Student-Teacher Racial Match & Learning

Lawrence Hotchkiss
James Raths
Joeseph Glutting

University of Delaware

Presentation to the Delaware Chapter of the American Statistical Association, Thursday, October 19, 2006.

Motivation:

> Recommendations for the aggressive recruitment of minority teachers are based on hypothesized role-model effects for minority students as well as evidence of racial biases among nonminority teachers. ⋯ Models of student achievement indicate that assignment to an own-race teacher significantly increased the math and reading achievement of both black and white students.*

*Thomas S. Dee. "Teachers, race, and student achievement in a randomized experiment." *The Review of Economics and Statistics*. Feburary, 2004. Pp. 195-210. (p. 195, Abstract)

Data – Project STAR class-size experiment:

- Authorized and funded by the Tennessee state legislature, begun in 1985-1986 school year
- Test of 3 class sizes/configurations: (1) Large – 22 students, (2) small – 15 students, (3) large (22) with teacher aide
- Random assignment of teachers and students to classes, *within* schools
- Reading and math test scores: scaled scores from the Stanford Achievement Test
- Four grades: K thru 3rd, same students followed – repeated measurements
- Profiles for students (e.g., gender, race, school lunch) & teachers (e.g., race, experience, degree, career track).
- School classification into inner city, suburban, urban, rural.

Data Profiles:

Demographics

	Female	Male	N
Student gender	47%	53%	11366

	Black	White	N
Student race	37%	63%	11366
Teacher race			
Kindergarten	17%	83%	6230
First grade	17%	83%	6698
Second grade	20%	80%	6646
Third grade	21%	79%	6699

Test Scores

	Mean	Std Dev	N
Kindergarten			
Reading	436.7	31.7	5761
Math	485.4	47.7	5843
First grade			
Reading	520.7	55.13	6351
Math	530.5	43.1	6555
Second grade			
Reading	583.9	45.9	6024
Math	580.6	44.5	6012
Third grade			
Reading	615.3	38.5	5959
Math	617.8	39.7	6037

School Type ×Student Race

	Student Race							
	Kindergarten		1st Grade		2nd Grade		3rd Grade	
School Type	Black	White	Black	White	Black	White	Black	White
Inner City	66%	1%	59%	1%	60%	1%	57%	1%
Urban	4	12	4	12	2	10	2	10
Suburban	21	23	27	21	29	22	32	22
Rural	8	64	10	66	8	67	9	67
Total N	2058	4234	2221	4528	2336	4370	2254	4496

Attrition

K	1st	2nd	3rd	N	Percent
Present	Present	Present	Present	3073	27.0368
Present	.	.	.	1650	14.5170
.	.	.	Present	1255	11.0417
.	Present	Present	Present	1154	10.1531
.	.	Present	Present	1087	9.5636
Present	Present	.	.	921	8.1031
.	Present	.	.	727	6.3963
.	.	Present	.	483	4.2495
Present	Present	Present	.	461	4.0560
.	Present	Present	.	336	2.9562
Present	.	Present	Present	74	0.6511
Present	Present	.	Present	45	0.3959
Present	.	Present	.	38	0.3343
.	Present	.	Present	32	0.2815
Present	.	.	Present	30	0.2639
Total				11366	

Teacher Race × Student Race

	Student Race							
	Kindergarten		1st Grade		2nd Grade		3rd Grade	
Teacher Race	Black	White	Black	White	Black	White	Black	White
Black	40%	5%	44%	4%	44%	8%	51%	6%
White	60	96	56	96	56	92	49	94
Total N	2051	4179	2210	4488	2276	4370	2205	4494

In spite of random assignment of teachers within schools, which Dee emphasiszes and checks, teacher race and student race were fairly strongly related due to race clustering by school. Small percentages of white children were taught by black teachers.

Statistical Model and Data Layout

Wide Format. Data are supplied in the usual layout – one observation per student and a different variable for each test (reading, math) and grade (K - 3) and other variables that differ by grade e.g.,

```
 ID    Race variables           Test Scores
---    --------------   ------------------------------------
    G     T  T  T  T
 S  e  S  R  R  R  R   R    M    R    M    R    M    R    M  ...
 t  n  R  a  a  a  a   e    a    e    a    e    a    e    a  ...
 u  d  a  c  c  c  c   a    t    a    t    a    t    a    t  ...
 I  e  c  e  e  e  e   d    h    d    h    d    h    d    h  ...
 D  r  e  K  1  2  3   K    K    1    1    2    2    3    3  ...
-------------------------------------------------------------------
001 2  1  1  1  1  2  450  500  614  553  602  599  626  629  ...
002 1  2  2  1  1  2  460  506  622  557  611  590  620  645  ...
```

Long Format. Dee changed the test scores to percentiles within grade and test and rearranged the data into long format –

```
 ID  Gender SRace  TRace  TestPct  Grade  TestType ...
--------------------------------------------------
001    2      1      1      71.2      K     reading
001    2      1      1      63.7      K     math
001    2      1      1      72.9      1     reading
001    2      1      1      70.0      1     math
001    2      1      1      64.9      2     reading
001    2      1      1      66.6      2     math
001    2      1      2      61.0      3     reading
001    2      1      2      61.0      3     math
002    1      2      2      80.5      K     reading
002    1      2      2      68.8      K     math
002    1      2      1      77.7      1     reading
002    1      2      1      72.6      1     math
002    1      2      1      70.9      2     reading
002    1      2      1      59.2      2     math
002    1      2      2      55.1      3     reading
002    1      2      2      74.7      3     math
```

With this organization of the data, Dee's statistical model is comparatively simple –

$$\begin{aligned} TestPct &= \alpha + \beta_s Gender + \beta_r SRace + \\ &\quad \beta_m SR + \beta_g Grade + \boldsymbol{\gamma x} + u \end{aligned}$$

where[a] –

$$\begin{aligned} \alpha,\ \beta_j\ , \boldsymbol{\gamma} &= \text{coefficients } (\boldsymbol{\gamma} \text{ is a row vector}) \\ SR &= \text{student \& teacher same race} \\ &\quad (0\ =\ \text{No},\quad 1\ =\ \text{Yes}) \\ \boldsymbol{x} &= \text{vector additional "control" variables, e.g.,} \\ &\quad \text{class-size assignment, teacher} \\ &\quad \text{characteristics, school-of-entry} \\ &\quad \text{fixed effects, free lunch, entry wave} \\ u &= \text{error or "disturbance"} \end{aligned}$$

[a]Most of Dee's regressions are reported for separate race × gender groups. Gender and student race variables are excluded from these models.

Dee reports several versions of this model –

- Varying list of regressors
- OLS & 2SLS with "intent" to assign a same-race teacher serving as an "instrument" for actual same-race teacher
- Separate models by race and gender of students (mostly)
- Huber-White robust standard errors
- Typical effects of a same-race teacher: 3 to 5 percentile points depending on the list of regressors, race and gender of the student, and method of estimation (OLS vs 2SLS)
- Statistical tests of the same-race effect generally are significant at $\alpha = 0.05$ to $\alpha = 0.01$.

HRG Model

- Retain original "wide" layout of data
- Two-equation first-order linear difference equation – presumes reading affects math, and math affects reading *

$$
\begin{aligned}
read_g &= \alpha_r + \beta_{rr} read_{g-1} + \beta_{rm} math_{g-1} + \beta_{rs} SGender \\
&\quad + \beta_{r\,(SR)} SRace \\
&\quad + \beta_{r(TR_g)} TRace_g + \beta_{r(TR_g \times SR)} TRace_g \times SRace \\
&\quad + \beta_{r(TR_{g-1})} TRace_{g-1} + \beta_{r(TR_{g-1} \times SR)} TRace_{g-1} \times SRace \\
&\quad + \boldsymbol{\gamma}_r \boldsymbol{x} + u_r \\
math_g &= \alpha_r + \beta_{mr} read_{g-1} + \beta_{mm} math_{g-1} + \beta_{ms} SGender \\
&\quad + \beta_{m\,(SR)} SRace \\
&\quad + \beta_{m(TR_g)} TRace_g + \beta_{m(TR_g \times SR)} TRace_g \times SRace \\
&\quad + \beta_{m(TR_{g-1})} TRace_{g-1} + \beta_{m(TR_{g-1} \times SR)} TRace_{g-1} \times SRace \\
&\quad + \boldsymbol{\gamma}_m \boldsymbol{x} + u_m
\end{aligned}
$$

---

*All race (student and teacher) variables are represented by a single dummy indicator: 1 =black $0 =$ white.

Expansion of Student-Teacher Race-Match Model

The basic idea of the race-match hypothesis is –

$$\text{test score} = b\,(\text{student race}) + c\,(\text{same race}) + \cdots$$

But this assumes the effect of a white teacher on white students is the same as the effect of a black teacher on black students. A full interaction model like –

$$\begin{aligned}\text{test score} &= b_1\,(\text{student race}) \\ &\quad + b_2\,(\text{teacher race}) \\ &\quad + b_3\,(\text{student race})\times(\text{teacher race}) + \cdots\end{aligned}$$

is consistent with the race-match hypothesis when –

$$\begin{aligned}0 &= 2b_2 + b_3 \\ 0 &= 2\text{ main effect} + \text{interaction effect}\end{aligned}$$

Otherwise, black students may benefit more from a black teacher than white students benefit from a white teacher, or vice versa. We test linear constraints like these to assess the perfect-match hypothesis.*

*When separate estimations are reported by race (& gender), as Dee mostly does, these tests are irrelevant.

Statistical estimation

- Combine race and gender groups – Chow tests not significant (mostly)
- Separate models for each grade interval: kindergarten – 1st grade, 1st grade – 2nd grade, and 2nd grade – 3rd grade.
- Separate models needed because
    (1) Sample attrition
    (2) Tests not scaled properly
    (3) Different effects at different times
- Heckman correction for attrition
- Correct standard errors for school-level clustering
- Autoregressive error is highly likely. It generally produces nonzero covariance between lagged endogenous variables and the disturbances. (More on this later.)

Results

Effect of Black Teacher by Student Race on Reading and Mathematics Learning

Panel 1: Reading Test

Student Race	Teacher Race Base Yr	Teacher Race Current Yr	Grade Interval K - 1	Grade Interval 1 - 2	Grade Interval 2 - 3
Black	Black	White	2.674	0.043	1.559
White	Black	White	-11.130a	-7.148	2.320
Black	White	Black	8.163	-5.100	-1.025
White	White	Black	-10.273	-5.100	-1.288
Black	Black	Black	10.837	-4.321	0.534
White	Black	Black	-21.403b	-5.058	1.032
Base-year match test:			-8.456	-7.062	1.559
Current-year match test:			-2.110	-2.316	-1.025
Combined-year match test:			-10.566	-9.379	0.534

a p=0.05, b p=0.01

Panel 2: Math Test

Student Race	Teacher Race Base Yr	Teacher Race Current Yr	---- Grade K - 1	Interval 1 - 2	---- 2 - 3
Black	Black	White	-2.803	-5.906	1.693
White	Black	White	-9.531	-0.075	1.379
Black	White	Black	13.773a	0.454	-3.503
White	White	Black	-13.006b	-4.919	-3.614
Black	Black	Black	10.970	-5.451	-1.810
White	Black	Black	-22.537b	-4.995	-2.235
Base-year match test:			-12.335	-5.981	3.072
Current-year match test:			0.767	-4.465	-7.118
Combined-year match test:			-11.567	-10.446	-4.045

a p=0.05, b p=0.01

In this model –

$$\begin{aligned}\text{test score} &= a + b_1\text{Student Race}\\ &\quad + b_2\text{Teacher Race}\\ &\quad + b_3\text{Teacher Race} \times \text{Student Race}\\ &\quad + \cdots\end{aligned}$$

Effect of a black teacher on a white student is the "main effect" of teacher race (b_2), and the effect of a black teacher on a black student is the sum of the "main effect" plus the "interaction effect" ($b_2 + b_3$).

The match tests are tests of the linear constraints $2b_2 + b_3 = 0$, which correspend to the "perfect match" model where the teacher race and interaction terms are replaced by a dummy indicator of same race (as above).

Summary:

- The estimated negative effects of a black teacher in both kindergarten and first grade on a white child are noteworthy on both reading (-21.403) and math (-22.537) learning. (These are the same as a positive effects of a white teacher in both grades.) However, the p-values are not small. So these estimates may not be very stable.
- The positive effects of a black teacher on black children during the K – 1 interval are smaller than the estimated effects of a white teacher on white children, but none of the perfect-match tests on the bottom margin of the two panels is significant.
- After the K – 1 interval, all the effect estimates are of negligible magnitude, and none is significant.

Conclusion: Evidence for recruiting minority teachers is weak at best.

Notes:

- Control for school type is important. Without it, black teachers do not have positive effects on black children, and some of the match tests are significant.
- Correction for clustering by school is important. Without it, standard errors are noticieably smaller, and more effect estimates are significant, but none after the K – 1 interval.
- Small number of white students with a black teacher may produce instability.
- Why do teacher-race effects decline so much from K – 1 to 2 – 3? This is hard to explain and seems a little suspicious.

Model Interpretation & Estimation

- The difference-equation model is equivalent to –

$$\begin{aligned}\Delta read_g &= \alpha_r + (\beta_{rr} - 1)\, read_{g-1} + \beta_{rm} math_{g-1} + \cdots \\ \Delta math_g &= \alpha_r + \beta_{mr} read_{g-1} + (\beta_{mm} - 1)\, math_{g-1} + \cdots\end{aligned}$$

- Change version is basis for interpretation as a model of learning
- Contrast to Dee model which we interpret as a model of current knowledge and distinguish it from a model of learning
- Reflects idea that what one knows facilitates further learning
- Matrix representation –

$$\begin{aligned}\boldsymbol{y}_{t+1} &= \mathbf{B} y_t + \boldsymbol{\Gamma x} + u_t \qquad \text{or} \\ \Delta \boldsymbol{y} &= (\mathbf{B} - \mathbf{I})\, y_t + \boldsymbol{\Gamma x} + u_t\end{aligned}$$

- Measurement interval exceeds reaction interval. Estimation therefore is based on accumulation of effects –

$$\boldsymbol{y}_{t+\Delta t} \;=\; \mathbf{B}^{*} y_t + \boldsymbol{\Gamma}^{*} \boldsymbol{x} + v_t$$

where the $\mathbf{B}^{*}$, $\boldsymbol{\Gamma}^{*}$ and v_t result from the solution to the difference equations –

$$\begin{aligned} \mathbf{B}^{*} &= \mathbf{B}^{\Delta t} \\ \boldsymbol{\Gamma}^{*} &= \left(\mathbf{I} - \mathbf{B}^{\Delta t}\right) \left(\mathbf{I} - \mathbf{B}\right)^{-1} \boldsymbol{\Gamma} \\ v_t &= \sum_{\tau=t}^{S} \mathbf{B}^{S-\tau} u_{\tau} \qquad (S \;=\; t + \Delta t - 1) \end{aligned}$$

This still is a linear model, but the implied estimation does not directly estimate the "fundamental" structure.

- Indefinite growth vs equilibrium depends on eigenvalues of $\mathbf{B}$. If the largest absolute value of λ is less than 1, mean equilibrium exists. Otherwise the system grows indefinitely –

$$\begin{aligned}
\max|\lambda| &< 1 \rightarrow \text{equilibrium} \\
& lim_{\Delta t \rightarrow \infty} E\boldsymbol{y} = (\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\Gamma}\boldsymbol{x} \\
\max|\lambda| &= 1 \rightarrow \text{linear growth} \\
\max|\lambda| &> 1 \rightarrow \text{exponential growth}
\end{aligned}$$

- Change model without lagged endogenous variables often reported and defended. But it implies infinite growth –

$$\begin{aligned}
\Delta y_1 &= \beta_{12} y_{2\,t} + \boldsymbol{\gamma}_1 \boldsymbol{x} + u_1 \\
\Delta y_2 &= \beta_{21} y_{1\,t} + \boldsymbol{\gamma}_2 \boldsymbol{x} + u_2 \\
y_{1\,t+1} &= \ 1\times y_{1\,t} + \beta_{12} y_{2\,t} + \boldsymbol{\gamma}_1 \boldsymbol{x} + u_1 \\
y_{2\,t+1} &= \ \beta_{21} y_{1\,t} + 1\times y_{2\,t} + \boldsymbol{\gamma}_2 \boldsymbol{x} + u_2 \\
\mathbf{B} &= \begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix} \\
\lambda &= 1 \pm \sqrt{\beta_{12}\beta_{21}} \quad \rightarrow \quad \max|\lambda| \geq 1
\end{aligned}$$

We argue the infinite-growth question should be settled empirically, not á priori. And therefore, this change model is not justified.

Also, notice that even if $\mathbf{B} = \begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{pmatrix}$, $\mathbf{B}^*$ will not contain ones in the diagonal (unless $\beta_{12}\beta_{21} = 0$). And $\mathbf{B}^*$ surely is what is really being estimated.

- Autoregressive error generally produces nonzero covariance between lagged endogenous variables and the disturbances.

  Example: One-equation, first order autoregressive error –

$$\begin{aligned} y_{t+1} &= \alpha + \beta y_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_{t-1} \rightarrow \mathrm{Cov}\,(y_t, u_t) = \frac{\rho\sigma_u^2}{1 - \rho\beta} \end{aligned}$$

  Since lagged y is a regressor in the estimation, this is a big problem. For this simple model, the OLS estimate of β is inflated by the regression coefficient of u_t on y_t –

$$E\,\hat{\beta}_{ols} = \beta + b_{u_t, y_t}$$

Example simulated data:

$$\alpha = 2$$

$$\left.\begin{array}{rcl}\beta & = & 0.75\\ \rho & = & 0.60\\ \sigma_\varepsilon^2 & = & 1.1\end{array}\right\} \Longrightarrow \left\{\begin{array}{ccc}\sigma_y^2 & = & 10.3571\\ \sigma_u^2 & = & 1.7188\\ Cov\,(u_t,\, y_t) & = & 1.875\\ \text{OLS bias} & = & 1.875/10.3571 \;=\; 0.1810\\ E\,\hat{\beta}_{ols} & = & 0.9310\end{array}\right.$$

$$n = 5000$$

Sample results OLS:

```
proc reg data=autoReg(where=(t=&t));
  model y = lag_y;
run;
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.60217	0.04376	13.76	<.0001
lag_y	1	0.92998	0.00513	181.18	<.0001

Sample results proc mixed:

```
proc mixed noclprint data=autoReg;
  class persID;
  model y = lag_y / solution;
  repeated / sub=persID type=AR(1) R=1;
run;
                              Standard
 Effect       Estimate           Error       DF    t Value    Pr > |t|

 Intercept      0.7809         0.03636     4999      21.48      <.0001
 lag_y          0.9045        0.004251     4999     212.78      <.0001
```

Variance-covariance estimates:

Estimated R Matrix
for persID 1

Row	Col1	Col2
1	1.4214	0.6155
2	0.6155	1.4214

Sample results: proc nlin:

Use following estimating equation –

$$y_{t+2} = \alpha(1-\rho) + (\beta+\rho)\,y_{t+1} - \rho\beta y_t + \varepsilon_t$$

(Linear in variables but nonlinear in parameters.)

```
proc nlin data=autoReg(where=(t=&t));
  parms a=1, b=0.7, rho=0.5;
  model y = a*(1-rho) + (b+rho)*lag_y -b*rho*lag2_y;
run;
   Parameter      Estimate    Std Error      Approximate 95% CL

   a                2.0825       0.2760      1.5414      2.6236
   b                0.7504       0.0333      0.6851      0.8158
   rho              0.5964       0.0406      0.5169      0.6760
```

First Difference Method

When repeated measures are available, a method sometimes called "first-difference" estimation has been used in a number of instances to control for unmeasured variables that do not change from one measurement to the next. Authors who use it tend to highly praise it as a way to eliminate effects of many unobserved variables.

One version of this model is[*] –

$$\begin{aligned} Y_{i1} &= \alpha + \gamma G_i + \epsilon_{i1} \\ Y_{i2} &= \alpha + \tau + \gamma G_i + \delta X_i + \epsilon_{i2} \\ \Delta Y = Y_{i2} - Y_{i1} &= \tau + \delta X_i + \epsilon^* \end{aligned}$$

where –

$$\begin{aligned} Y_{it} &= \text{outcome measure for observation } i \text{ at time } t \\ \alpha,\ \gamma,\ \tau,\ \delta &= \text{coefficients} \\ G_i &= \text{binary indicator of those who} \\ & \quad \text{choose to experience the "treatment", } X_i \\ X_i &= \text{binary indicator of those who actually experienced the treatment} \\ \epsilon_{it} &= \text{error for observation } i \text{ at time } t \\ \epsilon^* &= \epsilon_{i2} - \epsilon_{i1} \end{aligned}$$

[*] Paul D. Allison. (1990). "Change scores as dependent variables in regression analysis." in *Sociological Methodology*. Clifford C. Clogg (ed) Vol 20. Pp. 93 - 114.

Including G is designed to control for self selection.

However, if the observations are generated by a linear difference-equation process, then –

$$\begin{aligned} Y_{i1} &= \alpha + \beta Y_{i0} + \gamma G_i + \varepsilon_{i0} \\ Y_{i2} &= \alpha + \beta Y_{i1} + \gamma G_i + \delta X_i + \varepsilon_{i1} \\ Y_{i2} - Y_{i1} &= \beta \left(Y_{i1} - Y_{i0} \right) + \delta X_i + \left(\varepsilon_{i1} - \varepsilon_{i0} \right) \end{aligned}$$

Here, the regressor is the lagged differenced dependent variable $(Y_{i1} - Y_{i0})$, and its covariance with the differenced error $(\varepsilon_{i1} - \varepsilon_{i0})$ obviously is not zero.

Conclusions

- The case for learning benefits of racial match of teachers to students remains open. It is not completely discredited, but it remains very much in doubt.
- The research on the race-match hypothesis to date does not support an aggressive policy to recruit minority teachers.
- The correct specification of a statistical model of learning remains in doubt.

Meta Method Musings

- In the absence of random-assignment experiments, a structural model should be devised to represent the process that generates the observations. (This implies a built-in forecast method based on the theory.)
- Statistical estimation should be derived from the structural model.
- Derivation of the statistical model must recognize that the time gap between observations generally far exceeds the length of the "reaction" interval. In

the case of linear difference or differential equations, the solutions remain linear. But the algebraic form of the structural model and the estimating equations otherwise do not match.

- Generally, closed-form estimating equations are not feasible. Partly for this reason, linear models appear to be the best initial form.
- The connection between the data-generation process and statistical estimation in most published work is either unclear or too rudimentary to be taken seriously.
- The difference-equation model of learning is an initial model of the learning process. It has obvious problems but may serve as a springboard.
- Autoregressive error is a major problem in estimating parameters of a dynamic process. It deserves a lot of thought.