

Predictive effects of teachers and schools on test scores, college attendance, and earnings

Gary E. Chamberlain¹

Department of Economics, Harvard University, Cambridge, MA 02138

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2011.

Contributed by Gary E. Chamberlain, August 20, 2013 (sent for review April 18, 2013)

I studied predictive effects of teachers and schools on test scores in fourth through eighth grade and outcomes later in life such as college attendance and earnings. For example, predict the fraction of a classroom attending college at age 20 given the test score for a different classroom in the same school with the same teacher and given the test score for a classroom in the same school with a different teacher. I would like to have predictive effects that condition on averages over many classrooms, with and without the same teacher. I set up a factor model that, under certain assumptions, makes this feasible. Administrative school district data in combination with tax data were used to calculate estimates and do inference.

education production function | unmeasured inputs | teacher effects

The outcome data are based on elementary and middle school classrooms, grades four through eight. For a classroom, there is an average score based on a math or reading test given near the end of the school year. There are also later outcome measures for that classroom. These measures include the fraction of the classroom that is attending college at age 20 and the average earnings of the classroom at age 28. The classrooms can be grouped by schools, and, within a school, can be grouped by teacher.

The goal of the paper is to provide predictive effects of teachers and schools on these outcomes. For example, predict the fraction of a classroom attending college at age 20 given the test score for a different classroom in the same school with the same teacher and given the test score for a classroom in the same school with a different teacher. Or predict the fraction of a classroom attending college at age 20 given the fraction attending college for a different classroom with the same teacher and given the fraction attending college for a classroom in the same school with a different teacher. I would like to have predictive effects that condition on averages over many classrooms, with and without the same teacher. I set up a factor model that, under certain assumptions, makes this feasible. Then I can define teacher and school factors based on test score data and measure the predictive effect of the teacher factor on college attendance. More directly, I can define teacher and school factors based on the college attendance data and measure the predictive effect of the teacher factor on college attendance.

These predictive effects can be based on residuals, where first we form predictions based on observed variables (X) such as class size, years of teacher experience, lagged test scores, and parent characteristics. The residuals are the prediction errors. Then the teacher and school effects that I measure in these residuals correspond to unmeasured (latent) variables or, more precisely, to the parts of those latent variables that are not predictable using the observed variables in X . I am interested in these latent variables because they may be related to unmeasured characteristics of teachers that have a causal effect on outcomes, in the sense of unmeasured inputs in a production function. After setting up the factor model, I discuss how it could be related, under random assignment assumptions, to a production function.

Rivkin et al. (1) noted that students and parents refer often to differences in teacher quality and act to ensure placement in classes with specific teachers. Existing empirical evidence, however, does

not find a strong role for measured characteristics of teachers—such as teacher experience, education, and test scores of teachers—in the determination of academic achievement of students. This lack of a strong role for measured characteristics motivates interest in unmeasured characteristics of teachers that have a causal effect on academic achievement. Related literature on estimating teacher effects on test scores includes refs. 2–10. A typical finding is that a 1-SD increase in the teacher factor corresponds to an increase in individual scores on the order of 0.1, where the units are SDs in the distribution of scores for individual students.

In the Tennessee Student/Teacher Achievement Ratio experiment, known as Project STAR, children entering kindergarten were randomly assigned to class types, which were randomly assigned to teachers. The random assignment was within schools (e.g., ref. 11). It may be plausible to assume that the double random assignment of students and teachers applied also to specific classrooms (12, 13). Chetty et al. (13) were able to obtain data on later outcomes for these children, such as college attendance and earnings, which could be combined with the test score data in Project STAR. These data make it possible to study classroom effects (including teacher effects and peer effects) on later outcomes. The advantage of the random assignment is that prekindergarten characteristics of children are not correlated within a kindergarten class. In the STAR data, however, each kindergarten teacher is only observed teaching a single kindergarten class, making it difficult to separate out the part of the classroom effect due to the teacher. A strength of the data used in my paper is that teachers are observed in multiple classrooms. However, we do not have the random assignment, so there is a concern that within a classroom, there is correlation across the students in characteristics that existed before the class. A teacher effect may in part reflect sorting of students to teachers, with persistent differences across teachers in characteristics of the students entering their classes. A motivation for using residuals is that it is more plausible to make random assignment assumptions within a school when working with residuals. I recognize that the available control variables may not be adequate to justify “as if” random assignment within schools; for example, the parent characteristics do not include parents’ education. Nevertheless, it is useful to ask

Significance

This study measures the predictive effect of teachers on adult outcomes. The data are based on elementary and middle school classrooms, grades four through eight. For a classroom, there is an average score based on a math or reading test given near the end of the school year. There are also later outcome measures for that classroom. These measures include the fraction of the classroom that is attending college at age 20 and the average earnings of the classroom at age 28. The predictive effects are based on observing multiple classrooms with the same teacher.

Author contributions: G.E.C. wrote the paper.

The author declares no conflict of interest.

¹E-mail: gary_chamberlain@harvard.edu.

what would be identified under within-school random assignment, and that analysis provides some guidance in presenting and interpreting the predictive effects.

To anticipate my results, using the full set of controls in X , when the factors are constructed using test score data, the predictive effect on college attendance of a 1-SD increase in the teacher factor is 0.13 percentage points. When the factors are constructed using data on college attendance, the predictive effect of a 1-SD increase in the teacher factor is 0.79 percentage points. Under the assumption (for residuals) of random assignment of students and teachers within schools, the 0.79 estimate has a structural interpretation based on a production function, and the 0.13 estimate provides a lower bound.

Methods

Let $Y_{ij,h}$ denote outcome h for classroom j in school i . Let X_{ij} denote a $K \times 1$ vector of predictor variables such as class size, years of teacher experience, and an average of test scores from a previous year for members of the classroom. We shall work with residuals of the form $U_{ij,h} = Y_{ij,h} - X_{ij}'\alpha_h$, where α_h is defined to solve a prediction problem, which will be discussed below. Let U_{ij} denote the $H \times 1$ vector formed from the outcome residuals for classroom j in school i . Components of U_{ij} are the residuals based on outcomes such as classroom average test score (ts), the fraction of the classroom attending college at age 20 (co), and the average earnings of the classroom at age 28 (ea).

I treat the schools as if they were a random sample from some unknown distribution, so that the schools are exchangeable. I only use a school i if there is at least one pair of classrooms with the same teacher and at least one pair of classrooms with different teachers. Within school i , form the set of classrooms such that for each one there is at least one other with the same teacher. Assign equal probability to each of these classrooms, choose one at random, and denote it by A . Assign equal probability to each of the other classrooms that have the same teacher as A . Choose one at random and denote it by B . Assign equal probabilities to all of the classrooms that have teachers different from that of classroom A . Choose one at random and denote it by C . The prediction problems I consider fit into the following framework:

$$\theta = \arg \min_{d \in \mathbb{R}^I} E[W_i g(U_{iA}, U_{iB}, U_{iC}, d)], \quad [1]$$

where g is a given function. For example

$$g(U_{iA}, U_{iB}, U_{iC}, d) = (U_{iA,co} - d_0 - d_1 U_{iB,ts} - d_2 U_{iB,ts}^2 - d_3 U_{iC,ts} - d_4 U_{iC,ts}^2), \quad [2]$$

with $U_{iA,co}$ equal to the residual corresponding to attending college at age 20 and $U_{iB,ts}$ equal to the residual corresponding to the test score. Then, θ gives the coefficients in the (weighted) minimum mean-square-error linear predictor

$$E^*(U_{iA,co} | 1, U_{iB,ts}, U_{iB,ts}^2, U_{iC,ts}, U_{iC,ts}^2) = \theta_0 + \theta_1 U_{iB,ts} + \theta_2 U_{iB,ts}^2 + \theta_3 U_{iC,ts} + \theta_4 U_{iC,ts}^2. \quad [3]$$

An alternative could use the absolute value of the error instead of the squared error in Eq. 2, in which case θ would give the coefficients in the (weighted) minimum mean absolute error linear predictor. The nonnegative scalar W_i allows for a weight in forming the moments. $W_i = 0$ unless school i has at least two classrooms with the same teacher and at least two classrooms with different teachers, so that the random vector (A, B, C) is well defined. The nonzero values of W_i could, for example, be the number of classrooms in school i with teachers who have at least two classrooms.

My estimator for θ is a sample counterpart of the minimization problem in Eq. 1. To make this explicit, let $\mathcal{N} = \{1, 2, \dots\}$ denote the positive integers, and let $S_i \subset \mathcal{N}$ denote the set of classrooms in school i . For each classroom $a \in S_i$, there is a teacher, denoted by $q(a) \in \mathcal{N}$. We can partition S_i into subsets S_{it} with the same teacher: $S_i = \cup_{t \in \mathcal{N}} S_{it}$, where $S_{it} = \{a \in S_i : q(a) = t\}$. Use iterated expectations to evaluate the expectation in Eq. 1 and simplify notation by dropping the i subscript:

$$E[Wg(U_A, U_B, U_C, d)] = E[E[Wg(U_A, U_B, U_C, d) | W, U, S]].$$

The outer expectation corresponds to our treatment of the schools as a random sample from some unknown distribution [so that (W_i, U_i, S_i) is independent and identically distributed from some unknown distribution]. We shall evaluate explicitly the inner expectation, which is over classes within the same school, given outcomes for each of the classes. Conditional on $(W, U, S) = (w, u, s)$:

$$E[Wg(U_A, U_B, U_C, d) | (W, U, S) = (w, u, s)] = E[m(A, B, C) | (W, U, S) = (w, u, s)],$$

with $m(A, B, C) = wg(U_A, U_B, U_C, d)$.

$$E[m(A, B, C) | q(A) = t, (W, U, S) = (w, u, s)] = \frac{1}{|S_t|} \sum_{a \in S_t} \left[\sum_{b \in S_t - \{a\}} \sum_{c \in S - S_t} m(a, b, c) / [(|S_t| - 1)(|s| - |S_t|)] \right],$$

where $|s|$ denotes the number of elements in the set s , so that $|S_t|$ is the number of classes taught by teacher t . Only condition on values for t such that $|S_t| > 1$. Only condition on values for s such that there is at least one pair of classrooms with different teachers, so that $|s| - |S_t| > 0$.

Apply iterated expectations:

$$E[m(A, B, C) | (W, U, S) = (w, u, s)] = \left(\sum_{t: |S_t| > 1} |S_t| \right)^{-1} \sum_{t: |S_t| > 1} |S_t| E[m(A, B, C) | q(A) = t, (W, U, S) = (w, u, s)] = \left(\sum_{t: |S_t| > 1} |S_t| \right)^{-1} \sum_{t: |S_t| > 1} \sum_{a \in S_t} \sum_{b \in S_t - \{a\}} \sum_{c \in S - S_t} m(a, b, c) / [(|S_t| - 1)(|s| - |S_t|)].$$

Now we can use these results to form our estimator. Let α_h be defined to solve a prediction problem such as

$$\alpha_h = \arg \min_{d \in \mathbb{R}^K} E \left(\sum_{j \in S_i} (Y_{ij,h} - X_{ij}'d)^2 \right) \quad (h = 1, \dots, H). \quad [4]$$

The sample analog for Eq. 4 is

$$\hat{\alpha}_h = \arg \min_{d \in \mathbb{R}^K} \frac{1}{I} \sum_{i=1}^I \left(\sum_{j \in S_i} (Y_{ij,h} - X_{ij}'d)^2 \right) \quad (h = 1, \dots, H), \quad [4']$$

providing the estimated residuals $\hat{U}_{ij,h} = Y_{ij,h} - X_{ij}'\hat{\alpha}_h$. The sample analog for Eq. 1 is

$$\hat{\theta} = \arg \min_{d \in \mathbb{R}^I} \frac{1}{I} \sum_{i=1}^I W_i \left(\sum_{t: |S_{it}| > 1} |S_{it}| \right)^{-1} \times \sum_{t: |S_{it}| > 1} \sum_{a \in S_{it}} \sum_{b \in S_{it} - \{a\}} \sum_{c \in S_i - S_{it}} g(\hat{U}_{ia}, \hat{U}_{ib}, \hat{U}_{ic}, d) / [(|S_{it}| - 1)(|s_i| - |S_{it}|)].$$

Computation shows how the computation simplifies in a special case, which includes Eqs. 2 and 3. For inference, I shall use bootstrap methods, based on treating the schools as a random sample from some unknown distribution. This approach does not impose any structure on the covariances within a school.

Within a school, we can form a partition of the classrooms, $S_i = \cup_{l=1}^L S_i^l$, for example by subject and grade. We can apply our analysis separately within each cell of the partition. It may be useful to have a compact summary of the results. One way to do this is to define (A^l, B^l, C^l) for each cell $l = 1, \dots, L$ of the partition. Assign a nonnegative weight W_i^l to cell l in school i , which is zero unless S_i^l contains at least one pair of classrooms with the same teacher and one pair of classrooms with different teachers. For the nonzero values, we could use

$$W_i^l = \sum_{t: |S_{it}^l| > 1} |S_{it}^l|. \quad [5]$$

Only use a school i if $\sum_{l=1}^L W_i^l > 0$. If $W_i^l > 0$, form the set of classrooms in S_i^l such that for each one there is at least one other with the same teacher. Assign equal probability to each of these classrooms, choose one at random, and denote it by A^l . Assign equal probability to each of the other classrooms in S_i^l that have the same teacher as A^l . Choose one at random and denote it by B^l . Assign equal probabilities to all of the classrooms in S_i^l that have teachers different from that of classroom A^l . Choose one at random and denote it by C^l . $[(A^l, B^l, C^l)$ is undefined if $W_i^l = 0$]. The new prediction problem is

$$\theta = \arg \min_{d \in \mathbb{R}^I} E \left[\sum_{l=1}^L W_i^l g(U_{iA^l}, U_{iB^l}, U_{iC^l}, d) \right]. \quad [1']$$

Factor Model. These predictive effects condition on a single score for a different classroom with the same teacher and a single score for a classroom

with a different teacher. A factor model can provide predictive effects that condition on averages over many classrooms, with and without the same teacher, and can provide a limit as the number of such classrooms tends to infinity. This factor model has the advantage of getting rid of the noise that comes from using data on only a few classrooms. Let $Z_{iA,T}$ denote unmeasured characteristics of the teacher for classroom A in school i , and let $Z_{i,S}$ denote unmeasured characteristics of the school. Define

$$F_{in} + G_{in} = E[h_n(U_{iA})|Z_{iA,T}, Z_{i,S}], \quad F_{in} = E[h_n(U_{iA})|Z_{i,S}] \quad (n = 1, \dots, N),$$

where $h_n(\cdot)$ is a given function. For example, $h_n(\cdot)$ could select a component of U_{iA} and raise it to the power n : $h_n(U_{iA}) = U_{iA,ts}^n$. Assume that

$$E[h_n(U_{iA})|Z_{iA,T}, Z_{i,S}] = E[h_n(U_{iB})|Z_{iA,T}, Z_{i,S}], \quad E[h_n(U_{iA})|Z_{i,S}] = E[h_n(U_{iC})|Z_{i,S}].$$

This assumption follows from the assumption of exchangeability across schools and the random selection of classrooms A , B , and C .

Assume that U_{iA} and U_{iB} are independent conditional on the latent variables $Z_{iA,T}$ and $Z_{i,S}$. A motivation for this assumption is that, without conditioning on additional information, we can regard the random variables corresponding to different classrooms for the same teacher (within a school) as exchangeable. If they could be embedded in an infinite sequence of exchangeable random variables, then conditional independence would follow from de Finetti's theorem (14). A richer analysis could exploit additional information, such as the temporal ordering of the classrooms for a given teacher, where patterns of serial correlation could emerge. I shall not pursue that here. The conditional independence implies that

$$\begin{aligned} \text{Cov}(h_n(U_{iA}), h_p(U_{iB})) &= E[\text{Cov}(h_n(U_{iA}), h_p(U_{iB})|Z_{iA,T}, Z_{i,S})] \\ &\quad + \text{Cov}(E[h_n(U_{iA})|Z_{iA,T}, Z_{i,S}], E[h_p(U_{iA})|Z_{iA,T}, Z_{i,S}]) \\ &= \text{Cov}(F_{in} + G_{in}, F_{ip} + G_{ip}) \quad (n, p = 1, \dots, N). \end{aligned}$$

Likewise, assume that U_{iA} and U_{iC} are independent conditional on $Z_{i,S}$, which implies that

$$\text{Cov}(h_n(U_{iA}), h_p(U_{iC})) = \text{Cov}(F_{in}, F_{ip}) \quad (n, p = 1, \dots, N).$$

Note that

$$E(F_{in} + G_{in}|Z_{i,S}) = E[E[h_n(U_{iA})|Z_{iA,T}, Z_{i,S}]|Z_{i,S}] = E[h_n(U_{iA})|Z_{i,S}] = F_{in},$$

so that $E[G_{in}|Z_{i,S}] = 0$, which implies that

$$\text{Cov}(G_{in}, F_{ip}) = 0 \quad (n, p = 1, \dots, N).$$

Therefore, we can obtain the moments $\text{Cov}(F_{in}, F_{ip})$ and $\text{Cov}(G_{in}, G_{ip})$ from

$$\text{Cov}(h_n(U_{iA}), h_p(U_{iB})) \quad \text{and} \quad \text{Cov}(h_n(U_{iA}), h_p(U_{iC})).$$

Let M be a subset of $\{1, \dots, N\}$. Note that

$$\begin{aligned} E^* [h_n(U_{iA})|1, \{F_{ip}, G_{ip}\}_{p \in M}] &= E^* [E[h_n(U_{iA})|Z_{iA,T}, Z_{i,S}]|1, \{F_{ip}, G_{ip}\}_{p \in M}] \\ &= E^* [F_{in} + G_{in}|1, \{F_{ip}, G_{ip}\}_{p \in M}] \\ &= E^* [F_{in}|1, \{F_{ip}\}_{p \in M}] + E^* [G_{in}|1, \{G_{ip}\}_{p \in M}]. \end{aligned}$$

Therefore, the slope coefficients in the linear predictor $E^*[h_n(U_{iA})|1, \{G_{ip}, F_{ip}\}_{p \in M}]$ can be obtained from $\text{Cov}(h_n(U_{iA}), h_p(U_{iB}))$ and $\text{Cov}(h_n(U_{iA}), h_p(U_{iC}))$ for $p \in M$.

Production Function. There are connections between the factor model and a production function, under random assignment assumptions. To be specific, consider the college attendance outcome $U_{iA,co}$, and let g denote the production function

$$U_{iA,co} = g(Z_{iA,CL}, Z_{iA,T}, Z_{i,S}).$$

The inputs $Z_{iA,T}$ and $Z_{i,S}$ are, as above, unmeasured characteristics of the teacher and the school for classroom A in school i . There is an additional input, $Z_{iA,CL}$, which corresponds to unmeasured characteristics of the students in classroom A . Simplify notation by writing the function as

$$U_{co} = g(Z_{CL}, Z_T, Z_S).$$

Let $Z = Z_1 \times Z_2 \times Z_3$ denote the domain of the input arguments. We shall condition on $Z_5 = z_3$ and consider counterfactual outcomes $g(z_1, z_2, z_3)$ as (z_1, z_2) varies over $Z_1 \times Z_2$. At any such point, $g(z_1, z_2, z_3)$ is a random variable with

$$E[g(z_1, z_2, z_3)|Z_5 = z_3] = \bar{g}(z_1, z_2|z_3).$$

Still conditioning on $Z_5 = z_3$, consider counterfactual outcomes as z_2 varies over Z_2 , averaging over the conditional distribution of Z_{CL} given $Z_5 = z_3$:

$$\bar{g}_1(z_2|z_3) = E[\bar{g}(Z_{CL}, z_2|z_3)|Z_5 = z_3] \quad \text{for } z_2 \in Z_2.$$

There is a structural function interpretation for $\bar{g}_1(\cdot|z_3)$: within a school with $Z_5 = z_3$, we can obtain potential expected output for various assigned values of the teacher input z_2 , holding constant the distribution of classroom characteristics (at the conditional distribution of Z_{CL} given $Z_5 = z_3$).

If, within schools, students and teachers are randomly assigned to classrooms (as in Project STAR), then (Z_{CL}, Z_T) is independent of $\{g(z_1, z_2, z_3)\}_{(z_1, z_2) \in Z_1 \times Z_2}$ conditional on $Z_5 = z_3$. In that case

$$E(U_{co}|Z_{CL} = z_1, Z_T = z_2, Z_5 = z_3) = E[g(z_1, z_2, z_3)|Z_5 = z_3] = \bar{g}(z_1, z_2|z_3).$$

In addition, random assignment within schools implies that Z_{CL} and Z_T are independent conditional on Z_5 , so that

$$E(U_{co}|Z_T = z_2, Z_5 = z_3) = E[\bar{g}(Z_{CL}, z_2|z_3)|Z_5 = z_3] = \bar{g}_1(z_2|z_3).$$

As above, define the factor $F_{co} + G_{co} = E(U_{co}|Z_T, Z_5)$. We have shown that if students and teachers are randomly assigned to classrooms within schools, then

$$F_{co} + G_{co} = E(U_{co}|Z_T, Z_5) = \bar{g}_1(Z_T|Z_5),$$

providing a connection between this factor and the production function. A motivation for the choice of X in forming residuals U is to make this random assignment assumption more plausible.

Because the random assignment assumption is within schools, I am interested in the variation in expected output that corresponds to the variation in the teacher input within a school. A convenient summary measure is

$$\gamma_G^{co} = [E[\text{Var}[\bar{g}_1(Z_T|Z_5)|Z_5]]]^{1/2}.$$

As above, define the factor $F_{co} = E(U_{co}|Z_5)$. Then

$$G_{co} = \bar{g}_1(Z_T|Z_5) - E[\bar{g}_1(Z_T|Z_5)|Z_5]$$

and

$$\gamma_G^{co} = [\text{Var}(G_{co})]^{1/2}.$$

Likewise, a convenient summary measure for cross-school variation is

$$\gamma_F^{co} = [\text{Var}(F_{co})]^{1/2}.$$

With random assignment only within schools, γ_F^{co} does not have a structural interpretation.

Now suppose that data on later outcomes are not (yet) available for a teacher, but data on test scores for multiple classrooms with that teacher are available. How can we connect $U_{co} = g(Z_{CL}, Z_T, Z_S)$ to the test score data? Define the factors

$$F_n + G_n = E[h_n(U_{ts})|Z_T, Z_S], \quad F_n = E[h_n(U_{ts})|Z_S] \quad (n = 1, \dots, J),$$

where $h_n(\cdot)$ is a given function, such as $h_n(U_{ts}) = U_{ts}^n$. Then the linear predictor of U_{co} given these factors equals the linear predictor of $\bar{g}_1(Z_T|Z_5)$:

$$\begin{aligned} E^*(U_{co}|1, \{F_n, G_n\}_{n=1}^J) &= E^*[E(U_{co}|Z_T, Z_5)|1, \{F_n, G_n\}_{n=1}^J] \\ &= E^*(F_{co} + G_{co}|1, \{F_n, G_n\}_{n=1}^J) \\ &= E^*[\bar{g}_1(Z_T|Z_5)|1, \{F_n, G_n\}_{n=1}^J]. \end{aligned}$$

This result provides a connection between the production function for U_{co} and a linear predictor based on factors derived from test scores.

This linear predictor is flexible in that we can choose a variety of functions $h_n(\cdot)$ in defining the factors. This flexibility suggests finding a lower bound on the mean square error for linear prediction of U_{co} from factors based on test scores. For notation, use

$$E^*(U_{co} | 1, \{F_n, G_n\}_{n=1}^J) = \gamma_0 + \sum_{n=1}^J \gamma_n (F_n + G_n) + \sum_{n=1}^J \gamma_{J+n} F_n.$$

Define

$$h_{F+G}^{(j)}(U_{ts}) = \sum_{n=1}^j \gamma_n h_n(U_{ts}), \quad h_F^{(j)}(U_{ts}) = \sum_{n=1}^j \gamma_{J+n} h_n(U_{ts}).$$

Then

$$E^*(U_{co} | 1, \{F_n, G_n\}_{n=1}^J) = \gamma_0 + E[h_{F+G}^{(j)}(U_{ts}) | Z_T, Z_S] + E[h_F^{(j)}(U_{ts}) | Z_S].$$

This result implies the following lower bound:

$$\begin{aligned} \text{MSE}^{(j)} &= \min_{d \in \mathcal{R}^{2j+1}} E \left[U_{co} - d_0 - \sum_{n=1}^j d_n (F_n + G_n) - \sum_{n=1}^j d_{J+n} F_n \right]^2 \\ &\geq \min_{r_{F+G}(\cdot), r_F(\cdot)} E [U_{co} - E[r_{F+G}(U_{ts}) | Z_T, Z_S] - E[r_F(U_{ts}) | Z_S]]^2 = \text{MSE}^*. \end{aligned}$$

The second minimization is over (square-integrable) functions r_{F+G} and r_F . Under suitable assumptions, we can construct a sequence of functions $\{h_{j,n}\}_{n=1}^j$ so that $\text{MSE}^{(j)} \rightarrow \text{MSE}^*$ as $J \rightarrow \infty$.

Note that

$$E^*(U_{co} | 1, \{G_n\}_{n=1}^J) = E^*(F_{co} + G_{co} | 1, \{G_n\}_{n=1}^J) = E(U_{co}) + \sum_{n=1}^J \gamma_n G_n,$$

which implies that

$$E^*(G_{co} | 1, \{G_n\}_{n=1}^J) = \sum_{n=1}^J \gamma_n G_n$$

[because F_{co} and F_m are uncorrelated with G_n for $n, m = 1, \dots, J$ and $E(G_{co}) = E(G_n) = 0$]. Therefore,

$$(\gamma_G^{co})^2 = \text{Var}(G_{co}) \geq \text{Var}\left(\sum_{n=1}^J \gamma_n G_n\right).$$

In the empirical work, I shall focus on γ_G^{co} and on

$$\gamma_{G,ts}^{co} = \left[\text{Var}\left(\sum_{n=1}^J \gamma_n G_n\right) \right]^{1/2},$$

which has a structural role in providing a lower bound for γ_G^{co} . A convenient summary measure based on cross-school variation is

$$\gamma_{F,ts}^{co} = \left[\text{Var}\left(\sum_{n=1}^J (\gamma_n + \gamma_{J+n}) F_n\right) \right]^{1/2}.$$

With random assignment only within schools, $\gamma_{F,ts}^{co}$ does not have a structural interpretation.

Empirical Results

The work of Chetty et al. (15) is the first to measure teacher effects on later outcomes such as college attendance and earnings. They combine two databases: administrative school district records and information on those students and their parents from US tax records. The school records are for a large, urban school district, covering the school years 1988–1989 through 2008–2009 and grades 3–8. Test scores are available for English language arts and for math from spring 1989 to spring 2009. The

scores are normalized within the year and grade to have a mean of 0 and SD of 1. The student records are linked to classrooms and teachers. Individual earnings data are obtained from W-2 forms, which are available from 1999 to 2010. College attendance is based on 1098-T forms, which colleges and other postsecondary institutions are required to file for reporting tuition payments and scholarships for every student.

Chetty et al. conducted most of their analysis of long-term impacts using a dataset collapsed to class means. This dataset with class means was used to obtain the results below. $Y_{ij,co}$ is the average test score for the class. $Y_{ij,ea}$ is the percent of the classroom attending college at age 20, and $Y_{ij,ea}$ is the average earnings of the classroom at age 28, expressed in 2010 dollars.

I shall use (weighted) minimum mean square error linear predictors, as in Eqs. 2 and 3. The partition in Eq. 1' is by subject (math and reading) and grade (4–8), giving $L = 2 \times 5 = 10$ cells, with weights W_i^j as in Eq. 5. In the lower grades, students may have the same teacher for math and reading, so putting math and reading classes in separate cells helps to ensure that different classes do not have students in common. Likewise, different classes could have students in common because, for example, there is overlap between a fourth grade class in one year and a fifth grade class in the following year. We avoid this overlap by only making comparisons for classrooms within the same subject and grade.

There are 118,439 classrooms in 917 schools. Of these schools, 866 satisfy the condition that $\sum_{i=1}^{10} W_i^j > 0$. Consider the linear predictor for college attendance in Eq. 3:

$$\begin{aligned} E^*(U_{iA,co} | 1, U_{iB,ts}, U_{iB,ts}^2, U_{iC,ts}, U_{iC,ts}^2) &= \theta_0 + \theta_1 U_{iB,ts} + \theta_2 U_{iB,ts}^2 \\ &\quad + \theta_3 U_{iC,ts} + \theta_4 U_{iC,ts}^2. \end{aligned}$$

If X_{ij} includes only a constant ($X_{ij} = 1$), then the estimates (with SEs in parentheses) are

$$\begin{aligned} \hat{\theta}_1 &= 13.34 \quad (0.37), \quad \hat{\theta}_2 = 2.26 \quad (0.31), \quad \hat{\theta}_3 = 7.84 \quad (0.31), \\ \hat{\theta}_4 &= 0.64 \quad (0.22). \end{aligned}$$

Dropping the quadratic terms, the coefficients (SEs) are 13.86 (0.38) on $U_{iB,ts}$ and 7.97 (0.31) on $U_{iC,ts}$. I shall rely on the factor model (below) for my discussion of the magnitudes of predictive effects.

The coefficients θ are defined as solutions to the minimization problem in Eq. 1'. The minimized value of the objective function provides a population value for mean square error. Likewise, there is a mean square error using just a constant to form the linear predictor $E^*(U_{iA,co} | 1)$. Let $1 - R_{co}^2$ denote the ratio of these mean square errors, so that R_{co}^2 gives the proportional reduction in mean square error due to including a quadratic in $U_{iB,ts}$ and a quadratic in $U_{iC,ts}$ in the linear predictor for $U_{iA,co}$. The estimate (with SE) is $\hat{R}_{co}^2 = 0.30 \quad (0.015)$.

Now let X_{ij} be the baseline control vector used by Chetty et al. It was chosen following previous work, in particular that of Kane and Staiger (6). It includes the following classroom-level variables: school year and grade indicators, class-type indicators (honors, remedial), class size, indicators for teacher experience, and cubics in class and school-grade means of lagged test scores in math and English each interacted with grade. It also includes class and school-year means of the following student characteristics: ethnicity, sex, age, lagged suspensions, lagged absences, and indicators for grade repetition, special education, and limited English. This baseline control vector gives

$$\hat{\theta}_1 = 1.28 \text{ (0.20)}, \quad \hat{\theta}_2 = -2.42 \text{ (0.51)}, \quad \hat{\theta}_3 = 0.92 \text{ (0.16)}, \\ \hat{\theta}_4 = -2.42 \text{ (0.40)},$$

with $\hat{R}_{co}^2 = 0.002$ (0.0004). Dropping the quadratic terms, the coefficients (SEs) are 1.26 (0.21) on $U_{iB,ts}$ and 0.86 (0.16) on $U_{iC,ts}$.

The controls matter a lot. This sensitivity relates to the difficulty in attaching causal interpretations to these results. This point has been emphasized in Rothstein (16). The issue has been addressed in Kane and Staiger (6), using a dataset with random assignment of teachers to classrooms, and in Chetty et al. (15), who look at effects based on changes in teaching staff.

These results condition on a single score for a different classroom with the same teacher and a single score for a classroom with a different teacher. I would like to have predictive effects that condition on averages over many classrooms, with and without the same teacher, and consider a limit as the number of such classrooms tends to infinity. This goal is feasible under the assumptions of the factor model. For notation, let

$$E^*(U_{iA,co} | 1, F_{i1}, G_{i1}, F_{i2}, G_{i2}) = \gamma_0 + \gamma_1(F_{i1} + G_{i1}) + \gamma_2(F_{i2} + G_{i2}) \\ + \gamma_3 F_{i1} + \gamma_4 F_{i2}, \quad [6]$$

where

$$F_{i1} + G_{i1} = E(U_{iB,ts} | Z_{iA,T}, Z_{i,S}), \quad F_{i1} = E(U_{iC,ts} | Z_{i,S}), \\ F_{i2} + G_{i2} = E(U_{iB,ts}^2 | Z_{iA,T}, Z_{i,S}), \quad F_{i2} = E(U_{iC,ts}^2 | Z_{i,S}), \quad [7]$$

$Z_{iA,T}$ denotes characteristics of the teacher of classroom A , and $Z_{i,S}$ denotes characteristics of the school of classroom A . As in the production function discussion, I construct an index corresponding to variation in teacher inputs within a school

$$\text{Index}_{iG,ts}^{co} = \gamma_1 G_{i1} + \gamma_2 G_{i2},$$

and use it to obtain a predictive effect in SD units:

$$\gamma_{G,ts}^{co} = \left[\text{Var}(\text{Index}_{iG,ts}^{co}) \right]^{1/2}.$$

Likewise, I construct an index corresponding to variation across schools

$$\text{Index}_{iF,ts}^{co} = (\gamma_1 + \gamma_3)F_{i1} + (\gamma_2 + \gamma_4)F_{i2},$$

and use it to obtain a predictive effect in SD units:

$$\gamma_{F,ts}^{co} = \left[\text{Var}(\text{Index}_{iF,ts}^{co}) \right]^{1/2}.$$

With the baseline controls in X , the factor model estimates give

$$\hat{\gamma}_1 = 1.70 \text{ (0.72)}, \quad \hat{\gamma}_2 = 1.56 \text{ (3.17)}, \quad \hat{\gamma}_3 = 9.98 \text{ (2.54)}, \\ \hat{\gamma}_4 = -60.68 \text{ (10.79)},$$

with predictive effects

$$\hat{\gamma}_{G,ts}^{co} = 0.16 \text{ (0.059)}, \quad \hat{\gamma}_{F,ts}^{co} = 1.19 \text{ (0.14)},$$

and $\hat{R}_{co}^2 = 0.013$ (0.003). An SD increase in the teacher factor based on the test score index implies a predicted increase in college attendance for each student in class A of 0.16 percentage points. If X includes only a constant, then this estimate increases from 0.16 percentage points to 5.81 percentage points.

Thus far, we used a (quadratic) function of the test score in predicting college attendance. We can also use college atten-

dance for other classes, and the factor model provides a way to condition on averages over many classrooms, with and without the same teacher. For notation, let

$$F_{i,co} + G_{i,co} = E(U_{iB,co} | Z_{iA,T}, Z_{i,S}), \quad F_{i,co} = E(U_{iC,co} | Z_{i,S}).$$

Then $F_{i,co} + G_{i,co}$ corresponds to an average of $U_{iB,co}$ over many classrooms other than A that share a teacher with A , and $F_{i,co}$ corresponds to an average of $U_{iC,co}$ over many classrooms that do not share a teacher with A but are in the same school. The optimal linear predictor for college attendance is

$$E^*(U_{iA,co} | 1, F_{i1}, G_{i1}, F_{i2}, G_{i2}, F_{i,co}, G_{i,co}) = F_{i,co} + G_{i,co}.$$

With the production function interpretation

$$F_{co} + G_{co} = \bar{g}_1(Z_T | Z_S), \quad G_{co} = \bar{g}_1(Z_T | Z_S) - E[\bar{g}_1(Z_T | Z_S) | Z_S].$$

The predictive effects in SD units are

$$\gamma_G^{co} = [\text{Var}(G_{i,co})]^{1/2}, \quad \gamma_F^{co} = [\text{Var}(F_{i,co})]^{1/2}.$$

With the baseline controls in X , the factor model estimates imply the predictive effects

$$\hat{\gamma}_G^{co} = 0.99 \text{ (0.22)}, \quad \hat{\gamma}_F^{co} = 3.71 \text{ (0.11)},$$

and $\hat{R}_{co}^2 = 0.134$ (0.007). An SD increase in $G_{i,co}$ corresponds to an increase of 0.99 percentage points for college attendance of class A . It is clear that basing the predictions for college attendance just on the test scores loses a great deal of information.

In parallel with the optimal linear predictor of college attendance, the optimal linear predictor for the test score is

$$E^*(U_{iA,ts} | 1, F_{i1}, G_{i1}, F_{i2}, G_{i2}, F_{i,co}, G_{i,co}) = F_{i1} + G_{i1}.$$

The predictive effects are

$$\gamma_G^{ts} = [\text{Var}(G_{i1})]^{1/2}, \quad \gamma_F^{ts} = [\text{Var}(F_{i1})]^{1/2}.$$

With the baseline controls in X , the estimates are

$$\hat{\gamma}_G^{ts} = 0.087 \text{ (0.002)}, \quad \hat{\gamma}_F^{ts} = 0.052 \text{ (0.002)},$$

and $\hat{R}_{ts}^2 = 0.260$ (0.006). An SD increase in G_{i1} corresponds to a predicted increase in score for each student in class A of 0.087, where the score units are SDs in the distribution of scores for individual students.

Now consider using the quadratic specification in Eq. 3 to obtain a linear predictor for $U_{iA,ea}$, the residuals corresponding to earnings at age 28:

$$E^*(U_{iA,ea} | 1, U_{iB,ts}, U_{iB,ts}^2, U_{iC,ts}, U_{iC,ts}^2) = \theta_0 + \theta_1 U_{iB,ts} + \theta_2 U_{iB,ts}^2 \\ + \theta_3 U_{iC,ts} + \theta_4 U_{iC,ts}^2.$$

With the baseline controls in X , the estimates are

$$\hat{\theta}_1 = 697 \text{ (270)}, \quad \hat{\theta}_2 = -430 \text{ (586)}, \quad \hat{\theta}_3 = 383 \text{ (179)}, \\ \hat{\theta}_4 = -955 \text{ (284)},$$

with $\hat{R}_{ea}^2 = 0.002$ (0.001). Dropping the quadratic terms, the coefficients (SEs) are 688 (269) on $U_{iB,ts}$ and 308 (176) on $U_{iC,ts}$. These results are based on fewer classrooms, 14,236 instead of 118,439, because only some of the students reached the age of 28 by 2010. There are 524 schools, of which 364 satisfy the condition that $\sum_{i=1}^{10} W_i^i > 0$.

For notation in the factor model, let

$$E^*(U_{iA,ea} | 1, F_{i1}, G_{i1}, F_{i2}, G_{i2}) = \gamma_0 + \gamma_1(F_{i1} + G_{i1}) + \gamma_2(F_{i2} + G_{i2}) + \gamma_3 F_{i1} + \gamma_4 F_{i2}, \quad [8]$$

where the factors are based on the test score, as in Eq. 7. With the baseline controls in X , the factor model estimates give

$$\hat{\gamma}_1 = 586 \quad (1, 277), \quad \hat{\gamma}_2 = 4,424 \quad (5, 885), \quad \hat{\gamma}_3 = 2,457 \quad (2, 242), \quad \hat{\gamma}_4 = -16,027 \quad (7, 961),$$

with predictive effects

$$\hat{\gamma}_{G,ts}^{ea} = 186 \quad (111), \quad \hat{\gamma}_{F,ts}^{ea} = 400 \quad (85).$$

An SD increase in the teacher factor based on the test score index implies a predicted increase in earnings of \$186. This estimate, however, lacks precision. This lack of precision becomes more serious when I try to define a teacher factor based directly on the earnings data, and I shall not pursue that here.

Chetty et al. linked students to their parents by finding the earliest 1040 form from 1996 to 2010 on which the student was claimed as a dependent. They constructed an index of parent characteristics by using fitted values from a regression of test scores 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's household income. A second index is constructed in the same way, using fitted values from a regression of college attendance on parent characteristics. Repeating the analysis above with these two measures of parent characteristics added to the baseline control vector gives the following predictive effects for college attendance based on test scores

$$\hat{\gamma}_{G,ts}^{co} = 0.13 \quad (0.055), \quad \hat{\gamma}_{F,ts}^{co} = 0.87 \quad (0.10),$$

which are somewhat lower than the results above using the baseline controls. The predictive effects for earnings are

$$\hat{\gamma}_{G,ts}^{ea} = 196 \quad (95), \quad \hat{\gamma}_{F,ts}^{ea} = 282 \quad (75).$$

Compared with the results using the baseline controls, the teacher effect of \$196 is about the same (before: \$186), but the school effect of \$282 is substantially lower (before: \$400).

With the parent characteristics added to the baseline control vector, the predictive effects for college attendance based on the college attendance of other classes are

$$\hat{\gamma}_G^{co} = 0.79 \quad (0.23), \quad \hat{\gamma}_F^{co} = 2.70 \quad (0.08),$$

and $\hat{R}_{co}^2 = 0.080 \quad (0.005)$. The teacher effect is reduced from 0.99 to 0.79 percentage points. There are substantial reductions in the school effect and in \hat{R}_{co}^2 . The predictive effects for test scores based on the test scores of other classes are

$$\hat{\gamma}_G^{ts} = 0.087 \quad (0.002), \quad \hat{\gamma}_F^{ts} = 0.052 \quad (0.002),$$

and $\hat{R}_{ts}^2 = 0.261 \quad (0.006)$. Here the results are not affected by adding the parent characteristics.

I have repeated the analysis without using the quadratic terms, so that the linear predictor for $U_{iA,co}$ conditions on G_{i1} and F_{i1} , dropping G_{i2} and F_{i2} . With the baseline controls in X , this gives

$$\hat{\gamma}_{G,ts}^{co} = 0.16 \quad (0.060), \quad \hat{\gamma}_{F,ts}^{co} = 0.74 \quad (0.13).$$

Therefore, the teacher effect is still 0.16 percentage points. (The school effect is lower: 0.74 vs. 1.19 percentage points.)

Now consider a partition in Eq. 1' just by subject (math and reading) instead of by subject and grade. There are $L=2$ cells with weights W_i^l as in Eq. 5. With the baseline controls in X and without using the quadratic terms, this gives

$$\hat{\gamma}_{G,ts}^{co} = 0.30 \quad (0.056), \quad \hat{\gamma}_{F,ts}^{co} = 0.44 \quad (0.19).$$

This partition gives a substantially higher teacher effect: 0.30 vs. 0.16 percentage points (and a lower school effect). I prefer the estimates that partition on subject and grade.

Finally, consider predictive effects in the factor model that do not partial on the school. Therefore, in predicting college attendance

$$E^*(U_{iA,co} | 1, F_{i1} + G_{i1}) = \gamma_0 + \gamma_1(F_{i1} + G_{i1}), \quad \gamma_{F+G,ts}^{co} = [\text{Var}(\gamma_1(F_{i1} + G_{i1}))]^{1/2}.$$

With the baseline controls in X , without the quadratic terms, with the partition on subject and grade, this gives

$$\hat{\gamma}_{F+G,ts}^{co} = 0.51 \quad (0.083).$$

The predictive effect on college attendance of 0.51 percentage points is considerably larger than the effect based on within school variation: $\hat{\gamma}_{G,ts}^{co} = 0.16$ percentage points. I prefer the estimate of 0.16 percentage points.

Conclusion

With the baseline controls, using the factor model, an SD increase in the teacher factor based on test scores has a predictive effect on college attendance of 0.16 percentage points. With parent characteristics added to the baseline controls, the predictive effect is 0.13 percentage points. These estimates are lower bounds on the predictive effect of an SD increase in the teacher factor (G_{co}) based directly on college attendance. With the baseline controls, the predictive effect for G_{co} on college attendance is 0.99 percentage points. The R^2 estimate is 0.13, whereas basing the predictions just on test scores gives an R^2 estimate of 0.01. The teacher effect of 0.99 percentage points could reflect skills that are relevant for college attendance but are not measured by the test scores. These skills could be some combination of skills students bring to the class (not captured in X) and skills developed during the class, in part due to the contribution of the teacher. With the parent characteristics added to the baseline controls, the predictive effect is 0.79 percentage points.

The factor model provides a predictive effect for individual test scores of a 1-SD increase in the teacher factor (G_1) based directly on test scores. This effect is 0.087, where the units are SDs in the distribution of scores for individual students. This result is not affected by adding the parent characteristics to the baseline controls. The result is consistent with the related literature (1–10), where a typical finding is that a 1-SD increase in the teacher factor corresponds to an increase in individual scores on the order of 0.1 SDs (in the distribution of scores for individual students).

Computation

Suppose that Eq.1 has the following form:

$$\theta = \arg \min_{d \in \mathcal{R}^l} E \left[W_i [r_1(U_{iA}) - r_2(U_{iB}, U_{iC})'d]^2 \right],$$

where r_1 and r_2 are given functions. For example, $r_1(U_{iA}) = U_{iA,co}$ and $r_2(U_{iB}, U_{iC})'d$ is a quadratic polynomial. Then θ satisfies the linear equation

$$E \left[W_i r_2(U_{iB}, U_{iC}) r_2(U_{iB}, U_{iC})' \right] \theta = E \left[W_i r_2(U_{iB}, U_{iC}) r_1(U_{iA}) \right]. \quad [9]$$

Now suppose that the components of r_2 have the form

$$r_{2p}(U_{iB}, U_{iC}) = r_{2p1}(U_{iB}) \cdot r_{2p2}(U_{iC}) \quad (p = 1, \dots, P). \quad [10]$$

This form holds if $r_2(U_{iB}, U_{iC})'d$ is a polynomial. In this case, the expectations in Eq. 9 require evaluating terms of the form

$$E(W_i V_{1iA} V_{2iB} V_{3iC}), \quad [11]$$

where $V_{1iA} = q_1(U_{iA})$, $V_{2iB} = q_2(U_{iB})$, $V_{3iC} = q_3(U_{iC})$, and the q_s are given functions. The sample analog for a term of this form is

$$\frac{1}{I} \sum_{i=1}^I W_i \left(\sum_{t: |S_{it}| > 1} |S_{it}| \right)^{-1} \times \sum_{t: |S_{it}| > 1} \sum_{a \in S_{it}} \sum_{b \in S_{it} - \{a\}} \sum_{c \in S_{it} - S_{it}} \hat{V}_{1ia} \hat{V}_{2ib} \hat{V}_{3ic} / [(|S_{it}| - 1)(|S_{it}| - |S_{it}|)]$$

[with, for example, $\hat{V}_{1ia} = q_1(\hat{U}_{ia})$]. The triple sum over (a, b, c) can be simplified as

$$\begin{aligned} & \sum_{a \in S_{it}} \sum_{b \in S_{it} - \{a\}} \sum_{c \in S_{it} - S_{it}} \hat{V}_{1ia} \hat{V}_{2ib} \hat{V}_{3ic} \\ &= \sum_{a \in S_{it}} \hat{V}_{1ia} \left[\left(\sum_{a \in S_{it}} \hat{V}_{2ia} \right) - \hat{V}_{2ia} \right] \cdot \left[\sum_{a \in S_{it}} \hat{V}_{3ia} - \sum_{a \in S_{it}} \hat{V}_{3ia} \right] \\ &= \left[\sum_{a \in S_{it}} \hat{V}_{1ia} \sum_{a \in S_{it}} \hat{V}_{2ia} - \sum_{a \in S_{it}} \hat{V}_{1ia} \hat{V}_{2ia} \right] \cdot \left[\sum_{a \in S_{it}} \hat{V}_{3ia} - \sum_{a \in S_{it}} \hat{V}_{3ia} \right]. \end{aligned}$$

ACKNOWLEDGMENTS. I thank Alberto Abadie, Moshe Buchinsky, Raj Chetty, John Friedman, Bryan Graham, Guido Imbens, Maximilian Kasy, Michael Rothschild, and Jesse Rothstein for comments and discussions. I thank Raj Chetty for assistance in implementing the methods developed here to estimate teacher effects on the administrative data constructed at the Internal Revenue Service.

- Rivkin S, Hanushek E, Kain J (2005) Teachers, schools, and academic achievement. *Econometrica* 73(2):417–458.
- McCaffrey DF, Lockwood JR, Koretz D, Louis TA, Hamilton L (2004) Models for value-added modeling of teacher effects. *J Educ Behav Stat* 29(1):67–101.
- Nye B, Konstantopoulos S, Hedges L (2004) How large are teacher effects? *Educ Eval Policy Anal* 26(3):237–257.
- Rockoff J (2004) The impact of individual teachers on student achievement: Evidence from panel data. *Am Econ Rev* 94(2):247–252.
- Aaronson D, Barrow L, Sander W (2007) Teachers and student achievement in the Chicago public high schools. *J Labor Econ* 25(1):95–135.
- Kane T, Staiger D (2008) Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper* (National Bureau of Economic Research, Cambridge, MA).
- Kane T, Rockoff J, Staiger D (2008) What does certification tell us about teacher effectiveness? Evidence from New York City. *Econ Educ Rev* 27(6):615–631.
- Jacob B, Lefgren L, Sims D (2010) The persistence of teacher-induced learning. *J Hum Resour* 45(4):915–943.
- Hanushek E, Rivkin S (2010) Generalizations about using value-added measures of teacher quality. *Am Econ Rev* 100(2):267–271.
- Staiger D, Rockoff J (2010) Searching for effective teachers with imperfect information. *J Econ Perspect* 24(3):97–118.
- Krueger A (1999) Experimental estimates of education production functions. *Q J Econ* 114(2):497–532.
- Graham B (2008) Identifying social interactions through conditional variance restrictions. *Econometrica* 76(3):643–660.
- Chetty R, et al. (2011) How does your kindergarten classroom affect your earnings? Evidence from Project Star. *Q J Econ* 126(4):1593–1660.
- Billingsley P (1995) *Probability and Measure* (Wiley, New York).
- Chetty R, Friedman J, Rockoff J (2011) The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *NBER Working Paper* (National Bureau of Economic Research, Cambridge, MA).
- Rothstein J (2010) Teacher quality in educational production: Tracking, decay, and student achievement. *Q J Econ* 125(1):175–214.