ECON 398 – Professor Underwood

Emperical Project Data and Methods

Kien Vu

**1. Conceptual Model:**

This paper examinesteachers’ turnover in grade schools and aims to understand how teacher turnover affects student achievement. Using panel data on some Massachusetts school districts, we provide more evidence to suggest that *when controlling for the teachers’ quality and the student’s characteristics, changes in school environment contribute to a significant relationship between low turnover rates and good student performance.*

The existing literature proposes two mechanisms driving this relationship, one based on the change in the compositional makeup of the teachers, while another accounts for the disruptive effects on performance even with comparable teaching cohorts. The model here, while also controlling for certain measures of teacher quality, includes year and school fixed effects so as to account for time-invariant school characteristics and school-invariant shocks over time. We aim to further isolate the effects of teacher retention on students by controlling for other variables affecting specific schools over time than teacher retention, in this case including many variables of teaching environment such as leadership turnover, changes in class or student characteristics, and financial support.

**2. Data:**

We use a longitudinal data from the Massachusetts Department of Elementary and Secondary Education accessed through their School and District Profiles online reports. The data includes 360 observations representing 90 school districts in Massachusetts yearly, over a four-year period from the 2012-2013 school year to 2015-2016 school year. For each observation, the data includes information within 19 variables regarding district staff retention, Mathematics class size, composition, and teacher quality, Mathematics MCAS exam results, and monetary data regarding teacher salary and funding for schools in the district. Table 1 provides descriptions and summary statistics for these variables below

**Table 1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Obs | Unique | Mean | Min | Max | Label |
| distname | 360 | 91 | . | . | . | District Name |
| distcode | 360 | 90 | 4564222 | 10000 | 9150000 | District Code |
| year | 360 | 4 | 2014.5 | 2013 | 2016 | Year of interest |
| retperc\_p | 360 | 29 | 81.52528 | 0 | 100 | Principal % Retained |
| retperc\_t | 360 | 133 | 89.3875 | 76.4 | 97.9 | Teacher % Retained |
| classsize | 360 | 105 | 17.35361 | 9.6 | 22.8 | Average Class Size |
| stfemale | 360 | 124 | 48.4225 | 31.4 | 71.5 | Female % |
| ellperc | 360 | 80 | 2.001944 | 0 | 14.1 | English Language Learner % |
| poorperc | 360 | 258 | 22.88361 | 1.8 | 83.7 | Economically Disadvantaged % |
| cpi | 360 | 197 | 85.13278 | 66.3 | 98.6 | CPI for MCAS Math |
| sgpmed | 360 | 59 | 51.24306 | 23.5 | 78 | Median SGP for MCAS Math |
| licensed | 360 | 38 | 99.31556 | 60.4 | 100 | % of Teachers Licensed in Teaching Assignment |
| corequality | 360 | 52 | 98.92389 | 55 | 100 | % of Core Academic Classes Taught by Teachers Who are Highly Qualified |
| discp | 360 | 360 | 4.455818 | 0.08312 | 21.2707 | % of students disciplined any offenses |
| sdiscp | 360 | 356 | 1.646699 | 0 | 8.05687 | % of students disciplined for drugs, violent or criminal-related offenses |
| lwage | 360 | 360 | 11.19982 | 10.9505 | 11.4585 | Log of Average Teacher Salary |
| lin\_exppp | 360 | 360 | 9.586814 | 9.14334 | 10.7464 | Log of In-District Expenditures per Pupil |
| ltot\_exppp | 360 | 360 | 9.617603 | 9.21668 | 10.4286 | Log of Total Expenditures per Pupil |

Most of the variables’ values are as anticipated, with wide variations and some skewness – examples are discipline, ELL/poverty rates skewing right, and the two quality measures *licensed* and *corequality* which skews left. The former two is predictable as certain schools are more segregated than the mean, and there are enough samples at the tails to include in the analysis. Meanwhile, the two quality variables have more extreme skews, which means that different groups are to be tested separately.

However, there are certain limitations in the data that might be obstacles to effectively estimating models. Firstly, there are not sufficient data and organization on schools and at each grade, thus the data are for the districts and at all grades. Not only are grouping all the grades and schools in a district detrimental to the significance of the results, the models also cannot separate the effects caused by the dynamics of turnover at each school. Moreover, the teacher salaries and expenditures measures are not for Mathematics specifically, which might mean unrealistic changes attributable to other subjects’ teacher or aspects of the school budget. Finally, having data from more years would be more useful in studying longer-term effects of sustained low retention.

**3. Model Specifications:**

Our dependent variables, which are the variables indicating student performance, are measures from the Mathematics MCAS results for each school district each year. This includes *cpi*, which is the average value of points, over 100, earned on the MCAS tests by students in that district. Another measure is *sgpmed*, which indicates the district median SGP value over 100, which is the change in a group of students’ Mathematics MCAS achievement over time, with values below 30 indicating decline and values above 50 as being on target. For each of these outcome variables, we have our regression equation as below:

can be either of the two measures *cpi* and *sgpmed* at district *i* at year of interest *s.* is our explanatory variable of interest, the percentage of teachers retained in a district *i* at year of interest *s* from the year before. is a vector of time-varying characteristics of the school districts, while are the school and year fixed effects. By using the fixed effects in the model, we can control for time-invariant factors and time-varying shocks to all the districts and estimate the coefficient of interest with only the time-varying, district-specific variations. The controls vector in the model include Math class composition variables (*classsize, stfemale, ellperc, poorperc*), teacher quality (*licensed, corequality*), funding measures (*lwage, lin\_exppp* or *ltot\_exppp*), principal retention (*retperc\_p*), and disciplinary measures (*discp* or *sdiscp*). Information on the variables in this equation are available in Table 1.

For the coefficient of interest , we can interpret its estimations as the net effects of increasing the retention rate of teachers in a district by one percentage point on the Math MCAS performance of the same district the next year. There has been plenty of evidence relating teacher turnover rate with student performance, with the majority showing that high turnover affects student achievements negatively (Sorensen and Ladd 2020; Ronfeld 2013). Even as district and year fixed effects are included, the negative relationship is likely to hold due to other disruptive effects of teacher change in a school such as peer spill-over effects or other dynamic factors (Jackson and Bruegmann 2009). Certain variables already controlled in the model can have a quite significant coefficient, however, as Sorensen and Ladd (2020) summarized from past research significant relationships of the variables of interest with factors of teacher credentials, class size, and student discipline as well.

Bibliography

Ronfeldt, M., S. Loeb, and J. Wyckoff. “How teacher turnover harms student achievement.” *American Educational Research Journal* 50, no. 1 (2013): 4-36.

Sorensen, L. C., and H. F. Ladd. "The Hidden Costs of Teacher Turnover." *AERA Open* 6, no. 1 (2020).

Jackson, C. K., and E. Bruegmann. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." American Economic Journal: Applied Economics 4, no. 1 (2009): 85–108.