RESPONSE TO ROTHSTEIN (2014) ON "REVISITING THE IMPACTS OF TEACHERS"

Raj Chetty, John Friedman, and Jonah Rockoff

October 2014

In a recent paper, Jesse Rothstein (2014) successfully replicates Chetty, Friedman, and Rockoff's [CFR] results on teacher's impacts using data from North Carolina. Bacher-Hicks, Kane, and Staiger (2014) also successfully replicate CFR's results using data from the Los Angeles Unified School District. Together, these studies show that value-added (VA) measures of teacher quality show very consistent properties across different settings. Such replications are extremely useful for social science and we are very grateful to Rothstein and Bacher-Hicks, Kane, and Staiger for their efforts in conducting these studies.

Rothstein (2014) additionally raises three concerns about the validity of CFR's methods. One of these concerns is new, while the other two were raised and addressed during the peer review publication process of the CFR papers. (Rothstein was one of the peer reviewers of the CFR papers – a fact that was disclosed during his testimony for the defense in Vergara v. California.)

While Rothstein raises important issues based on very careful work, we do not believe that any of them ultimately threaten the validity of CFR's methods or conclusions. We address Rothstein's new concern in detail in a recent note, which we summarize below:

- *Rothstein Concern #1: The "teacher-switching" quasi-experimental design used by CFR is invalid because it fails a placebo test by showing "effects" of changes mean teacher value-added on changes in prior test scores.*

  Response: *These apparent "placebo effects" stem from a mechanical feature of using lagged scores for the placebo test itself rather than a flaw in CFR's design.* Because teacher value-added (VA) is estimated using data from students in the same schools in previous years, teachers will tend to have high VA estimates when their students happened to do well in prior years. Regressing changes in prior test scores on changes in teacher VA effectively puts the same data on the left- and right-hand side of the regression, which can mechanically produce a positive coefficient even with a valid research design. We find that accounting for such mechanical effects – for instance, by controlling for school-subject-year level shocks – eliminates the correlation between changes in lagged scores and current teacher VA but does not affect the original estimate of forecast bias. Rothstein replicates this finding in a revised version of his paper (Appendix Table 6, Row 5). In addition, we find that placebo tests that do not directly re-use the same historical data used to estimate VA – for instance, estimating VA only using data from future years – uncover no evidence of a correlation with prior test scores.

  Bacher-Hicks, Kane, and Staiger (2014) also document the same set of patterns in data from the Los Angeles Unified School District. Unlike Rothstein, they conclude that the correlation with lagged scores is due to a mechanical bias in the placebo test rather than a failure of the quasi-experimental design. Based on this evidence, they conclude that teacher VA estimates exhibit little or no forecast bias in LA.

We now turn to Rothstein's two other points, drawing from the response letters we submitted to the American Economic Review to address the reviewers' comments prior to publication.

- *Rothstein Concern #2: The exclusion of teachers with missing VA estimates biases the conclusion that VA estimates are forecast unbiased.*

  Response: *The most definitive way to evaluate whether missing data are a concern is to focus on the subsample of school-grades where no data are missing. In this subsample, we find no evidence of forecast bias* (Column 4, Table 5, CFR I). Rothstein confirms this finding in his own data (Appendix Table 5, Column 4). Rothstein raises a set of issues about how to predict VA for teachers who only appear in the data for one year, drawing on information about other teachers in the school. We view this as an interesting exploration, but one that is unrelated to our goal of evaluating the degree of forecast bias in conventional teacher VA estimates that only use data from each teacher's past performance to predict his/her future performance.

- *Rothstein Concern #3: The "cross-classroom" method of estimating the long-term impacts of teachers relies on strong assumptions that may not hold in practice.*

  Response: *Our quasi-experimental design, which does not rely on any such assumptions, yields very similar estimates of the long-term impacts of teachers.* We fully agree with Rothstein's observation that our first design (Section III, CFR II), which is based on OLS regressions comparing outcomes across classrooms, relies on a strong selection-on-observables assumption. Indeed, we emphasize on page 2 of our second paper this is "a very strong assumption." Exploring the effects of alternative assumptions is an interesting exercise, but can never be definitive because one cannot be confident that one has correctly controlled for all potential confounding factors using this approach. This is why we returned to our quasi-experimental "teacher switching" research design to obtain estimates that do not rely on such assumptions (Section IV, CFR II). We find that the quasi-experimental method yields very similar estimates of teachers' long-term impacts for the outcomes for which we have adequate precision (e.g., college attendance). As noted in point #1 above, we continue to believe that this research design is valid as originally implemented, and hence our conclusion that teachers have significant long-term impacts is unchanged.

In summary, Rothstein (2014) makes significant progress in understanding the complex issues that can arise in evaluating VA measures. However, our interpretation of the Rothstein (2014) and Bacher-Hicks, Kane, and Staiger (2014) replications is that they ultimately reinforce the view that VA measures exhibit little forecast bias and predict teacher's long-term impacts. Nevertheless, as in our original papers, we caution that further research is needed to (1) determine whether using VA for high-stakes evaluation leads to an erosion of its signal quality and (2) compare VA metrics to other measures (e.g., classroom observations, principal evaluations) to determine the best way to measure teacher effectiveness.

Works Cited

Chetty, Raj, John Friedman, and Jonah Rockoff. 2014. "Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates" *American Economic Review* 104(9): 2593-2632.

Chetty, Raj, John Friedman, and Jonah Rockoff. 2014. "Measuring the Impact of Teachers II: The Long-Term Impacts of Teachers" *American Economic Review* 104(9): 2633-2679.

Bacher-Hicks, Andrew, Thomas J. Kane, and Douglas O. Staiger. 2014. "Validating Teacher Effect Estimates Using Changes in Teacher Assignments: A Replication of Chetty et al. in Los Angeles" Harvard University Working Paper.

Rothstein, Jesse. 2014. "Revisiting the Impacts of Teachers." UC-Berkeley Working Paper.