Does Class Size Reduction Close the Achievement Gap? Evidence from TIMSS 2011

Wei Li
Spyros Konstantopoulos
Michigan State University
Class Size and Student Achievement

Class size reduction as a way of improving student achievement
- Project STAR: Finn & Achilles, 1990; Krueger, 1999; Nye, Hedges, & Konstantopoulos, 2000

Class size reduction as a way of close achievement gap
- Assumption: Teachers spend more time on instructing students at-risk in smaller classes (Correa, 1993; Finn & Achilles, 1990; Nye, Hedges, & Konstantopoulos, 2002)
Class Size and Student Achievement Gap

One way to evaluate whether class size reduction closes achievement gap is to estimate the differential class size effects across student achievement distribution using quantile regression.

- **Project STAR**
  - High-achieving students benefited more from small classes in kindergarten, first grade and second grade (Ding & Lehrer, 2011; Jackson & Page, 2013; Konstantopoulos, 2008).
  - Low-achieving students benefited more from small classes in fourth and sixth grade (Konstantopoulos & Chung, 2009).

- **Differential class size effects in Dutch primary schools**
  - Levin (2001) found no class size effects for either low- or high-achieving students.
  - Ma and Koenker (2006) re-analyzed Levin’s data and found, for mathematics scores, lower-achieving students benefited more from smaller classes.

No recent study has used current data to evaluate whether class size reduction closes achievement gap between low- and high-achieving students.
Purpose of Our Study

- Explore the causal effects of class size on student mathematics achievement through quasi-experimental design that utilizes the maximum class size rules
- Evaluate class size effects across achievement distribution through quantile regression
- Compare class size effects between low- and high-achieving students and evaluate whether class size reduction enlarges or closes achievement gap
Data: TIMSS 2011

TIMSS: Trends in International Mathematics and Science Study
- Measures mathematics and science achievement at fourth and eighth grade
- Conducted every four years since 1995; the latest one is TIMSS 2011
- Stratified two-stage sampling, national probability sample
- Collects extensive information about students, teachers and schools
- 14 European countries that had clear rules about maximum class size limits for fourth graders in 2011
Measures

Dependent Variable:
- Fourth Grade Mathematics Achievement (five plausible values)

Independent Variables:
- Class size from teacher questionnaire
- Student variables
  - Gender, age, SES, language spoken at home, etc.
- Classroom/Teacher variables:
  - Teacher experiences, gender, education level, instruction time
  - Peer characteristics such as percent females, and mean-SES
- School variables from school survey:
  - Grade four enrollment, percent low-SES students, etc.
- Missing data flags
Quantile Regression

Allow researchers to evaluate class size effect across achievement distribution

For example, the class size coefficients at 75\textsuperscript{th}-quantile shows the correlation between class size and achievement for students at the 75\textsuperscript{th}-quantile of achievement distribution

\[
Score_i = \beta_0 + \beta_1 ClassSize_i + ST_i B_2 + CL_i B_3 + SC_i B_4 + \varepsilon_i \quad (1)
\]
Challenge: Omitted Variables

Class size effects produced using observational data may be biased because student and teacher allocation to classes within schools is likely non-random.

Variation of class size might reflect differences of students’ background such as student ability, motivation, parents’ education, family income, etc., which were not provided in TIMSS.
Instrument or Instrumental Variable (IV)

- Highly correlated with reported class size
- Independent of unobserved variable captured by the error term in equation (1)
Maximum Class Size Rules

One promising instrument is to take advantage of the maximum class size rule that is available in most European countries (Angrist & Lavy, 1999).

In particular, we computed the school and grade specific average class size based on the maximum class size requirement as our instrument variable.
An Example

- For example, if the maximum class size rule is 30
  - Enrollment = 20, one class with 20 students
  - Enrollment = 30, one class with 30 students
  - Enrollment = 40, two classes with 20 students
Two-Stage Estimation

- We adopted the control function approach proposed by Lee (2007) to get quantile-specific IV estimates.
- The control function approach is also a two-stage estimation method that is similar to two-stage-least square (2SLS).
  - The first stage is to estimated the proxy of omitted variables- a control variable.
  - The second stage is to add the control variable (or a function of control variable) into equation (1) to get the quantile-specific IV estimates.
- The basic idea is to add a control variable to equation (1) such that, once we condition on this variable, the teacher reported class size will be independent of omitted variables (see Wooldridge, 2010).
First-Stage Linear Regression

We use OLS regression in the first-stage analysis.

\[
\text{ClassSize}_i = \pi_0 + \pi_1 IV_i + ST_i \Pi_2 + CL_i \Pi_3 + SC_i \Pi_4 + u_i \quad (3)
\]

\[
\hat{u}_i = \text{ClassSize}_i - \text{ClassSize}_i, \text{ the control variable}
\]
Second-Stage Quantile Regression

\[ \text{Score}_i = \delta_0 + \delta_1 \text{ClassSize}_i + \lambda(\hat{u}_i) + \text{ST}_i \Delta_2 + \text{CL}_i \Delta_3 + \text{SC}_i \Delta_4 + \xi_{ii} \quad (4) \]

where \( \lambda(\hat{u}_i) \) is a fifth order polynomial of \( \hat{u}_i \)

- Student weights were used to make projections to the population.
- Bootstrap standard errors were used to correct for the two-stage approach and the possible clustering effects (students nested in schools).
- We firstly did the analysis separately for each plausible value. Then, we combined estimates and SEs across the five analysis using multiple imputation formulas by Schafer (1999)
Did schools follow the rule?

Computed Class Size \rightarrow Class Size \rightarrow Score

\[ \text{ClassSize}_i = \pi_0 + \pi_1 IV_i + ST_i \Pi_2 + CL_i \Pi_3 + SC_i \Pi_4 + u_i \]  (3)
<table>
<thead>
<tr>
<th>Countries with Strong IV</th>
<th>Countries with Weak IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria 5.23</td>
<td>Czech Republic 7.94</td>
</tr>
<tr>
<td>Germany 5.06</td>
<td>Spain 8.96</td>
</tr>
<tr>
<td>Hungary 5.28</td>
<td>Lithuania 5.49</td>
</tr>
<tr>
<td>Potrugal 3.38</td>
<td>Romania 7.38</td>
</tr>
<tr>
<td>Slovak Republic 6.62</td>
<td>Slovenia 5.52</td>
</tr>
<tr>
<td>Coratia 0.72</td>
<td>Denmark 2.54</td>
</tr>
<tr>
<td>Italy 1.70</td>
<td>Malta 1.71</td>
</tr>
</tbody>
</table>
Did the IV Influence Student Achievement only through Class Size?

Possible ways of manipulation
- Parents manipulated the rules: Angrist & Lavy (1999)
- School manipulated the rules: Urquiola & Verhoogen (2009)
One advantage of our IV method is that it could also be seen as a regression discontinuity design (Angrist & Lavy, 1999).

Regression Discontinuity Design

Austria Rule=25

Italy Rule=26
Regression Discontinuity Design

Regression discontinuity (RD) design is a quasi-experimental design

- A valid RD design approximates a local randomized experiment
- Schools with fourth grade enrollment just below the cutoffs are good comparisons to those just above the cutoffs
- In other words, schools close to the cut-offs should be quite similar in terms of school, teacher or student characteristics
We could test the validity of the RD design and the validity of the IV through examining if the observed covariates are "locally" balanced on either side of the cut-offs.

- If parents or schools manipulated the maximum class size rule, schools’ characteristics and their students’ characteristics should be systematically different.

We restricted the sample to students in schools at most five students away from the cut-offs.

- If the rule is 30, the +/- 5 subsample is schools with fourth graders enrollment between [26,35], [56,65], [86, 95], etc.
Results: Checking Local Balance

We checked the local balance of 46 variables (23 covariates and 23 missing data flags)

- For most of the countries—twelve out of fourteen, only four or less variables were unbalanced around cutoffs
- Malta: 10 variables were unbalanced
- Spain: 13 variables were unbalanced
Countries with Strong and Valid IV

- Countries with weak IV: Croatia, Denmark, Italy, and Malta
- Countries with invalid IV: Malta and Spain
- Other countries: Austria, the Czech Republic, Germany, Hungary, Lithuania, Portugal, Romania, the Slovak Republic, and Slovenia
Table 2. 2SLS and Quantile Regression Estimates and Standard Errors of Class Size

<table>
<thead>
<tr>
<th>Country</th>
<th>2SLS</th>
<th>Quantile</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10th</td>
<td>25th</td>
<td>50th</td>
<td>75th</td>
<td>90th</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>-1.82</td>
<td>0.35</td>
<td>-2.07</td>
<td>-2.26</td>
<td>-2.51</td>
<td>-1.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(3.37)</td>
<td>(3.04)</td>
<td>(2.04)</td>
<td>(2.15)</td>
<td>(2.79)</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.24</td>
<td>0.83</td>
<td>0.67</td>
<td>0.75</td>
<td>-0.10</td>
<td>-0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(1.74)</td>
<td>(1.48)</td>
<td>(1.28)</td>
<td>(1.56)</td>
<td>(1.48)</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>1.33</td>
<td>2.25</td>
<td>2.28</td>
<td>2.26</td>
<td>1.58</td>
<td>-0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(2.28)</td>
<td>(1.97)</td>
<td>(1.68)</td>
<td>(2.16)</td>
<td>(2.42)</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>0.45</td>
<td>1.31</td>
<td>-0.50</td>
<td>0.70</td>
<td>1.19</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
<td>(2.65)</td>
<td>(2.24)</td>
<td>(1.78)</td>
<td>(1.92)</td>
<td>(2.44)</td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>-1.25</td>
<td>-0.06</td>
<td>-0.15</td>
<td>-0.79</td>
<td>-2.15</td>
<td>-2.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.28)</td>
<td>(3.27)</td>
<td>(1.65)</td>
<td>(1.75)</td>
<td>(1.71)</td>
<td>(2.41)</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>-3.80</td>
<td>-2.84</td>
<td>-2.89</td>
<td>-2.89</td>
<td>-3.05</td>
<td>-4.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(4.63)</td>
<td>(3.10)</td>
<td>(2.78)</td>
<td>(3.10)</td>
<td>(3.69)</td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>-4.84*</td>
<td>-5.46</td>
<td>-5.52</td>
<td>-5.72*</td>
<td>-4.86</td>
<td>-4.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.28)</td>
<td>(3.63)</td>
<td>(3.71)</td>
<td>(2.64)</td>
<td>(2.92)</td>
<td>(3.46)</td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>-4.40*</td>
<td>-4.42*</td>
<td>-3.59</td>
<td>-4.10</td>
<td>-4.68*</td>
<td>-4.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(2.24)</td>
<td>(2.57)</td>
<td>(2.19)</td>
<td>(2.16)</td>
<td>(2.78)</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>-1.87</td>
<td>-1.13</td>
<td>-1.69</td>
<td>-2.03</td>
<td>-2.65</td>
<td>-2.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td>(2.69)</td>
<td>(2.09)</td>
<td>(1.86)</td>
<td>(2.31)</td>
<td>(2.52)</td>
<td></td>
</tr>
</tbody>
</table>

*p ≤ .05

Note: Bootstrap standard errors are in parentheses.
Conclusion and Limitation

- We investigated the differential effects of class size at different levels of mathematics achievement for fourth graders in 14 European countries.
  - In general, no differential class size effects were found.
  - Significant and negative class size effects were found in Romania and the Slovak Republic.
  - No evidence to show if class size reduction would enlarge or close achievement gap.

- Limitation:
  - Weak IV
  - Measurement error of fourth grade enrollment.
Thank you!

Comments are welcome:
Wei Li
liwei8@msu.edu