Beyond Exclusion: The Role of High-Stake Testing on Attendance

Magdalena Bennett [1]

🐦 @maibennett
✉ m.bennett@austin.utexas.edu

Christopher Neilson[2] Nicolás Rojas[3]

[1] McCombs School of Business, The University of Texas at Austin
[2] Economics Department, Princeton University
[3] Teachers College, Columbia University

Introduction

Non-representative patterns of attendance can skew how useful test scores measures are for accomplishing their goal. The **main objectives** of this paper are the following:

1. Understand the **average effect of testing on school attendance** across grades and performance

2. Identify schools that **incentivize non-representative patterns** of attendance by **combining causal inference methods and machine learning**

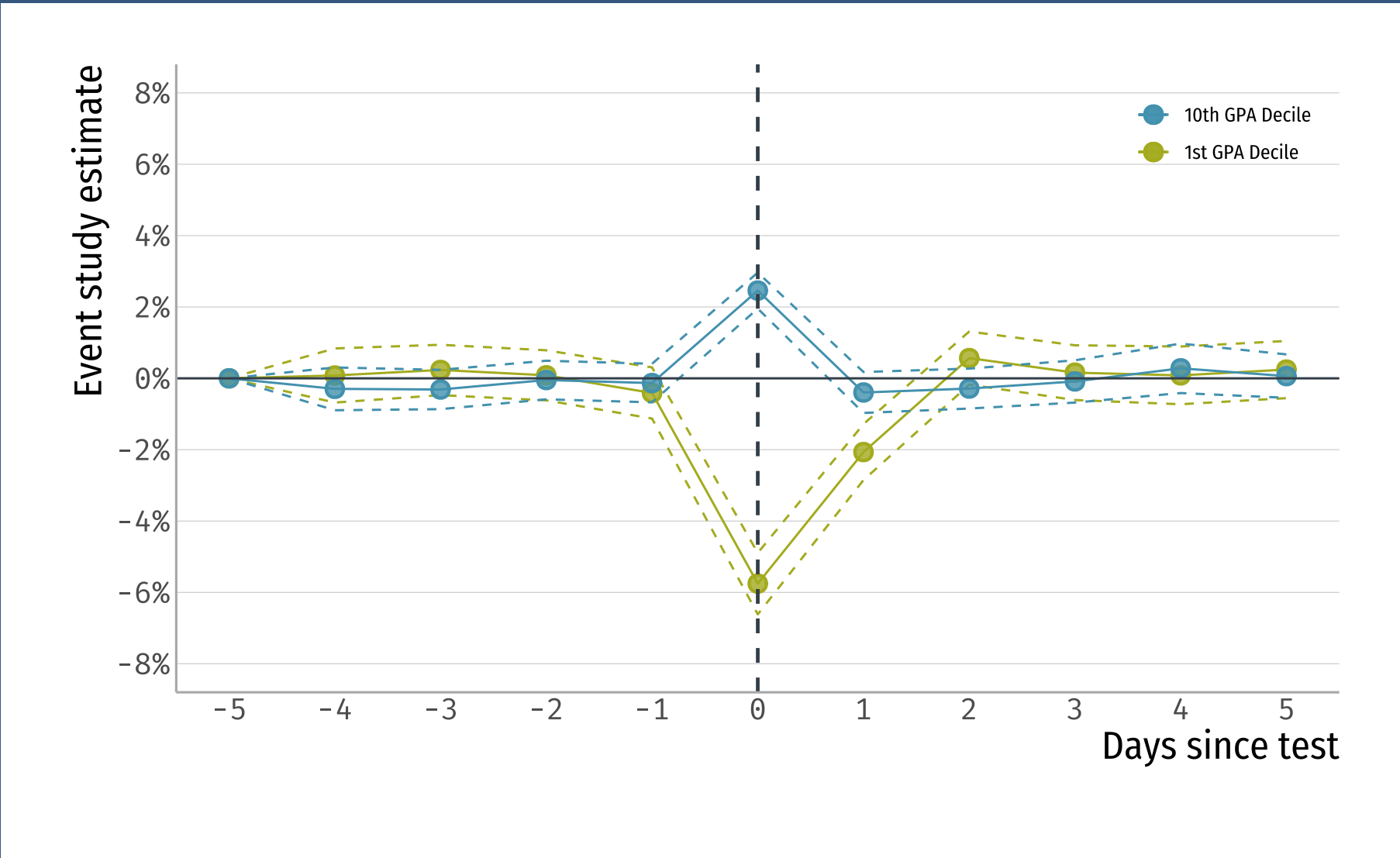3. Help improve current **imputation methods**

Methods

Event Study

$$Y_{ipsgt} = \sum_{P=1}^{5} \sum_{T=-4}^{5} \tau^{PT} D_{ipsgt}^{PTG^*} + \gamma_{pt} + \alpha_i + \epsilon_{ipsgt}$$

- Y_{ipsgt}: Attendance (1,0) for student i, from GPA group p, in school s and grade g for day t.

Not only **low-performers attend less** on the day of the test, but **high-performers attend more**

Using machine learning methods we can also **identify those schools more likely to incentivize low-attendance on bottom performers**


- $D_{ipsgt}^{PTG^*}$: Indicator variable where G^* is the tested grade.

Prediction of counterfactual attendance

- **XGBoost** for large student panel of daily attendance

  - Includes FE by day of the week, school, grade, and student. Also includes sybling's attendance (if any) and attendance lag.

- Identify types of schools by clustering on $Obs\ Att - Predicted\ Att$ using **K-means**.

- Test different **imputation policies** and its consequences.

Results

Students skip school on the day of the test. In lower grades, **lower-performers** attend less and **higher-performers** attend more, compared to a regular day. In higher grades, we **only observe action at the top of the distribution**
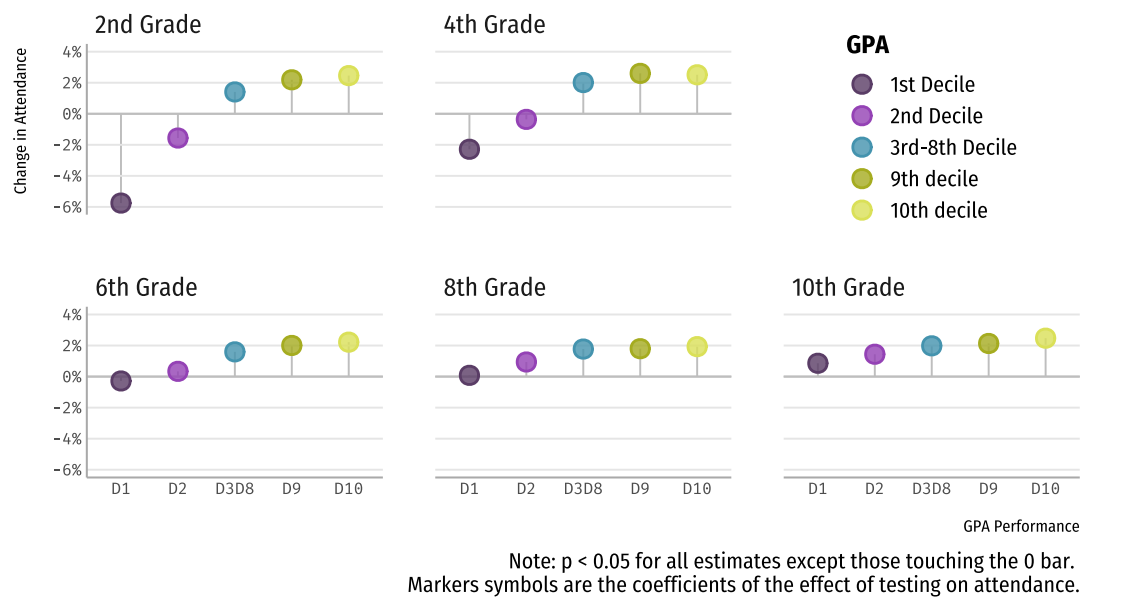
There is **important heterogeneity between schools**.

Using **K-means** analysis to identify **clusters of schools** according to their **difference between predicted and observed attendance distribution**. We find **two main clusters**, where one of them incentivize the **exclusion of lower-performers**. Those schools are more vulnerable and have overall lower performance.