Sending Value-Added into Tailspin: A Simulation Study of Measurement Error and Nonrandom Sorting

Cassandra Guarino
Eun Hye Ham
Mark Reckase
Brian Stacy
Jeffrey Wooldridge

Oct 11, 2013

Acknowledgment: This work was supported in part by a Pre-Doctoral Training Grant from the Institute for Education Sciences, U.S. Department of Education (Award #R305B090011) to Michigan State University and an IES Statistical Research and Methodology Grant (Award #R305100028). The opinions expressed are those of the authors and do not represent views of the IES or the U.S. Department of Education.

Introduction

- Accurate teacher value-added measures in high demand
- Bias caused by measurement error increasingly a concern
- Some districts and states moving toward estimators that try to correct for measurement error

Research Questions:

1. How does measurement error affect teacher evaluation measures?
2. Which teachers most affected by bias ?
3. Do measurement error corrections reduce bias in VAMs?

Findings:

- Measurement error can generate noticeable bias, particularly for teachers with students in tails
- Measurement error correction techniques work well when assignment based on true scores
- Work less well when assignment based on observed scores
- Test score ceilings and floors are a less important issue

Background on Measurement Error in Test Scores

- Measurement Error
  - State achievement tests composed of finite number of test items (40-50 questions)
  - With small number of test items, have imperfect measures of student ability
  - There are other sources of measurement error as well
- This measurement error can cause bias in value-added estimates

Data Generating Process

$$A^*_{i3} = \lambda A^*_{i2} + \beta_{i3} + c_i + u_{i3}$$

- DGP based on true scores
- A^*_{ig} - true achievement level of student i in grade g
- β_{ig} - teacher effect
- c_i - student learning heterogeneity
- u_{ig} - idiosyncratic error term independent across students and time

Model and Empirical Approach

Value-Added Model and Estimators

Considering Model:

$$A_{ig} = \lambda A_{ig-1} + T_{ig}\beta + c_i + u_{ig} + v_{ig} - \lambda v_{ig-1}$$

Examine Performance of:

- DOLS: OLS estimation
- EIVReg: Errors in Variables Regression - uses average measurement error variance
- Colorado Growth Model

Simulation Design

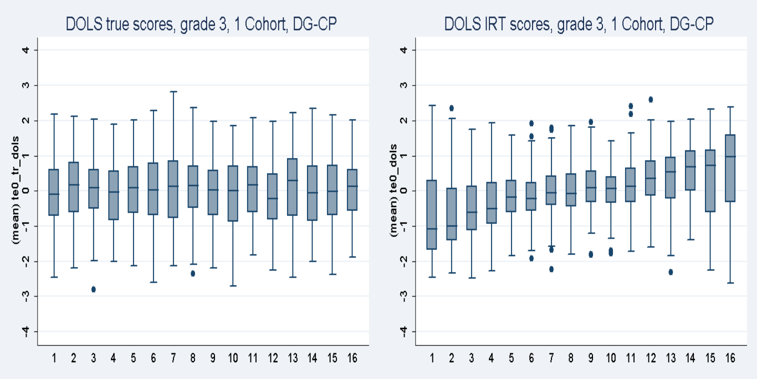
Simulation Parameters

- Grade 3 plus a base year
- 320 Students
- 16 teachers
- 1 School
- Class size fixed at 20
- 100 Simulation Reps

Test Score Generation

- Test scores generated based on 3 parameter Logistic IRT model
- Item parameters come from large, diverse southern state's test
- Estimate A_{ig} by MLE

DOLS

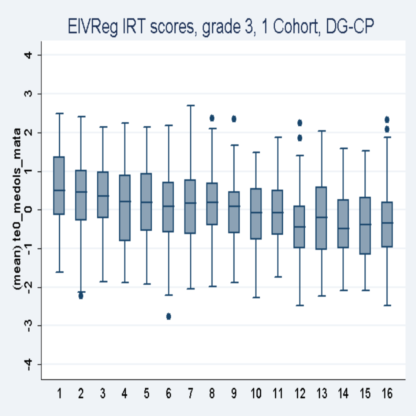


- Box and Whisker Plots:
  - Middle of Box - Median from 100 reps
  - Lower Part of Box - 25th Percentile
  - Top of Box- 75th Percentile
- For additional simplicity, we assume every teacher has identical effect of 0
- Classes sorted by prior year true achievement level or test score
- Teacher on left side of graph receive class with lowest prior score

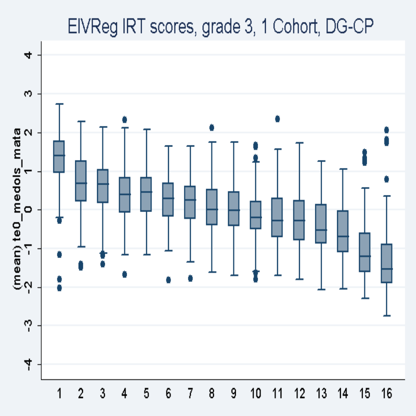
Measurement Error Corrections

- Measurement error corrections can give bias free estimates if some assumptions are met
  - Use information on measurement error variances to adjust estimates for bias
  - Requires that measurement error uncorrelated with true scores and other covariates
    - Not true when assignment based on observed scores or with test ceilings and floors
    - Plausible cause of violation is if principals assign observed scores
    - Then non-random assignment of classrooms to teachers creates correlation between teacher assignment and measurement error

EIVReg Estimator



Assignment using True Scores

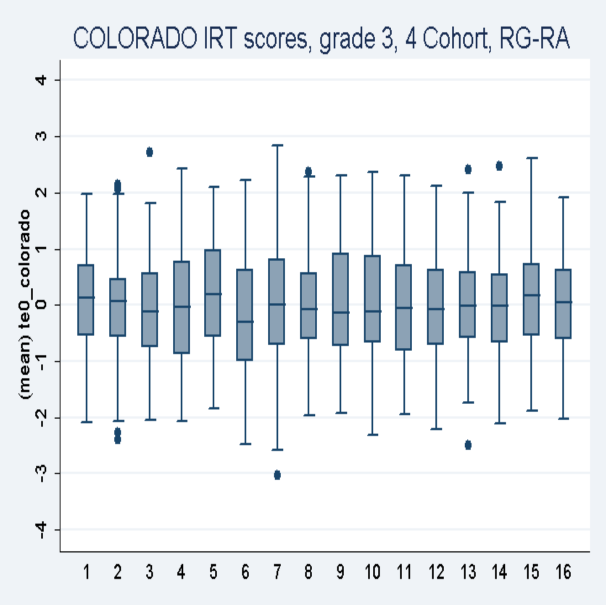


Assignment based on IRT Scores

Colorado Growth Model

- With measurement error, Colorado growth model approach may be problematic
  - Difficult to make measurement error adjustsments in quantile regression
  - Heteroskedastic measurement error in dependent variable can cause bias in quantile regression

CGM Boxplots



Conclusion

Conclusions

- Under some conditions of non-random assignment and measurement error, VAMs can mischaracterize teachers
- Estimators that correct for measurement error don't work when assignment based on observed score
- Colorado Growth Model looks surprisingly good, but we need to study more
- Test score floors and ceilings can add even more potential for bias, although in simulations did not see big differences

END

Assignment and Grouping Based on Observed Scores

- Principals could group students on observed scores
- Then non-random assignment of classrooms to teachers creates correlation between teacher assignment and measurement error
- Can create bias even for EIVReg and other measurement error corrected estimators
- No longer theoretically clear whether bias greater or less than VAMs that ignore measurement error