July 2012
Revised February 18, 2013

PREDICTIVE EFFECTS OF TEACHERS AND SCHOOLS ON TEST SCORES, COLLEGE ATTENDANCE, AND EARNINGS

Gary Chamberlain*
Harvard University

ABSTRACT

I study 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. The predictive effects have the following form: 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 which, under certain assumptions, makes this feasible. Administrative school district data in combination with tax data were used to calculate estimates and do inference.

*I am grateful to Alberto Abadie, Moshe Buchinsky, Raj Chetty, John Friedman, 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. See the SOI White Paper entitled "New Evidence on the Long-Term Impacts of Tax Credits" for the statistics underlying the estimates reported here.

PREDICTIVE EFFECTS OF TEACHERS AND SCHOOLS ON TEST SCORES, COLLEGE ATTENDANCE, AND EARNINGS

1. INTRODUCTION

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. My predictive effects have the following form: 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 which, under certain assumptions, makes this feasible.

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 and schools 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.

2. 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 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:

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

where g is a given function. For example,

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

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

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

An alternative could use the absolute value of the error instead of the squared error in (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 (1). To make this explicit, let $\mathcal{N} = \{1, 2, \ldots\}$ 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 (1), and simplify notation by dropping the i subscript:

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

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 i.i.d. 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 $(W, U, S) = (w, u, s)$:

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

with $m(A, B, C) = w g(u_A, u_B, u_C, d)$.

$$\begin{aligned} &E[m(A, B, C) \,|\, q(A) = t, (W, U, S) = (w, u, s)] \\ &\quad = \frac{1}{|s_t|} \sum_{a \in s_t} \Bigg[\sum_{b \in s_t - \{a\}} \sum_{c \in s - s_t} m(a, b, c) / [(|s_t| - 1)(|s| - |s_t|)] \Bigg], \end{aligned}$$

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:

$$
\begin{aligned}
&E[m(A,B,C)\,|\,(W,U,S)=(w,u,s)] \\
&\quad = \Big(\sum_{t:|s_t|>1} |s_t|\Big)^{-1} \sum_{t:|s_t|>1} |s_t| E[m(A,B,C)\,|\,q(A)=t,(W,U,S)=(w,u,s)] \\
&\quad = \Big(\sum_{t:|s_t|>1} |s_t|\Big)^{-1} \sum_{t:|s_t|>1} \sum_{a\in s_t} \sum_{b\in s_t-\{a\}} \sum_{c\in s-s_t} m(a,b,c)/[(|s_t|-1)(|s|-|s_t|)].
\end{aligned}
$$

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

$$
\alpha_h = \arg\min_{d\in\mathcal{R}^K} E\Big(\sum_{j\in S_i} (Y_{ij,h} - X'_{ij}d)^2\Big) \qquad (h=1,\dots,H). \tag{4}
$$

The sample analog for (4) is

$$
\hat{\alpha}_h = \arg\min_{d\in\mathcal{R}^K} \frac{1}{I}\sum_{i=1}^{I}\Big(\sum_{j\in S_i} (Y_{ij,h} - X'_{ij}d)^2\Big) \qquad (h=1,\dots,H), \tag{4$'$}
$$

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

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

The Appendix shows how the computation simplifies in a special case, which includes (2) and (3). For inference, I shall use bootstrap methods, based on treating the schools as a random sample from some unknown distribution. This 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 $l = 1, \ldots, L$ 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

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

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

$$\theta = \arg\min_{d \in \mathcal{R}^J} E[\sum_{l=1}^{L} W_i^l g(U_{iA^l}, U_{iB^l}, U_{iC^l}, d)], \tag{1'}$$

2.1 *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. 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

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

where $h_n(\,\cdot\,)$ is a given function, such as $h_n(U_{iA}) = U_{iA,ts}^n$. Assume that

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

Assume that U_{iA} and U_{iB} are independent conditional on the latent variables $Z_{iA,T}$, $Z_{i,S}$. Then

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

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

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

Note that

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

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

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

So we can obtain the moments $\mathrm{Cov}(F_{in}, F_{ip})$ and $\mathrm{Cov}(G_{in}, G_{ip})$ from $\mathrm{Cov}(h_n(U_{iA}), h_p(U_{iB}))$ and $\mathrm{Cov}(h_n(U_{iA}), h_p(U_{iC}))$

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

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

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

2.2 *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 $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \mathcal{Z}_3$ denote the domain of the input arguments, so we can consider counterfactual outcomes $g(z)$ for any point $z = (z_1, z_2, z_3) \in \mathcal{Z}$. At any such point, $g(z)$ is a random variable with

$$Eg(z_1, z_2, z_3) = \bar{g}(z_1, z_2, z_3).$$

Let $Z = (Z_{CL}, Z_T, Z_S)$. If Z is independent of $\{g(z)\}_{z \in \mathcal{Z}}$, then $Eg(Z) = E\bar{g}(Z)$. Define

$$\bar{g}_1(z_2, z_3) = E\bar{g}(Z_{CL}, z_2, z_3) \quad \text{for} \quad (z_2, z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3.$$

As above, define the factor $F_{co} + G_{co} = E(U_{co} \,|\, Z_T, Z_S)$. If Z_{CL} is independent of (Z_T, Z_S), then

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

providing a connection between this factor and the production function. Define

$$\bar{g}_{1,2}(z_3) = E\bar{g}(Z_{CL}, Z_T, z_3) \quad \text{for} \quad z_3 \in \mathcal{Z}_3.$$

As above, define the factor $F_{co} = E(U_{co} \,|\, Z_S)$. If (Z_{CL}, Z_T) is independent of Z_S, then

$$F_{co} = E(U_{co} \,|\, Z_S) = \bar{g}_{1,2}(Z_S),$$

providing a connection between this factor and the production function.

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

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

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 $\bar{g}_1(Z_T, Z_S)$:

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

This provides a connection between the production function for U_{co} and a linear predictor based on factors derived from 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 implies the following lower bound on the mean-square error for linear prediction of U_{co} from factors based on test scores:

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

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 $\text{MSE}^{(J)} \to \text{MSE}^*$ as $J \to \infty$.

3. EMPIRICAL RESULTS

The work of Chetty, Friedman, and Rockoff (2011) is pathbreaking in measuring 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 U.S. 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 year and grade to have mean 0 and standard deviation 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 reporting tuition payments and scholarships for every student.

Chetty, Friedman, and Rockoff conduct 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. $Y_{ij,ts}$ is the average test score for the class. $Y_{ij,co}$ is the per cent 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 (2) and (3). The partition in ($1'$) is by subject (math and reading) and grade (4, 5, 6, 7, 8), giving $L = 2 \times 5 = 10$ cells, with weights W_i^l as in (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

$\sum_{l=1}^{10} W_i^l > 0$. Consider the linear predictor for college attendance in (3):

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

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

$$\hat{\theta}_1 = 13.34\ (.37), \quad \hat{\theta}_2 = 2.26\ (.31), \quad \hat{\theta}_3 = 7.84\ (.31), \quad \hat{\theta}_4 = .64\ (.22).$$

I construct an index using a quadratic function of the test score from another class with the same teacher:

$$\text{Index}_{iB,ts}^{co} = \theta_1 U_{iB,ts} + \theta_2 U_{iB,ts}^2,$$

and use it to obtain a predictive effect in standard deviation units:

$$\theta_{B,ts}^{co} = (\text{Var}(\text{Index}_{iB,ts}^{co}))^{1/2}.$$

Likewise, the index for another class with a different teacher, same school, is

$$\text{Index}_{iC,ts}^{co} = \theta_3 U_{iC,ts} + \theta_4 U_{iC,ts}^2,$$

with predictive effect in standard deviation units

$$\theta_{C,ts}^{co} = (\text{Var}(\text{Index}_{iC,ts}^{co}))^{1/2}.$$

The estimates of these predictive effects are

$$\hat{\theta}_{B,ts}^{co} = 9.00\ (.29), \quad \hat{\theta}_{C,ts}^{co} = 5.16\ (.22).$$

The predictive effect for same teacher is larger than that for same school: a 9.0 percentage point increase in college attendance versus 5.2 percentage points.

The coefficients θ are defined as solutions to the minimization problem in (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^2_{co} 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 standard error) is $\hat{R}^2_{co} = .30$ (.015).

Now let X_{ij} be the baseline control vector used by Chetty, Friedman, and Rockoff. 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, gender, age, lagged suspensions, lagged absences, and indicators for grade repetition, special education, and limited English. This gives

$$\hat{\theta}_1 = 1.28 \ (.20), \quad \hat{\theta}_2 = -2.42 \ (.51), \quad \hat{\theta}_3 = .92 \ (.16), \quad \hat{\theta}_4 = -2.42 \ (.40),$$

with predictive effects in standard deviation units:

$$\hat{\theta}^{co}_{B,ts} = .31 \ (.04), \quad \hat{\theta}^{co}_{C,ts} = .27 \ (.03),$$

and $\hat{R}^2_{co} = .002$ (.0004).

The controls matter a lot. This relates to the difficulty in attaching causal interpretations to these predictive effects. This has been emphasized in Rothstein (2010). The issue has been addressed in Kane and Staiger (2008), using a dataset with random assignment of teachers to classrooms, and in Chetty, Friedman, and Rockoff (2011), who look at effects based on changes in teaching staff. The predictive effect on college attendance of a one standard deviation increase in the index for a different class with the same teacher is .31 percentage points. The predictive effect of a one standard deviation increase in the index for another class with a different teacher, same school, is .27 percentage points. So the observable control variables account for much of the predictive effects.

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. 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 is feasible under the assumptions of the factor model. For notation, let

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

where

$$\begin{aligned} F_{i1} + G_{i1} &= E(U_{iB,ts}\,|\,Z_{iA,T}, Z_{i,S}), \quad F_{i1} = E(U_{iC,ts}\,|\,Z_{i,S}) \\ F_{i2} + G_{i2} &= E(U^2_{iB,ts}\,|\,Z_{iA,T}, Z_{i,S}), \quad F_{i2} = E(U^2_{iC,ts}\,|\,Z_{i,S}), \end{aligned} \tag{7}$$

$Z_{iA,T}$ denotes characteristics of the teacher of classroom A, and $Z_{i,S}$ denotes characteristics of the school of classroom A. I construct an index using the (limiting) average over other classes with the same teacher of a quadratic function of the test score:

$$\text{Index}^{co}_{i,F+G|F,ts} = \gamma_1(F_{i1} + G_{i1}) + \gamma_2(F_{i2} + G_{i2}),$$

and use it to obtain a predictive effect in standard deviation units:

$$\gamma^{co}_{F+G|F,ts} = (\text{Var}(\text{Index}^{co}_{i,F+G|F,ts}))^{1/2}.$$

Likewise, using the (limiting) average over classes with a different teacher, same school:

$$\text{Index}^{co}_{i,F|F+G,ts} = \gamma_3 F_{i1} + \gamma_4 F_{i2},$$

with predictive effect in standard deviation units:

$$\gamma^{co}_{F|F+G,ts} = (\text{Var}(\text{Index}^{co}_{i,F|F+G,ts}))^{1/2}.$$

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

$$\hat{\gamma}_1 = 1.70\ (.72), \quad \hat{\gamma}_2 = 1.56\ (3.17), \quad \hat{\gamma}_3 = 9.98\ (2.54), \quad \hat{\gamma}_4 = -60.68\ (10.79),$$

with predictive effects

$$\hat{\gamma}^{co}_{F+G|F,ts} = .18\ (.067), \quad \hat{\gamma}^{co}_{F|F+G,ts} = 1.16\ (.16),$$

and $\hat{R}^2_{co} = .013$ (.003). Consider an average of the test score index across classes that share a teacher with class A. A standard deviation increase in this average corresponds to a predicted increase in college attendance for each student in class A of .18 percentage points. This is a partial predictive effect, holding constant the average of the score index across classes, in the same school, which do not share a teacher with class A. The (partial) predictive effect for the average over classes with a different teacher is a considerably larger increase of 1.16 percentage points.

An alternative measure of predictive effects can be based on an index formed from the differences within a school:

$$G_{i1} = E(U_{iB,ts} \,|\, Z_{iA,T}, Z_{i,S}) - E(U_{iC,ts} \,|\, Z_{i,S}), \tag{8}$$
$$G_{i2} = E(U^2_{iB,ts} \,|\, Z_{iA,T}, Z_{i,S}) - E(U^2_{iC,ts} \,|\, Z_{i,S}).$$

The alternative indices are

$$\mathrm{Alt\_Index}^{\mathrm{co}}_{\mathrm{iG,ts}} = \gamma_1 \mathrm{G}_{\mathrm{i1}} + \gamma_2 \mathrm{G}_{\mathrm{i2}}, \quad \mathrm{Alt\_Index}^{\mathrm{co}}_{\mathrm{iF,ts}} = (\gamma_1 + \gamma_3)\mathrm{F}_{\mathrm{i1}} + (\gamma_2 + \gamma_4)\mathrm{F}_{\mathrm{i2}},$$

with predictive effects in standard deviation units:

$$\gamma^{co}_{G,ts} = (\mathrm{Var}(\mathrm{Alt\_Index}^{\mathrm{co}}_{\mathrm{iG,ts}}))^{1/2}, \quad \gamma^{\mathrm{co}}_{\mathrm{F,ts}} = (\mathrm{Var}(\mathrm{Alt\_Index}^{\mathrm{co}}_{\mathrm{iF,ts}}))^{1/2}.$$

The estimates of these alternative predictive effects are

$$\hat{\gamma}^{co}_{G,ts} = .16\ (.059), \quad \hat{\gamma}^{co}_{F,ts} = 1.19\ (.14),$$

quite close to the estimates using the original definitions. This closeness is because most of the test score variation is within schools (as we shall see below), and because $(\hat{\gamma}_1 + \hat{\gamma}_3)$ and $(\hat{\gamma}_2 + \hat{\gamma}_4)$ are relatively close to $\hat{\gamma}_3$ and $\hat{\gamma}_4$.

So far we have used a (quadratic) function of the test score in predicting college attendance. We can also use college attendance 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

$$F_{i3} + G_{i3} = E(U_{iB,co} \,|\, Z_{iA,T}, Z_{i,S}), \quad F_{i3} = E(U_{iC,co} \,|\, Z_{i,S}).$$

Then $F_{i3} + G_{i3}$ corresponds to an average of $U_{iB,co}$ over many classrooms other than A that share a teacher with A, and F_{i3} 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

$$E^*(U_{iA,co} \,|\, 1, \{F_{in}, G_{in}\}_{n=1}^3) = F_{i3} + G_{i3}.$$

The predictive effects in standard deviation units are

$$\gamma_{G,co}^{co} = (\mathrm{Var}(G_{i3}))^{1/2}, \quad \gamma_{F,co}^{co} = (\mathrm{Var}(F_{i3}))^{1/2}$$

(with $\gamma_{F+G,co}^{co} = [(\gamma_{G,co}^{co})^2 + (\gamma_{F,co}^{co})^2]^{1/2}$).

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

$$\hat{\gamma}_{G,co}^{co} = .99 \ (.22), \quad \hat{\gamma}_{F,co}^{co} = 3.71 \ (.11),$$

and $\hat{R}^2_{co} = .134 \ (.007)$ (with $\hat{\gamma}_{F+G,co}^{co} = 3.84 \ (.12)$). If we could average the college attendance over many classes that share a teacher with class A, then a one standard deviation increase would correspond to a very substantial increase of 3.84 percentage points for college attendance of class A. If we base the predictive effect on the difference within a school:

$$G_{i3} = E(U_{iB,co} \,|\, Z_{iA,T}, Z_{i,S}) - E(U_{iC,co} \,|\, Z_{i,S}),$$

then a standard deviation increase in G_{i3} corresponds to an increase of .99 percentage points for college attendance of class A. It is clear that basing the predictions for college attendance just on the test scores looses a great deal of information.

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

$$E^*(U_{iA,ts} \,|\, 1, \{F_{in}, G_{in}\}_{n=1}^3) = F_{i1} + G_{i1}.$$

The predictive effects are

$$\gamma^{ts}_{G,ts} = (\mathrm{Var}(G_{i1}))^{1/2}, \quad \gamma^{ts}_{F,ts} = (\mathrm{Var}(F_{i1}))^{1/2}.$$

With the baseline controls in X, the estimates are

$$\hat{\gamma}^{ts}_{G,ts} = .087\ (.002), \quad \gamma^{ts}_{F,ts} = .052\ (.002),$$

and $\hat{R}^2_{ts} = .260\ (.006)$ (with $\hat{\gamma}^{ts}_{F+G,ts} = .101\ (.002)$). Consider an average of the test score index across other classes that share a teacher with class A. A standard deviation increase in this average corresponds to a predicted increase in score for each student in class A of .101, where the units are standard deviations in the distribution of scores for individual students. If we base the predictive effect on the difference within a school:

$$G_{i1} = E(U_{iB,ts} \,|\, Z_{iA,T}, Z_{i,S}) - E(U_{iC,ts} \,|\, Z_{i,S}),$$

then a standard deviation increase in this difference corresponds to a predicted increase in score for each student in class A of .087.

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

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

With the baseline controls in X, the estimates are

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

with predictive effects (in standard deviation units):

$$\hat{\theta}^{ea}_{B,ts} = 149\ (54), \quad \hat{\theta}^{ea}_{C,ts} = 118\ (32),$$

and $\hat{R}^2_{ea} = .002$ (.001). These results are based on fewer classrooms, 14236 instead of 118439, because only some of the students have reached the age of 28 by 2010. There are 524 schools, of which 364 satisfy the condition that $\sum_{l=1}^{10} W_i^l > 0$. The results are less precise, but the point estimates give a predictive effect of \$149 for a standard deviation increase in the test score index for a different class with the same teacher. The predictive effect is \$118 for a standard deviation increase in the test score index for another class with a different teacher, same school.

For notation in the factor model, let

$$E^*(U_{iA,ea} \,|\, 1, F_{i1}, G_{i1}, F_{i2}, G_{i2}) = \gamma_0 + \gamma_1(F_{i1} + G_{i1}) + \gamma_2(F_{i2} + G_{i2}) + \gamma_3 F_{i1} + \gamma_4 F_{i2}, \qquad (9)$$

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

$$\hat{\gamma}_1 = 586 \ (1277), \quad \hat{\gamma}_2 = 4424 \ (5885), \quad \hat{\gamma}_3 = 2457 \ (2242), \quad \hat{\gamma}_4 = -16027 \ (7961),$$

with predictive effects

$$\hat{\gamma}^{ea}_{F+G|F,ts} = 218 \ (165), \quad \hat{\gamma}^{ea}_{F|F+G,ts} = 473 \ (189),$$

and $\hat{R}^2_{ea} = .009$ (.004). There is a predictive effect of \$218 for a standard deviation increase in the average of the test score index over different classes with the same teacher. The predictive effect is \$473 for a standard deviation increase in the average of the test score index over classes with different teachers, same school. The alternative predictive effects, based on differencing within a school as in (8), are

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

Chetty, Friedman, and Rockoff link students to their parents by finding the earliest 1040 form from 1996–2010 on which the student was claimed as a dependent. They construct 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:

$$\hat{\gamma}^{co}_{F+G|F,ts} = .14\ (.063), \quad \hat{\gamma}^{co}_{F|F+G,ts} = .83\ (.13), \quad \hat{\gamma}^{co}_{G,ts} = .13\ (.055), \quad \hat{\gamma}^{co}_{F,ts} = .87\ (.10),$$

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

$$\hat{\gamma}^{ea}_{F+G|F,ts} = 224\ (127), \quad \hat{\gamma}^{ea}_{F|F+G,ts} = 317\ (153), \quad \hat{\gamma}^{ea}_{G,ts} = 196\ (95), \quad \hat{\gamma}^{ea}_{F,ts} = 282\ (75).$$

Compared with the results using the baseline controls, the teacher components of \$224 and \$196 are about the same (before: \$218 and \$186), but the school components of \$317 and \$282 are substantially lower (before: \$473 and \$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

$$\hat{\gamma}^{co}_{G,co} = .79\ (.23), \quad \hat{\gamma}^{co}_{F,co} = 2.70\ (.08),$$

and $\hat{R}^2_{co} = .080\ (.005)$ (with $\hat{\gamma}^{co}_{F+G,co} = 2.81\ (.10)$). There are substantial reductions in the predictive effects and in $\hat{R}^2_{co}$. The predictive effects for test scores based on the test scores of other classes are

$$\hat{\gamma}^{ts}_{G,ts} = .087\ (.002), \quad \hat{\gamma}^{ts}_{F,ts} = .052\ (.002),$$

and $\hat{R}^2_{ts} = .261\ (.006)$. Here the results are not affected by adding the parent characteristics.

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

$$\hat{\gamma}^{co}_{F+G|F,ts} = .18\ (.069), \quad \hat{\gamma}^{co}_{F|F+G,ts} = .65\ (.14), \quad \hat{\gamma}^{ea}_{F+G|F,ts} = 145\ (98), \quad \hat{\gamma}^{ea}_{F|F+G,ts} = 193\ (125).$$

The school components are lower: .65 versus 1.16 percentage points and \$193 versus \$473.

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

$$\hat{\gamma}^{co}_{F+G|F,ts} = .32\ (.058), \quad \hat{\gamma}^{co}_{F|F+G,ts} = .34\ (.18), \quad \hat{\gamma}^{ea}_{F+G|F,ts} = 251\ (89), \quad \hat{\gamma}^{ea}_{F|F+G,ts} = 145\ (126).$$

This gives substantially higher teacher components and lower school components, both in predicting college attendance and earnings.

There are corresponding results for the linear predictors of $U_{iA,co}$ and $U_{iA,ea}$ conditional on $U_{iB,ts}$ and $U_{iC,ts}$, dropping the quadratic terms $U^2_{iB,ts}$ and $U^2_{iC,ts}$. With the baseline controls in X, with the partition on subject and grade, this gives

$$\begin{aligned} &\hat{\theta}^{co}_{B,ts} = .25\ (.041), \quad \hat{\theta}^{co}_{C,ts} = .18\ (.033), \quad \hat{\theta}^{co}_{B,ts} - \hat{\theta}^{co}_{C,ts} = .075\ (.029), \\ &\hat{\theta}^{ea}_{B,ts} = 148\ (57), \quad \hat{\theta}^{ea}_{C,ts} = 68\ (36), \quad \hat{\theta}^{ea}_{B,ts} - \hat{\theta}^{ea}_{C,ts} = 80\ (57). \end{aligned}$$

With the partition just on subject,

$$\begin{aligned} &\hat{\theta}^{co}_{B,ts} = .20\ (.038), \quad \hat{\theta}^{co}_{C,ts} = .062\ (.027), \quad \hat{\theta}^{co}_{B,ts} - \hat{\theta}^{co}_{C,ts} = .14\ (.027), \\ &\hat{\theta}^{ea}_{B,ts} = 121\ (53), \quad \hat{\theta}^{ea}_{C,ts} = 26\ (22), \quad \hat{\theta}^{ea}_{B,ts} - \hat{\theta}^{ea}_{C,ts} = 95\ (64). \end{aligned}$$

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

$$E^*(U_{iA,co} \,|\, 1, F_{i1} + G_{i1}) = \gamma_0 + \gamma_1(F_{i1} + G_{i1}), \quad \gamma^{co}_{F+G,ts} = \gamma_1(\mathrm{Var}(F_{i1} + G_{i1}))^{1/2},$$

with a similar definition for $\gamma^{ea}_{F+G,ts}$. With the baseline controls in X, without the quadratic terms, with the partition on subject and grade, this gives

$$\hat{\gamma}^{co}_{F+G,ts} = .51\ (.083), \quad \hat{\gamma}^{ea}_{F+G,ts} = 254\ (95).$$

The predictive effect on college attendance is considerably larger than the partial effect that controls for school: .51 versus $\hat{\gamma}^{co}_{F+G|F,ts} = .18$ percentage points. The predictive effect on earnings is also larger than the partial effect: \$254 versus $\hat{\gamma}^{ea}_{F+G|F,ts} = \$145$. However, if the partition is just on subject, then $\hat{\gamma}^{co}_{F+G,ts} = .42$ (.080), $\hat{\gamma}^{ea}_{F+G,ts} = 203$ (91). So here the predictive effect for earnings is smaller than the partial predictive effect: \$203 versus $\hat{\gamma}^{ea}_{F+G|F,ts} = \$251$ (89).

4. DISCUSSION

The predictive effects are substantially reduced when the baseline controls are used in X to form the residuals (compared with using $X = 1$). With the baseline controls, the predictive effect for college attendance of a standard deviation increase in the test score index for a different class with the same teacher is .31 percentage points; for a class with a different teacher, same school, it is .27 percentage points. Under the factor model, the comparable results using (limiting) averages of the test score index over different classes with the same teacher and over classes with different teachers, same school, are .18 and 1.16 percentage points. With the parent characteristics added to the baseline controls, the predictive effects are .14 and .83 percentage points.

I have also constructed predictive effects based on the college attendance of other classes in the same school, with and without the same teacher. The factor model provides a predictive effect for college attendance of a one standard deviation increase in the (limiting) average of college attendance for other classes with the same teacher. This gives 3.84 percentage points. A one standard deviation increase in the average of college attendance for classes with a different teacher, same school, is 3.71 percentage points. If we difference these averages within schools, a standard deviation increase in the difference gives .99 percentage points. The R^2 estimate is .13 whereas basing the predictions just on test scores gives R^2 estimates of .01. The teacher effect of .99 percentage points could reflect skills that are relevant for college attendance but are not measured by the test scores. These 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 corresponding results are 2.81, 2.70, and .79 percentage points, with an R^2 estimate of .08. So including parent characteristics gives a substantial reduction in predictive effects based on the college attendance of other classes.

The factor model provides a predictive effect for individual test scores of a one standard deviation increase in the average scores for other classes with the same teacher. This is .101, where the units are standard deviations in the distribution of scores for individual students. A one standard deviation increase in the average score for classes with a different teacher, same school, gives an increase of .052 in individual scores. If we difference these averages within schools, a standard deviation increase in the difference gives a predicted increase in individual scores of .087. These results are not affected by adding the parent characteristics to the baseline controls.

The predictive effect for earnings of a standard deviation increase in the test score index for a different class with the same teacher is $149; for a class with a different teacher, same school, it is $118. Under the factor model, the comparable results using averages of the test score index over different classes with the same teacher and over classes with different teachers, same school, are $218 and $473. The standard errors here are substantial: $165 and $189.

Much of the related literature uses factor models based on test scores to estimate teacher effects. Examples include McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004), Nye, Konstantopoulos, and Hedges (2004), Rockoff (2004), Rivkin, Hanushek, and Kain (2005), Aaronson, Barrow, and Sander (2007), Kane and Staiger (2008), Kane, Rockoff, and Staiger (2008), Jacob, Lefgren, and Sims (2010), Hanushek and Rivken (2010), and Staiger and Rockoff (2010). A typical finding is that a one standard deviation increase in the teacher factor corresponds to an increase in individual scores on the order of .1, where the units are standard deviations in the distribution of scores for individual students. My results agree with that.

Chetty, Friedman, and Rockoff (2011) provide estimates of teacher effects on college attendance and earnings. In their Table 5 with percent attending college at age 20, the coefficient (standard error) on teacher value added is 4.92 (0.65). The standard deviation of teacher value added is about .1, giving a teacher effect of .492 (.065) percentage points. In Appendix Table 7, allowing for school-year fixed effects gives .26 (.05) percentage points. The comparable results for me are the ones that partition just by subject (math and reading) instead of by subject and grade. This gives

$\hat{\gamma}^{co}_{F+G,ts} = .42$ (.080) percentage points when I do not partial on school, and $\hat{\gamma}^{co}_{F+G|F,ts} = .32$ (.058) percentage points when I do. In their Table 6 with earnings at age 28, the coefficient (standard error) on teacher value added is $1,815$ (729), giving a teacher effect of \$182 (73). In Appendix Table 7, allowing for school-year fixed effects gives a teacher effect of \$194 (67). I obtain $\hat{\gamma}^{ea}_{F+G,ts}$ = \$203 (91) when I do not partial on school and $\hat{\gamma}^{ea}_{F+G|F,ts}$ = \$251 (89) when I do.

APPENDIX

Suppose that (1) has the following form:

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

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

$$E[W_i r_2(U_{iB}, U_{iC}) r_2(U_{iB}, U_{iC})']\theta = E[W_i r_2(U_{iB}, U_{iC}) r_1(U_{iA})]. \tag{10}$$

Now suppose that the components of r_2 have the form

$$r_{2p}(U_{iB}, U_{iC}) = r_{2p1}(U_{iB}) \cdot r_{2p2}(U_{iC}) \qquad (p = 1, \ldots, P). \tag{11}$$

This holds if $r_2(U_{iB}, U_{iC})'d$ is a polynomial. In this case, the expectations in (10) require evaluating terms of the form

$$E(W_i V_{1iA} V_{2iB} V_{3iC}), \tag{12}$$

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

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

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

$$\sum_{a \in S_{it}} \sum_{b \in S_{it}-\{a\}} \sum_{c \in S_i - S_{it}} \hat{V}_{1ia}\hat{V}_{2ib}\hat{V}_{3ic}$$
$$= \sum_{a \in S_{it}} \hat{V}_{1ia}\Bigg[\Bigg(\sum_{a \in S_{it}} \hat{V}_{2ia}\Bigg) - \hat{V}_{2ia}\Bigg] \cdot \Bigg[\sum_{a \in S_i} \hat{V}_{3ia} - \sum_{a \in S_{it}} \hat{V}_{3ia}\Bigg]$$
$$= \Bigg[\sum_{a \in S_{it}} \hat{V}_{1ia} \sum_{a \in S_{it}} \hat{V}_{2ia} - \sum_{a \in S_{it}} \hat{V}_{1ia}\hat{V}_{2ia}\Bigg] \cdot \Bigg[\sum_{a \in S_i} \hat{V}_{3ia} - \sum_{a \in S_{it}} \hat{V}_{3ia}\Bigg].$$

REFERENCES

Aaronson, D., L. Barrow, and W. Sander (2007): "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 25, 95–135.

Chetty, R., J. Friedman, and J. Rockoff (2011): "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," National Bureau of Economic Research Working Paper No. 17699.

Hanushek, E. and S. Rivkin (2010): "Generalizations about Using Value-Added Measures of Teacher Quality," *American Economic Review: Papers & Proceedings*, 100, 267–271.

Jacob, B., L. Lefgren, and D. Sims (2010): "The Persistence of Teacher-Induced Learning," *The Journal of Human Resources*, 45, 915–943.

Kane, T. and D. Staiger (2008): "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," National Bureau of Economic Research Working Paper No. 14607.

Kane, T., J. Rockoff, and D. Staiger (2008): "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review*, 27, 615–631.

McCaffrey, D., J. Lockwood, D. Koretz, T. Louis, and L. Hamilton (2004): "Models for Value-Added Modeling of Teacher Effects," *Journal of Educational and Behavioral Statistics*, 29, 67–101.

Nye, B., S. Konstantopoulos, and L. Hedges (2004): "How Large are Teacher Effects?", *Educational Evaluation and Policy Analysis*, 26, 237–257.

Rivkin, S., E. Hanushek, and J. Kain (2005): "Teachers, Schools, and Academic Achievement," *Econometrica*, 73, 417–458.

Rockoff, J. (2004): "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *The American Economic Review: Papers & Proceedings*, 94, 247–252.

Rothstein, J. (2010): "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, 125, 175–214.

Staiger, D. and J. Rockoff (2010): "Searching for Effective Teachers with Imperfect Information," *Journal of Economic Perspectives*, 24, 97–118.