Bias in Students' Evaluations

Which factors influence students' evaluations?

A.Burzacchi, D.Falco, M.Teodori
Supervised by Prof.ssa A.Guglielmi, Dott. R.Corradin

Politecnico di Milano

19/02/2020

When we met before

We found some problems in the models that we presented:

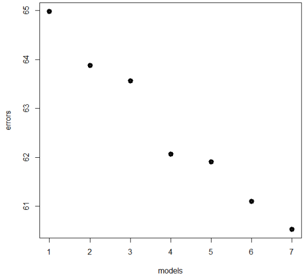
- EN and SSVS were too permissive
- Too many covariates in our models (both linear and hierarchical)
- Problems in the convergence

$\Rightarrow$ One step back to adjust them!

Linear Model

$$
\begin{cases}
Y_i \mid \mu_i, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_i,\ \sigma^2) & i = 1:N \\
\mu_i = \beta_0 + X_i^t \boldsymbol{\beta} & \\
\beta_0 \mid \tau_0^2 \ \sim\ \mathcal{N}(0,\ \tau_0^2) & \\
\boldsymbol{\beta} \mid \tau_1^2, \ldots, \tau_p^2 \ \sim\ \mathcal{N}_p\left(\mathbf{0},\ \begin{bmatrix} \tau_1^2 & & 0 \\ & \cdots & \\ 0 & & \tau_p^2 \end{bmatrix}\right) & \\
\sigma \ \sim\ \mathcal{U}\mathrm{nif}(0, \sigma_0) &
\end{cases}
$$

Results from CV: error in different models



The selected covariates are (out of 19):

escs, learning_time_science, gender,

enjoy_science, motivat, cultural_possessions,

study_after_school, disciplinary_climate,

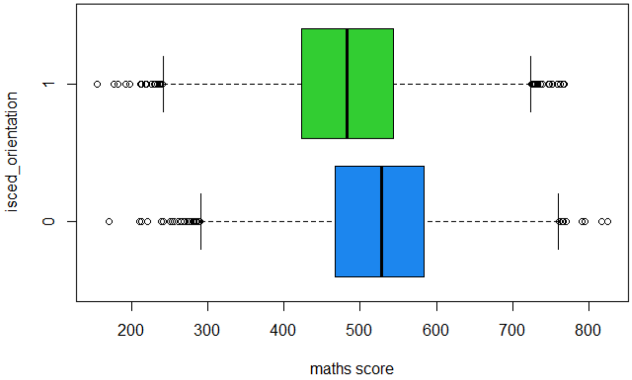
interest_broad_science

video_games, test_anxiety, study_before_school

Linear Mixed Effects Model

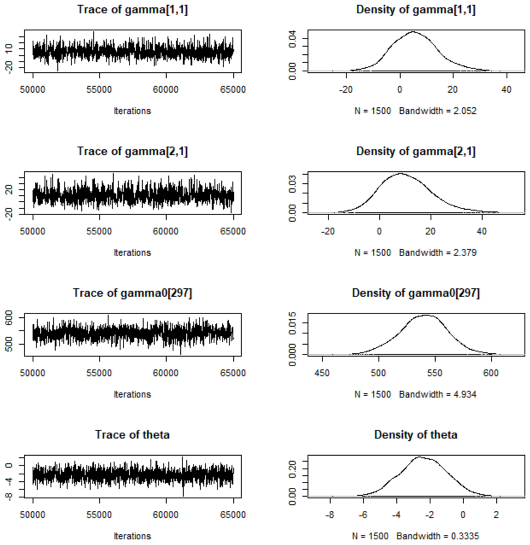
$$
\begin{cases}
Y_{ij} \mid \mu_{ij}, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_{ij},\ \sigma^2) & i=1:N;\ j=1:ng \\
\mu_{ij} = \gamma_{0j} + X_j^t \boldsymbol{\theta} + Z_{ij}^t \boldsymbol{\gamma}_j & \\
\gamma_{0j} \mid \hat{Y}, \tau_0^2 \overset{iid}{\sim} \mathcal{N}(\hat{Y},\ \tau_0^2) & j=1:ng \\
\boldsymbol{\theta} \mid \tau_\theta^2 \sim \mathcal{N}_p\left(\mathbf{0}, \tau_\theta^2\, \mathbb{I}_p\right) & \\
\boldsymbol{\gamma}_j \mid \tau_1^2, \ldots, \tau_v^2 \overset{iid}{\sim} \mathcal{N}_v\left(\mathbf{0},\ \begin{bmatrix} \tau_1^2 & & 0 \\ & \cdots & \\ 0 & & \tau_v^2 \end{bmatrix}\right) & j=1:ng \\
\sigma \mid a_1, a_2 \sim \text{Inv-}\mathcal{G}(a_1,\ a_2) & \\
\tau_0^2, \tau_l^2 \mid b_1, b_2 \overset{iid}{\sim} \text{Inv-}\mathcal{G}(b_1,\ b_2) & l=1:v; \\
1/\tau_\theta^2 \mid \lambda \sim \mathcal{E}(\lambda) + 0.5 &
\end{cases}
$$

Covariate selection with Elastic Net



The method does not select any schools' covariate
We use ISCED_orientation.

Univariate hierarchical model



Bivariate LM

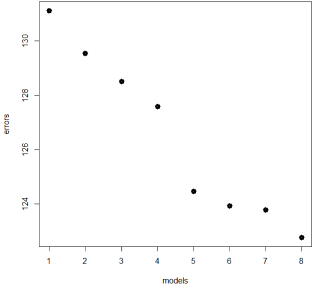
$$\begin{cases} \boldsymbol{Y}_i \mid \boldsymbol{\mu}_i, \Sigma \stackrel{ind}{\sim} \mathcal{N}_2(\boldsymbol{\mu}_i,\, \Sigma) \qquad i = 1:N \\ \boldsymbol{\mu}_i = \begin{bmatrix} \beta_{0,1} + \boldsymbol{X}_i^t \boldsymbol{\beta}_1 \\ \beta_{0,2} + \boldsymbol{X}_i^t \boldsymbol{\beta}_2 \end{bmatrix} \\ \boldsymbol{\beta}_0 \mid \tau_0^2 \sim \mathcal{N}_2(\mathbf{0},\, \tau_0^2\, \mathbb{I}_2) \\ \boldsymbol{\beta}_1 \mid \tau_1^2 \sim \mathcal{N}_p(\mathbf{0},\, \tau_1^2\, \mathbb{I}_p) \\ \boldsymbol{\beta}_2 \mid \tau_2^2 \sim \mathcal{N}_p(\mathbf{0},\, \tau_2^2\, \mathbb{I}_p) \\ \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \\ \sigma_1, \sigma_2 \stackrel{iid}{\sim} \mathcal{U}\text{nif}(0, 120) \\ \rho \sim \mathcal{U}\text{nif}(-1, 1) \end{cases}$$

Covariate selection: EN

$$\left\{\begin{array}{lr}
\boldsymbol{Y}_i \mid \boldsymbol{\mu}_i, \Sigma \overset{ind}{\sim} \mathcal{N}_2(\boldsymbol{\mu}_i,\ \Sigma) & i = 1 : N \\
\boldsymbol{\mu_i} = \boldsymbol{\beta_0} + X_{i,1}\boldsymbol{\beta_1} + \ldots + X_{i,p}\boldsymbol{\beta_p} & i = 1 : N \\
\boldsymbol{\beta}_0 \mid \tau_0^2, \sim \mathcal{N}_2(\mathbf{0},\ \tau_0^2 \mathbb{I}_2) & \\
\boldsymbol{\beta}_j \mid \tau_j, a_2 \overset{ind}{\sim} \mathcal{N}_2(\mathbf{0},\ \frac{\tau_j - 1}{\tau_j a_2} \mathbb{I}_2) & j = 1 : p \\
\tau_j \mid a_1, a_2 \overset{iid}{\sim} \text{trunc-}\mathcal{G}(\frac{1}{2},\ \frac{a_1^2}{8a_2},\ 1, \infty) & j = 1 : p \\
a_1, a_2 \overset{iid}{\sim} \mathcal{E}(0.1) + 0.5 & \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} & \\
\sigma_1, \sigma_2 \overset{iid}{\sim} \mathcal{U}\text{nif}(0, 100) & \\
\rho \sim \mathcal{U}\text{nif}(-1, 1) &
\end{array}\right.$$

Comparison between models

Results from CV: error in different models



The selected covariates are:

escs, learning_time_science, gender(math),

enjoy_science, motivat, cultural_possessions,

study_after_school, interest_broad_science

video_games, test_anxiety, gender(read)

study_before_school

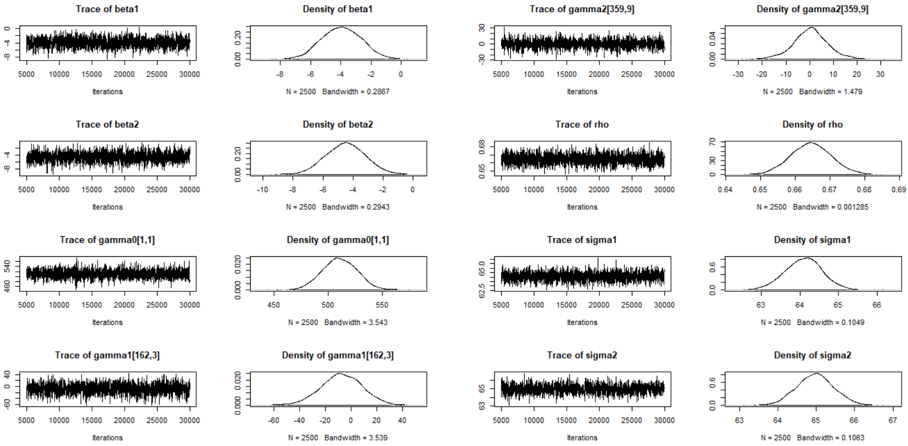
Bivariate LMM (1)

$$
\begin{cases}
\boldsymbol{Y}_{ij} \mid \boldsymbol{\mu}_{ij}, \Sigma \overset{ind}{\sim} \mathcal{N}_2(\boldsymbol{\mu}_{ij},\ \Sigma) & i = 1:N;\ j = 1:ng \\
\boldsymbol{\mu}_{ij} = \begin{bmatrix} \gamma_{0j,1} + \boldsymbol{Z}_{ij}^t \gamma_{1j} + \boldsymbol{X}_j^t \beta_1 \\ \gamma_{0j,2} + \boldsymbol{Z}_{ij}^t \gamma_{2j} + \boldsymbol{X}_j^t \beta_2 \end{bmatrix} & \\
\gamma_{0j} \mid \hat{\boldsymbol{Y}}, \tau_0^2 \overset{ind}{\sim} \mathcal{N}_2(\hat{\boldsymbol{Y}},\ \tau_0^2 \mathbb{I}_2) & j = 1:ng \\
\gamma_{1j,k} \mid \tau_{1,k}^2 \overset{ind}{\sim} \mathcal{N}(0,\ \tau_{1,k}^2) & j = 1:ng;\ k = 1:v \\
\gamma_{2j,k} \mid \tau_{2,k}^2 \overset{ind}{\sim} \mathcal{N}(0,\ \tau_{2,k}^2) & j = 1:ng;\ k = 1:v \\
\beta_1 \mid \omega_1^2 \sim \mathcal{N}_p(\mathbf{0},\ \omega_1^2 \mathbb{I}_p) & \\
\beta_2 \mid \omega_2^2 \sim \mathcal{N}_p(\mathbf{0},\ \omega_2^2 \mathbb{I}_p) &
\end{cases}
$$

Bivariate LMM (2)

$$
\begin{cases}
\tau_0^2,\ \tau_{1,1}^2, \ldots, \tau_{1,v}^2,\ \tau_{2,1}^2, \ldots, \tau_{2,v}^2 \overset{iid}{\sim} \text{Inv-}\mathcal{G}(2, 50) \\
1/\omega_1^2,\ 1/\omega_2^2 \overset{iid}{\sim} \mathcal{E}(100) + 0.5 \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \\
\sigma_1, \sigma_2 \overset{iid}{\sim} \mathcal{U}\text{nif}(0, 120) \\
\rho \sim \mathcal{U}\text{nif}(-1, 1)
\end{cases}
$$

with $\boldsymbol{Z}_{ij}$ covariates of the bivariate linear model,
$\boldsymbol{X}_j = \text{ISCED\_orient}$



Traceplot and posterior density plot of some parameters

Predictive distribution of a new student from an existing school

$\boldsymbol{Y}_j^{new}$ evaluations; $\boldsymbol{X}^{new}, \boldsymbol{Z}^{new}$ data
Set $\Theta_j = \{\gamma_{0j}, \gamma_{1j}, \gamma_{2j}\}$

Predictive distribution (related to bivariate LMM):

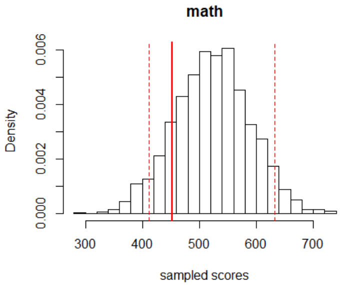
$$\mathcal{L}(\boldsymbol{Y}_j^{new} \mid \text{data}) = \int \mathcal{L}(\boldsymbol{Y}_j^{new} \mid \Theta_j, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Sigma, \boldsymbol{X}^{new}, \boldsymbol{Z}^{new})$$
$$\times \mathcal{L}(d\Theta_j, d\boldsymbol{\beta}_1, d\boldsymbol{\beta}_2, d\Sigma \mid \text{data})$$

Predictive distribution of a new student from an existing school

$\boldsymbol{Y}_j^{new}$ evaluations; $\boldsymbol{X}^{new}, \boldsymbol{Z}^{new}$ data
Set $\Theta_j = \{\gamma_{0j}, \gamma_{1j}, \gamma_{2j}\}$

Predictive distribution (related to bivariate LMM):

$$\mathcal{L}(\boldsymbol{Y}_j^{new} \mid \text{data}) = \int \mathcal{L}(\boldsymbol{Y}_j^{new} \mid \Theta_j, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Sigma, \boldsymbol{X}^{new}, \boldsymbol{Z}^{new}) \times \mathcal{L}(d\Theta_j, d\boldsymbol{\beta}_1, d\boldsymbol{\beta}_2, d\Sigma \mid \text{data})$$



Predictive distribution of a new student from a new school

$\boldsymbol{Y}_s^{new}$ evaluations; $\boldsymbol{X}^{new}, \boldsymbol{Z}^{new}$ data
Set $\Theta_j = \{\gamma_{0j}, \gamma_{1j}, \gamma_{2j}\} \quad \forall j = 1, \ldots, ng, s$
Set $\mathrm{T} = \{\tau_0^2, \tau_{1,1}^2, \ldots, \tau_{1,p}^2, \tau_{2,1}^2, \ldots, \tau_{2,p}^2\}$

Predictive distribution (related to bivariate LMM):

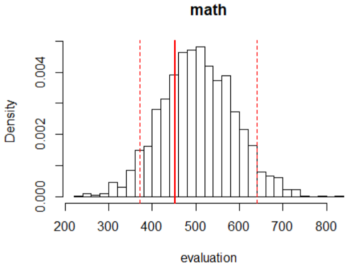
$$\mathcal{L}(\boldsymbol{Y}_s^{new}, \Theta_s \mid \text{data}) = \int \mathcal{L}(\boldsymbol{Y}_s^{new} \mid \Theta_s, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{X}^{new}, \boldsymbol{Z}^{new}, \Sigma)\, \mathcal{L}(\Theta_s \mid \mathrm{T}) \\ \times \mathcal{L}(d\Theta_1, \ldots, d\Theta_{ng}, d\boldsymbol{\beta}_1, d\boldsymbol{\beta}_2, d\mathrm{T}, d\Sigma \mid \text{data})$$

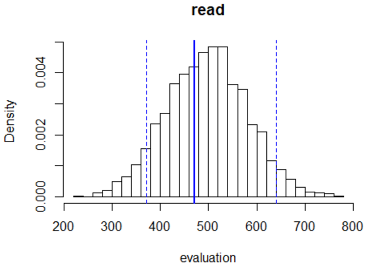
Predictive distribution of a new student from a new school

$\boldsymbol{Y}_s^{new}$ evaluations; $\boldsymbol{X}^{new}$, $\boldsymbol{Z}^{new}$ data
Set $\Theta_j = \{\gamma_{0j}, \gamma_{1j}, \gamma_{2j}\} \quad \forall j = 1, \ldots, ng, s$
Set $\mathrm{T} = \{\tau_0^2, \tau_{1,1}^2, \ldots, \tau_{1,p}^2, \tau_{2,1}^2, \ldots, \tau_{2,p}^2\}$

Predictive distribution (related to bivariate LMM):

$$\mathcal{L}(\boldsymbol{Y}_s^{new}, \Theta_s \mid \text{data}) = \int \mathcal{L}(\boldsymbol{Y}_s^{new} \mid \Theta_s, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{X}^{new}, \boldsymbol{Z}^{new}, \Sigma)\, \mathcal{L}(\Theta_s \mid \mathrm{T})$$
$$\times\, \mathcal{L}(d\Theta_1, \ldots, d\Theta_{ng}, d\boldsymbol{\beta}_1, d\boldsymbol{\beta}_2, d\mathrm{T}, d\Sigma \mid \text{data})$$



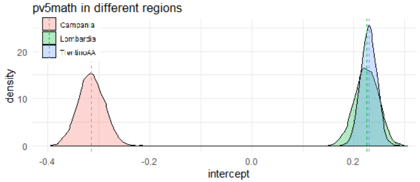


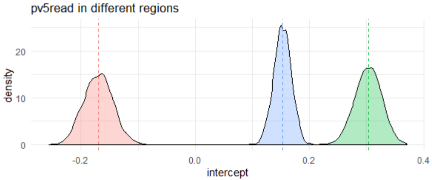
ANOVA

$$
\begin{cases}
\boldsymbol{Y}_{ij} \mid \gamma_j, \Sigma \overset{ind}{\sim} \mathcal{N}_2(\gamma_j,\ \Sigma) & i = 1:N; j = 1:3 \\
\gamma_j \mid \tau^2 \overset{iid}{\sim} \mathcal{N}_2(\mathbf{0},\ \tau^2\,\mathbb{I}_2) & j = 1:3 \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} & \\
\sigma_1, \sigma_2 \overset{iid}{\sim} \mathcal{U}\mathrm{nif}(0,1) & \\
\rho \sim \mathcal{U}\mathrm{nif}(-1,1) &
\end{cases}
$$

ANOVA

$$
\begin{cases}
\boldsymbol{Y}_{ij} \,|\, \gamma_j, \Sigma \overset{ind}{\sim} \mathcal{N}_2(\gamma_j,\ \Sigma) & i = 1:N; j = 1:3 \\
\gamma_j \,|\, \tau^2 \overset{iid}{\sim} \mathcal{N}_2(\mathbf{0},\ \tau^2\, \mathbb{I}_2) & j = 1:3 \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} & \\
\sigma_1, \sigma_2 \overset{iid}{\sim} \mathcal{U}\mathrm{nif}(0,1) & \\
\rho \sim \mathcal{U}\mathrm{nif}(-1,1) &
\end{cases}
$$



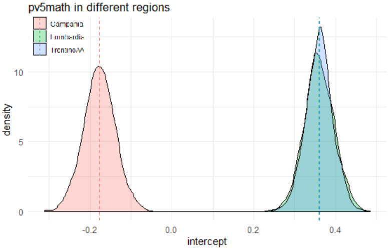


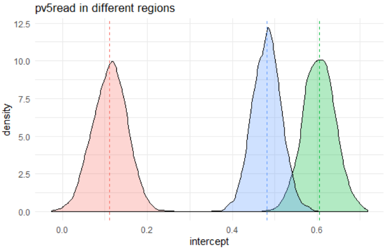
ANOVA 2

$$
\begin{cases}
\boldsymbol{Y}_{ij} \mid \boldsymbol{\mu}_{ij}, \Sigma \overset{ind}{\sim} \mathcal{N}_2(\boldsymbol{\mu}_{ij},\ \Sigma) \quad i = 1:N; j = 1:3 \\
\boldsymbol{\mu}_{ij} = \boldsymbol{\gamma}_j + \begin{bmatrix} X_i^t \boldsymbol{\beta}_1 \\ X_i^t \boldsymbol{\beta}_2 \end{bmatrix} \\
\boldsymbol{\gamma}_j \mid \tau^2 \overset{iid}{\sim} \mathcal{N}_2(\mathbf{0},\ \tau^2\, \mathbb{I}_2) \quad j = 1:3 \\
\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \mid \omega^2 \overset{iid}{\sim} \mathcal{N}_p(\mathbf{0},\ \omega^2\, \mathbb{I}_p) \\
\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \\
\sigma_1, \sigma_2 \overset{iid}{\sim} \mathcal{U}\mathrm{nif}(0,1) \\
\rho \sim \mathcal{U}\mathrm{nif}(-1,1)
\end{cases}
$$

ANOVA 2

On the left maths evaluations, on the right reading evaluations in the different regions





Credible interval for γ_{j1} (math) of level 0.95:

$$\begin{cases} [-0.253;\ -0.099] & \text{for Camp.} \\ [0.289;\ 0.432] & \text{for Lomb.} \\ [0.295;\ 0.4258] & \text{for Trent.} \end{cases}$$

Credible interval for γ_{j2} (read) of level 0.95:

$$\begin{cases} [-0.253;\ -0.099] & \text{for Camp.} \\ [0.289;\ 0.432] & \text{for Lomb.} \\ [0.295;\ 0.4258] & \text{for Trent.} \end{cases}$$

Hypothesis testing 1

H0: "Campania has worse evaluations than Lombardia and Trentino"

H1: "Campania does not have worse evaluations than Lombardia and Trentino"

Considering model ANOVA2:

$$\text{H0: } \begin{cases} \gamma_{\text{camp}} < \gamma_{\text{lomb}} \\ \gamma_{\text{camp}} < \gamma_{\text{trent}} \end{cases}$$

H1: otherwise

$$BF = \frac{P(\text{H0} \mid \text{data})}{P(\text{H1} \mid \text{data})} \frac{P(\text{H1})}{P(\text{H0})}$$

$\Rightarrow$ There is very strong evidence in favour of H0

Hypothesis testing 2

H0: "Lombardia has better evaluations than Trentino"

H1: otherwise

Considering model ANOVA2:

H0: $\gamma_{\text{trent}} < \gamma_{\text{lomb}}$

H1: otherwise

$$2\log(BF) = 2\log\left(\frac{P(\text{H0} \mid \text{data})}{P(\text{H1} \mid \text{data})}\frac{P(\text{H1})}{P(\text{H0})}\right) = 2.15045$$

Thus, there is weak evidence in favour of H0

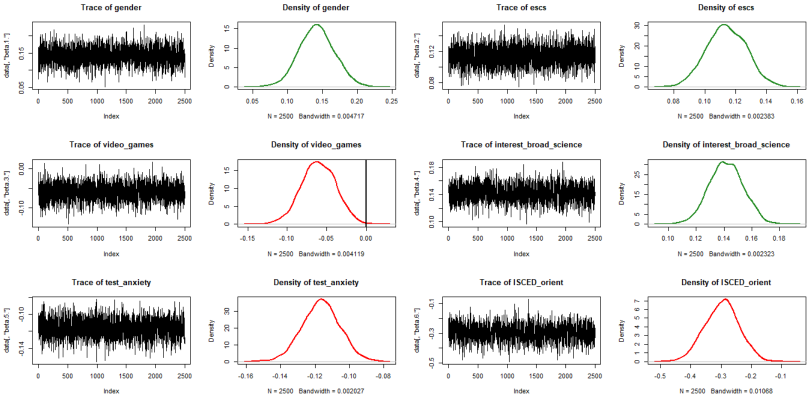
LMM with random effects iid from DP

Gibbs sampler strategy with JAGS:

$$\begin{cases} Y_i \mid p_i \overset{ind}{\sim} \mathcal{N}(p_i,\ 1) & i = 1 : N \\ p_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + b_{j[i]} & i = 1 : N; j = 1 : ng \\ \boldsymbol{\beta} \perp \{b_j,\ j = 1 : ng\} & \\ \boldsymbol{\beta} \sim \mathcal{N}_6(\mathbf{0},\ 1000\mathbb{I}_6) & \\ b_1, \ldots, b_{ng} \mid P \overset{iid}{\sim} P & \\ P \sim DP(\alpha, P_0) & \end{cases}$$

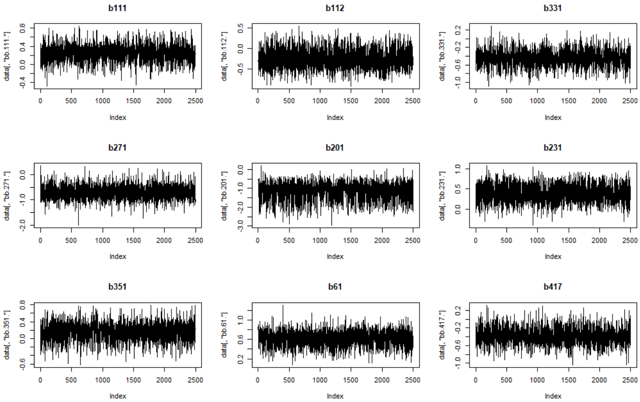
For each iteration $m = 1 : M$, unique values in $(b_1^{(m)}, \ldots, b_{ng}^{(m)})$, coming from the posterior distribution, identify a partition ρ of the schools.

Traceplots and distributions of the coefficients: all of them are significant



Positive coefficients: gender, escs, interest_broad_science
Negative coefficients: video_games, test_anxiety, ISCED_orient.

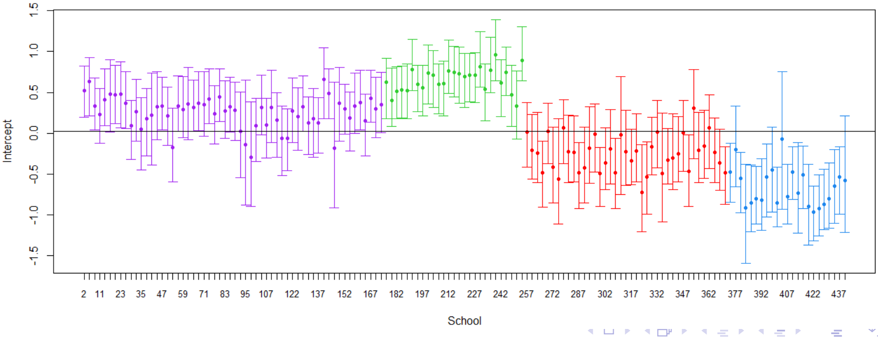
Traceplots of the random effects of some schools



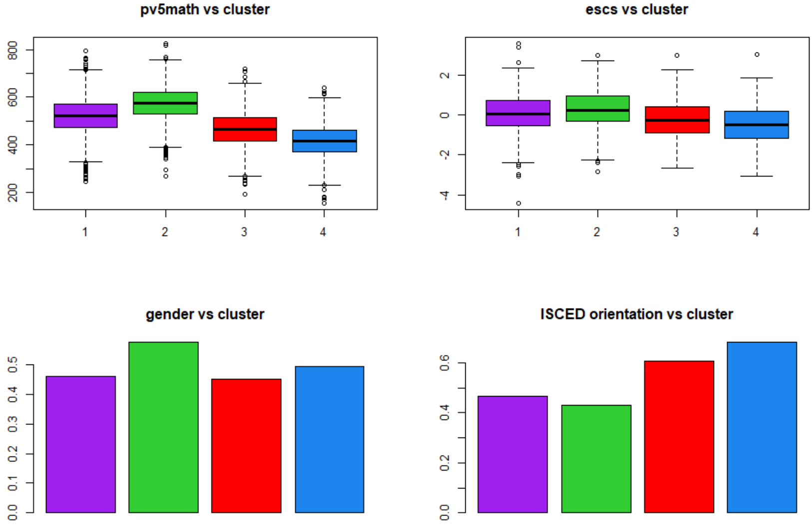
Given a vector of allocation variables $(c_1^{(m)}, \ldots, c_{ng}^{(m)})$ such that $b_i^{(m)} = b_j^{(m)} \Leftrightarrow c_i^{(m)} = c_j^{(m)}$, the best partition $\hat{\rho}$ is the one minimizing the Binder's loss function.

Cluster estimate

$$
\begin{cases}
\hat{m} = \underset{m}{\operatorname{argmin}}(LF_m) & \\
\pi_{ij} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{1}_{\{c_i^{(m)} = c_j^{(m)}\}} & i,j = 1:ng,\, i<j \\
LF_m = \sum_{i<j}(K - \pi_{ij})\mathbf{1}_{\{c_i^{(m)} = c_j^{(m)}\}} & i,j = 1:ng,\, i<j
\end{cases}
$$



Analysis of clusters



Bibliography

- OCED, "PISA Data Analysis Manual", *SPSS, Second Edition*, 2009
- Gelman, "Prior distributions for variance parameters in hierarchical models", *Bayesian Analysis*, 2006
- Gelman, Hill, "Data Analysis Using Regression and Multilevel/Hierarchical Models", *Cambridge University Press*, 2006
- Gelman et Al., "Bayesian Data Analysis", *CRC Press*, 2013
- Park, Casella, "The Bayesian Lasso", *Journal of the American Statistical Association*, 2008
- Mueller, Quintana, Jara, Hanson, "Bayesian Nonparametric Data Analysis", *Springer*, 2015
- Jackman, "Bayesian analysis for the Social Sciences", *Wiley*, 2009