RAND EDUCATION

Value-Added Models: Analytic Issues

Daniel F. McCaffrey & J.R. Lockwood

November 14, 2008

The Goal of Value-Added Modeling

- Use longitudinal administrative student test score data and complex statistical models to estimate the effects of educational inputs on achievement
- Estimates ideally would be
  - Unconfounded by characteristics of the students, or their families, communities, or prior educational experiences
  - Precise (very little sampling error)
- We focus primarily on estimating the effects of individual teachers

Simple Modeling Approaches Are Less than Ideal

- General agreement that very simple models are likely to fail to remove bias
- E.g. linear regression on a single prior score is particularly susceptible to confounding by student inputs
- McCaffrey et al. (2008) found this method was much more likely to identify teachers as significantly better than average than other methods
  - And the identified teachers tended to be teaching advanced courses with students with high prior achievement
- Simple methods can make inefficient use of data, often excluding student records that other methods include
  - Comparisons by McCaffrey et al. (2008) found this to be true of simple average gain scores but not of simple regression models

No Consensus on Which Complex Method Is Best

- Two main perspectives: statistical and econometric
- Developed somewhat independently in separate literatures
- Estimators from both perspectives:
  - Can be unbiased even when students are heterogeneously grouped to classes provided other assumptions hold
  - Can have low precision
  - Generally have been found to be unstable across years
- Complex methods from both perspectives have additional challenges of implementation and transparency, which can negatively impact scale-up and buy-in

Statistical Approach to VAM Describes Achievement Trajectories Using Accumulation of Random Components

- Simple example for achievement trajectory $(A_1, A_2, \ldots, A_4)$ for a single student over four years:

$$\begin{aligned} A_1 &= \mu_1 + \theta_{1,1} + \epsilon_1 \\ A_2 &= \mu_2 + \theta_{1,2} + \theta_{2,2} + \epsilon_2 \\ A_3 &= \mu_3 + \theta_{1,3} + \theta_{2,3} + \theta_{3,3} + \epsilon_3 \\ A_4 &= \mu_4 + \theta_{1,4} + \theta_{2,4} + \theta_{3,4} + \theta_{4,4} + \epsilon_4 \end{aligned}$$

- $\theta_{g,t}$ captures commonalities of scores at grade $t >= g$ for students sharing a teacher in grade g

Particular Specifications of Model Make Different Assumptions about Random Components

- $\theta_{g,t}$:
  - **"Complete persistence"** $\theta_{g,t} = \theta_g$ **(Sanders et al.; Raudenbush and Bryk)**
  - **"Variable persistence"** $\theta_{g,t} = \alpha_{g,t}\theta_g$ **(McCaffrey et al.; Lockwood et al.)**
  - **"Generalized persistence"** $\theta_{g,t}$ **unstructured (Mariano et al.)**
- ϵ_t:
  - **Unstructured covariance (Sanders et al.; McCaffrey et al.; Lockwood et al.)**
  - **Random growth model (Raudenbush and Bryk)**

Statistical Models and Heterogenous Groupings

- Model is not structural - identifies only shared sources of variance
- Assumes residual errors in student outcomes are independent of other students except for teacher effects
- Assumes teacher effects are independent of student residuals
  - Appears to confound teacher effects with student heterogeneity across classes
- However, correlation among repeated measures on the same student can mitigate this confounding (Lockwood and McCaffrey, 2007)
  - Implicit regression adjustment
  - Requires selection to depend on low-dimensional vector of time invariant student factors but factors can be differentially related to different tests
  - Requires possibly large number of tests

Strengths and Weaknesses of Statistical Models

Strengths

- Relatively flexible assumptions about persistence of past inputs and test score scaling provide good fit to data
- By default use shrinkage estimates which can improve precision of estimates
- Expanding the model does not require new estimation procedures

Weaknesses

- At best can identify effects of assignment to a classroom because they do not attempt to disentangle true individual teacher effects from classroom context
- Assessing the properties of estimators can be difficult and heuristic statements about estimators can be misleading

Econometric Approach Starts with Cumulative Education Production Function

$A_t = f_t(\textbf{family history}, \textbf{schooling history}, \textbf{innate ability}, \textbf{residual error})$

- **Make sequence of assumptions that allow estimation of teacher effects using standard least-squares estimators**
- **E.g. assuming all inputs to achievement decay at the same rate λ leads to simplified version:**

$$A_t = (\boldsymbol{S}_t, \boldsymbol{T}_t, \boldsymbol{P}_t, \boldsymbol{Z}_t)'\boldsymbol{\beta} + \phi + \lambda A_{t-1} + \xi_t$$

where $(\boldsymbol{S}_t, \boldsymbol{T}_t, \boldsymbol{P}_t, \boldsymbol{Z}_t)$ are school, teacher, peer and other classroom variables at time t, ϕ is a student fixed effect and ξ_t is a residual error

- **Often assumed that $\lambda = 1$ (no decay)**

Strengths and Weaknesses of Econometric Models

Strengths

- Model is structural - provided that assumptions hold, can identify causal effects
- Can disentangle effects of teachers from certain contextual factors

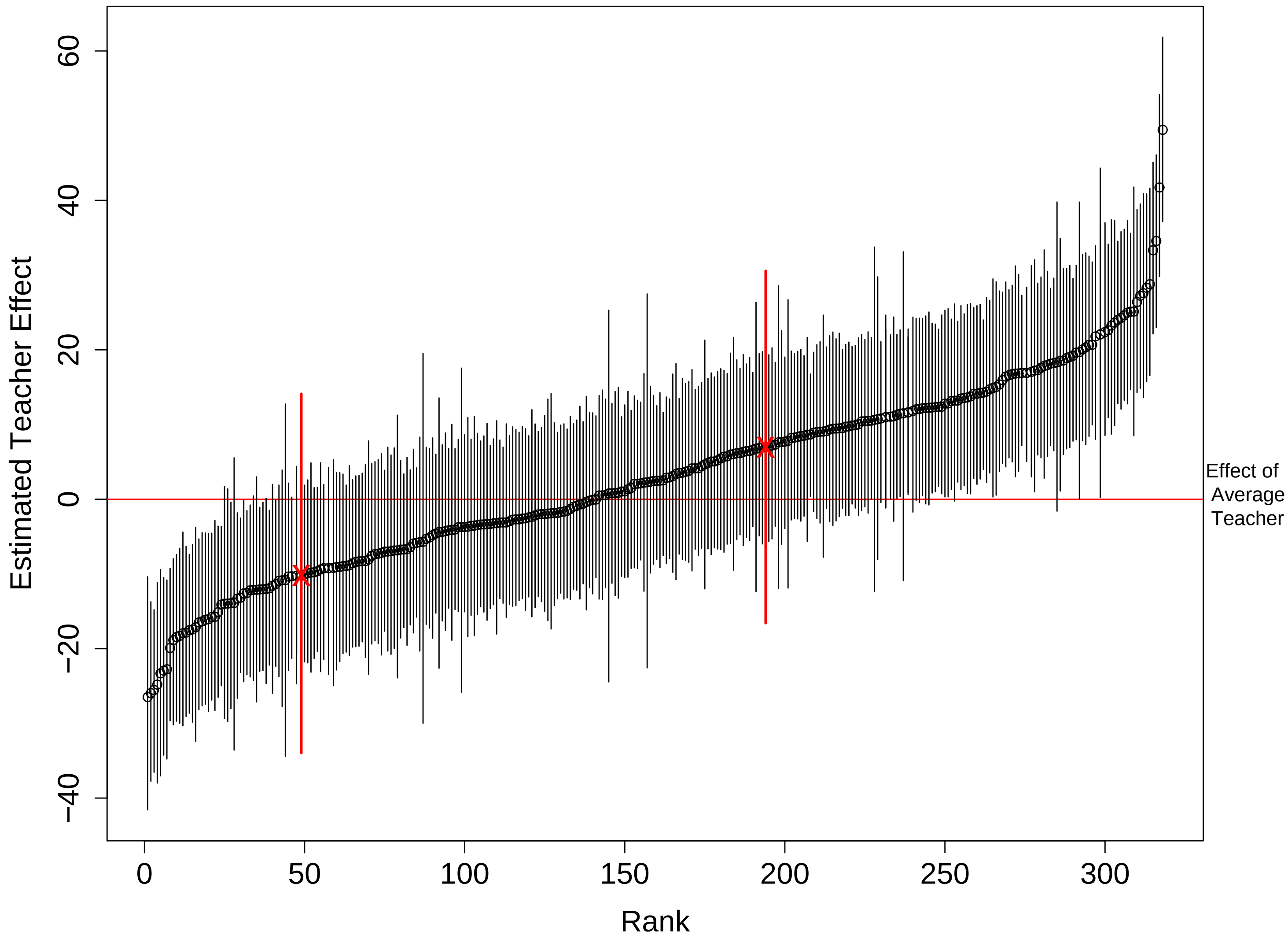
Weaknesses

- Relatively rigid assumptions about persistence of past inputs and test score scaling required for estimators to identify effects
- There is little known about quality of estimates when the model is misspecified

Empirical Findings about Complex Models are Mixed

- Kane et al. find measures correlate with experimental results
- Rivkin et al. reported smaller teacher variance for classes that were apparently randomly assigned, indicating potential for bias more generally
- Rothstein finds that students are not randomly assigned to teachers and fixed effects on gains will not correct for the selection bias
- McCaffrey et al. find that estimates from a wide range of models correlated with prior student inputs
- McCaffrey et al. find in a different dataset that tracking of students to courses in middle schools is a potentially large source of bias
- Estimates are generally robust to the inclusion of student-level covariates but classroom or school-level variables are problematic

Precision of Annual Estimates Is Insufficient to Distinguish Most Teachers from the Mean



Also Evidence of Intra-Temporal Instability

- Studies consistently find that only about a third of teacher effects in the top or bottom quintile in one year remain so in the next year (Aaronson et al.; Ballou)
- Studies also find only moderate correlations (0.2 - 0.5) of estimates over time for the same teacher (Aaronson et al.; McCaffrey et al.)
- McCaffrey et al. find that elementary teacher effects are more variable over time than middle school teacher effects, but this could reflect a stable source of bias in middle school
- Interpreting instability is challenging
  - True stability of teachers is unknown but assumed greater than estimates
  - Stable sources of bias can make estimates with greater bias more stable across time than other estimates

Summary

- General agreement that complex approaches are necessary to remove bias
- There is little agreement on which complex approaches are best and the evidence is mixed about whether even these are effective
  - Complex methods were developed from two paradigms: statistical and econometric
  - Contrary to common beliefs, VA estimators from both the econometric and statistical paradigms can be unbiased even when students are heterogeneously grouped among classes or schools, provided other assumptions hold
  - Existence or size of bias cannot be empirically measured, but heterogeneous grouping is widely observed and specification tests suggest that necessary assumptions of models are not met in some data
- Precision and stability remain outstanding concerns, possibly exacerbated by increased complexity required for unbiasedness