OBMEP when compared to state or municipal schools. Also, the presence of a library or a laboratory, increases the probability of presence of a school in any year in the second phase of OBMEP. The 95% posterior credible interval of the coefficient for NEL includes zero for all educational levels and years. This suggests that the presence of more than one educational level in a school does not influence the probability of presence in the second phase, after adjusting for all the other covariates.

Figure 4 shows the posterior summary of the coefficients of the covariates in the mean of the beta distribution, μ_{tij}. The covariate with the highest effect on the $logit\, \mu_{tij}$ is the administrative level (ADM) followed by HDI, LIB, LAB. That is, federal schools have their mean scores increased across the years. HDI has a positive effect for all the years and educational levels. The coefficient of the proportion of boys (BOYS) is estimated at relatively small values, being different from zero for some years and educational levels. The posterior summary of the coefficient for the presence of more than one educational level in a school (NEL) includes zero for all years and educational levels, suggesting this covariate is not important to explain the $logit\, \mu_{tij}$.

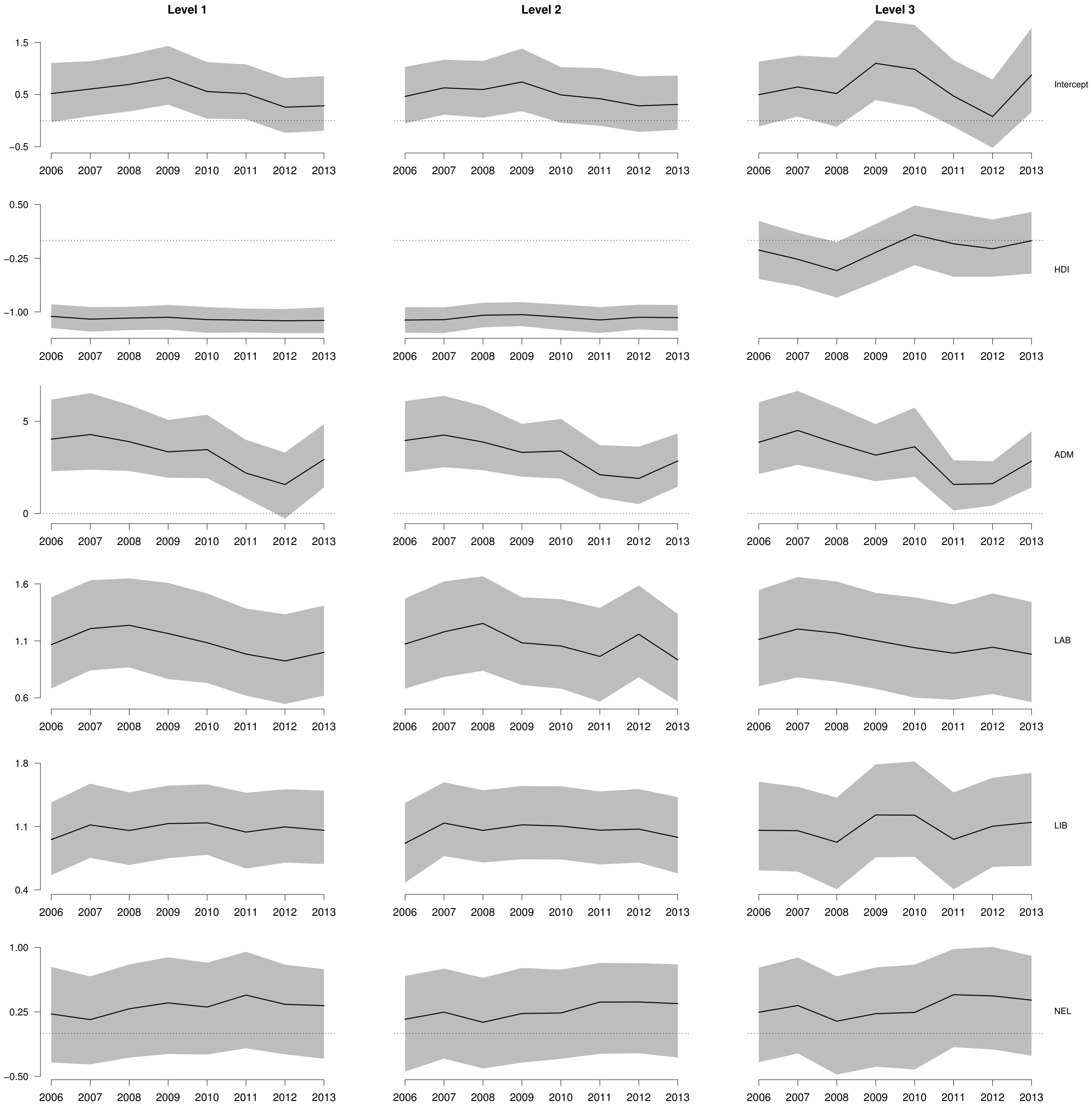

Figure 3. Posterior summary (mean: solid line, and limits of the 95% credible intervals: shaded area) of the coefficients β^{1}_{lit}, for the intercept, HDI, ADM, LAB, LIB, and NEL (rows) by educational level (columns 1, 2 and 3) $l = 0, 1, 2, 3, 4, 5$. The horizontal dotted line when present, represents the value 0.

Panels of Figure 5 show the posterior summary of the coefficients of the covariates in the logarithm of the precision ϕ_{tij}, grouped by educational level (columns 1 to 3), together with the overall effect $\boldsymbol{\alpha}^{3}_{t}$ (4th column). Clearly, the number of students in the second phase of OBMEP has a positive effect on the logarithm of the precision for all years and educational levels. Also, regardless of the educational level, the precision of y_{tij} is estimated at high values, as the intercept is estimated at relatively high values.

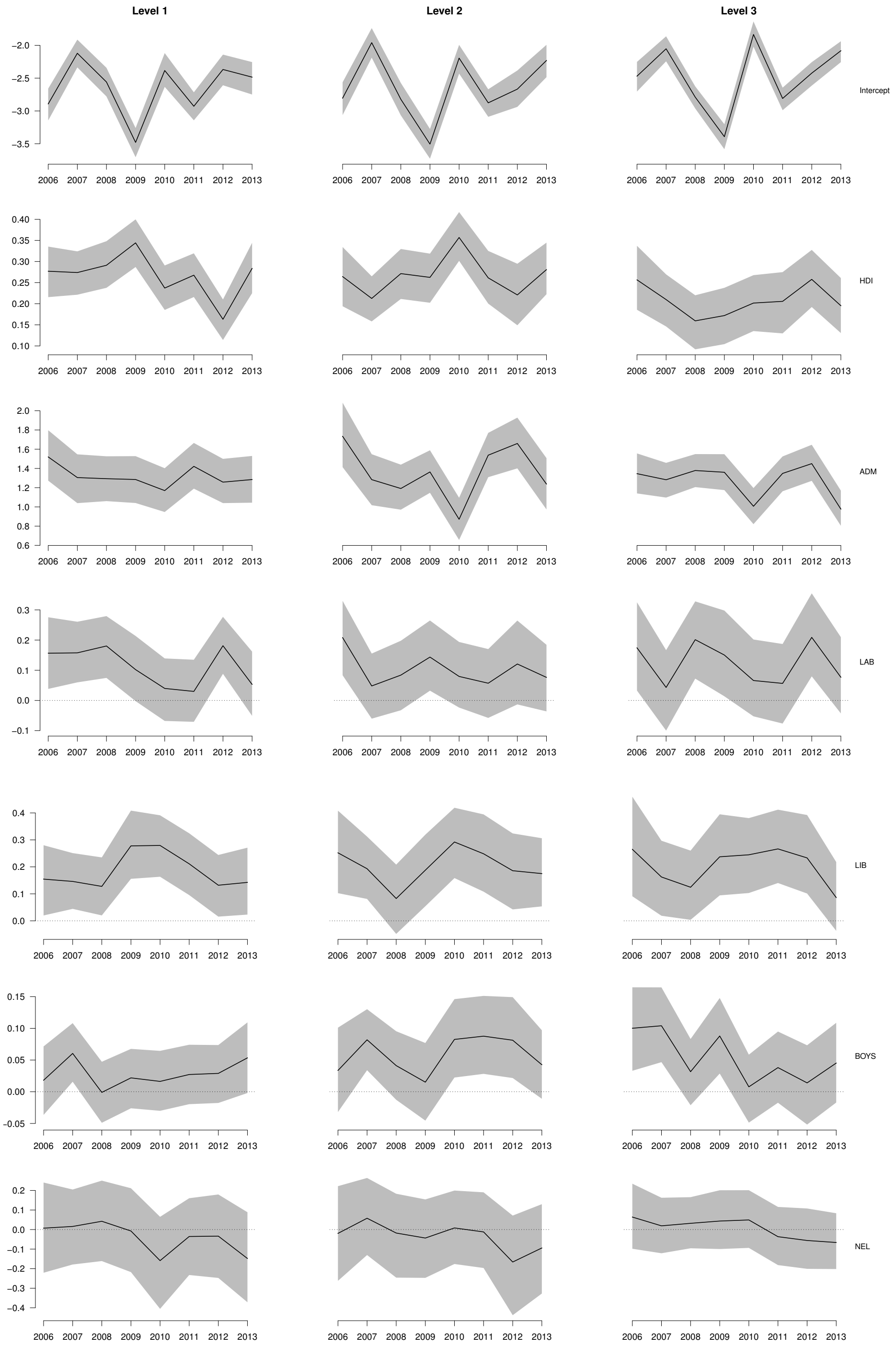

Figure 4. Posterior summary (mean: solid line, and limits of the 95% credible intervals: shaded area) of the coefficients β^2_{lit}, of the covariates (intercept, HDI, ADM, LAB, LIB, BOYS, and NEL, rows) in the mean of y_{tij}, by educational level (columns 1, 2 and 3) $l = 0, 1, 2, 3, 4, 5, 6$. The horizontal dotted line when present, represents the value 0.

Panels of Figure 6 show the posterior summary of the coefficients of the covariates in the lower level of hierarchy for the probability of presence in the second phase, $\boldsymbol{\alpha}_t^1$ (see equation (2)). HDI has an overall negative effect, whereas ADM, LAB, and LIB have a positive effect. And, for all years, 0 falls within the 95% posterior credible interval of the coefficient for NEL, suggesting that it does not make a difference in the probability of presence if a school has available more than one educational level or not.

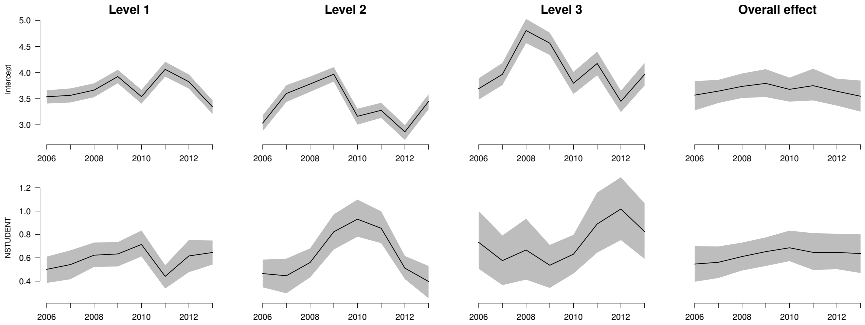

Figure 5. Posterior summary (mean: solid line, and limits of the 95% credible intervals: shaded area) of the coefficients of the covariates in the precision, $\boldsymbol{\beta}^3_{ti}$, the intercept, *nstudent*. The horizontal dotted line when present, represents the value 0.

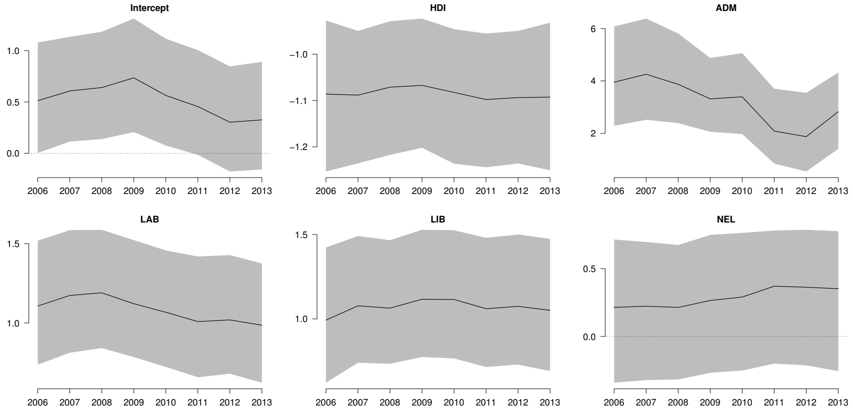

Figure 6. Posterior summary (mean: solid line, and limits of the 95% credible intervals: shaded area) of the overall coefficients ($\boldsymbol{\alpha}_t^1$) in the probability of presence, for the intercept, HDI, ADM, LAB, LIB, and NEL. The horizontal dotted line when present, represents the value 0.

Panels of Figure 7 show the posterior summary of the coefficients of the covariates in the lower level of hierarchy for the mean, coefficients $\boldsymbol{\alpha}_t^2$. Clearly the coefficients for ADM, HDI, LIB, assume positive values across the years, whereas the posterior summary of the coefficient for LAB includes zero for 2010 and 2012. On the other hand, the posterior summary for the coefficients of $BOYS$ and NEL include zero for all years.

Panels of Figure 8 show the posterior summary of the fitted values (mean: black solid line; ranges of the 95% posterior credible interval: dotted lines) together with the posterior summary (mean: gray cross, 95% credible interval: gray vertical solid line) of the probability of presence for 9 schools, chosen to depict specific situations observed in the sample. The solid circle in each panel represents the observed value when the school actually took part in the second phase of the OBMEP. If the solid circle is missing for a particular school in a year, then that school did not take part in the second phase of OBMEP. Overall, it is clear that the probability of presence is able to capture the presence of the school in the second phase. In general, the probability of presence correctly decreases if the school did not take part in the second phase of a given year. Also, the fitted values are close to the observed ones. This panel makes it clear that even if the school did not take part in the second phase of OBMEP, our model provides an estimate of its performance (see, e.g., panel for School 963). This is because the summaries of the predictive distribution are computed based on the posterior samples for which we obtained $z^{l}_{tij} = 1$, with l representing the l-th sampled value of $z^{(l)}_{tij}$.

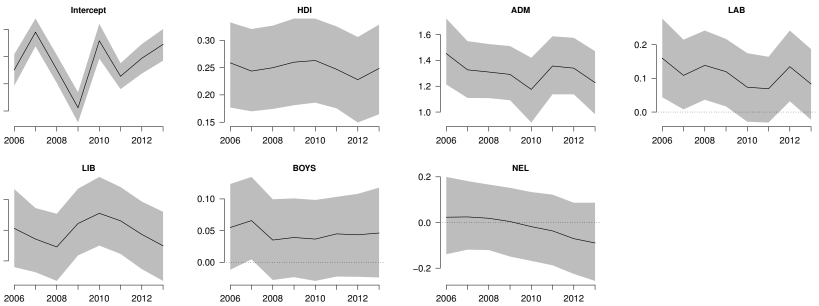

Figure 7. Posterior summary (mean: solid line, and limits of the 95% credible intervals: shaded area) of the overall coefficients ($\boldsymbol{\alpha}^2_t$), for the intercept, HDI, ADM, LAB, LIB, BOYS, and NEL.

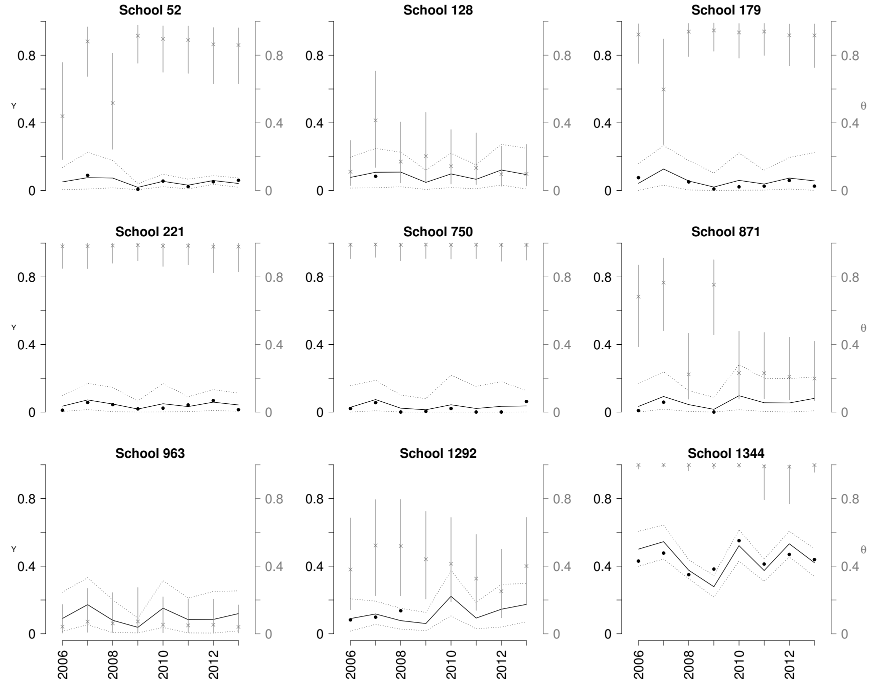

Figure 8. Posterior summary (mean: black solid line, and limits of the 95% credible intervals: dotted lines) of the fitted values of the observed scores (solid circles), together with the posterior summary of the probability of presence (mean: gray cross, and limits of the 95% credible intervals: gray vertical lines) of 9 schools in the second phase of OBMEP. When a solid circle is missing for a particular school, within a particular year, it is because that school did not take part in the second phase of OBMEP.

5. Discussion and future work

This paper proposes a hierarchical dynamic beta mixture model for the analysis of the performance of schools across eight editions of the second phase of the OBMEP, from 2006 until 2013. As not necessarily all the schools in the sample took part in the second phase of the OBMEP across these years, we propose a mixture dynamic hierarchical model that allows to estimate the probability of presence of a school in year t. And given that the school is present, it provides estimates of the mean and precision of the distribution of the score for a particular year and educational level. All the coefficients of the probability of presence, the mean and precision of the underlying distribution of the score are allowed to evolve with time according to a hierarchical dynamic model.

We fit different versions of the proposed model. Inference procedure is performed under the Bayesian paradigm and uncertainty about parameters' estimates are obtained in a straightforward fashion. A sample from the posterior distribution of the parameters was obtained through MCMC. Implementation of the MCMC algorithm was done using the software `JAGS`. Model comparison criteria suggest model M3 (which includes school's random effects both in the probability of presence and in the mean) is the best among those fitted.

Important conclusions are drawn from this study. Overall, the mean performance of schools tend to be low. Results show that, in general, federal schools perform better than state or municipal ones. In general, the difference between federal, and state and municipal schools tend to be greater in the first and second levels of the OBMEP. The human development index (HDI) also has a positive effect on the mean of the scores, but smaller than that of the administrative level. This suggests that federal schools in regions with higher values of HDI tend to perform slightly better than those with smaller values of the HDI, when all the other covariates are held fixed. The proportion of boys present in the second phase of the OBMEP has a very small positive effect for some years for all educational levels. The possible difference in performance of boys and girls in mathematics exams has been the object of interest in different studies, see e.g. Hyde and Mertz (2009), Liu (2009), and references therein. The analysis of the results of PISA 2009 show that boys outperformed girls in mathematics in 35 out of the 65 countries and economies that took part in PISA 2009. On the other hand, for 25 countries no significant difference was observed between the genders, whereas for 5 countries girls outperformed boys in the mathematics exam of PISA 2009 (OECD, 2011).

Our model naturally provides forecasts of the scores for schools that were not included in the sample. Under the Bayesian framework this is done through the posterior predictive distribution, and as shown in equation (3) it naturally accounts for the uncertainty in estimating the parameter vector $\boldsymbol{\Theta}$. Although it is not our interest to perform any prediction, we showed the results of the prediction for the scores of 90 schools that were not included in the model fitting step but were actually present in the second phase of OBMEP. The results were quite satisfactory, as only approximately 7% of the schools did not have their observed score falling within their respective 95% posterior predictive credible interval (Figure 2).

Note that the coefficients of the covariates in the mean and precision of the beta distribution (Figures 4 and 5) do not have a smooth pattern across the years. This is probably happening because we do not have any information on the level of difficulty of the different exams across the years. It would have been better if organizers of the OBMEP used some tool from item response theory (van der Linden and Hambleton, 2013) to standardize the level of difficulty of the exams across years. This is an issue that should be tackled in the next editions of the OBMEP.

Our current interest is to investigate what kind of impact the OBMEP has on the

Brazilian educational system. Biondi et al. (2012) quantify the effects of the 2007 edition of the OBMEP on the average math scores of the ninth-graders participating in *Prova Brasil*, which is a national exam applied by *INEP* to all Brazilian students in the 8^{th} and 9^{th} grades of publich schools. We plan to focus on students in the last year of high school. Considering different years we plan to use causal inference and propensity score methods (Hirano and Imbens, 2004) to investigate the effect of the OBMEP on the performance of students in different editions of the High School Brazilian National Exam (*Exame Nacional do Ensino Médio*, ENEM). Every year, results of the ENEM are used by nearly 500 universities in Brazil as a selection criterion for admission to higher education.

Acknowledgements

This work is part of Moraes' M.Sc. Dissertartion under the supervision of A.M. Schmidt and H. S. Migon. The authors thank the financial support from CNPq (Schmidt and Migon) and FAPERJ (Schmidt), Brazil, and CAPES, Brazil (scholarship to Moraes). The authors also thank Professor Claudio Landim (IMPA, Brazil), Monica Souza (OBMEP) and Luiz L. R. da Conceição (OBMEP) for introducing the problem and making the dataset available.

References

Bayes, C. L., Bazán, J. L., García, C., 2012. A new robust regression model for proportions. Bayesian Analysis 7 (4), 841–866.

Biondi, R. L., Vasconcellos, L., Menezes-Filho, N., 2012. Evaluating the impact of the Brazilian Public School Math Olympics on the quality of education. Economía 12, 143–175.

Branscum, A. J., Johnson, W. O., Thurmond, M. C., 2007. Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. Australian New Zealand Journal of Statistics 49, 287–301.

Campbell, J. R., Walberg, H. J., 2010. Olympiad studies: Competitions provide alternatives to developing talents that serve national interests. Roeper Review 33 (1), 8–17.

Cepeda-Cuervo, E., Núñez Antón, V., 2013. Spatial double generalized beta regression models: extensions and applications to study quality of education in Colombia. Journal of Educational and Behavioral Statistics 38 (6), 604–628.

Czado, C., Gneiting, T., Held, L., 2009. Predictive model assessment for count data. Biometrics 65, 1254–1261.

Da-Silva, C. Q., Migon, H. S., 2016. Hierarchical dynamic beta model. Revstat Statistical Journal 14 (1), 49–73.

Da-Silva, C. Q., Migon, H. S., Correia, L. T., 2011. Dynamic Bayesian beta models. Computational Statistics & Data Analysis 55 (6), 2074–2089.

Dalenius, T., Hodges, J. L. J., 1959. Minimum variance stratification. Journal of the American Statistical Association 54, 88–101.

Fernandes, M. V., Schmidt, A. M., Migon, H. S., 2009. Modelling zero-inflated spatio-temporal processes. Statistical Modelling 9 (1), 3–25.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics 31 (7), 799–815.

Gamerman, D., Migon, H. S., 1993. Dynamic hierarchical models. Journal of the Royal Statistical Society, Series B 55 (3), 629–642.

Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science, 457–472.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), Bayesian Statistics 4 : Proceedings of the Fourth Valencia Meeting. Oxford, UK, pp. 169–193.

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction and estimation. Journal of the American Statistical Association 102, 359–378.

Gschlößl, S., Czado, C., 2007. Spatial modelling of claim frequency and claim size in non-life insurance. Scandinavian Actuarial Journal 3, 202–225.

Gschlößl, S., Czado, C., 2008. Modelling count data with overdispersion and spatial effects. Statistical Papers 49, 531–552.

Hirano, K., Imbens, G. W., 2004. The propensity score with continuous treatments. In: Applied Bayesian modeling and causal inference from incomplete-data perspectives. Gelman, A. and Meng, X. L.(eds). Wiley, Oxford, U.K., pp. 73–84.

Hyde, J. S., Mertz, J. E. a., 2009. Gender, culture, and mathematics performance. Proceedings of the National Academy of Sciences of the United States of America 106, 8801–8807.

Liu, O. L., 2009. An investigation of factors affecting gender differences in standardized math performance: Results from U.S. and Hong Kong 15 year olds. International Journal of Testing 9 (3), 215–237.

Losada, M., Rejali, A., 2015. The role of mathematical competitions and other challenging contexts in the teaching and learning of mathematics. In: The Proceedings of the 12th International Congress on Mathematical Education. Springer, pp. 563–568.

Lunn, D. J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing 10 (4), 325–337.

OECD, 2011. How do girls compare to boys in mathematics skills? In: PISA 2009 at a Glance. OECD Publishing, Paris, France.
URL :http://dx.doi.org/10.1787/9789264095250-8-en

Paolino, P., 2001. Maximum likelihood estimation of models with beta-distributed dependent variables. Political Analysis 9, 325–346.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. Vol. 124. Vienna, p. 125.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News 6 (1), 7–11.
URL :http://CRAN.R-project.org/doc/Rnews/

Smithson, M., Merkle, E. C., Verkuilen, J., 2011. Beta regression finite mixture models of polarization and priming. Journal of Educational and Behavioral Statistics 36 (6), 804–831.

Smithson, M., Verkuilen, J., 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological methods 11 (1), 54–71.

Spiegelhalter, D., Best, N., Carlin, B., Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 583–639.

Thompson, S. K., 2012. Sampling. Wiley Series in Probability and Statistics, 3rd edition.

van der Linden, W. J., Hambleton, R. K., 2013. Handbook of modern item response theory. Springer Science & Business Media.

Verkuilen, J., Smithson, M., 2012. Mixed and mixture regression models for continuous bounded responses using the beta distribution. Journal of Educational and Behavioral Statistics 37 (1), 82–113.

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, 2nd Edition. Springer Series in Statistics.

Zimprich, D., 2010. Modeling change in skewed variables using mixed beta regression models. Research in Human Development 7 (1), 9–26.

Appendix A. Directed acyclic graph of the proposed model

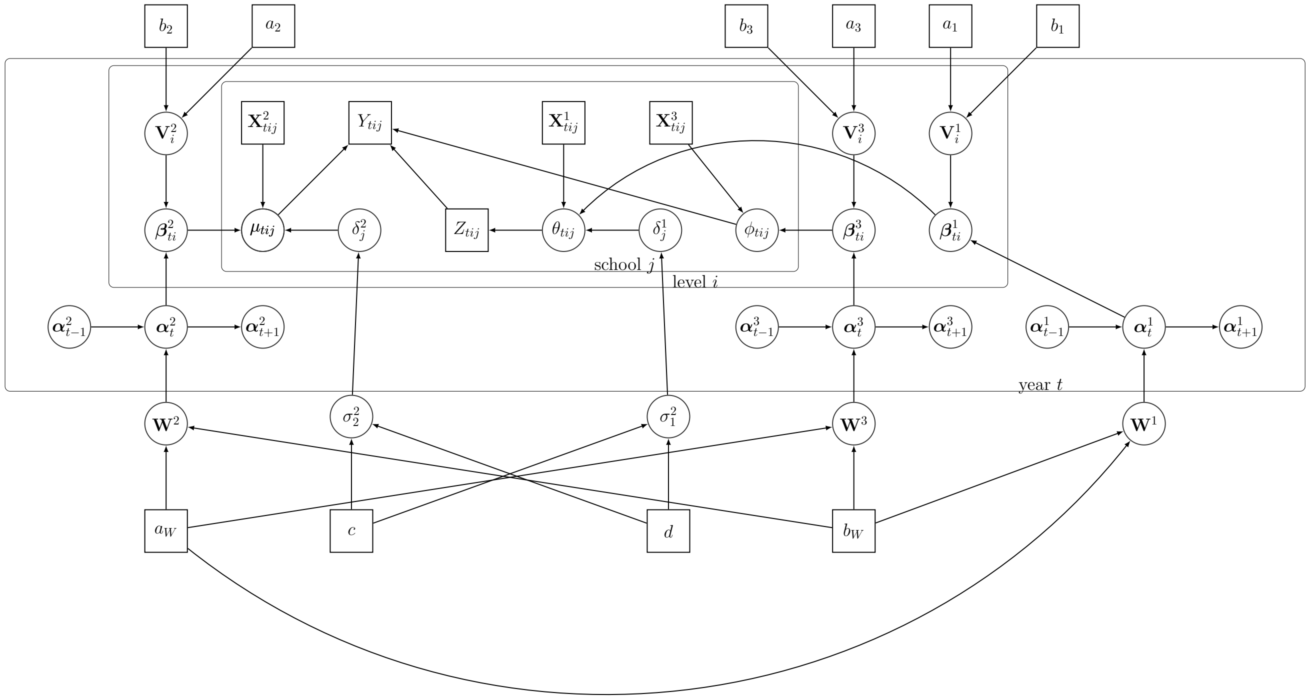

Figure A1. Directed acyclic graph of the hierarchical model proposed in equations (1) and (2), without school random effect in the precision equation of $p(.|\mu_{tij}, \phi_{tij})$, that is $\delta_j^3 = 0, \forall j$.