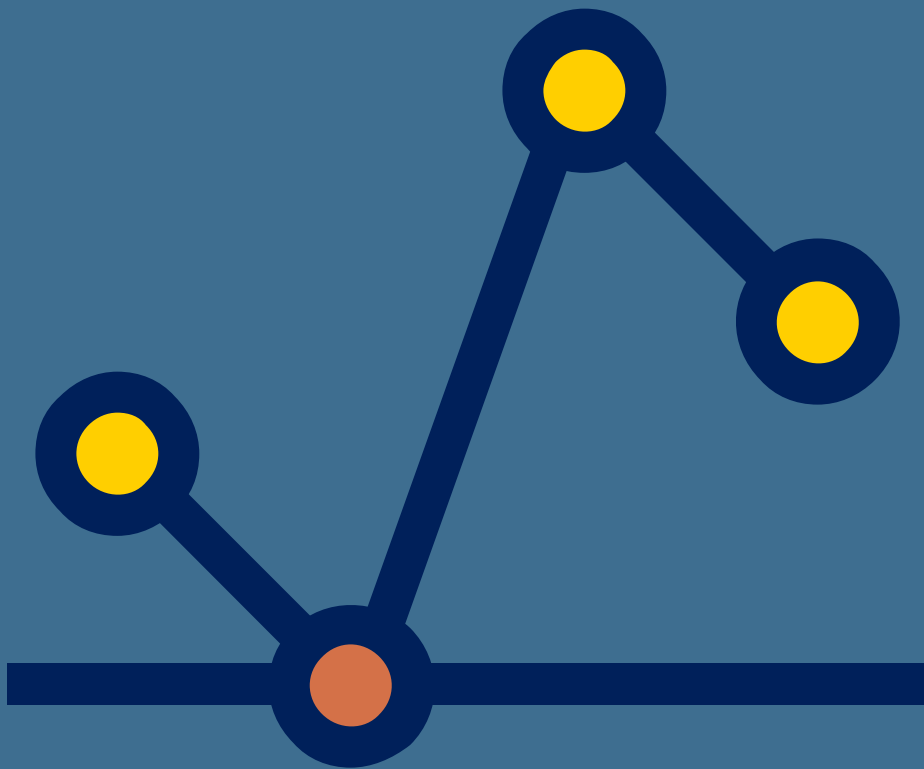


---

Edited by  
Paola Cerchiello · Arianna Agosto  
Silvia Osmetti · Alessandro Spelta

# Proceedings of the Statistics and Data Science Conference



PaviaUniversityPress



*Copertina*: Cristina Bernasconi, Milano

Copyright © 2023 EGEA S.p.A.  
Via Salasco, 5 - 20136 Milano  
Tel. 02/5836.5751 - Fax 02/5836.5753  
egea.edizioni@unibocconi.it - www.egeaeditore.it

Quest'opera è rilasciata nei termini della Creative Commons Attribution 4.0 International Licence (CC BY-NC-SA 4.0), eccetto dove diversamente indicato, che impone l'attribuzione della paternità dell'opera e ne esclude l'utilizzo a scopi commerciali. Sono consentite le opere derivate purché si applichi una licenza identica all'originale. Il testo completo è disponibile alla pagina web <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it>.

Date le caratteristiche di Internet, l'Editore non è responsabile per eventuali variazioni di indirizzi e contenuti dei siti Internet menzionati.

Pavia University Press  
info@paviauniversitypress.it – www.paviauniversitypress.it

Prima edizione: maggio 2023  
ISBN volume 978-88-6952-170-6

## Preface

The development of large-scale data analysis and statistical learning methods for data science is gaining more and more interest, not only among statisticians, but also among computer scientists, mathematicians, computational physicists, economists, and, in general, all experts in different fields of knowledge who are interested in extracting insight from data.

Cross-fertilization between the different scientific communities is becoming crucial for progressing and developing new methods and tools in data science.

In this respect, the Statistics & Data Science group of the Italian Statistical Society has organized an international conference held in Pavia on the 27 and 28 of April 2023, attended by over 70 researchers from different scientific fields.

A collection of the presented papers is available in the present Proceedings showing a huge variety of approaches, methods, and data-driven problems, always tackled according to a rigorous and robust scientific paradigm.

The Statistics & Data Science group



# Contents

<b>Fractional random weight bootstrap in presence of asymmetric link functions</b> . . . . .	1
La Rocca Michele, Niglio Marcella, Restaino Marialuisa	
<b>Innovation patterns within a regional economy through consensus community detection on labour market network</b> . . . . .	6
Morea Fabio, De Stefano Domenico	
<b>Sparse Inference in functional conditional Gaussian Graphical Models under Partial Separability</b> . . . . .	12
Fici Rita, Sottile Gianluca , Augugliaro Luigi	
<b>A Conformal Approach to Model Explainability</b> . . . . .	18
Mata Naranjo Juan, Brutti Pierpaolo	
<b>A S.A.F.E. approach for Sustainable, Accurate, Fair and Explainable Machine Learning Models</b> . . . . .	24
Raffinetti Emanuela , Giudici Paolo	
<b>Do we really care about data ethics?</b> . . . . .	30
Ferrara Alfio	
<b>Ethical concepts of data ethics between public and private interests</b> . . . . .	36
Durante Massimo	
<b>Being a statistician in the big data era: A controversial role?</b> . . . . .	42
Manzi Giancarlo	
<b>Forecasting relative humidity using LoRaWAN indicators and autoregressive moving average approaches</b> . . . . .	47
Rojas Guerra Renata, Vizziello Anna, Gamba Paolo	

<b>Interpretability of Machine Learning algorithms: how these techniques can correctly guess the physical laws?</b> .....	53
De Corato Marzio, Ferrara Alfio, Salini Silvia	
<b>The role of BERT in Neural Network sentiment scoring for Time Series Forecast</b> .....	55
Basili Roberto, Croce Danilo, Iezzi Domenica, Monte Roberto	
<b>Diagnostics for topic modelling. The dubious joys of making quantitative decisions in a qualitative environment</b> .....	61
Sciandra Andrea, Trevisani Matilde, Tuzzi Arjuna	
<b>Mapping the thematic structure of Data Science literature with an embedding strategy</b> .....	67
Irpino Antonio, Misuraca Michelangelo, Giordano Giuseppe	
<b>Critical Visual Explanations. On the Use of Example-Based Strategies for Explaining Artificial Intelligence to Laypersons</b> .....	73
Gobbo Beatrice	
<b>Visualising unstructured social media data: a chart-based approach</b> .....	77
Aversa Elena	
<b>From teaching Statistics to designers to teaching Statistics through design</b> .....	85
Mauri Michele, Vantini Simone	
<b>Forecasting Spatio-Temporal Data with Bayesian Neural Networks</b> .....	90
Ravenda Federico, Cesarini Mirko, Peluso Stefano, Mira Antonietta	
<b>Oracle-LSTM: a neural network approach to mixed frequency time series prediction</b> .....	96
Bitetto Alessandro, Cerchiello Paola	
<b>Streamlined Variational Inference for Modeling Italian Educational Data</b>	102
Gioia Di Credico, Claudia Di Caterina, Francesco Santelli	
<b>The use of magnetic resonance images for the detection and classification of brain cancers with D-CNN</b> .....	108
Mascolo Davide, Plini Leonardo, Pecchini Alessandro, Antonicelli Margaret	
<b>Modeling and clustering of traffic flows time series in a flood prone area</b> .	113
Zuccolotto Paola, De Luca Giovanni, Metulini Rodolfo, Carpita Murizio	
<b>Global mobility trends from smartphone app data. The MobMeter dataset</b> .....	119
Finazzi Francesco	

Contents

<b>Spatio-temporal statistical analyses for risk evaluation using big data from mobile phone network</b> . . . . .	124
Perazzini Selene, Metulini Rodolfo, Carpita Maurizio	
<b>A Robust Approach to Profile Monitoring</b> . . . . .	130
Capezza Christian, Centofanti Fabio, Lepore Antonio, Palumbo Biagio	
<b>The FDA contribution to Health Data Science</b> . . . . .	133
Ieva Francesca	
<b>A new topological weighted functional regression model to analyse wireless sensor data</b> . . . . .	139
Romano Elvira, Irpino Antonio, Andrea Diana	
<b>Clustering for rotation-valued functional data</b> . . . . .	145
Stamm Aymeric, Bellanger Lise	
<b>Giudici Paolo InstanceSHAP: An instance-based estimation approach for Shapley values</b> . . . . .	151
Babei Golnoosh, Giudici Paolo	
<b>A new paradigm for Artificial Intelligence based on Group Equivariant Non-Expansive Operators (GENEOs) applied to protein pocket detection</b> . . . . .	152
Bocchi Giovanni, Micheletti Alessandra, Frosini Patrizio, Pedretti Alessandro, Gratteri Carmen, Lunghini Filippo, Beccari Andrea Rosario, Talarico Carmine	
<b>Clustering Italian medical texts: a case study on referrals</b> . . . . .	158
Torri Vittorio, Ercolanoni Michele, Bortolan Francesco, Leoni Olivia, Ieva Francesca	
<b>Classification of Recommender systems using Deep Learning based generative models</b> . . . . .	164
Filali-Zegzouti Sanae, Banouar Oumayma, Benslimane Mohamed	
<b>Sparse Inference in Gaussian Graphical Models via Adaptive Non-Convex Penalty Function</b> . . . . .	170
Cuntrera Daniele, Muggeo Vito M.R., Augugliaro Luigi	
<b>Bayesian causal inference from discrete networks</b> . . . . .	177
Castelletti Federico, Consonni Guido	
<b>Sign-Flip tests for Spatial Regression with PDE regularization</b> . . . . .	182
Cavazzutti Michele, Arnone Eleonora, Ferraccioli Federico, Finos Livio, Sangalli Laura M.	
<b>A novel sequential testing procedure for selecting the number of changepoints in segmented regression models</b> . . . . .	187
Priulla Andrea, D'Angelo Nicoletta	

<b>On the numerical stability of the efficient frontier</b> . . . . .	193
Fassino Claudia, Uberti Pierpaolo	
<b>Spatial regression with differential regularization over linear networks</b> . .	196
Clemente Aldo, Arnone Eleonora, Mateu Jorge, Sangalli Laura M.	
<b>An Estimation Tool for Spatio-Temporal Events over Curved Surfaces</b> . . .	201
Panzeri Simone, Begu Blerta, Arnone Eleonora, Sangalli Laura M.	
<b>Gromov-Wasserstein barycenters for optimal portfolio allocation</b> . . . . .	207
Spelta Alessandro, Pecora Nicolò, Maggi Mario	
<b>Online Job Advertisements: toward the quality assessment of classification algorithms for the occupation and the activity sector</b> . . . . .	214
Catanese Elena, Inglese Francesca, Lucarelli Annalisa, Righi Alessandra, Ruocco Giuseppina	
<b>Linear Programming for Wasserstein Barycenters</b> . . . . .	220
Auricchio Gennaro, Bassetti Federico, Gualandi Stefano, Veneroni Marco	
<b>A multi-channel convolution approach for forecast reconciliation</b> . . . . .	224
Marcocchia Andrea, Arima Serena, Brutti Pierpaolo	
<b>Hedging global currency risk with factorial machine learning models</b> . . . .	230
Giudici Paolo, Pagnottoni Paolo, Spelta Alessandro	
<b>Predicting musical genres from Spotify data by statistical machine learning</b> . . . . .	236
Biazzo Federica, Farné Matteo	
<b>The use of Bradley-Terry comparisons in statistical and machine learning models to predict football results</b> . . . . .	242
Macri Demartino Roberto, Torelli Nicola, Egidio Leonardo	
<b>A new approach for quantum phase estimation based algorithms for machine learning</b> . . . . .	248
Ouedrhiri Oumayma, Banouar Oumayma, El Hadaj Salah, Raghay Said	
<b>A comparison of ensemble algorithms for item-weighted Label Ranking</b> .	254
Albano Alessandro, Sciandra Mariangela, Plaia Antonella	
<b>Unsupervised Learning of Option Price in a Controlled Environment: a Neural Network Approach</b> . . . . .	260
Gatta Federico, Schiano Di Cola Vincenzo, Piccialli Francesco, Cuomo Salvatore	
<b>SEMgraph: An R Package for Causal Network Inference of High-Throughput Data with Structural Equation Models</b> . . . . .	266
Grassi Mario, Tarantino Barbara	



Contents

<b>Dynamic models based on stochastic differential equations for biomarkers and treatment adherence in heart failure patients</b> . . . . .	271
Gregorio Caterina, Rares Franco Nicola, Ieva Francesca	
<b>Detecting anomalies in time series categorical data: a conformal prediction approach</b> . . . . .	277
Landrò Matteo, Stamm Aymeric, Vantini Simone	
<b>The structural behavior of Santa Maria del Fiore Dome: an analysis with machine learning techniques</b> . . . . .	282
Masini Stefano, Bacci Silvia, Cipollini Fabrizio, Bertaccini Bruno	
<b>Statistics and Data Science for Arts and Culture: an Application to the City of Brescia</b> . . . . .	288
Ricciardi Riccardo, Carpita Maurizio, Perazzini Selene, Zuccolotto Paola, Manisera Marica	
<b>Detecting Stance in Online Discussions about Vaccines</b> . . . . .	294
Francesco Pierri, Pizzo Fabio, Brambilla Marco	
<b>Towards the specification of a self-exciting point process for modelling crimes in Valencia</b> . . . . .	300
Chiodi Marcello, D'Angelo Nicoletta, Adelfio Giada, Mateu Jorge	
<b>A Clusterwise regression method for distributional data</b> . . . . .	306
Balzanella Antonio, Verde Rosanna, de Carvalho Francisco de A.T.	
<b>Increasing accuracy in classification models for the identification of plant species based on UAV images</b> . . . . .	311
Simonetto Anna, Tariku Girma, Gilioli Gianni	
<b>Travel time to university as determinant on students' performances</b> . . . . .	317
Burzacchi Arianna, Rossi Lidia, Agasisti Tommaso, Paganoni Anna Maria, Vantini Simone	
<b>The FAITH project: integrated tools and methodologies for digital humanities</b> . . . . .	323
Ferrara Alfio, Picascia Sergio, Rocchetti Elisabetta, Varese Gaia	
<b>Assessing the quality of Automatic Passenger Counter data for the analysis of mobility flows of local public transport systems</b> . . . . .	328
Urbano Valeria Maria, Burzacchi Arianna, Cherubini Francesco, Arena Marika, Azzone Giovanni, Secchi Piercesare, Vantini Simone	

# Streamlined Variational Inference for Modeling Italian Educational Data

Gioia Di Credico, Claudia Di Caterina, Francesco Santelli

**Abstract** The streamlined version of the mean field variational Bayes (MFVB) algorithm for linear mixed models with crossed random effects allows simplifying calculations but may require one group's dimension to be moderate. Data collecting high school students' first term evaluations and INVALSI scores for Italian and Maths subjects perfectly comply with this setting: students are a vast random sample of those who enrolled at the university in 2019/20, while the number of tests is limited to 6. Three different MFVB product restrictions with incremental complexity are evaluated. All of them are convenient with respect to classic MCMC solutions from both a computational and a memory storage viewpoint. Results and interpretation of model coefficients are in line with the literature on educational data.

**Key words:** Crossed random effects, INVALSI, Mean field variational Bayes.

## 1 Introduction

Linear mixed-effects models are commonly used to analyze data with a continuous Gaussian outcome and multiple levels of variability arising from a grouped data structure. In order to account for the variability introduced by the nested or crossed structure of the observations, it may be convenient to include random effects treated as random variables in the model.

In the following, we focus on a crossed-data application where two levels of variability exist. Crossed designs imply that each combination of the group levels is represented in the data [1]. Our data refer to a random sample of students who enrolled in a Bachelor program at an Italian university in the academic year 2019/2020. Out-

---

Gioia Di Credico  
Università degli studi di Trieste, Trieste, e-mail: [gioia.dicredico@deams.units.it](mailto:gioia.dicredico@deams.units.it)

Claudia Di Caterina  
Università degli studi di Verona, Verona, e-mail: [claudia.dicaterina@univr.it](mailto:claudia.dicaterina@univr.it)

Francesco Santelli  
Università degli studi di Trieste, Trieste, e-mail: [fsantelli@units.it](mailto:fsantelli@units.it)

comes consist of students' evaluations during their 10th and 13th high school grades. Here, students and the type of tests define our two-group crossed-data design. As the groups size increases, model estimation gets slower and may even become unfeasible. Streamlined variational inference has recently been applied to overcome these estimation difficulties in random-effects models, e.g. by [3] for the nested group structure and by [4] for the crossed one. The key idea relies on the sparseness of the matrix to be inverted, which enables a quicker computation and less storage capacity. Nested data imply a so-called arrowhead block structure for this matrix [5], and non-zero sub-blocks can be easily identified to simplify the calculation of its inverse. In the crossed-data case studied here the matrix is less sparse, however, considering the most accurate restriction, the streamlined approach offers advantages when one group is moderate in size. Our motivating data exhibit such a feature: the number of students involved is very large (around 7000), while the tests whose mark is recorded on each student are limited to 6.

## 2 Methods

For each  $i$ th student, we assume the scores  $\mathbf{y}_{i i'}$  on test  $i'$  follows a linear mixed model with two crossed random effects:

$$\begin{aligned} \mathbf{y}_{i i'} | \beta, \mathbf{u}_i, \mathbf{u}'_{i'}, \sigma^2 &\stackrel{\text{ind.}}{\sim} N(\mathbf{X}_{i i'} \beta + \mathbf{Z}_{i i'} \mathbf{u}_i + \mathbf{Z}'_{i i'} \mathbf{u}'_{i'}, \sigma^2 \mathbf{I}), \quad i = 1, \dots, m, \\ \mathbf{u}_i | \Sigma &\stackrel{\text{ind.}}{\sim} N(0, \Sigma), \quad \mathbf{u}'_{i'} | \Sigma' \stackrel{\text{ind.}}{\sim} N(0, \Sigma'), \quad i' = 1, \dots, m', \end{aligned} \quad (1)$$

where  $\mathbf{X}_{i i'}$  is the  $n_{i i'} \times p$  design matrix,  $\mathbf{Z}_{i i'}$  and  $\mathbf{Z}'_{i i'}$ , respectively of dimension  $n_{i i'} \times q$  and  $n_{i i'} \times q'$ , are the random effects matrices,  $\beta$  is the  $p$ -vector of fixed-effect coefficients,  $\mathbf{u}_i$  and  $\mathbf{u}'_{i'}$ , respectively  $q \times 1$  and  $q' \times 1$ , are the vectors of random effects,  $\Sigma$  and  $\Sigma'$  are their  $q \times q$  and  $q' \times q'$  respective covariance matrices and  $\sigma^2$  is the error variance.

The joint *a priori* density of the  $p$  fixed effects is  $\beta \sim N_p(\mu_\beta, \Sigma_\beta)$ . For the error variance  $\sigma^2$  and the random effects covariance matrices  $\Sigma$  and  $\Sigma'$ , we consider the following family of marginally non-informative prior distributions [2]:

$$\begin{aligned} \sigma^2 | a_{\sigma^2} &\sim \text{Inverse-}\chi^2(v_{\sigma^2}, 1/a_{\sigma^2}), \quad a_{\sigma^2} \sim \text{Inverse-}\chi^2(1, 1/(v_{\sigma^2} s_{\sigma^2}^2)), \\ \Sigma | \mathbf{A}_\Sigma &\sim \text{Inverse-G-Wishart}(G_{\text{full}}, v_\Sigma + 2q - 2, \mathbf{A}_\Sigma^{-1}), \\ \Sigma' | \mathbf{A}_{\Sigma'} &\sim \text{Inverse-G-Wishart}(G_{\text{full}}, v_{\Sigma'} + 2q' - 2, \mathbf{A}_{\Sigma'}^{-1}), \\ \mathbf{A}_\Sigma &\sim \text{Inverse-G-Wishart}(G_{\text{diag}}, 1, \Lambda_{\mathbf{A}_\Sigma}), \quad \Lambda_{\mathbf{A}_\Sigma} = \{v_\Sigma (s_{\Sigma,1}^2, s_{\Sigma,2}^2)\}^{-1}, \\ \mathbf{A}_{\Sigma'} &\sim \text{Inverse-G-Wishart}(G_{\text{diag}}, 1, \Lambda_{\mathbf{A}_{\Sigma'}}), \quad \Lambda_{\mathbf{A}_{\Sigma'}} = \{v_{\Sigma'} (s_{\Sigma',1}^2, s_{\Sigma',2}^2)\}^{-1}. \end{aligned} \quad (2)$$

In our application, the first group of random effects  $\mathbf{u}_i$  ( $i = 1, \dots, m = 7005$ ) corresponds to students enrolled at an Italian university in 2019/2020, and the second group  $\mathbf{u}'_{i'}$  ( $i' = 1, \dots, m' = 6$ ) corresponds to scores from assessments of Italian and

Math skills. Specifically, for each student, a written and oral score was recorded at the end of the first term and one written standardized INVALSI score (see Section 3) was recorded at the end of the final term. Each combination student/test, corresponding to the pair  $(i, i')$ , is observed  $n_{ii'} = n = 2$  times, namely in the 10th and 13th grades of high school. The design matrix  $\mathbf{X}_{ii'}$  has  $p = 32$  columns, including the intercept. Moreover,  $q = q' = 2$  because we consider both random intercepts and random slopes for the two groups, meaning

$$\mathbf{Z}_{ii'} = \mathbf{Z}'_{ii'} = [1 \ x_{1,ii'j}]_{j=1,2},$$

where  $x_{1,ii'j} = 1, 2$  is the year indicator encoding the two high school grades. According to (1) and this set-up, the two scores of the  $i$ th student on the  $i'$ th test are modeled to be

$$y_{ii'j} | \beta, u_{0i}, u_{1i}, u'_{0i'}, u'_{1i'}, \sigma^2 \stackrel{\text{ind.}}{\sim} N(\beta_0 + u_{0i} + u_{0i'} + (\beta_1 + u_{1i} + u_{1i'})x_{1,ii'j} + \sum_{k=1}^{31} \beta_k x_{k,ii'j}, \sigma^2)$$

for  $j = 1, 2$ . The formula above shows that this modelling strategy allows for a different intercept and slope for every student/test combination. Heterogeneities among intercepts and slopes are defined by appropriate entries of  $\Sigma$  and  $\Sigma'$ .

We consider three product restrictions on the mean field approximation of the joint conditional density function of all parameters in (1) with covariance priors (2) ([4], Sect. 3):

$$q(\beta, \mathbf{u}, \mathbf{u}', \sigma^2, \Sigma, \Sigma') = \begin{cases} q(\beta)q(\mathbf{u})q(\mathbf{u}')q(\sigma^2, \Sigma, \Sigma'), & \text{restriction I,} \\ q(\beta, \mathbf{u})q(\mathbf{u}')q(\sigma^2, \Sigma, \Sigma'), & \text{restriction II,} \\ q(\beta, \mathbf{u}, \mathbf{u}')q(\sigma^2, \Sigma, \Sigma'), & \text{restriction III.} \end{cases} \quad (3)$$

Product restriction I has the simplest streamlined implementation and scales well to very large problems, but may produce small posterior variances as it sets all posterior correlations between  $\beta$ ,  $\mathbf{u}$  and  $\mathbf{u}'$  to zero. Conversely, product restriction III allows for a full joint posterior covariance matrix of  $(\beta, \mathbf{u}, \mathbf{u}')$ , leading to higher inferential accuracy but challenging computing that can be streamlined for limited  $m'$ . A compromise is given by product restriction II, which includes posterior correlations between  $\beta$  and  $\mathbf{u}$ , for  $\mathbf{u}$  larger than  $\mathbf{u}$ . For all the product restrictions, the prior distributions specification 2 leads to a fully factorization of the  $q$ -densities related to the covariance matrix and auxiliary variables [4].

The  $q$ -density parameters can be obtained using a coordinate ascent iterative algorithm. However, if applied naively, the potentially prohibitively large matrix  $\Sigma_{q(\beta, \mathbf{u}, \mathbf{u})}$  requires storage and inversion. Product restrictions I, II and III lead to streamlined mean field variational Bayes (MFVB) algorithms with varying degrees of storage and computational overhead (see [4], Sect. 4).

### 3 Italian students' proficiency data

Data drawn from the Italian 'Anagrafe Nazionale della Formazione Superiore' has been processed according to the research project 'From high school to the job market: analysis of the university careers and the university North-South mobility' carried out by the University of Palermo (head of the research program), the Italian 'Ministero Università e Ricerca', and INVALSI. The dataset is known as MOBYSU. In the Italian School System, the scholastic assessment is in charge of the "Italian national institute for the evaluation of the school system" (INVALSI), which uses a set of standardized tests to evaluate the proficiency of students attending different schools at different years. Several domains are tested, and the main domains are Mathematical skills, English language, Italian language, and Science.

Our data regard the cohort of pupils that finished high school, achieved the Diploma in 2018/19, and then enrolled at university in 2019/20. Such students are more than 240000. To be included, students must have never failed a scholastic year, and must attend high school for the first time in the Italian school system. The response variable refers to students marks: four recorded at the end of the first term (Italian and Math, written and oral) and two throughout the INVALSI test (Italian and Math, written), during their 10th and 13th high school grades. Predictors involved in the analysis are listed in Tab.1. They include information on the socio-economic background, parental occupation and demographics.

**Table 1** Model predictors, their description and reference categories.

Variable	Description	Reference
Gender	Male, Female	Female
Age	Reception (one year ahead), Regular	Regular
Nation	Foreigner, Italian	Italian
Student escs (EscsStud)	student socio-economic level	
School escs (EscsSch)	school socio-economic level	
School type (SchTy)	13 categories	Classical Lyceum
Work Mother (Work.M)	5 categories	Unemployed
Work Father (Work.F)	5 categories	Unemployed
Year	School year	
NUTS2 classification (NUTS2)	5 areas	Center
School	Private, Public	Public

### 4 Analysis and results

First term and INVALSI scores were centered to the national means and INVALSI were also scaled to standardize them to a common range and adapt to the prior distributions setting. Furthermore, we excluded students with missing information

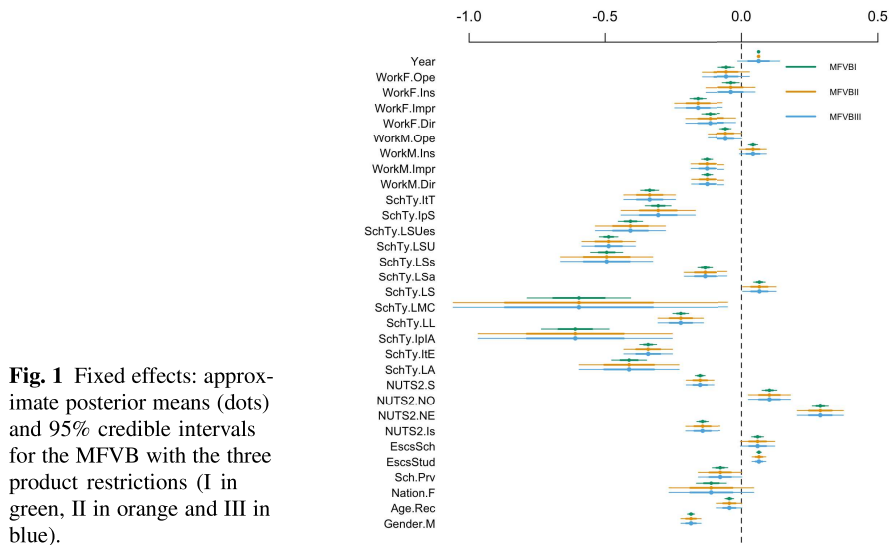
so that valid and complete data refer to 21228 students. The final model is fitted on a random sample of 33% of the units, corresponding to 7005 total pupils.

As hyperparameters, we set  $\mu_\beta = \mathbf{0}$ ,  $\Sigma_\beta = 10^{10}\mathbf{I}$ ,  $v_{\sigma^2} = 1$ ,  $v_\Sigma = v_{\Sigma'} = 2$ ,  $s_{\sigma^2} = s_{\Sigma,1}^2 = s_{\Sigma,2}^2 = s_{\Sigma',1}^2 = s_{\Sigma',2}^2 = 10^5$ . For each product restriction, we run the MFVB algorithm for 100 iterations. Computational times were about 8 minutes for the MFVB with product restriction I, 11 and 21 times longer for the MFVB with product restriction II and III. The model 1 with prior distributions as in 2 was also simulated through MCMC in Stan. In particular, 4 chains of 2000 iterations each (1000 warm-up; 1000 sampling) were simulated. In the following, MCMC inference is based on the sampling step draws. The running time for the MCMC setup was of 19 hours.

As expected, variational inference on the random effects and error variance components are relatively affected by the MFVB product restriction used, giving very similar results (Tab. 2). Differences between MFVB and MCMC on the tests random effects variability are likely due to a slow convergence of the MCMC chains, advised by a low effective sample size on the  $\Sigma'$  parameters. The product restriction impact is evident on the estimated variability of the fixed effects. While approximate posterior means of the fixed effects are the same across product restrictions, the least accurate (I) strongly underestimates the variability, while restrictions II and III lead to very similar results on all the coefficients, except for the year variable (see Fig. 1). On average, fixed marginal effects suggest that male students and those from islands and Southern Italy regions perform worse, while the Northeast is the area with best proficiency. The socio-economic status dimension has a significant positive effect, as expected, both at the individual and school levels. Lyceums record the best scores on average, with Scientific lyceum overperforming all the others. When parents are less involved in demanding jobs, students usually perform better. No clear effects are found for Italian nationality and students one year ahead. In both random intercept and slope, the variability carried by the student group is slightly larger than the item group one.

**Table 2** Random effects standard deviation estimates (approximate posterior mean). Square root of diagonal entries of  $\hat{\Sigma}$  ( $\hat{\Sigma}'$ ) are denoted by  $\hat{\sigma}_1$  ( $\hat{\sigma}'_1$ ) and  $\hat{\sigma}_2$  ( $\hat{\sigma}'_2$ ). Correlation between random intercept and slope is denoted by  $\hat{\rho}$  ( $\hat{\rho}'$ ).

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}$	$\hat{\sigma}'_1$	$\hat{\sigma}'_2$	$\hat{\rho}'$	$\hat{\sigma}$
MFVB I	0.677	0.114	0.252	0.540	0.087	-0.578	0.931
MFVB II	0.678	0.114	0.251	0.540	0.087	-0.578	0.931
MFVB III	0.678	0.114	0.251	0.591	0.095	-0.523	0.931
MCMC	0.679	0.114	0.251	0.883	0.189	-0.201	0.931



**Fig. 1** Fixed effects: approximate posterior means (dots) and 95% credible intervals for the MFVB with the three product restrictions (I in green, II in orange and III in blue).

## 5 Conclusions

The work analysed Italian students' proficiency data using the streamlined MFVB algorithm based on three product restrictions. The code is not optimized and computational times are reported for comparative purposes only. Even so, the MFVB algorithms are much faster than standard MCMC solutions. Comparing the three MFVB product restrictions, the second appears to be an excellent compromise balancing estimation speed and results accuracy.

## References

1. Baayen, R. H., Davidson, D. J., Bates, D. M.: Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* (2008) doi: 10.1016/j.jml.2007.12.005
2. Huang, A., Wand, M. P.: Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* (2013) doi: 10.1214/13-BA815
3. Lee, C. Y. Y., Wand, M. P.: Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical J.* (2016) doi: 10.1002/bimj.201500007
4. Menictas, M., Di Credico, G., Wand, M. P.: Streamlined variational inference for linear mixed models with crossed random effects. *J. Comput. Graph. Stat.* (2022) doi: 10.1080/10618600.2022.2096622
5. Nolan, T. H., Wand, M. P.: Streamlined solutions to multilevel sparse matrix problems. *ANZIAM J.* (2020) doi: 10.1017/S1446181120000061