Chetty, Friedman and Rockoff (2014): Methodology

Matthew Naven

University of California, Davis

November 14, 2017

Contents

1 Statistical Model

- School principals assign each student i in school year t to a classroom $c = c(i, t)$.

- Principals then assign a teacher $j(c)$ to each classroom c. For simplicity, assume that each teacher teaches one class per year, as in elementary schools.

- Let $j = j(c(i,t))$ denote student i's teacher in year t and μ_{jt} represent the teacher's value-added in year t (i.e., teacher j's impact on test scores). We scale teacher VA so that the average teacher has value-added $\mu_{jt} = 0$ and the effect of a one-unit increase in teacher VA on end-of-year test scores is 1.

- Student i's test score in year t, A^*_{it}, is given by
  - $A^*_{it} = \boldsymbol{\beta} \boldsymbol{X}_{it} + \nu_{it}$
  - where $\nu_{it} = \mu_{jt} + \theta_c + \tilde{\varepsilon}_{it}$
  - Here, $\boldsymbol{X}_{it}$ denotes observable determinants of student achievement, such as lagged test scores and family characteristics.

- We decompose the error term ν_{it} into three components: teacher value-added μ_{jt}, exogenous class shocks θ_c, and idiosyncratic student-level variation $\tilde{\varepsilon}_{it}$. Let $\varepsilon_{it} = \theta_c + \tilde{\varepsilon}_{it}$ denote the unobserved error in scores unrelated to teacher quality. Student characteristics $\boldsymbol{X}_{it}$ and ε_{it} may be correlated with μ_{jt}. Accounting for such selection is the key challenge in obtaining unbiased estimates of μ_{jt}.

- The model in (1) permits teacher quality μ_{jt} to fluctuate stochastically over time. We do not place any restrictions on the stochastic processes that μ_{jt} and ε_{it} follow except for the following assumption.
  - ASSUMPTION 1 (Stationarity): Teacher VA and student achievement follow a stationary process:
  - $E[\mu_{jt}|t] = E[\varepsilon_{it}|t] = 0$ for all t.
  - $cov(\mu_{jt}, \mu_{j,t+s}) = \sigma_{\mu_s}$ for all t.
  - $cov(\varepsilon_{it}, \varepsilon_{i,t+s}) = \sigma_{\varepsilon_s}$ for all t.
  - Assumption 1 requires that
    1. Mean teacher quality does not vary across calendar years.
    2. The correlation of teacher quality, class shocks, and student shocks across any pair of years depends only on the amount of time which elapses between those years.
  - This assumption simplifies the estimation of teacher VA by reducing the number of parameters to be estimated. Note that the variance of teacher effects, $\sigma^2_\mu = var(\mu_{jt})$, is constant across periods under stationarity.

2 Estimating Teacher Value-Added

- We develop an estimator for teacher value-added in year t (μ_{jt}) based on mean test scores in prior classes taught by teacher j.
  - To maximize statistical precision, we use data from all other years - both in the past and future - to predict VA in year t in our empirical implementation. To simplify notation, we present the derivation in this section for the case in which we only use prior data to predict VA.

- To simplify exposition, we derive the estimator for the case in which data on test scores is available for t years for all teachers, where all classes have n students, and where each teacher teaches one class per year. In online Appendix A, we provide a step-by-step guide to implementation (along with corresponding Stata code) which accounts for differences in class size, multiple classrooms per year, and other technical issues which arise in practice.

- We construct our estimator in three steps.
  1. First, we regress test scores A^*_{it} on $\boldsymbol{X}_{it}$ and compute test score residuals adjusting for observables.
  2. Next, we estimate the best linear predictor of mean test score residuals in classrooms in year t based on mean test score residuals in prior years, using a technique analogous to an OLS (ordinary least squares) regression.
  3. Finally, we use the coefficients of the best linear predictor to predict each teacher's VA in year t.

1. Let the residual student test score after removing the effect of observable characteristics be denoted by $A_{it} = A^*_{it} - \boldsymbol{\beta}\boldsymbol{X}_{it} = \mu_{jt} + \varepsilon_{it}$.
   - We estimate $\boldsymbol{\beta}$ using variation across students taught by the same teacher using an OLS regression of the form $A^*_{it} = \alpha_j + \boldsymbol{\beta}\boldsymbol{X}_{it}$, where α_j is a teacher fixed effect.
   - Our approach of estimating $\boldsymbol{\beta}$ using within-teacher variation differs from prior studies, which typically use both within- and between-teacher variation to estimate $\boldsymbol{\beta}$ (e.g., Kane, Rockoff, and Staiger 2008; Jacob, Lefgren, and Sims 2010). If teacher VA is correlated with $\boldsymbol{X}_{it}$, estimates of $\boldsymbol{\beta}$ in a specification without teacher fixed effects overstate the impact of the $\boldsymbol{X}$s because part of the teacher effect is attributed to the covariates.
     - Teacher fixed effects account for correlation between $\boldsymbol{X}_{it}$ and mean teacher VA. If $\boldsymbol{X}_{it}$ is correlated with fluctuations in teacher VA across years due to drift, then one may still understate teachers' effects even with fixed effects. We show in Table 6 below that dropping

teacher fixed effects when estimating (5) yields VA estimates which have a correlation of 0.98 with our baseline estimates because most of the variation in $\boldsymbol{X}_{it}$ is within classrooms. Since sorting to teachers based on their average impacts turns out to be quantitatively unimportant in practice, sorting based on fluctuations in those impacts is likely to have negligible effects on VA estimates.

- For example, suppose $\boldsymbol{X}$ includes school fixed effects. Estimating $\boldsymbol{\beta}$ without teacher fixed effects would attribute all the test score differences across schools to the school fixed effects, leaving mean teacher quality normalized to be the same across all schools. With school fixed effects, estimating $\boldsymbol{\beta}$ within teacher requires a set of teachers to teach in multiple schools, as in Mansfield (2013). These switchers allow us to identify the school fixed effects independent of teacher effects and obtain a cardinal global ranking of teachers across schools.

2. Let $\bar{A}_{jt} = \frac{1}{n}\sum_{i\in\{i:j(i,t)=j\}} A_{it}$ denote the mean residual test score in the class teacher j teaches in year t. Let $\boldsymbol{A}_j^{-t} = (\bar{A}_{j1}, \ldots, \bar{A}_{j,t-1})'$ denote the vector of mean residual scores prior to year t in classes taught by teacher j. Our estimator for teacher j's VA in year t ($E[\mu_{jt}|\boldsymbol{A}_j^{-t}]$) is the best linear predictor of $\bar{A}_{jt}$ based on prior scores ($E^*[\bar{A}_{jt}|\boldsymbol{A}_j^{-t}]$), which we write as $\hat{\mu}_{jt} = \sum_{s=1}^{t-1} \psi_s \bar{A}_{js}$.

   - We choose the vector of coefficients $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{t-1})'$ to minimize the mean-squared error of the forecasts of test scores: $\boldsymbol{\psi} = \arg\min_{\{\psi_1,\ldots,\psi_{t-1}\}} \sum_j (\bar{A}_{jt} - \sum_{s=1}^{t-1} \psi_s \bar{A}_{js})^2$.
   - The resulting coefficients $\boldsymbol{\psi}$ are equivalent to those obtained from an OLS regression of $\bar{A}_{jt}$ on $\boldsymbol{A}_j^{-t}$. In particular, $\boldsymbol{\psi} = \sum_A^{-1} \boldsymbol{\gamma}$ where $\boldsymbol{\gamma} = (cov(\bar{A}_{jt}, \bar{A}_{j1}), \ldots, cov(\bar{A}_{jt}, \bar{A}_{j,t-1}))'$ is the vector of auto-covariances of mean test scores for classes taught by a given teacher and Σ_A is the variance-covariance (VCV) matrix of $\boldsymbol{A}_j^{-t}$. The diagonal elements of Σ_A are the variance of mean class scores σ_A^2. The off-diagonal elements are the covariance of mean test scores of different classes taught by a given teacher $cov(\bar{A}_{jt}, \bar{A}_{j,t-s})$. Note that $cov(\bar{A}_{jt}, \bar{A}_{j,t-s}) = \sigma_{A_s}$ depends only on the time lag s between the two periods under the stationarity assumption in (3).

3. Finally, using the estimates of $\boldsymbol{\psi}$, we predict teacher j's VA in period t as $\hat{\mu}_{jt} = \boldsymbol{\psi}'\boldsymbol{A}_j^{-t}$.

   - Note that the VA estimates in (7) are leave-year-out (jackknife) measures of teacher quality, similar to those in Jacob, Lefgren, and Sims (2010), but with an adjustment for drift. This is important for our analysis below because regressing student outcomes in year t on teacher VA estimates without leaving out the data from year t when estimating VA would introduce the same estimation errors on both the left- and right- hand side of the regression and produce biased estimates of teachers' causal impacts.

- This is the reason that Rothstein (2010, p. 199) finds that, "...fifth grade teachers whose students have above average fourth grade gains have systematically lower estimated value-added than teachers whose students underperformed in the prior year." Students who had unusually high fourth grade gains due to idiosyncratic, nonpersistent factors (e.g., measurement error) will tend to have lower than expected fifth grade gains, making their fifth grade teacher have a lower VA estimate.

- Importantly, $\hat{\mu}_{jt} = E^*[\bar{A}_{jt}|\boldsymbol{A}_j^{-t}]$ represents simply the best linear predictor of the future test scores of students assigned to teacher j in observational data. This prediction does not measure necessarily the expected causal effect of teacher j on students' scores in year t, $E[\mu_{jt}|\boldsymbol{A}_j^{-t}]$, because part of the prediction could be driven by systematic sorting of students to teachers.

2.1 Special Cases

1. First, consider predicting a teacher's impact in year t using data from only the previous year $t-1$. In this case, $\gamma = \sigma_{A,1}$ and $\Sigma_A^{-1} = \frac{1}{\sigma_A^2}$. Hence, ψ simplifies to $\frac{\sigma_{A,1}}{\sigma_A^2}$ and $\hat{\mu}_{jt} = \psi \bar{A}_{j,t-1}$.

   - The shrinkage factor ψ in this equation incorporates two forces which determine ψ in the general case. First, because past test scores are a noisy signal of teacher quality, the VA estimate is shrunk toward the sample mean ($\mu_{jt} = 0$) to reduce mean-squared error. Second, because teacher quality drifts over time, the predicted effect differs from past performance. For instance, if teacher quality follows a mean-reverting process, past test scores are further shrunk toward the mean to reduce the influence of transitory shocks to teacher quality.

2. Second, consider the case where teacher quality is fixed over time ($\mu_{jt} = \mu_j$ for all t) and the student- and class-level errors are i.i.d. This is the case considered by most prior studies of value-added. Here, $cov(\bar{A}_{jt}, \bar{A}_{j,t-s}) = cov(\mu_j, \mu_j) = \sigma_\mu^2$ for all $s \neq t$ and $\sigma_A^2 = \sigma_\mu^2 + \sigma_\theta^2 + \frac{\sigma_{\tilde{\varepsilon}}^2}{n}$. In this case, (7) simplifies to $\hat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{(\sigma_\theta^2 + \frac{\sigma_{\tilde{\varepsilon}}^2}{n})}{(t-1)}}$, where $\bar{A}_j^{-t}$ is the mean residual test score in classes taught by teacher j in years prior to t and $\psi = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{(\sigma_\theta^2 + \frac{\sigma_{\tilde{\varepsilon}}^2}{n})}{(t-1)}}$ is the *reliability* of the VA estimate.

   - This formula coincides with equation (5) in Kane and Staiger (2008) in the case with constant class size.

- Kane and Staiger (2008) derive (9) using an Empirical Bayes approach instead of a best linear predictor. If teacher VA, class shocks, and student errors follow independent Normal distributions, the posterior mean of μ_{jt} coincides with (9). Analogously, (7) can be interpreted as the posterior expectation of μ_{jt} when teacher VA follows a multivariate Normal distribution whose variance-covariance matrix controls the drift process.

- Here, the signal-to-noise ratio ψ does not vary across years because teacher performance in any year is equally predictive of performance in year t. Because years are interchangeable, VA depends purely on mean test scores over prior years, again shrunk toward the sample mean to reduce mean-squared error.

3 Definition of Bias

- An intuitive definition of bias is to ask whether VA estimates $\hat{\mu}_{jt}$ predict differences accurately in the mean test scores of students who are randomly assigned to teachers in year t, as in Kane and Staiger (2008). Consider an OLS regression of residual test scores A_{it} in year t on $\hat{\mu}_{jt}$ (constructed from observational data in prior years) in such an experiment: $A_{it} = \alpha_t + \lambda\hat{\mu}_{jt} + \zeta_{it}$.

- Because $E[\varepsilon_{it}|\hat{\mu}_{jt}] = 0$ under random assignment of students in year t, the coefficient λ measures the relationship between true teacher effects μ_{jt} and estimated teacher effects $\hat{\mu}_{jt}$: $\lambda \equiv \frac{cov(A_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})} = \frac{cov(\mu_{jt},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})}$.

- We define the degree of bias in VA estimates based on this regression coefficient as follows.
  - DEFINITION 1: The amount of forecast bias in a VA estimator $\hat{\mu}_{jt}$ is $B(\hat{\mu}_{jt}) = 1 - \lambda$.
  - Forecast bias determines the mean impact of changes in the estimated VA of the teaching staff. A policy that increases estimated teacher VA $\hat{\mu}_{jt}$ by 1 SD raises student test scores by $(1-B)\sigma(\hat{\mu}_{jt})$, where $\sigma(\hat{\mu}_{jt})$ is the standard deviation of VA estimates scaled in units of student test scores. If $B = 0$, $\hat{\mu}_{jt}$ provides an unbiased forecast of teacher quality in the sense that an improvement in estimated VA of $\Delta\hat{\mu}_{jt}$ has the same causal impact on test scores as an increase in true teacher VA $\Delta\mu_{jt}$ of the same magnitude.
    - Ex post, any estimate of VA can be made forecast-unbiased by redefining VA as $\hat{\mu}'_{jt} = (1-B)\hat{\mu}_{jt}$. However, the causal effect of a policy which raises the estimated VA of teachers is unaffected by such a rescaling, as the effect of a 1 SD improvement in $\hat{\mu}'_{jt}$ is still $(1-B)\sigma(\hat{\mu}_{jt})$.

Hence, given the standard deviation of VA estimates, the degree of forecast bias is informative about the potential impacts of improving estimated VA.

- Two issues should be kept in mind when interpreting estimates of forecast bias. First, VA estimates can be forecast-unbiased even if children with certain characteristics (e.g., those with higher ability) are assigned systematically to certain teachers. Forecast unbiasedness requires only that the observable characteristics $\boldsymbol{X}_{it}$ used to construct test score residuals A_{it} are sufficiently rich that the remaining unobserved heterogeneity in test scores ε_{it} is balanced across teachers with different VA estimates. Second, even if VA estimates $\hat{\mu}_{jt}$ are forecast unbiased, potentially one can improve forecasts of μ_{jt} by incorporating other information beyond past test scores. For example, data on principal ratings or teacher characteristics potentially could reduce the mean-squared error of forecasts of teacher quality. Hence, the fact that a VA estimate is forecast-unbiased does not necessarily imply that it is the optimal forecast of teacher quality given all the information which may be available to school districts.

4 Data

- School District Data
  - We obtain information on students test scores and teacher assignments from the administrative records of a large urban school district.
  - These data span the school years from 1988-1989 to 2008-2009 and cover roughly 2.5 million children in grades 3-8.
  - For simplicity, we refer below to school years by the year in which the spring term occurs (e.g., the school year 1988-1989 is 1989).
  - The data include approximately 18 million test scores.
  - Test scores are available for English language arts and math for students in grades 3-8 in every year from the spring of 1989 to 2009, with the exception of seventh grade English scores in 2002.
    * We also have data on math and English test scores in grade 2 from 1991 to 1994, which we use only when evaluating sorting on lagged test score gains. Because these observations constitute a very small fraction of our sample, excluding them has little impact on our results.

- The testing regime varies over the 20 years we study. In the early- and mid-1990s, all tests were specific to the district. Starting at the end of the 1990s, the tests in grades 4 and 8 were administered as part of a statewide testing system, and all tests in grades 3-8 became statewide in 2006 as required under the No Child Left Behind law. All tests were administered in late April or May during the early- and mid-1990s. Statewide tests were sometimes given earlier in the school year (e.g., February) during the latter years of our data.
- Because of this variation in testing regimes, we follow prior work by normalizing the official scale scores from each exam to have mean zero and standard deviation one by year and grade.
- The dataset contains information on ethnicity, gender, age, receipt of special education services, and limited English proficiency for the school years from 1989 to 2009.
- The database used to code special education services and limited English proficiency changed in 1999, creating a break in these series which we account for in our analysis by interacting these two measures with a post-1999 indicator.
- Information on free and reduced price lunch is available starting in school year 1999.
- The dataset links students in grades 3-8 to classrooms and teachers from 1991 to 2009.
  * Five percent of students switch classrooms or schools in the middle of a school year. We assign these students to the classrooms in which they took the test to obtain an analysis dataset with one observation per student-year-subject. However, when defining class and school-level means of student characteristics (such as fraction eligible for free lunch), we account for such switching by weighting students by the fraction of the year they spent in that class or school.
- This information is derived from a data management system which was phased in over the early 1990s, so not all schools are included in the first few years of our sample.
- In addition, data on course teachers for middle school and junior high school students - who, unlike students in elementary schools, are assigned different teachers for math and English - are more limited. Course teacher data are unavailable prior to the school year 1994, then grow in coverage to roughly 60 percent by 1998, and stabilize at approximately 85 percent after 2003. To ensure that our estimates are not biased by the missing data, we show that our conclusions remain very similar in a subsample of school-grade-subject cells with no missing data (see Table 5).

• Tax Data

- We obtain information on parent characteristics from US federal income tax returns spanning 1996-2011.
  - Here and in what follows, the year refers to the tax year: i.e., the calendar year in which income is earned. In most cases, tax returns for tax year t are filed during the calendar year $t + 1$.
- The school district records were linked to the tax data using an algorithm based on standard identifiers (date of birth, state of birth, gender, and names) described in online Appendix C, after which individual identifiers were removed to protect confidentiality.
- Students were then linked to parents based on the earliest 1040 form filed between tax years 1996-2011 on which the student was claimed as a dependent.
- We define parental household income as mean Adjusted Gross Income (capped at $117,000, the ninety-fifth percentile in our sample) between 2005 and 2007 for the primary filer who first claimed the child; measuring parent income in other years yields very similar results (not reported). For years in which parents did not file a tax return, they are assigned an income of zero. We measure income in 2010 US$, adjusting for inflation using the Consumer Price Index.
- We define marital status, home ownership, and 401(k) saving as indicators for whether the first primary filer who claims the child ever files a joint tax return, makes a mortgage interest payment (based on data from 1040s for filers and 1099s for non-filers), or makes a 401(k) contribution (based on data from W-2s) between 2005 and 2007.
- We define mother's age at child's birth using data from Social Security Administration (SSA) records on birth dates for parents and children. For single parents, we define the mother's age at child's birth using the age of the filer who first claimed the child, who is typically the mother but is sometimes the father or another relative. We set the mother's age at child's birth to missing for 78,007 observations in which the implied mother's age at birth based on the claiming parent's date of birth is below 13 or above 65, or where the date of birth is missing entirely from SSA records. When a child cannot be matched to a parent, we define all parental characteristics as zero, and we always include a dummy for missing parents in regressions that include parent characteristics.

5 Sample

- First, because our estimates of teacher value-added always condition on prior test scores, we restrict our sample to grades 4-8, where prior test scores are available.

- Second, we exclude the 6 percent of observations in classrooms where more than 25 percent of students are receiving special education services, as these classrooms may be taught by multiple teachers or have other special teaching arrangements.
- We also drop the 2 percent of observations where the student is receiving instruction at home, in a hospital, or in a school serving disabled students solely.
- Third, we drop classrooms with less than 10 students or more than 50 students as well as teachers linked with more than 200 students in a single grade, because such students are likely to be mislinked to classrooms or teachers (0.5 percent of observations).
- Finally, when a teacher is linked to students in multiple schools during the same year, which occurs for 0.3 percent of observations, we use only the links for the school where the teacher is listed as working according to human resources records and set the teacher as missing in the other schools.
- The linked school district and tax record analysis dataset has one row per student per subject (math or English) per school year, as illustrated in online Appendix Table 1.
- Each observation in the analysis dataset contains the student's test score in the relevant subject test, demographic information, teacher assignment, and time-invariant parent characteristics. We organize the data in this format so that each row contains information on a treatment by a single teacher conditional on predetermined characteristics.
- We account for the fact that each student appears multiple times in the dataset by clustering standard errors as described in Section III.
- After imposing the sample restrictions described above, the linked dataset contains 10.7 million student-year-subject observations. We use this core sample of 10.7 million observations to construct quasi-experimental estimates of forecast bias, which do not require any additional controls; 9.1 million records in the core sample have information on teacher assignment and 7.6 million have information on teachers, lagged test scores, and the other controls needed to estimate our baseline VA model.

6 Specification

- We estimate teacher VA using the methodology in Section IB in three steps:

  1. Construct student test score residuals.
  2. Estimate the autocovariance of scores across classes taught by a given teacher.

3. Predict VA for each teacher in each year using test score data from other years.

1. Within each subject (math and English) and school-level (elementary and middle), we construct test score residuals A_{it} by regressing raw standardized test scores A^*_{it} on a vector of covariates $\boldsymbol{X}_{it}$ and teacher fixed effects, as in (5).

   - We control for prior test scores using a cubic polynomial in prior-year scores in math and a cubic in prior-year scores in English, and we interact these cubics with the student's grade level to permit flexibility in the persistence of test scores as students age.

     - We exclude observations with missing data on current or prior scores in the subject for which we are estimating VA. We also exclude classrooms which have fewer than seven observations with current and lagged scores in the relevant subject (2 percent of observations) to avoid estimating VA based on very few observations. When prior test scores in the other subject are missing, we set the other subject prior score to zero and include an indicator for missing data in the other subject interacted with the controls for prior own-subject test scores.

   - We also control for students' ethnicity, gender, age, lagged suspensions and absences, and indicators for special education, limited English proficiency, and grade repetition.

   - We also include the following class- and school-level controls: (i) cubics in class and school-grade means of prior-year test scores in math and English (defined based on those with non-missing prior scores) each interacted with grade; (ii) class and school-year means of all the other individual covariates; (iii) class size and class-type indicators (honors, remedial); and (iv) grade and year dummies.

7 Specification/Falsification Tests

- Specification Test

  - Under the stationarity assumption in (3), an OLS regression of A_{it} on $\hat{\mu}_{jt}$ - the best linear predictor of A_{it} - should yield a coefficient of 1 by construction.

  - We confirm that this is the case in column 1 of Table 3, which reports estimates from a univariate OLS regression of test score residuals A_{it} on $\hat{\mu}_{jt}$ in the sample used to estimate the VA model.

  - We include fixed effects for subject (math versus English) by school-level (elementary versus middle) in this and all subsequent regressions to obtain a single regression coefficient which is identified purely from variation within the subject-by-school-level cells.

- We cluster standard errors at the school-by-cohort level (where cohort is defined as the year in which a child entered kindergarten) to adjust for correlated errors across students within classrooms and the multiple observations for each student in different grades and subjects.

- Forecast Bias
  - To begin, observe that regressing test score residuals A_{it} on $\hat{\mu}_{jt}$ in observational data yields a coefficient of 1 because $\hat{\mu}_{jt}$ is the best linear predictor of A_{it}: $\frac{cov(A_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})} = \frac{cov(\mu_{jt},\hat{\mu}_{jt})+cov(\varepsilon_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})} = 1$.
  - It follows from the definition of forecast bias that in observational data, $B(\hat{\mu}_{jt}) = \frac{cov(\varepsilon_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})}$.
  - Intuitively, the degree of forecast bias can be quantified by the extent to which students are sorted to teachers based on unobserved determinants of achievement ε_{it}.
  - Although we cannot observe ε_{it}, we can obtain information on components of ε_{it} using variables that predict test score residuals A_{it} but were omitted from the VA model, such as parent income.
  - Let $\boldsymbol{P}^{*}_{it}$ denote a vector of such characteristics and $\boldsymbol{P}_{it}$ denote the residuals obtained after regressing the elements of $\boldsymbol{P}^{*}_{it}$ on the baseline controls $\boldsymbol{X}_{it}$ in a specification with teacher fixed effects, as in (5). Decompose the error in score residuals $\varepsilon_{it} = \boldsymbol{\rho}\boldsymbol{P}_{it} + \varepsilon'_{it}$ into the component which projects onto $\boldsymbol{P}_{it}$ and the remaining (unobservable) error ε'_{it}. To estimate forecast bias using $\boldsymbol{P}$, we make the following assumption.
    * ASSUMPTION 2 (Selection on Excluded Observables): *Students are sorted to teachers purely on excluded observables* $\boldsymbol{P}$: $E[\varepsilon'_{it}|j] = E[\varepsilon'_{it}]$.
    * Under this assumption, $B = \frac{cov(\boldsymbol{\rho}\boldsymbol{P}_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})}$. As in (5), we estimate the coefficient vector $\boldsymbol{\rho}$ using an OLS regression of A_{it} on $\boldsymbol{P}_{it}$ with teacher fixed effects: $A_{it} = \alpha_j + \boldsymbol{\rho}\boldsymbol{P}_{it}$.
    * This leads to the feasible estimator $B_p = \frac{cov(A^p_{it},\hat{\mu}_{jt})}{var(\hat{\mu}_{jt})}$ where $A^p_{it} = \hat{\boldsymbol{\rho}}\boldsymbol{P}_{it}$ is estimated using (13). Equation (14) shows that forecast bias B_p can be estimated from an OLS regression of predicted scores A^p_{it} on VA estimates under Assumption 2.
  - Parent Characteristics
    * We define a vector of parent characteristics $\boldsymbol{P}^{*}_{it}$ which consists of the following variables: mother's age at child's birth, indicators for parent's 401(k) contributions and home ownership, and an indicator for the parent marital status interacted with a quartic in parent household income.
      · We code the parent characteristics as zero for the 12.3 percent of students whom we are unable to match to a parent either because we could not match the student to the tax data

(10.1 percent) or because we could not find a parent for a matched student (2.2 percent). We include indicators for missing parent data in both cases. We also code mother's age at child's birth as zero for observations where we match parents but do not have valid data on the parent's age, and include an indicator for such cases.

  * We construct residual parent characteristics $\boldsymbol{P}_{it}$ by regressing each element of $\boldsymbol{P}^{*}_{it}$ on the baseline control vector $\boldsymbol{X}_{it}$ and teacher fixed effects, as in (5).
  * We then regress A_{it} on $\boldsymbol{P}_{it}$, again including teacher fixed effects, and calculate predicted values $A^{p}_{it} = \hat{\boldsymbol{\rho}}\boldsymbol{P}_{it}$. We fit separate models for each subject and school level (elementary and middle) as above when constructing the residuals $\boldsymbol{P}_{it}$ and predicted test scores A^{p}_{it}.
  * In column 2 of Table 3, we regress A^{p}_{it} on $\hat{\mu}_{jt}$, including subject-by-school-level fixed effects as above.
  * Another intuitive way to assess the degree of selection on parent characteristics - which corresponds to familiar methods of assessing omitted variable bias - is to control for $\boldsymbol{P}^{*}_{it}$ when estimating the impact of $\hat{\mu}_{jt}$ on test scores, as in Kane and Staiger (2008, Table 6). To implement this approach using within-teacher variation to construct residuals, we first regress raw test scores A^{*}_{it} on the baseline control vector used in the VA model $\boldsymbol{X}_{it}$, parent characteristics $\boldsymbol{P}^{*}_{it}$, and teacher fixed effects, as in (5). Again, we fit a separate model for each subject and school level (elementary and middle). We then regress the residuals from this regression (adding back the teacher fixed effects) on $\hat{\mu}_{jt}$, including subject-by-school-level fixed effects as in column 1.

- Prior Test Scores
  * Another natural set of variables to evaluate bias is prior test scores (Rothstein 2010). Value-added models typically control for $A_{i,t-1}$, but one can evaluate sorting on $A_{i,t-2}$ (or, equivalently, on lagged gains, $A_{i,t-1} - A_{i,t-2}$). The question here is effectively whether controlling for additional lags affects substantially VA estimates once one controls for $A_{i,t-1}$. For this analysis, we restrict attention to the subsample of students with data on both lagged and twice-lagged scores, essentially dropping fourth grade from our sample. We reestimate VA $\hat{\mu}_{jt}$ on this sample to obtain VA estimates on exactly the sample used to evaluate bias.
  * We assess forecast bias due to sorting on lagged score gains using the same approach as with parent characteristics. Column 4 of Table 3 replicates column 2, using predicted score residuals based on $A_{i,t-2}$, which we denote by A^{l}_{it}, as the dependent variable.

- Table 3

- Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses.
- The regressions are run on the sample used to estimate the baseline VA model, restricted to observations with a non-missing leave-out teacher VA estimate.
- There is one observation for each student-subject-school year in all regressions.
- Teacher VA is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years, following the procedure in Sections IB and III.
- Teacher VA is estimated using the baseline control vector, which includes: a cubic in lagged own- and cross-subject scores, interacted with the student's grade level; student-level characteristics including ethnicity, gender, age, lagged suspensions and absences, and indicators for grade repetition, special education, and limited English; class size and class-type indicators; cubics in class and school-grade means of lagged own- and cross-subject scores, interacted with grade level; class and school-year means of all the student-level characteristics; and grade and year dummies.
- When prior test scores in the other subject are missing, we set the other subject prior score to zero and include an indicator for missing data in the other subject interacted with the controls for prior own-subject test scores.
- In columns 1 and 3, the dependent variable is the student's test score in a given year and subject.
- In column 2, the dependent variable is the predicted value generated from a regression of test score on mother's age at child's birth, indicators for parent's 401(k) contributions and home ownership, and an indicator for the parent's marital status interacted with a quartic in parent household income, after residualizing all variables with respect to the baseline control vector.
- In column 4, the dependent variable is the predicted value generated in the same way from twice-lagged test scores.
- See Section IVB for details of the estimating equation for predicted scores.

8 Stata *create_schd_work_final.do* File

8.1 Control Variables

- PREZ*
  - Test grade interacted with a cubic in same subject prior test score.

- `i.test_grade#(c.l_score##c.l_score##c.l_score)`

- PREO*
  - Test grade interacted with a cubic in other subject prior test score, with missing other subject test scores replaced as zero.
  - `i.test_grade#(c.l_other##c.l_other##c.l_other)`
- MPREO*
  - Missing other subject prior test score.
  - `MPREO = l_other==.`
  - Missing other subject prior test score interacted with a cubic in same subject prior test score.
  - `i.MPREO#(c.l_score##c.l_score##c.l_score)`
- CL*
  -
- SG*
  -
- SY*
  -
- female
  -
- age_current
  -
- lunch
  -
- miss_lunch

–

- ETHN*
  –
- sped
  –
- limited_english
  –
- repeat_grade
  –
- repeat_test
  –
- absences
  –
- suspensions
  –
- post99*
  –
- hon
  –
- rem
  –
- enrich

–

- mod

  –

- foreign

  –

9 Stata *analysis_final.do* File

9.1 ${datadir}/schd_work_trim_${SCHD_worktrimdate}.dta

1. `use ${base_vars} ${controls_to_load} using ${datadir}/schd_work_${SCHD_workdate} if score~=.&class ~=.&teacher~=.&sample_to_keep==1, clear`
2. `keep if inrange(grade,4,8)`
3. `egen miss = rowmiss(score PREZl* PREOl*)`
4. `g nomiss = miss==0`
5. `bys class: egen n_tested = sum(nomiss)`
6. `keep if n_tested>=7`
7. `keep if miss==0`
8. `xi, prefix(YEAR)i.year`
9. `xi, prefix(GRDE)i.grade`
10. `egen sch_yr = group(sch year)`
11. `g cohort=year-grade`
12. `egen sch_cohort = group(sch cohort)`
13. `save ${datadir}/schd_work_trim_${SCHD_worktrimdate}.dta, replace`

9.2 Table 3

9.2.1 Column 1

1. `use ${datadir}/tfx/score_r_${tfxdate}, clear`
2. `gen middle_math=middle*math`
3. `merge m:1 teacher year math using ${datadir}/tfx/tfx_${tfxdate}_tym.dta, assert(3)nogen keepusing (tv)`
4. `reg score_r tv middle math middle_math, cluster(sch_cohort)`

9.2.2 Column 2

1. `use ${datadir}/analysis_irsdate${IRSdate}_tfxdate${tfxdate}.dta, clear`
2. `reg diff_score_rminuspar tv middle math middle_math [w=diff_score_rminuspar_count], cluster(sch_cohort )`

9.2.3 Column 3

1. `use ${datadir}/analysis_irsdate${IRSdate}_tfxdate${tfxdate}.dta, clear`
2. `reg score_r_withpar tv middle math middle_math [w=score_r_withpar_count], cluster(sch_cohort)`

9.2.4 Column 4

1. `use ${datadir}/schd_work_trim_${SCHD_worktrimdate}.dta, clear`
2. `keep if l2_score~=.`
3. `vam score, teacher(teacher)year(year)class(class)output_addvars(sch class_id math) controls(${baseline_VAM_controls})output(${datadir}/tfx/tfx_l2_score_notmiss_${tfxdate}) tfx_resid(teacher)by(middle math)driftlimit(7) data(merge tv score_r)`
4. `postvam_save_l2score, controls(${baseline_VAM_controls})output(${datadir}/tfx/score_r_l2_score_notmiss_tfxdat {tfxdate})score_r_l2 l2_score_r`

(a) `areg l2_score ${baseline_VAM_controls}, a(teacher)`

(b) `predict l2_score_r if e(sample), dr`

(c) `areg score l2_score ${baseline_VAM_controls}, a(teacher)`

(d) `predict score_r_l2 if e(sample), dr`

- Do this by `middle` and `math`, i.e. `if middle==1 & math==0`

5. `gen diff_score_l2score = score_r - score_r_l2`

6. `reg diff_score_l2score tv middle math middle_math, cluster(sch_cohort)`

10 Stata *vam.ado* File

This section details the steps used in the vam.ado file that implements the procedure outlined in (Chetty, Friedman and Rockoff, 2014).

The program takes the following values: *depvar* (A^*_{it}), *teacher* ($_j$), *year* ($_t$), *class* ($_c$), *controls* ($\mathbf{X}_{it}$), and *absorb* or *tfx_resid* (α_j). The program is by-able and accepts a *byvar*.

1. Drop observations if missing *teacher*, *year*, or *class*.

2. Keep observations for the specific value of *byvar*.

3. Create residuals: *score_r* (A_{it}).

   (a) If neither *absorb* nor *tfx_resid* are specified then residuals $A_{it} = A^*_{it} - \hat{\beta}\mathbf{X}_{it}$ are created by running a regression of $A^*_{it} = \beta\mathbf{X}_{it}$.

   (b) If *absorb* is specified, then residuals $A_{it} = A^*_{it} - \hat{\alpha}_j - \hat{\beta}\mathbf{X}_{it}$ are created by running an areg of $A^*_{it} = \alpha_j + \beta\mathbf{X}_{it}$ with α_j absorbed. The residuals do not add back in the absorbed effects.

   (c) If *tfx_resid* is specified, then residuals $A_{it} = A^*_{it} - \hat{\beta}\mathbf{X}_{it}$ are created by running an areg of $A^*_{it} = \alpha_j + \beta\mathbf{X}_{it}$ with α_j absorbed. The residuals add back in the absorbed effects as done with teacher fixed effects in (Chetty, Friedman and Rockoff, 2014).

4. Save the number of observations used in the regression: *num_obs* (N).

5. Save the number of parameters: *num_par*.

(a) If *absorb* is specified then *num_par* = model degrees of freedom (*df_m*, the number of βs) + degrees of freedom for absorbed effect (*df_a*) + 1.

(b) If *tfx_resid* is specified or neither *absorb* nor *tfx_resid* are specified then *num_par* = model degrees of freedom (*df_m*, the number of βs) + 1.

6. Compute total variance: $var\_total = var(A_{it}) \cdot \frac{N-1}{num\_obs - num\_par}$

7. Compute individual (within class) variance: var_ind (σ^2_ε).

   (a) Calculate the mean A_{it} by teacher by year by class: *class_mean* ($\bar{A}_{ct}$).

   (b) Calculate the within class deviation of A_{it}: $individual\_dev\_from\_class = A_{it} - \bar{A}_{ct}$.

   (c) Calculate the number of students tested by teacher by year by class: *n_tested* (n_{ct}).

   (d) Calculate the number of classes with nonzero *num_tested*: *num_class* (C).

   (e) $\hat{\sigma}^2_\varepsilon = var(individual\_dev\_from\_class) \cdot \frac{N-1}{N-C-num\_par+1}$

8. Collapse the data to the teacher by year by class level by keeping the first observation for each teacher by year by class.

9. If there are multiple classes per teacher by year then do the following:

   (a) Randomly sort the classes conditional on teacher and year, with the order within teacher by year given by *class_num*.

   (b) *tsset* the data with the teacher by year as the *panelvar* and *class_num* as the *timevar*. Teacher by year combinations are given by *identifier*.

   (c) Calculate the covariance of $\bar{A}_{ct}$ and $F.\bar{A}_{ct}$ weighted by $n_{ct}+F.n_{ct}$ for all observations with multiple classes per teacher by year: *cov_sameyear* ($\hat{\sigma}^2_{A0}$).

   (d) Calculate the correlation of $\bar{A}_{ct}$ and $F.\bar{A}_{ct}$ weighted by $n_{ct}+F.n_{ct}$ for all observations with multiple classes per teacher by year: $corr\_sameyear = \frac{\hat{\sigma}^2_{A0}}{\sqrt{var(\bar{A}_{ct})} \cdot \sqrt{var(F.\bar{A}_{ct})}}$.

   (e) Calculate the number of observations used to calculate $\hat{\sigma}^2_{A0}$: *obs_sameyear*.

10. Compute the class-level variance component: var_class $\hat{\sigma}^2_\theta = var\_total - \hat{\sigma}^2_\varepsilon - \hat{\sigma}^2_{A0}$.

11. Compute the weight for each classroom: $weight$ $h_{ct} = \frac{1}{\hat{\sigma}^2_\theta + \frac{\hat{\sigma}^2_\varepsilon}{n_{ct}}}$.

12. Generate a dummy *excess_weight* if an observation is missing h_{ct}. Replace $h_{ct} = 1$ if the observation is missing h_{ct}.

13. Collapse the data to the teacher by year by *byvar* level. Calculate the average $\bar{A}_{ct}$ weighted by h_{ct} ($\bar{A}_{jt}$), and the raw sum of h_{ct}, n_{ct}, and *excess_weight*. For teachers with one classroom per year then $\bar{A}_{ct}$ is unchanged.

14. Subtract the raw sum of *excess_weight* from *weight*.

15. Calculate the covariance of years t and $t+s$ ($\hat{\sigma}_{As}, s \in \{1,2,\dots\}$) for every s and store in a vector: $\boldsymbol{m}$. $\boldsymbol{m}$ is a $dim \times 1$ vector where dim is equal to the maximum difference in years in the dataset: *maxspan*.

    (a) *tsset* the data with the teacher as the *panelvar* and year as the *timevar*.

    (b) Calculate the covariance of $\bar{A}_{jt}$ and $Fs.\bar{A}_{jt}$ ($\hat{\sigma}_{As}$) weighted by $n_{ct}+Fs.n_{ct}$. This is the $[s,1]$ entry in the $dim \times 4$ matrix $\boldsymbol{CC}$.

    (c) Calculate the correlation of $\bar{A}_{jt}$ and $Fs.\bar{A}_{jt}$ weighted by $n_{ct}+Fs.n_{ct}$ as $\frac{\hat{\sigma}_{As}}{\sqrt{var(\bar{A}_{jt})}\cdot\sqrt{var(Fs.\bar{A}_{jt})}}$. This is the $[s,2]$ entry in the $dim \times 4$ matrix $\boldsymbol{CC}$.

    (d) Calculate the number of observations used to calculate the covariance of $\bar{A}_{jt}$ and $Fs.\bar{A}_{jt}$. This is the $[s,3]$ entry in the $dim \times 4$ matrix $\boldsymbol{CC}$.

    (e) Run a regression of $\bar{A}_{jt} = \beta_s Fs.\bar{A}_{jt}$ weighted by $n_{ct}+Fs.n_{ct}$ with standard errors clustered by teacher. Calculate the t-statistic of the coefficient on $Fs.\bar{A}_{jt}$ as $t-stat_s = \frac{\hat{\beta}_s}{s.e.(\hat{\beta}_s)}$. Calculate the standard error of the covariance of $\bar{A}_{jt}$ and $Fs.\bar{A}_{jt}$ as $|\frac{\hat{\sigma}_{As}}{t-stat_s}|$. This is the $[s,4]$ entry in the $dim \times 4$ matrix $\boldsymbol{CC}$.

    (f) If $dim = 7$ then:

$$\mathbf{CC} = \begin{bmatrix} \hat{\sigma}_{A1} & corr(\bar{A}_{jt}, F1.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F1.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A1}}{t-stat_1}| \\ \hat{\sigma}_{A2} & corr(\bar{A}_{jt}, F2.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F2.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A2}}{t-stat_2}| \\ \hat{\sigma}_{A3} & corr(\bar{A}_{jt}, F3.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F3.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A3}}{t-stat_3}| \\ \hat{\sigma}_{A4} & corr(\bar{A}_{jt}, F4.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F4.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A4}}{t-stat_4}| \\ \hat{\sigma}_{A5} & corr(\bar{A}_{jt}, F5.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F5.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A5}}{t-stat_5}| \\ \hat{\sigma}_{A6} & corr(\bar{A}_{jt}, F6.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F6.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A6}}{t-stat_6}| \\ \hat{\sigma}_{A7} & corr(\bar{A}_{jt}, F7.\bar{A}_{jt}) & obs(\bar{A}_{jt}+F7.\bar{A}_{jt}) & |\frac{\hat{\sigma}_{A7}}{t-stat_7}| \end{bmatrix}$$

    (g) If *drift_limit* is not specified, then the vector $\boldsymbol{m}$ is equal to

    $\mathbf{m} = [\hat{\sigma}^2_{A0}, \hat{\sigma}_{A1}, \hat{\sigma}_{A2}, \dots, \hat{\sigma}_{Adim}]$.

    (h) If *drift_limit* is specified, then the vector $\boldsymbol{m}$ is equal to

    $\mathbf{m} = [\hat{\sigma}^2_{A0}, \hat{\sigma}_{A1}, \dots, \hat{\sigma}_{Adrift\_limit}, \dots, \hat{\sigma}_{Adrift\_limit}]$.

16. Compute teacher value added

(a) Calculate the number of observations per teacher: *obs_teacher*.

(b) For each teacher $j = \iota$ by year $t = \tau$ observation do the following:

i. Keep the observations for all years corresponding to teacher ι.

ii. Let T denote a rescaling of t such that, conditional on teacher j, $T = t - min(t)_j + 1$. Thus $T = 1$ for the first year of data for each teacher.

iii. Drop the observations for teacher ι corresponding to year τ.

iv. Drop the observations for teacher ι that are missing $\bar{A}_{jt}$.

v. If there are actually observations in other years:

A. Create a $dim \times dim$ symmetric matrix $\boldsymbol{M}$ where the $[i, k]$ entry of the lower triangle is $\hat{\sigma}^2_{A_{i-k}}$.

$$\mathbf{M} = \begin{bmatrix} \hat{\sigma}^2_{A0} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A2} & \cdots & \hat{\sigma}^2_{Adim} \\ \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A0} & \hat{\sigma}^2_{A1} & & \vdots \\ \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A1} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \hat{\sigma}^2_{Adim-1} \\ \hat{\sigma}^2_{Adim} & \hat{\sigma}^2_{Adim-1} & \cdots & \cdots & \hat{\sigma}^2_{A0} \end{bmatrix}$$

B. Create a $N_{jt} \times max(T)_j$ matrix ***temp*** such that for rows $i = 1, \ldots, N_{jt}$ row i contains a value of 1 in column T_i and 0s in all other columns, where T_i indicates the ith value of all possible values T (excluding year τ) for teacher ι and N_{jt} is the number of valid years for teacher ι excluding year τ.

For example, if teacher has values $T = \{1, 3, 4, 5, 6\}$ and we are calculating value added for year $T = 4$ then $N_{jt} = 4$ and row 3 of matrix *temp* will have a 1 in column 5.

$$\text{i.e. } temp = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

C. Create a $N_{jt} \times dim$ matrix $\boldsymbol{A}$ equal to ***temp*** left appended to an $N_{jt} \times dim - max(T)_j$ matrix of all 0s.

Using the above example where $dim = 7$:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

D. Compute a matrix ***vcv*** $(\Sigma_{A_{jt}})$ which is equal to $A \times M \times A'$ added to a diagonal matrix with $\frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}}$ as the elements of the diagonal. For teachers with one classroom per year then $\sum_{c\in\{c:j(c)=j\}} h_{ct} = h_{ct}$ and thus $\frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} = \hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\varepsilon^2}{n_{ct}}$ because they have the same number of observations at the teacher by year by classroom level as at the teacher by year level.

Using the above example where $dim = 7$:

$$\mathbf{\Sigma}_{A_{jt}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} & \hat{\sigma}_{A1}^2 & \hat{\sigma}_{A2}^2 & \cdots & \hat{\sigma}_{Adim}^2 \\ \hat{\sigma}_{A1}^2 & \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} & \hat{\sigma}_{A1}^2 & & \vdots \\ \hat{\sigma}_{A2}^2 & \hat{\sigma}_{A1}^2 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \hat{\sigma}_{Adim-1}^2 \\ \hat{\sigma}_{Adim}^2 & \hat{\sigma}_{Adim-1}^2 & \cdots & \cdots & \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{\Sigma}_{A_{jt}} = \begin{bmatrix} \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} & \hat{\sigma}_{A2}^2 & \hat{\sigma}_{A4}^2 & \hat{\sigma}_{A5}^2 \\ \hat{\sigma}_{A2}^2 & \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} & \hat{\sigma}_{A2}^2 & \hat{\sigma}_{A3}^2 \\ \hat{\sigma}_{A4}^2 & \hat{\sigma}_{A2}^2 & \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} & \hat{\sigma}_{A1}^2 \\ \hat{\sigma}_{A5}^2 & \hat{\sigma}_{A3}^2 & \hat{\sigma}_{A1}^2 & \hat{\sigma}_{A0}^2 + \frac{1}{\sum_{c\in\{c:j(c)=j\}} h_{ct}} \end{bmatrix}$$

E. Compute a $1 \times N_{jt}$ matrix *phi* (γ_{jt}) which is equal to the Tth row of $\boldsymbol{M}$ multiplied by ***A***'

Using the above example where $dim = 7$:

$$\gamma_{jt} = \begin{bmatrix} \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A0} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A3} \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\gamma_{jt} = \begin{bmatrix} \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A2} \end{bmatrix}$$

F. Create a vector of weights equal to $\gamma_{jt} \times \mathbf{\Sigma}^{-1}_{A_{jt}}$.

Using the above example where $dim = 7$:

$$\begin{bmatrix} \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A2} \end{bmatrix}$$

$$\times \begin{bmatrix} \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A4} & \hat{\sigma}^2_{A5} \\ \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A3} \\ \hat{\sigma}^2_{A4} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A1} \\ \hat{\sigma}^2_{A5} & \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} \end{bmatrix}^{-1}$$

G. Compute the teacher value added tv ($\hat{\mu}_{jt}$) as $\gamma_{jt} \times \mathbf{\Sigma}^{-1}_{A_{jt}}$ multiplied by the vector of valid $\bar{A}_{jt}$ excluding the $\bar{A}_{jt}$ from year τ ($\vec{A}^{-t}_{j}$).

Using the above example where $dim = 7$:

$$\hat{\mu}_{jt} = \begin{bmatrix} \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A2} \end{bmatrix}$$

$$\times \begin{bmatrix} \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A4} & \hat{\sigma}^2_{A5} \\ \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A3} \\ \hat{\sigma}^2_{A4} & \hat{\sigma}^2_{A2} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} & \hat{\sigma}^2_{A1} \\ \hat{\sigma}^2_{A5} & \hat{\sigma}^2_{A3} & \hat{\sigma}^2_{A1} & \hat{\sigma}^2_{A0} + \frac{1}{\sum_{c \in \{c:j(c)=j\}} h_{ct}} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \bar{A}_{j1} \\ \bar{A}_{j3} \\ \bar{A}_{j5} \\ \bar{A}_{j6} \end{bmatrix}$$

References

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *The American Economic Review*, 104(9): 2593–2632.