A COMPARISON OF GROWTH PERCENTILE AND VALUE-ADDED MODELS OF TEACHER PERFORMANCE

Cassandra Guarino
Mark Reckase
Brian Stacy
Jeffrey Wooldridge

October 10, 2013

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305D100028 and R305B090011 to Michigan State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Different approaches to computing teacher performance measures

- States and districts have an array of methods to choose from in computing teacher performance measures
- Often the choice is based on simply adopting another state or district's approach
- Growth percentile models have enjoyed popularity among decision makers in policy implementation, whereas value-added models tend to be more widely used by researchers
- This study offers a systematic comparison of the two types of approaches

Value-Added Models (VAMs)

- Value-added models generally estimate teacher effects using linear regression

- A common regression formulation is the following:

$$A_{it} = \lambda A_{i,t-1} + X_{it}\gamma + \beta_1 tch_1 + \ldots + \beta_J tch_J + \nu_{it}$$

- Because we include teacher dummy variables in the regression, we say that we are treating teacher effects as fixed

- The coefficients β_j on the teacher dummies are interpreted as representing the contribution of teacher j to student achievement

- An advantage of this formulation is that if students are grouped in classrooms based on their prior achievement ("tracked"), then we are controlling for the classroom assignment mechanism and "partialing" it out of the teacher effect estimates

Some popular value-added models do not control for assignment based on test scores

- Other formulations are often used in which the teacher effects are treated as random—i.e., uncorrelated with prior test scores or any other covariates, such as demographics
- For example, residual-based methods are very popular especially with the addition of shrinkage (e.g., MET project, Chetty et al. 2012)

$$A_{it} = \lambda A_{i,t-1} + X_{it}\gamma + \nu_{it}$$

- Simulations (Guarino, Reckase & Wooldridge, forthcoming) show that residual based methods are inferior to teacher fixed effects specifications when nonrandom assignment exists

Growth percentile models

- Growth models estimate the relative position of teachers in the distribution of effectiveness rather than teacher effects

- Effectiveness is defined as the mean or median conditional percentile growth of students in a teacher's class

- We consider two types of growth percentile models
  - Colorado Growth Model (Betebenner 2011)
  - Nearest Neighbor Matching Model (Barlevy & Neal 2012; Fryer, Levitt, List & Sadoff 2012)

Colorado Growth Model (CGM)

- Four main steps
  - Using student level data, run 100 quantile regressions of current achievement on prior achievement to obtain predicted growth values at each percentile for each student
  - Locate the predicted value closest to the actual value of current achievement and assign the student that growth percentile
  - For each teacher, compute the mean or median assigned growth percentile for his or her students
  - Rank the teachers
- Assumes teacher effects are random—i.e., uncorrelated with prior test scores

Nearest Neighbor Matching Model (NNMM)

- Four main steps
  - For each student find the nine closest matches in other schools based on prior test scores, using a Mahalanobis distance measure
  - Determine where the student ranks within the overall set of ten
  - Compute the average of the student ranks per teacher
  - Rank the teachers
- Assumes random teacher effects

Methodology to compare models

- We simulate student test scores based on known teacher effects
- Estimate the teacher effect as if we did not know it, using both value-added and growth models
- Then, for each estimation approach, we compare the estimates of the teacher effects to the true teacher effects

Data generating process to create student test scores

- Student test scores are generated as follows, where A_{i3} is achievement of child *i* in grade 3:

$$A_{i3} = \lambda A_{i2} + \beta_j + c_i + e_{i3}$$

- We generate the following from normal distributions:
  - A_{i2} base score for each student
  - β_j effect for teacher *j* (constant over time)
  - c_i student effect
  - e_{i2} random deviation for each student

Simulated data structure

- We create the following student-level data structure:
  - A single grade
  - 10 schools
  - 4 teachers per school
  - Class size of 20 students
  - Each teacher teaches 4 cohorts of students
  - Students and teachers stay in the same school – no crossover

We examine four estimation approaches

1) OLS-Lag – Lag-score equation with teacher dummies

2) AR – Average residuals per teacher from lag-score equation

3) CGM – Colorado Growth Model

4) NNMM – Nearest neighbor matching model

We examine different classroom assignment scenarios

- We vary the way students are grouped into classrooms and the way classrooms are assigned to teachers

- Grouping:
  - Random
  - Grouping based on prior test scores (“tracking”)

- Assignment:
  - Random
  - Nonrandom

We evaluate the various estimators

- For each scenario and estimator combination, we run 100 replications
- We then evaluate the various estimators by looking at the average Spearman rank correlation of the true and estimated teacher effects

Spearman rank correlations of estimates with true teacher effects

	OLS-Lag	AR	CGM	NNMM
Random grouping Random assignment	.88	.88	.83	.85
Tracked Nonrandom assignment	.88	.78	.71	.75

Sensitivity: Alternative error distribution

- Since quantile regression is less sensitive to outliers than OLS, we also ran a simulation in which the error term in the achievement equation was generated from a fatter tailed distribution
  - We used the t with 3 degrees of freedom instead of the normal

- Results were that under random assignment, the CGM did perform very slightly better than OLS-Lag

- However, it still performed substantially worse under nonrandom assignment

Analysis with data from anonymous district

- Data
  - Large diverse anonymous school district
  - 215,411 usable observations
  - Grades 5 and 6
  - 5,666 teacher-year observations

- Estimation
  - Estimate teacher effectiveness using OLS-Lag, EB-Lag, and CGM
  - Included two prior years of student test scores

Correlations of teacher effect estimates across estimators

	OLS-Lag	EB-Lag	CGM
OLS-Lag	1		
EB-Lag	0.96	1	
CGM	0.71	0.68	1

Tracking and non-tracking schools

- Simulations showed that VAMs and CGM gave similar teacher rankings under random grouping assignment but different rankings under non-random grouping and assignment

- In data, we implement MNL test of Dieterle et al. (2012) for tracking based on prior test scores

- Break sample up into teachers at schools with evidence of tracking and teachers at schools with no evidence of tracking

- Produce separate correlations for these two groups

Correlations across estimators in tracking and non-tracking schools

Schools with tracking

	OLS-Lag	EB-Lag	CGM
OLS-Lag	1		
EB-Lag	0.95	1	
CGM	0.70	0.65	1

Schools with random grouping

	OLS-Lag	EB-Lag	CGM
OLS-Lag	1		
EB-Lag	0.97	1	
CGM	0.75	0.72	1

Conclusion

- Under the simulation conditions investigated, we find:
  - Growth percentile models perform fairly similarly to value-added models under random assignment of students to teachers
  - Under nonrandom assignment, value-added models that treat teacher effects as fixed outperform models that treat teacher effects as random, including growth models

- In actual data from a large district, we find:
  - Correlations across estimators differed more in schools with tracking than in those without (as predicted by the simulations)

END

Distributions from which student and teacher effects were drawn

- Student base achievement scores are drawn from:
  - Normal(0,1)
- Student fixed effects drawn from:
  - Normal(.5,.5)
- Teacher effects are drawn from:
  - Normal(.5,.25)
- Random deviation for each student is drawn from:
  - Normal(0,1) --- also t_3

Correlations across estimators in tracking and non-tracking schools

Schools with tracking

Variables	OLS-Lag	EB-Lag	CGM
OLS-Lag	1		
EB-Lag	0.95	1	
CGM	0.70	0.65	1
Obs.	3,672	3,672	3,672

Schools with random grouping

Variables	OLS-Lag	EB-Lag	CGM
OLS-Lag	1		
EB-Lag	0.97	1	
CGM	0.75	0.72	1
Obs.	1,876	1,876	1,876