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

08/01/2020

When we met before

Our dataset collects the results of a questionnaire given to:

- 11 583 students
- 473 schools in Italy

Our goal is to study students' evaluations:

1. Linear model
2. Mixed effect linear model grouping by school
3. Prediction on new observations

The Model

$$
\begin{cases}
Y_i \mid \mu_i, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_i,\ \sigma^2) & i{=}1:N \\
\mu_i = \beta_0 + X_i^t \boldsymbol{\beta} & \\
\beta_0 \mid \widehat{\beta}_0, \tau_0^2 \sim \mathcal{N}(\widehat{\beta}_0,\ \tau_0^2) & \\
\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}, \tau_1^2, \dots, \tau_p{}^2 \sim \mathcal{N}_{\boldsymbol{p}}(\widehat{\boldsymbol{\beta}},\ \begin{bmatrix} \tau_1^2 & & 0 \\ & \cdots & \\ 0 & & \tau_p^2 \end{bmatrix}\ ) & \\
\sigma^2 \mid \alpha_1, \alpha_2 \sim \text{Inv-}\mathcal{G}(\alpha_1,\ \alpha_2) &
\end{cases}
$$

$\boldsymbol{Y}$ contains the students' evaluations in Maths,
while X is the design matrix of the covariates about the students.

Covariate selection

To select the interesting student covariates:

1. Penalized regression with Elastic Net

$$
\begin{cases}
\beta_0 \mid \tau_0^2 \sim \mathcal{N}(0,\ \tau_0^2) & \\
\beta_j \mid a_2, \tau_j \overset{ind}{\sim} \mathcal{N}(0,\ \frac{\tau_j - 1}{a_2 \tau_j}) & j{=}1:p \\
\tau_j \mid a_1, a_2 \overset{iid}{\sim} \text{tr-}\mathcal{G}(0.5, \frac{a_1^2}{8a_2},\ 1, \infty) & j{=}1:p \\
a_1, a_2 \overset{iid}{\sim} \mathcal{E}(\lambda) + 0.5 &
\end{cases}
$$

Covariate selection

To select the interesting student covariates:

1. Penalized regression with Elastic Net
2. Stochastic Search Variable Selection

$$
\begin{cases}
\beta_j \mid \tau_j \overset{ind}{\sim} \mathcal{N}(0,\ \tau_j) & j=0:p \\
\tau_j = c_1(1-\gamma_j) + c_2\gamma_j & \\
\gamma_j \mid \lambda \overset{iid}{\sim} \mathcal{B}(\lambda) & j=0:p
\end{cases}
$$

Covariate selection

To select the interesting student covariates:

1. Penalized regression with Elastic Net
2. Stochastic Search Variable Selection

```
"gender"                     "index_immigration_status" "cultural_possessions"
"escs"                       "video_games"              "internet_social"
"study_before_school"        "study_after_school"       "learning_time_math"
"learning_time_lang"         "learning_time_science"    "disciplinari_climate"
"enjoy_science"              "interest_broad_science"   "test_anxiety"
"education_parents_medium"   "education_parents_high"   "instrum_motivat"
"home_possessions"           "motivat"
```

In red, selected only by EN; in blue, selected only by SSVS.

Best model with WAIC

We compared 8 models with different covariates:

1. only always choosen covariates — $\text{WAIC}_1 = -10269.12$
2. + instrum-motivat — $\text{WAIC}_2 = -10267.62$
3. + home-poss — $\text{WAIC}_3 = -10265.5$
4. + motivat — $\text{WAIC}_4 = -10264.68$
5. + instrum-motivat + home-poss — $\text{WAIC}_5 = -10263.88$
6. + instrum-motivat + motivat — $\text{WAIC}_6 = -10264.07$
7. + home-poss + motivat — $\text{WAIC}_7 = -10260.49$
8. all the covariates — $\text{WAIC}_8 = -10260.09$

Best model with WAIC

We compared 8 models with different covariates:

1. only always choosen covariates	$\text{WAIC}_1 = -10269.12$
2. + instrum-motivat	$\text{WAIC}_2 = -10267.62$
3. + home-poss	$\text{WAIC}_3 = -10265.5$
4. + motivat	$\text{WAIC}_4 = -10264.68$
5. + instrum-motivat + home-poss	$\text{WAIC}_5 = -10263.88$
6. + instrum-motivat + motivat	$\text{WAIC}_6 = -10264.07$
7. + home-poss + motivat	$\text{WAIC}_7 = -10260.49$
8. all the covariates	$\text{WAIC}_8 = -10260.09$

$\rightarrow$ Best model is model 7

Final model

$$Y_i = \beta_0 + X_i^t \boldsymbol{\beta} + \epsilon_i \qquad i = 1, \dots, N$$

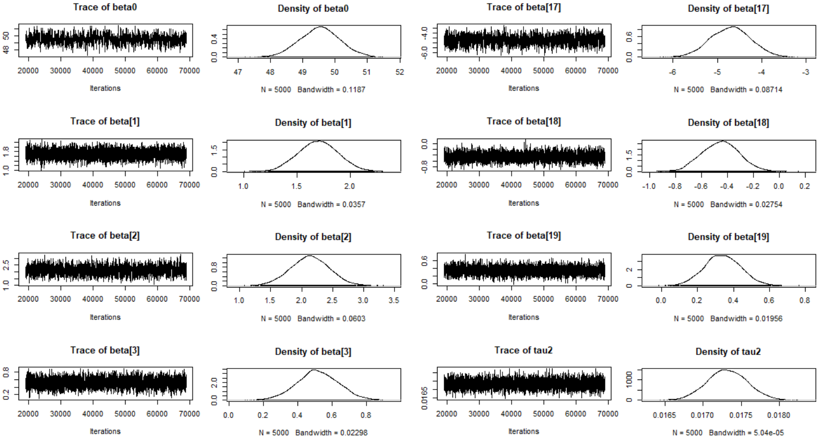
The selected covariates in X are:

```
"gender"                     "index_immigration_status" "cultural_possessions"
"escs"                       "video_games"              "internet_social"
"study_before_school"        "study_after_school"       "learning_time_math"
"learning_time_lang"         "learning_time_science"    "disciplinari_climate"
"enjoy_science"              "interest_broad_science"   "test_anxiety"
"education_parents_medium"   "education_parents_high"   "home_possessions"
"motivat"
```

The discarded covariates:

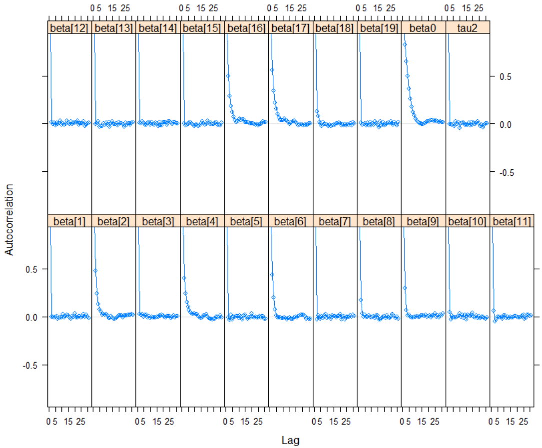
```
"home_edu_resources"    "sport_after_school"     "read_books"
"meet_friends"          "subjective_well_being"  "instrum_motivat"
```

Some plots



Traceplots and posterior density plots of the parameters

Some plots



Autocorrelation plot of the parameters

The model

$$
\begin{cases}
Y_{ij} \mid \mu_{ij}, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_{ij},\ \sigma^2) & i{=}1:N;\ j{=}1:ng \\
\mu_{ij} = \theta_0 + \gamma_{0j} + Z_j^t\boldsymbol{\theta} + X_{ij}^t\gamma_j & i{=}1:N;\ j{=}1:ng \\
\qquad \theta_0 \mid \omega_0^2 \sim \mathcal{N}(0,\ \omega_0^2) & \\
\qquad \gamma_{0j} \mid \tau_{0j}^2 \overset{ind}{\sim} \mathcal{N}(0,\ \tau_{0j}^2) & j{=}1:ng \\
\boldsymbol{\theta} \mid \omega_1^2, \ldots, \omega_q^2 \sim \mathcal{N}_q\left(\mathbf{0},\ \begin{bmatrix} \omega_1^2 & & 0 \\ & \cdots & \\ 0 & & \omega_q^2 \end{bmatrix}\right) & \\
\boldsymbol{\gamma}_j \mid \tau_{1j}^2, \ldots, \tau_{pj}^2 \overset{ind}{\sim} \mathcal{N}_p\left(\mathbf{0},\ \begin{bmatrix} \tau_{1j}^2 & & 0 \\ & \cdots & \\ 0 & & \tau_{pj}^2 \end{bmatrix}\right) & j{=}1:ng \\
\quad \sigma^2 \mid a_1, a_2 \sim \text{Inv-}\mathcal{G}(a_1,\ a_2) & \\
\omega_0^2, \omega_k^2, \tau_{0j}^2, \tau_{lj}^2 \overset{iid}{\sim} \text{Inv-}\mathcal{G}(b_1,\ b_2) & k{=}1:q;\ l{=}1:p;\ j{=}1:ng
\end{cases}
$$

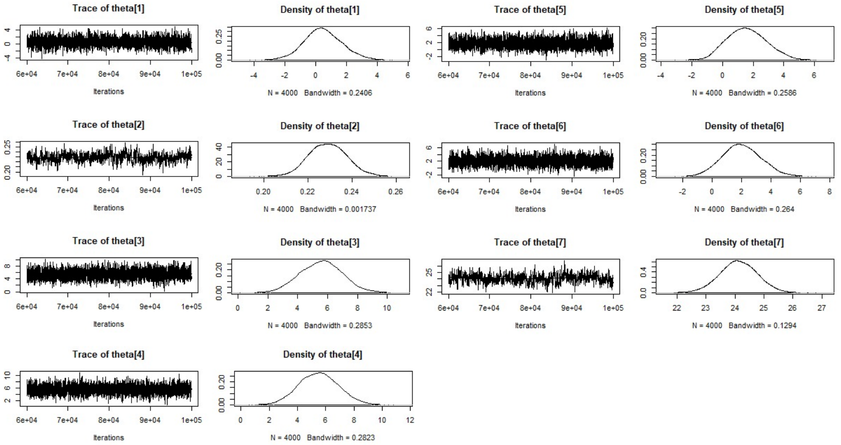
Covariate selection with Lasso

$$
\begin{cases}
Y_{ij} \mid \mu_{ij}, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_{ij},\ \sigma^2) & i{=}1:N;\ j{=}1:ng \\
\mu_{ij} = \gamma_{0j} + Z_j^t \boldsymbol{\theta} + X_{ij}^t \boldsymbol{\gamma}_j & i{=}1:N;\ j{=}1:ng \\
\gamma_{0j} \mid \tau_0^2 \overset{iid}{\sim} \mathcal{N}(0,\ \tau_0^2) & j{=}1:ng \\
\boldsymbol{\gamma}_{kj} \mid s_k^2 \overset{ind}{\sim} \mathcal{N}(0,\ s_k^2) & k{=}1:p;\ j{=}1:ng \\
s_k^2 \overset{iid}{\sim} \mathcal{U}\text{nif}(0,\ 100) & k{=}1:p \\
\theta_k \mid l_2 \overset{iid}{\sim} \text{double-}\mathcal{E}(0,\ l_2^{-1/2}) & k{=}1:q \\
l_2 \sim \mathcal{E}(\lambda) + 0.5 & \\
\sigma^2 \sim \text{Inv-}\mathcal{G}(\alpha_1, \alpha_2) &
\end{cases}
$$

Covariate selection with Elastic Net

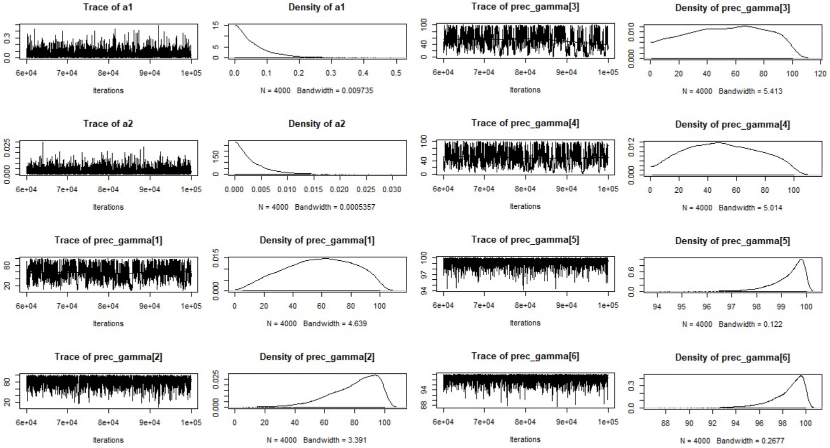
$$
\begin{cases}
Y_{ij} \mid \mu_{ij}, \sigma^2 \overset{ind}{\sim} \mathcal{N}(\mu_{ij},\ \sigma^2) & i{=}1:N;\ j{=}1:ng \\
\mu_{ij} = \gamma_{0j} + Z_j^t \boldsymbol{\theta} + X_{ij}^t \boldsymbol{\gamma}_j & i{=}1:N;\ j{=}1:ng \\
\gamma_{0j} \mid s_0^2 \overset{iid}{\sim} \mathcal{N}(0,\ s_0^2) & j{=}1:ng \\
\boldsymbol{\gamma}_{kj} \mid s_k^2 \overset{ind}{\sim} \mathcal{N}(0,\ s_k^2) & k{=}1:p;\ j{=}1:ng \\
s_k^2 \overset{iid}{\sim} \mathcal{U}\mathrm{nif}(0,\ 100) & k{=}1:p \\
\theta_k \mid \tau_k, a_2 \overset{ind}{\sim} \mathcal{N}\left(0,\ \frac{\tau_k - 1}{\tau_k a_2}\right) & k{=}1:q \\
\tau_k \mid a_1, a_2 \overset{iid}{\sim} \text{tr-}\mathcal{G}\left(0.5, \frac{a_1^2}{8\, a_2^2},\ 1, \infty\right) & k{=}1:q \\
a_1, a_2 \overset{iid}{\sim} \mathcal{E}(0.1) + 0.5 & \\
\sigma^2 \sim \text{inv-}\mathcal{G}(\alpha_1, \alpha_2) &
\end{cases}
$$

Some plots from Elastic Net



Traceplots and posterior density plots of $\boldsymbol{\theta}$

Some plots from Elastic Net



Traceplots and posterior densitiy plots of a_1, a_2 and s_k^2

Covariate selection: comparison

The covariates from Lasso are:

1. ISCED orientation
2. Certified teacher proportion
3. Student-teacher ratio
4. Science teacher proportion

The covariates from EN are:

1. ISCED orientation
2. Certified teacher proportion
3. Student-teacher ratio
4. School size

Covariate selection: comparison

The covariates from Lasso are:

1. ISCED orientation
2. Certified teacher proportion
3. Student-teacher ratio
4. Science teacher proportion

The covariates from EN are:

1. ISCED orientation
2. Certified teacher proportion
3. Student-teacher ratio
4. School size

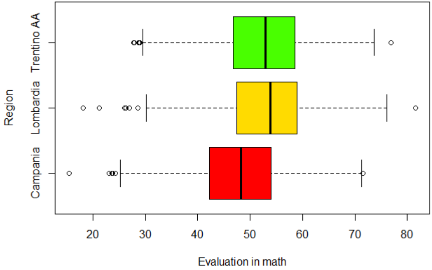
Final Model

$$Y_{ij} = \theta_0 + \gamma_{0j} + Z_j^t \boldsymbol{\theta} + X_{ij}^t \boldsymbol{\gamma}_j + \epsilon_{ij} \qquad i{=}1 : N;\ j{=}1 : ng$$

Finally, the significant covariates of the model are:

```
"home_possessions"            "cultural_possessions"       "escs"
"disciplinari_climate"        "learning_time_math"         "learning_time_lang"
"learning_time_science"       "enjoy_science"              "interest_broad_science"
"motivat"                     "test_anxiety"               "gender"
"education_parents_medium"    "education_parents_high"     "video_games"
"immigration_status"          "internet_social"            "study_before_school"
"study_after_school"          "ISCED_orient"               "certified_teacher_prop"
"student_teacher_ratio"       "school_size"
```

Hierarchical model grouping by region



- **PRO**: We solve the problem of having groups with only 1 observation (schools with 1 student)
- **CON**: We have only 3 regions: Campania, Lombardia, Trentino-Alto Adige
  $\rightarrow$ We loose a lot of information: 7524 students out of 11583 (65%)

Clustering

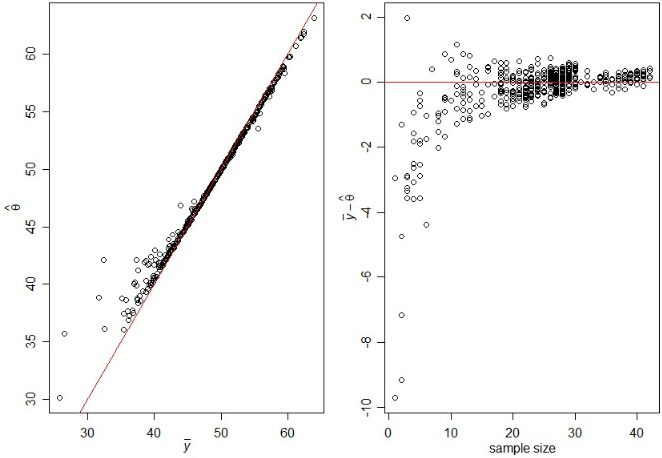
Goal: to find patterns among the characteristics of the schools, with respect to the mean evaluations in each school.

$$E\left(\theta_j \mid \overline{y}_j, \mu, \tau^2, \sigma^2\right) = \frac{n_j/\sigma^2}{n_j/\sigma^2 + 1/\tau^2}\overline{y}_j + \frac{1/\tau^2}{n_j/\sigma^2 + 1/\tau^2}\mu \qquad j{=}1 : ng$$

Where:

- $\overline{y}_j$ mean vote of each school for $j{=}1 : ng$
- $\mu \sim \mathcal{N}\left(\mu_0,\ \gamma_0^2\right)$ overall mean
- $\sigma^2 \sim \text{Inv-}\mathcal{G}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$ within group variance
- $\tau^2 \sim \text{Inv-}\mathcal{G}\left(\frac{\eta_0}{2}, \frac{\eta_0\tau_0^2}{2}\right)$ between group variance

Clustering



Shrinkage with respect to the grand mean

Clustering

$\Rightarrow$ Estimation a partition of the schools through a Gibbs sampler stratregy in a Mixture Model framework:

$$P \sim DP(a,\ P_0);\ P = \sum_{j=1}^{\infty} w_j \delta_{\theta_j^*}$$
$$\text{with} \quad \{\theta_j^*\} \overset{iid}{\sim} P_0$$
$$w_j = v_j \prod_{l<j}(1 - v_l)$$
$$v_j \overset{iid}{\sim} \mathcal{B}\text{eta}(1,\ a)$$

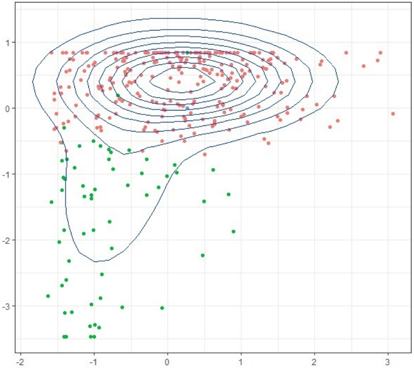
Approximated mixture:

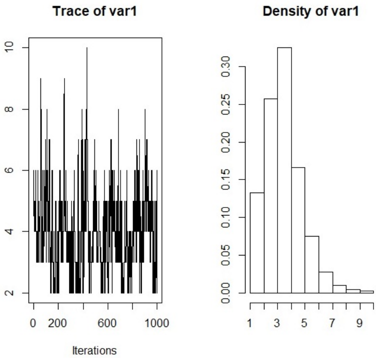
$$f(\boldsymbol{x}|\boldsymbol{w}, \boldsymbol{\theta}) = \sum_{j=1}^{M} w_j k(\boldsymbol{x}, \theta_j^*)$$
$$\text{with} \quad k(\boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}_p(\boldsymbol{x};\ \boldsymbol{\mu}, \Sigma)$$
$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \Sigma) \sim \mathcal{N}_p(\boldsymbol{\mu};\ \overline{\boldsymbol{x}}, \Sigma/k_0) \text{ x Inv-}\mathcal{W}(\Sigma;\ S_0^{-1}, \eta_0)$$

Clusters indexes: we obtain $\{z_i\}_{i=1:ng}$ such that:

$$P(z_i{=}j|\boldsymbol{w}, \boldsymbol{\theta}) = \sum_{l=1}^{M} 1_{\{z_i=j\}} w_l k(\boldsymbol{x}_i, \theta_l)$$

Clustering





A first trial with school size and certified teacher prop

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