Tracing the Trajectory of Math Proficiency in New York High Schools: A Hierarchical Bayesian Analysis (2011-2017)

Eugene Han and Elder Veliz

Yale University

Background

- Standardized testing and educational policies greatly influence student outcomes.
- Analysis of student performance is essential to understanding the efficacy of educational systems and guiding policy decisions.
- Our work explores how socioeconomic factors and resource allocation influenced math proficiency benchmarks across New York state's counties and schools.

Objectives

- Adopt a hierarchical Bayesian framework to explore how Title I eligibility, student-teacher ratios, and year-specific effects impacted student math proficiency in New York between 2011 and 2017.
- Account for school-level and county-level random effects in the modeling procedure to capture hierarchical structure in the data.
- Compare three linear models: (1) without random effects, (2) with school-level random effects, and (3) with school-level and county-level random effects.
- Interpret the results of our analysis and reveal descriptive insights into student math proficiency variations over time.

Overview of Data

- Data retrieved from the Urban Institute, aggregated from sources including the National Center for Education Statistics and the U.S. Department of Education.
- Data for 825 schools over 7 years across 61 counties, totaling 5775 observations.
- The dataset consisted of the following fields:
  - `year`: school year, between 2011 and 2017.
  - `ncessch`: National Center for Education Statistics (NCES) school identification number.
  - `county`: county code corresponding to each school.
  - `title_i_eligible`: dummy variable indicating Title I eligibility (low-income support).
  - `student_teacher_ratio`: engineered feature defined as (`teachers_fte_crdc` / `enrollment`).
  - `math_test_pct_prof_midpt`: midpoint of range reporting the share of students scoring proficient on a mathematics assessment, 0-100 scale.
- `teachers_fte_crdc` values were only observed in 2011, 2013, 2015, and 2017, so linear interpolation was used to impute missing values for 2012, 2014, and 2016.
- Values for `math_test_pct_prof_midpt` were divided by 100 and mapped to $\mathbb{R}$ using a logit transformation.

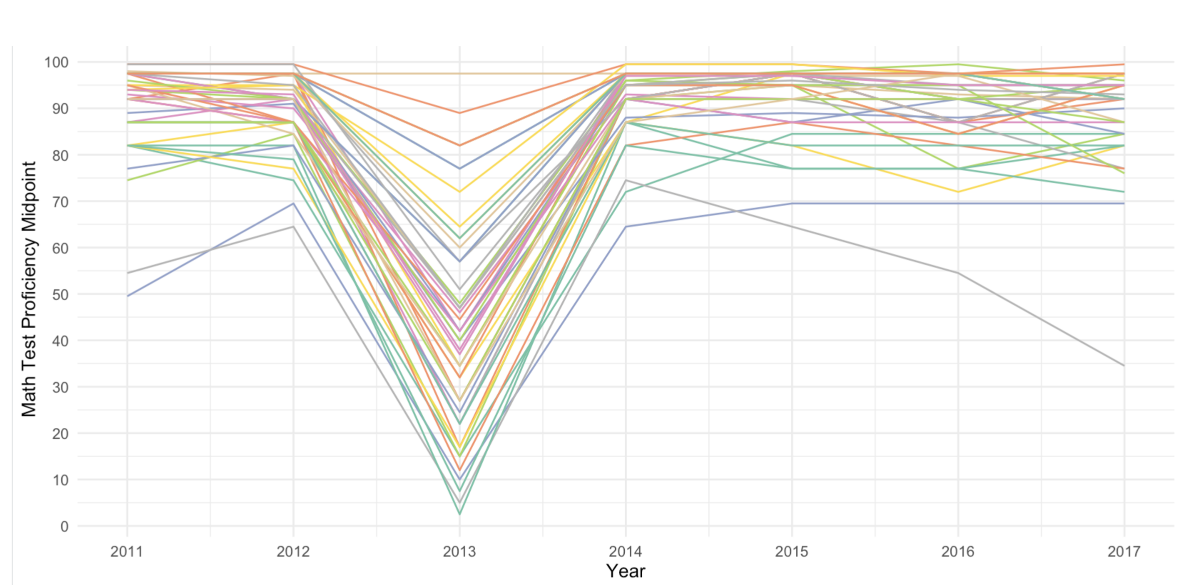


Figure 1. Math test proficiency midpoint for 50 randomly selected schools by year

Statistical Methods

1. Standard linear model (no random effects):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boxed{y_{it} \sim N(\boldsymbol{x}_{it}^T\boldsymbol{\beta}, \sigma_\varepsilon^2)}$$

2. Linear mixed effects model with school-level random effects:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

$$\boxed{y_{it} \sim N(\boldsymbol{x}_{it}^T\boldsymbol{\beta} + \alpha_i, \sigma_\varepsilon^2)}$$

3. Linear mixed effects model with school-level and county-level random effects:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\phi} + \boldsymbol{\varepsilon}$$

$$\boxed{y_{it} \sim N(\boldsymbol{x}_{it}^T\boldsymbol{\beta} + \alpha_i + \phi_{c(i)}, \sigma_\varepsilon^2)}$$

Priors:

$$\begin{gathered}\beta_j \sim N(0, 100^2)\\ \alpha_i \sim N(0, \sigma_\alpha^2)\\ \phi_{c(i)} \sim N(0, \sigma_\phi^2)\\ \sigma_\varepsilon^2, \sigma_\alpha^2, \sigma_\phi^2 \sim \text{Inv-Gamma}(0.01, 0.01)\end{gathered}$$

Results

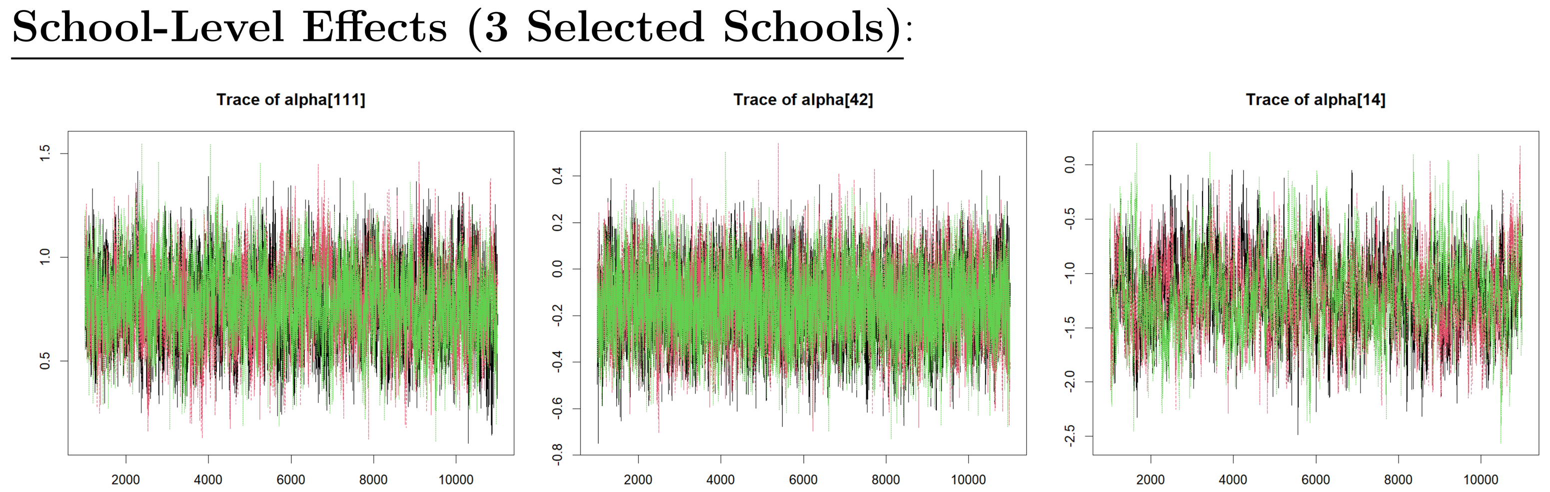
Below are the results of the model with **school- and county-level random effects**.

Fixed Effects:

Effect	Post. Mean	SD	Quantiles 0.025	Quantiles 0.975	ESS	G–R
(Intercept)	2.742	0.195	2.364	3.103	63.4	1.03
Year 2012	0.049	0.046	−0.042	0.139	4133.3	1.00
Year 2013	−3.098	0.046	−3.187	−3.007	4012.0	1.00
Year 2014	0.177	0.046	0.087	0.268	3382.7	1.00
Year 2015	0.044	0.048	−0.050	0.139	2132.3	1.00
Year 2016	−0.086	0.048	−0.180	0.008	2154.6	1.00
Year 2017	−0.221	0.049	−0.315	−0.125	2229.5	1.00
Title I Eligibility	−0.350	0.038	−0.424	−0.275	1839.6	1.00
Students/Teachers	0.201	0.017	0.169	0.234	10083.3	1.00

Table 1. Summary of posterior distributions for fixed effects with convergence diagnostics

School-Level Effects (3 Selected Schools):



(a) Bronx Collegiate Academy
Post. Mean = 0.775
Post. SD = 0.176

(b) Stuyvesant High School
Post. Mean = −0.148
Post. SD = 0.149

(c) Mattituck Junior-Senior HS
Post. Mean = −1.194
Post. SD = 0.347

Results (cont.)

County-Level Effects:

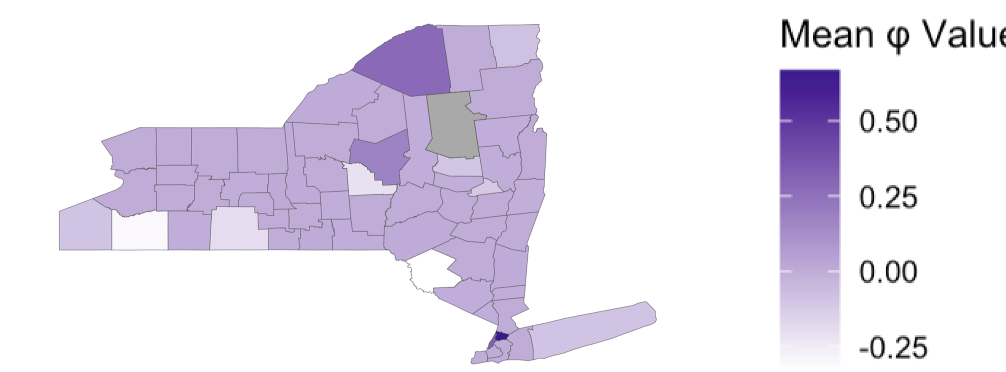


Figure 2. Heatmap of posterior mean county-level random effects

Model Comparison:

Model	Fixed 2013	Fixed Title I	Fixed S/T	Random (School) Bronx	Random (School) Mattituck
No random effects	−3.220	−0.844	−0.099	—	—
School-level	−3.188	−0.351	0.200	1.091	−1.622
School- and county-level	−3.187	−0.350	0.201	0.775	−1.194

Table 2. Comparison of posterior means for selected fixed and random effects across models

Model	DIC	pD
No random effects	17624	10.1
School-level	15641	123.2
School- and county-level	15640	122.6

Table 3. DIC analysis comparison

Conclusions

- The model with school-level and county-level random effects performed the best out of the three with respect to DIC.
- Bronx and New York County schools perform significantly better than expected of schools sharing similar characteristics.
- Common Core adoptions in 2013 negatively shocked midpoint math proficiency.
- The altered sign of the student-to-teacher ratio after including random effects implies its impact may be more related to individual school management rather than resource allocation differences.

Extensions and Further Questions

- Include additional socioeconomic predictors (e.g. teacher salaries, student household incomes, extracurricular activities) as fixed effects.
- Investigate spatial correlations to understand relationships between neighboring schools and counties. Is there evidence neighbors perform similarly?
- Analyze spatiotemporal relationships using historical academic performance data combined with geographic data. Do temporal dependencies alter our findings?

e.han@yale.edu, elder.veliz@yale.edu