Measuring the Impacts of Teachers: Reply to Rothstein

Raj Chetty, John Friedman, and Jonah Rockoff

Online Appendix A. Teacher Followers and Prior Scores

In this appendix, we provide further detail on why including teachers who follow students across grades can produce correlations between changes in current VA and prior test scores across cohorts.

Consider the relationship between changes in math test scores and math teacher VA from 1994 to 1995 in 5th grade in a given school. Suppose a teacher with high estimated VA leaves 5th grade after 1994 and is replaced by a teacher with average VA; assume there are no other changes in the teaching roster. We know that the high-VA teacher who departed did not teach the children who were in grade 5 in 1995 when they were in 4th grade in 1994 (because she taught 5th grade in 1994). However, she may have taught the children who were in grade 5 in 1994 when they were in 4th grade in 1993. As a result, the high VA of the departing teacher is positively correlated with lagged test scores of the cohort that reaches 5th grade in 1994, but not the test scores of those who reach 5th grade in 1995. This effect makes lagged (4th grade) test scores fall on average across the two cohorts. Since (by construction) teacher VA is also falling in this example, there is a positive correlation between changes in lagged (4th grade) scores across the two cohorts and mean teacher VA.

It is useful to distinguish between two separate channels that drive this correlation. The first channel is fluctuations in student test scores that are not related to the persistent component of teacher value-added, i.e., noise in student test scores. The teachers in 5th grade in 1994 could have higher estimated VA simply because her students in 4th grade test in 1993 did particularly well by chance (e.g., because the curriculum in the school happened to be well aligned with the test questions that year). This creates a mechanical correlation between lagged scores and VA estimates but has no bearing on our estimate of forecast bias using current test scores. Second, the correlation could be driven by teacher treatment effects. If the 5th grade teachers in 1994 were of truly high quality, they would affect the performance of 4th graders in 1993 (because some of them taught 4th grade in 1993), but not the 4th graders in 1994 (because we know they are teaching 5th grade in 1994). Note that, in contrast to the first channel, the direct treatment effect of teachers in prior grades could potentially bias our estimate of λ, as having better teachers in prior school years can increase current scores. The magnitude of bias depends upon the rate of fade-out in the sample where a teacher teaches the same child twice. The fact that the estimates of λ do not change when we exclude followers (Columns 1 and 2 of Table 4) shows that in practice, there is little or no bias in our estimate of λ from this latter channel.

Online Appendix B. Residualization Using Within-Teacher Variation

In a standard partial-regression implementation of a multivariate regression model, one must residualize *both* the left- and right-hand side variables with respect to the covariates to obtain a consistent estimate of the regression coefficient of interest. In this appendix, we show why one should not residualize the right hand size variable (teacher VA) with respect to covariates X_{it} when the effects of the covariates on long-term outcomes Y_{it}^* are estimated using *within-teacher* variation.

Suppose the statistical model for earnings is

$$Y_i^* = \alpha + \kappa m_j + \beta^Y X_{it} + \} \eta_{it}, \tag{1}$$

with η_{it} orthogonal to X_{it} and m_j defined as normalized teacher VA, as in equations (2) and (4) in CFR-II.

First, observe that if we knew β^Y, we could mechanically construct $Y_{it} = Y_i^* - \beta^Y X_{it}$ and then simply regress Y_{it} on m_j (without including any controls) to obtain an unbiased estimate of κ under the selection on observables assumption in CFR-II (Assumption 2). In this case, residualizing value-added m_j with respect to X_{it} would *not* yield a consistent estimate of κ because the true model is $Y_{it} = \alpha + \kappa m_j + \eta_{it}$.

Now suppose that we do not know β^Y and estimate it using an OLS regression (without teacher fixed effects) of the form

$$Y_i^* = a + b^Y X_{it} + \varepsilon_{it}.$$

Here b^Y does not provide a consistent estimate of β^Y if teacher VA m_j is correlated with X_{it}: b^Y converges to $\beta^Y + \kappa Cov(m_j, X_{it})/Var(X_{it})$. Since b^Y is not a consistent estimate of β^Y, one cannot simply regress $Y_i^* - b^Y X_{it}$ on m_j to obtain a consistent estimate of κ. Intuitively, the reason one must residualize both Y_{it} and m_j in a multivariate regression is that an OLS regression of Y_{it} on X_{it} does not produce a consistent estimate of the structural parameter β^Y in (1) because it partly picks up the effect of m_j, which is correlated with X_{it}. To correct for the incorrect estimate of β^Y, one must residualize the right-hand-side variable m_j with respect to X_{it} and then regress the earnings residuals Y_{it} on the VA residuals $\tilde{m}_{jt} = m_j - \gamma X_{it}$.

Now consider our approach, where we estimate β^Y using an OLS regression *with* teacher fixed effects a_j:

$$Y_i^* = a_j + b_f^Y X_{it} + \varepsilon_{it}^{'}.$$

Here, the coefficient b_f^Y is identified purely from within-teacher variation in X_{it} that is mechanically uncorrelated with variation in m_j. Therefore, under the model in (1), the coefficient b_f^Y converges to β^Y and hence regressing $Y_i^* - b_f^Y X_{it}$ on m_j yields a consistent estimate of κ, for the same reason that regressing $Y_i^* - \beta^Y X_{it}$ on m_j when β^Y is known yields a consistent estimate of κ in the first case considered above. In contrast, using residual VA $\tilde{m}_{jt} = m_j - \gamma X_{it}$ in the second regression would yield an inconsistent estimate of κ. Intuitively, when we use within-teacher variation to estimate β^Y, we immediately obtain a consistent estimate of the effect of X on earnings that is not contaminated by the correlation with teacher value-added. Hence, one simply has to regress the outcome residual on VA to estimate the effect of teacher VA in the second stage.

Online Appendix C. Stata Code for Simulations

Imputation of Missing Data

```
****************
* This simulation shows that imputing zeros reduces the coefficient in the regression of changes in scores on changes in mean VA when VA is correlated across teachers in a cell
****************

clear all

*1 Generate data at the school-year-teacher level
set obs 1000000
set seed 5071788
g year = mod(_n-1,2)+1
g teacher = ceil(_n/2)
g school = ceil(teacher/2)

*2 Generate correlated VA within cells
global corr = 0.2
tsset teacher year
g va = rnormal(0,.1) if year == 1
replace va = (${corr}*l.va + sqrt(1-${corr}^2)*rnormal(0,.1)) if year == 2

*3 Generate missing data
g rand = runiform()
g miss = rand<.2
g va_miss = va
replace va_miss = . if miss==1

*4 Generate scores
g score = va + rnormal(0,1)
g score_miss = score
replace score_miss = . if va_miss == .

*5 Imputation of 0 for missing data
g va_impute = va_miss
replace va_impute = 0 if va_miss == .

*6 Collapse to school-year and run regressions
collapse va va_miss va_impute score score_miss, by(school year)
tsset school year

*7 Results
log using imputation.smcl, replace

_eststo clear
_eststo Full_Sample: reg d.score d.va // Coefficient in full sample (ideal data)
_eststo No_Missing: reg d.score_miss d.va_miss // Coefficient on subsample with no missing data
_eststo Impute_0s: reg d.score d.va_impute // Coefficient with imputation
```

```
is downward-biased
esttab _all, mtitles title("Missing Data Imputation Simulation Results") se not
log close
```

Prior Test Scores (Table 3)

```
*This simulation shows that prior test score changes with be correlated with changes in current mean VA across cohorts when there are school-year level shocks
*Simulates data for one subject, so shocks should be interpreted as school-year-subject shocks
*The program simulates class-level data, incorporating teacher effects, class effects, student-level noise, and a school-year shock common to all classrooms.

clear all
set seed 717806
set more off

* Parameters governing simulation
global min_grade = 3 // Minimum grade level (for readability)
global min_year = 1992 // Start year (for readability)
global n_school = 10000 // Number of schools
global n_year = 6 // Number of years
global n_grade = 6 // Number of grades taught per school
global n_rooms = 4 // Number of classrooms per school and grade
global n_class = 25 // Number of students per class
global var_tot = 0.25 // Total variance of scores
global sd_va = 0.10 // Standard deviation of value added
global sd_class_shock = 0.08 // Standard deviation of classroom-level shocks
global sd_sy_shock = .08 // Standard deviation of school-by-year shocks
global rho_sy = 0.35 // Autocorrelation on school-by-year shocks

* Generate basic data
set obs `= ${n_school} * ${n_grade} * ${n_rooms} * ${n_year}'
g school = ceil(_n / (${n_grade} * ${n_rooms} * ${n_year}))
g grade = mod(ceil(_n / (${n_rooms} * ${n_year})) - 1 , ${n_grade}) + ${min_grade}
g teacher = ceil(_n / ${n_year})
g year = mod(_n - 1, ${n_year}) + ${min_year}
g id = rnormal()

* Replace some teachers in 1997
* Only in grades 5-8 (since others not used in experiment at end)
g replacement = mod(teacher , ${n_rooms}) < 1 if year == ${min_year}
replace replacement = replacement[_n-1] if year > ${min_year}
replace grade = grade - 1 if replace == 1 & year >= 1997
replace school = mod(school , ${n_school}) + 1 if grade == (${min_grade}+1) & year >= 1997 & replace == 1
replace grade = ${min_grade} + ${n_grade} - 1 if grade == (${min_grade}+1) & year >= 1997 & replace == 1
replace grade = grade + 1 if replace == 1 & year >= 1997 & grade < (
```

```
${min_grade}+1)

* Generate true VA and class shocks
g class_shock = rnormal(0, ${sd_class_shock})
g va_true = rnormal(0,${sd_va}) if year==${min_year}
replace va_true = va_true[_n-1] if va_true==.

* Generate average lagged true VA and average lagged class shocks as "double lags" of true VA and class shocks
sort school year grade id
g temp1 = va_true[_n - ${n_rooms} * (${n_grade} + 1)] if year > ${min_year} & grade > ${min_grade}
g temp2 = temp1[_n - ${n_rooms} * (${n_grade} + 1)] if year > ${min_year} & grade > ${min_grade}
g temp3 = class_shock[_n - ${n_rooms} * (${n_grade} + 1)] if year > ${min_year} & grade > ${min_grade}
g temp4 = temp3[_n - ${n_rooms} * (${n_grade} + 1)] if year > ${min_year} & grade > ${min_grade}
by school year grade: egen l_va_true = mean(temp1)
by school year grade: egen l2_va_true = mean(temp2)
by school year grade: egen l_class_shock = mean(temp3)
by school year grade: egen l2_class_shock = mean(temp4)
drop temp*

* Generate school-by-year shocks
g sy_shock = rnormal(0 , ${sd_sy_shock} * sqrt(1 - ${rho_sy}^2)) if mod(_n ,${n_rooms} * ${n_grade}) == 1
replace sy_shock = sy_shock / sqrt(1 - ${rho_sy}^2) if year == ${min_year} & ~missing(sy_shock)
replace sy_shock = sy_shock[_n - 1] if missing(sy_shock)
replace sy_shock = sy_shock + ${rho_sy} * sy_shock[_n - ${n_rooms} * ${n_grade}] if year > ${min_year}
g l_sy_shock = sy_shock[_n - ${n_rooms} * ${n_grade}] if year > ${min_year}
g l2_sy_shock = sy_shock[_n - 2 * ${n_rooms} * ${n_grade}] if year > (${min_year} + 1)

* Generate classroom average score, lagged, and twice-lagged scores
global sd_indv = sqrt((${var_tot} - ${sd_va}^2 - ${sd_sy_shock}^2 - ${sd_class_shock}^2) / ${n_class})
g l2_score = (l2_va_true + l2_sy_shock + l2_class_shock + rnormal(0,${sd_indv}))
g l_score = l_va_true + l_sy_shock + l_class_shock + rnormal(0,${sd_indv})
g score = va_true + sy_shock + class_shock + rnormal(0,${sd_indv})

*Make dataset balanced panel
replace l_score = . if l2_score == .
replace score = . if l_score == .

* Residualize scores using a single lag
sort teacher year
sum score
global tot_var = r(Var)
```

```
tsset teacher year
corr score l.score, c
global teach_var = r(cov_12)
global ind_var = (${n_class} / (${n_class} - 1)) * (${var_tot} - ${tot_var})
global class_var = ${var_tot} - ${ind_var} - ${teach_var}

* Construct leave-two-out shrinkage and VA estimate
g temp = ~inrange(year , 1996 , 1997) & ~missing(score)
by teacher: egen temp1 = sum(temp)
g shrinkage = ${teach_var} / (${teach_var} + ${class_var} / temp1 + ${ind_var} / (${n_class} * temp1))
g temp2 = score if ~inrange(year , 1996 , 1997) & ~missing(score)
by teacher: egen temp3 = mean(temp2)
g va = temp3 * shrinkage if inrange(year,1996,1997)
drop temp*

* Construct Rothstein (2016) leave-three-out shrinkage and VA estimate
g temp = ~inrange(year , 1995 , 1997) & ~missing(score)
by teacher: egen temp1 = sum(temp)
g shrinkage_3out = ${teach_var} / (${teach_var} + ${class_var} / temp1 + ${ind_var} / (${n_class} * temp1))
g temp2 = score if ~inrange(year , 1995 , 1997) & ~missing(score)
by teacher: egen temp3 = mean(temp2)
g va_3out = temp3 * shrinkage_3out if inrange(year,1995,1997)
drop temp*

* Construct prior shrinkage and VA estimate
g temp = (year < 1997) & ~missing(score)
bys teacher: egen temp1 = sum(temp)
g shrinkage_prior = ${teach_var} / (${teach_var} + ${class_var} / temp1 + ${ind_var} / (${n_class} * temp1))
g temp2 = score if (year < 1997) & ~missing(score)
bys teacher: egen temp3 = mean(temp2)
drop temp*

save vam_simulation, replace

* Collapse data to school-grade-year level to implement quasi-experimental analysis
keep if inrange(year,1996,1997)
collapse score l_score va va_3out va_true, by(school grade year)
egen sy = group(school year)
egen sg = group(school grade)
tsset sg year
save vam_simulation_collapse, replace

* Results
log using lagged_score_simulation.smcl, replace
eststo clear
_eststo d_score: qui reg d.score d.va
_eststo d_score_sy: qui reg d.score d.va , a(sy)
_eststo d_score_3out: qui reg d.score d.va_3out
_eststo lag_d_score: qui reg d.l_score d.va
```

```
_eststo lag_d_score_sy: qui reg d.l_score d.va , a(sy)
_eststo lag_d_score_3out: qui reg d.l_score d.va_3out
esttab _all, mtitles title("Quasi-Experimental Forecast Bias Estimates") se not
log close
```

Long-Term Effects of VA (Table 5)

```
*This simulation shows that estimates of long-term impacts are downward-biased in a multi-variable regression because VA is estimated with error and is correlated with X

clear all
set seed 7817806
*Specify target: true effect of 1 unit increase in va_score on earnings
*Note that this is equivalent to effect of 1 unit increase in mu (not m = mu/sd(mu))
global true_coeff = 100

************************PART 1*********************
*******************Generate data*******************
***************************************************

set obs 1000000
global n_class = 20
global classes_per_teach = 10
g class = ceil(_n/$n_class)
g teacher = ceil(_n/(${n_class}*${classes_per_teach}))

*Generate test-score VA (mu_j)
bys teacher: g temp = _n
g temp1 = rnormal(0,0.1) if temp == 1
bys teacher: egen va_score = mean(temp1)

*Generate pure earnings component of VA
g temp2 = rnormal(0,0.1) if temp == 1
bys teacher: egen va_earn = mean(temp2)
drop temp*

*Generate total earnings VA (tau_j)
g va_comb = va_score + va_earn

*Generate covariate X correlated with teacher's total earnings VA
global rho = 0.33
g x = (${rho}*va_comb + (1-${rho})*rnormal(0,0.1))/sqrt(${rho}^2+(1-${rho})^2)

*Generate scores and earnings
*Note that only va_score affects test scores, while both va_score and va_earn affect earnings
g score = va_score + x + rnormal(0,sqrt(1-.1^2))
g earn = ${true_coeff}*va_comb + 10*x + rnormal(0,10)

**********************PART 2. ********************
**Estimate VA using within-teacher residualization**
```

```
****************************************************

* Residualize scores using within teacher variation as in CFR (2014b)
qui areg score x, a(teacher)
predict score_res, dr

* Estimate teacher-level variance
preserve
collapse score_res, by(teacher class)
tsset teacher class
qui corr score_res l.score_res, c
global teach_var = r(cov_12)
restore

* Estimate residual variance and shrinkage
sum score_res
global tot_var = r(Var)
global ind_var = ${tot_var} - ${teach_var}
scalar shrinkage = ${teach_var}/(${teach_var} + ${ind_var}/(${n_class} * (${classes_per_teach}-1)))

* Estimate Leave-Out VA
bys teacher: egen temp = mean(score_res)
bys teacher class: egen temp1 = mean(score_res)
g va = (${classes_per_teach}*temp - temp1)/(${classes_per_teach}-1)*shrinkage
drop temp*

```



```
*Confirm that regressing test score residuals on VA gives a coeff of 1
reg score_res va

***********************PART 3***************************
**Alternative Estimators of Teachers' Long-Term Effects**
*********************************************************

log using long_term_controls_simulation.smcl, replace

***Column 1. Estimate long-term effects using two-step residualization as in CFR (2014b)
*Yields correct estimate of long-run effects as expected
qui areg earn x, a(teacher)
predict earn_res, dr
reg earn_res va, cl(teacher)

***Column 2. Estimate long-term effects using multivariable regression as in Rothstein (2016)
*Yields attenuated coefficient as expected
reg earn va x, cl(teacher)

*Column 3: Rothstein (2014) 2SLS estimator yields estimate similar to OLS in #3
ivreg earn (score_res = va) x

log close

```