

The Long-Term Impacts of Teachers: Teacher Value-Added and Students' Outcomes in Adulthood

Raj Chetty, Harvard
John N. Friedman, Harvard
Jonah Rockoff, Columbia

December 2011

Introduction: Teacher Value-Added

- Many people advocate improving quality of teaching, but how should teacher quality be measured and improved?
- One approach: “value-added” (VA) measures
 - Rate teachers based on their students’ test score gains
- School districts have started to use such VA measures, leading to considerable debate in policy circles
 - Washington D.C. lays off teachers and offers bonuses using a metric that puts 50% weight on VA measures
 - LA Times publishes VA of 12,000 public schools teachers by name
 - New York City school district lawsuits

Debate About Teacher Value-Added

- Debate about value-added stems primarily from two basic issues:
 1. Dispute about whether VA measures are biased
[Kane and Staiger 2008, Rothstein 2010]
 - Do differences in test-score gains across teachers capture causal impacts of teachers or are they driven by student sorting?
 - If VA estimates are biased, they will incorrectly reward or penalize teachers for the mix of students they get
 2. Lack of evidence on teachers' long-term impacts
 - Do teachers who raise test scores improve students' long-term outcomes or are they simply better at teaching to the test?

Objectives of This Paper

- This study answers the these questions by tracking one million children from childhood to early adulthood
 - Develop new quasi-experimental tests for bias in VA estimates
 - Test whether children who get high VA teachers have better outcomes in adulthood
- Results also shed light on broader issues in the economics of education
 - What are the long-run returns to investments in better teaching?
 - Are impacts on scores a good proxy for long-term impacts of educational interventions?

Dataset 1: School District Data

- Teacher and class assignments from 1991-2009 for 2.5 million children
- Test scores from 1989-2009
 - Scaled scores standardized by grade and subject (math/reading)
 - 18 million test scores, grades 3-8
- Exclude students in special ed. schools and classrooms (6% of obs.)

Dataset 2: United States Tax Data

- Selected data from U.S. federal income tax returns from 1996-2010
 - Includes non-filers via information forms (e.g. W-2's)
- Student outcomes: earnings, college, teenage birth, neighborhood quality
- Parent characteristics: household income, 401k savings, home ownership, marital status, age at child birth
 - Omitted variables from standard VA models
- Approximately 90% of student records matched to tax data
 - Data were analyzed as part of a broader project on tax policy
 - Research based purely on statistics aggregating over thousands of individuals, not on individual data

Data Structure

Student	Subject	Year	Grade	Class	Teacher	Test Score	Age 28 Earnings
			⋮				
Tom	Math	1992	4	1	Samuelson	0.5	\$22K
Tom	English	1992	4	1	Samuelson	1.3	\$22K
Tom	Math	1993	5	2	Solow	0.9	\$22K
Tom	English	1993	5	2	Solow	0.1	\$22K
Tom	Math	1994	6	3	Arrow	1.5	\$22K
Tom	English	1994	6	4	Stigler	0.5	\$22K
			⋮				

- One observation per student-subject-year

Summary Statistics

Variable	Mean (1)	S.D. (2)
<u>Student Data:</u>		
Class size (not student-weighted)	28.3	5.8
Teacher experience (years)	8.08	7.72
Test score (SD)	0.12	0.91
Female	50.3%	50.0%
Age (years)	11.7	1.6
Free lunch eligible (1999-2009)	76.0%	42.7%
Minority (Black or Hispanic)	71.8%	45.0%
English language learner	10.3%	30.4%
Special education	3.4%	18.1%
Repeating grade	2.7%	16.1%
Number of subject-school years per student	6.14	3.16
Student match rate to adult outcomes	89.2%	31.0%
Student match rate to parent chars.	94.6%	22.5%

Summary Statistics

Variable	Mean (1)	S.D. (2)
<u>Adult Outcomes:</u>		
Annual wage earnings at age 20	4,796	6,544
Annual wage earnings at age 25	15,797	18,478
Annual wage earnings at age 28	20,327	23,782
In college at age 20	36.2%	48.1%
In college at age 25	17.3%	37.8%
College Quality at age 20	24,424	12,834
Contribute to a 401(k) at age 25	14.8%	35.5%
ZIP code % college graduates at age 25	13.2%	7.1%
Had a child while a teenager (for women)	8.4%	27.8%
<u>Parent Characteristics:</u>		
Household income (child age 19-21)	35,476	31,080
Ever owned a house (child age 19-21)	32.5%	46.8%
Contributed to a 401k (child age 19-21)	25.1%	43.3%
Ever married (child age 19-21)	42.1%	49.4%
Age at child birth	27.6	7.4
Predicted Score	0.16	0.26

Conceptual Framework

- We begin from a structural model of the education production function for scores and earnings [Todd and Wolpin 2003]
 - Scores depend on history of teachers and inputs up to current grade
 - Earnings depend on the sequence of teachers and other inputs throughout years in school
- Teachers affect scores and earnings directly and through other inputs
 - Better teachers may track students to better future teachers
 - Parental investment may react to teacher quality
- Students and teachers assigned to classrooms based on history of chars.

Empirical Model: Scores

- We estimate value-added using this equation for student i in grade g :

$$A_{ig} = \pi A_{i,g-1} + \phi X_{ig} + \underbrace{\mu_{j(i,g)} + \varepsilon_{ig}}_{\text{Unobservable Components}}$$

- μ_j is the VA for teacher j , interpreted as the reduced form causal effect of randomly assigning a teacher to a classroom
- Estimate $\hat{\mu}_j$ as the mean residual across classes taught by teacher j
 - Identification Assumption 1 [Score Impacts]:

No sorting to teachers based on unobserved test-score innovations ε_{ig}

Empirical Model: Earnings

- Estimating equation for long-run impacts:

$$Y_i = \beta\mu_{j(i,g)} + \mu_{j(i,g)}^\perp + \pi^Y A_{i,g-1} + \phi^Y X_{ig} + \rho\varepsilon_{ig} + v_{ig}$$

Teacher effect on earnings
orthogonal to value-added

Student unobservables
orthogonal to scores

Empirical Model: Earnings

- Estimating equation for long-run impacts:

$$Y_i = \beta \mu_{j(i,g)} + \mu_{j(i,g)}^\perp + \pi^Y A_{i,g-1} + \phi^Y X_{ig} + \rho \varepsilon_{ig} + v_{ig}$$

- β = covariance of teacher effect on score and earnings
- Estimate β by regressing Y_i on $\hat{\mu}_j$
 - Identification Assumption 2 [Long-Term Impacts]:
Teacher value-added μ_j uncorrelated with unobservable determinants of earnings v_{ig}
 - Sorting to teachers on v_{ig} is fine if uncorrelated with teacher VA

Constructing Value-Added

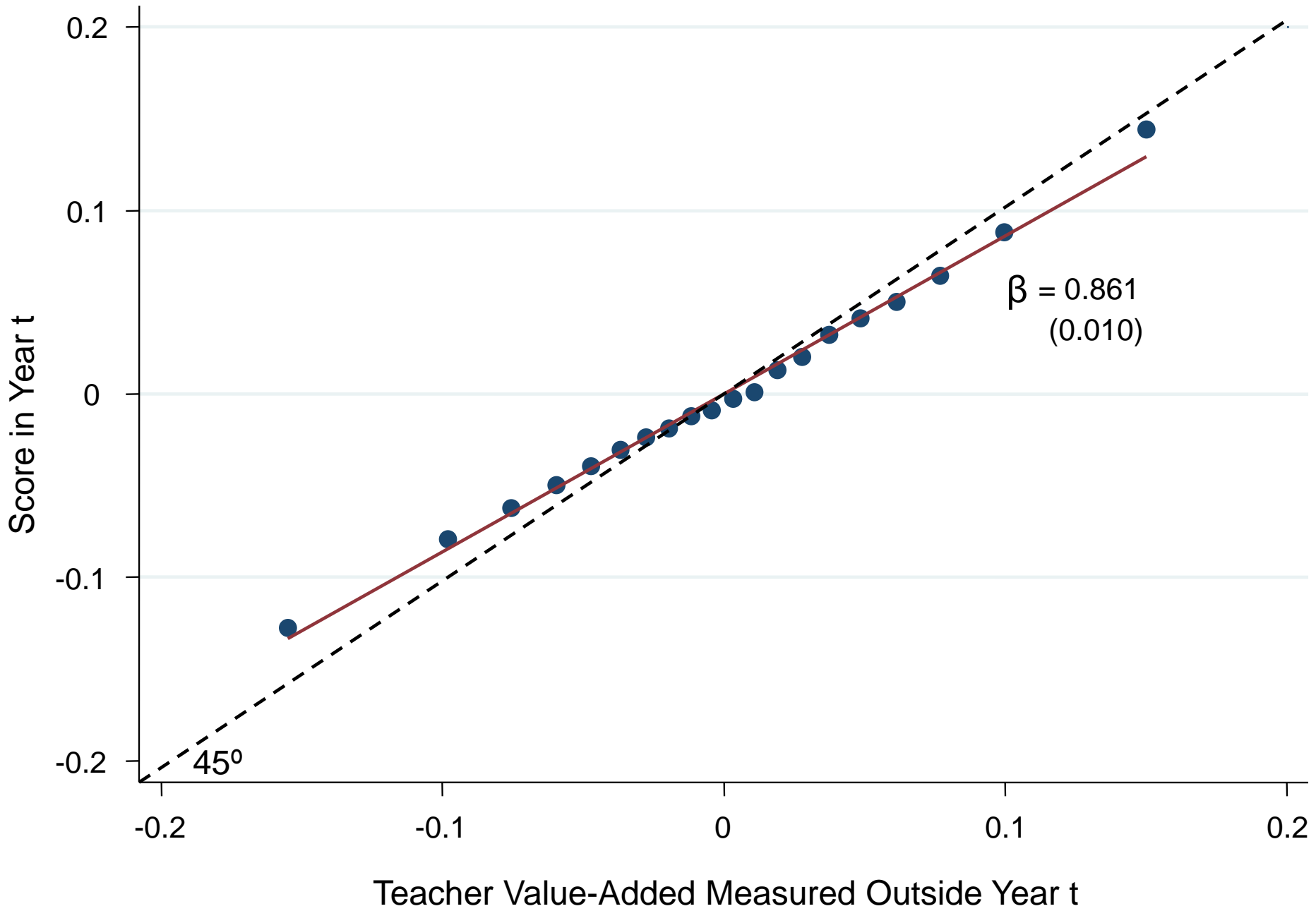
- Estimation of teacher VA (μ_j) following Kane and Staiger (2008)
 1. Regress student test scores on lagged scores and other pre-determined characteristics to form score residuals
 2. Compute mean of student residuals across classes taught by teacher t to estimate teacher's impact
 3. Account for noise in estimated teacher effect by shrinking estimate toward sample mean (0) based on number of students (Empirical Bayes)
- Use **leave-year-out mean** to measure μ_j (Jacob, Lefgren, and Sims 2010) to avoid mechanical bias from correlation between scores and earnings

Empirical Specification

$$Y_i = \beta \hat{\mu}_{j(i,g)} + \pi^Y(A_{i,g-1}) + \phi^Y X_{ig} + v_{ig}$$

- Cubic in lagged scores for both own-subject and cross-subject
 - Also include lagged mean scores at the class- and school-grade level
- Control vector (Kane, Rockoff, Staiger 2008):
 - Student lagged suspensions, lagged absences
 - Student demographics (ethnicity, gender, age)
 - Grade repetition, special education, limited English
 - Honors and remedial class indicators
 - Class size, and class-level means of all covariates
 - School x year means of covariates
 - Year, grade, teacher experience dummies
- Every regression + graph that follows conditions on these controls

Score in Year t vs. Teacher Value-Added



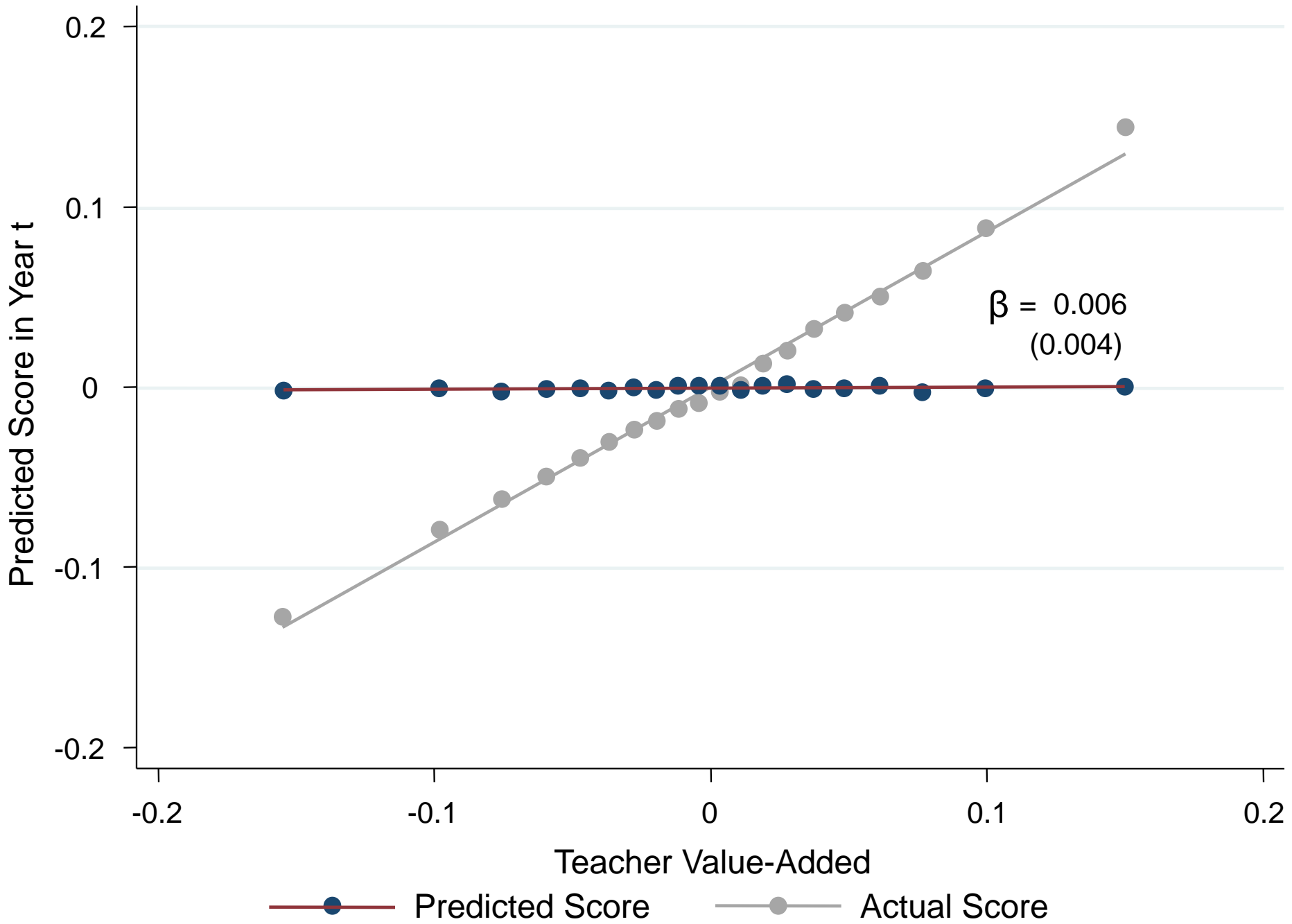
Question 1: Are VA Estimates Unbiased?

- Students are not randomly assigned to teachers
 - Teachers' estimated VA may pick up the types students they get rather than causal impacts of teacher
- Standard approach to deal with non-random assignment: control for prior year variables
- Is this sufficient to obtain consistent estimates of teacher impacts?
 - Recent studies have reached conflicting conclusions (e.g. Kane and Staiger 2008, Rothstein 2010)

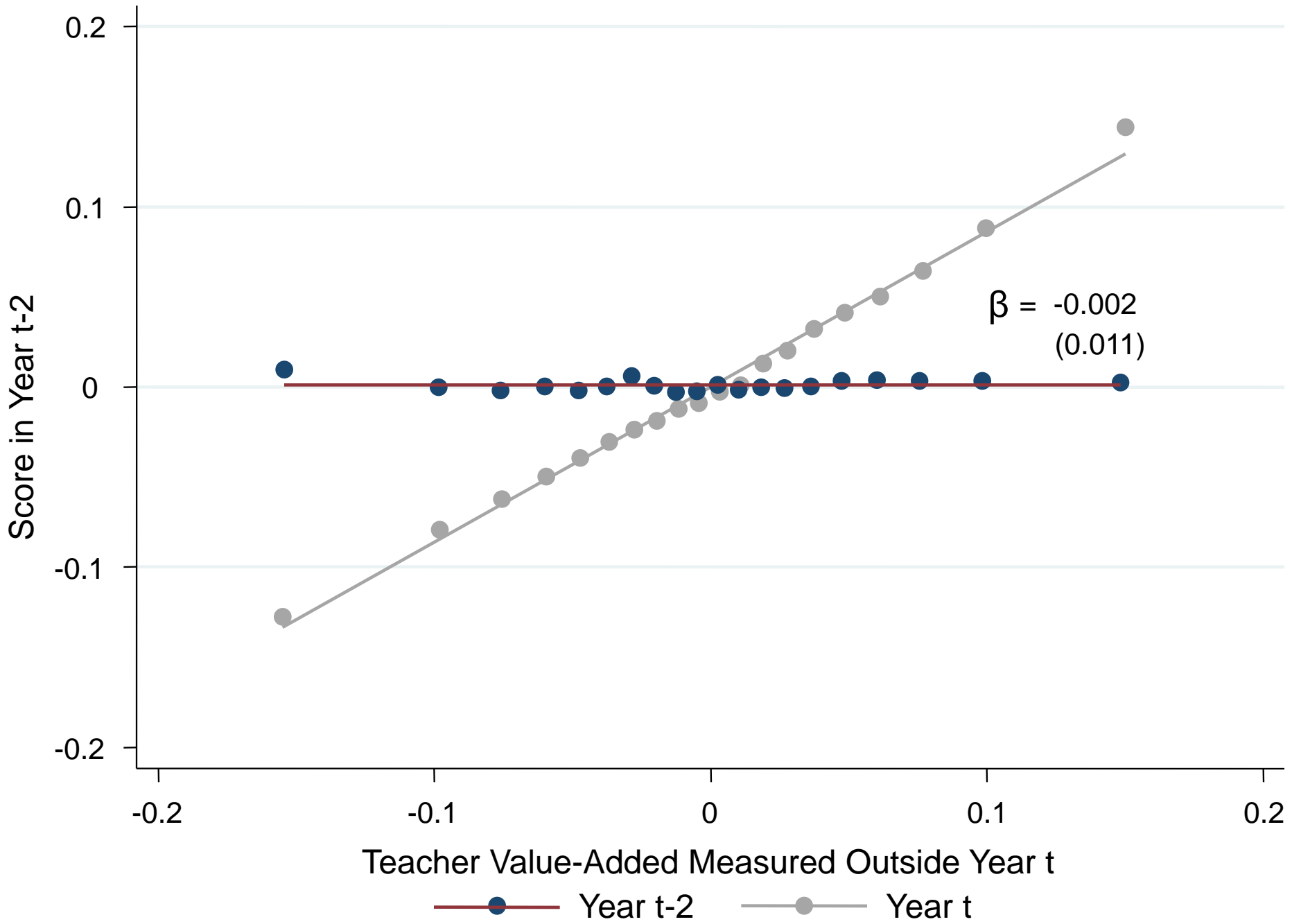
Question 1: Are VA Estimates Unbiased?

- We evaluate where VA estimates are biased in two ways
 1. Test for bias on observables:
 - Are parent characteristics correlated with teacher VA?
 - Are twice-lagged test scores correlated with teacher VA?
 2. Test for bias on unobservables:
 - Quasi-experimental “teacher switchers” variation
- Combine parental background characteristics into a single **predicted score** using cross-sectional regression

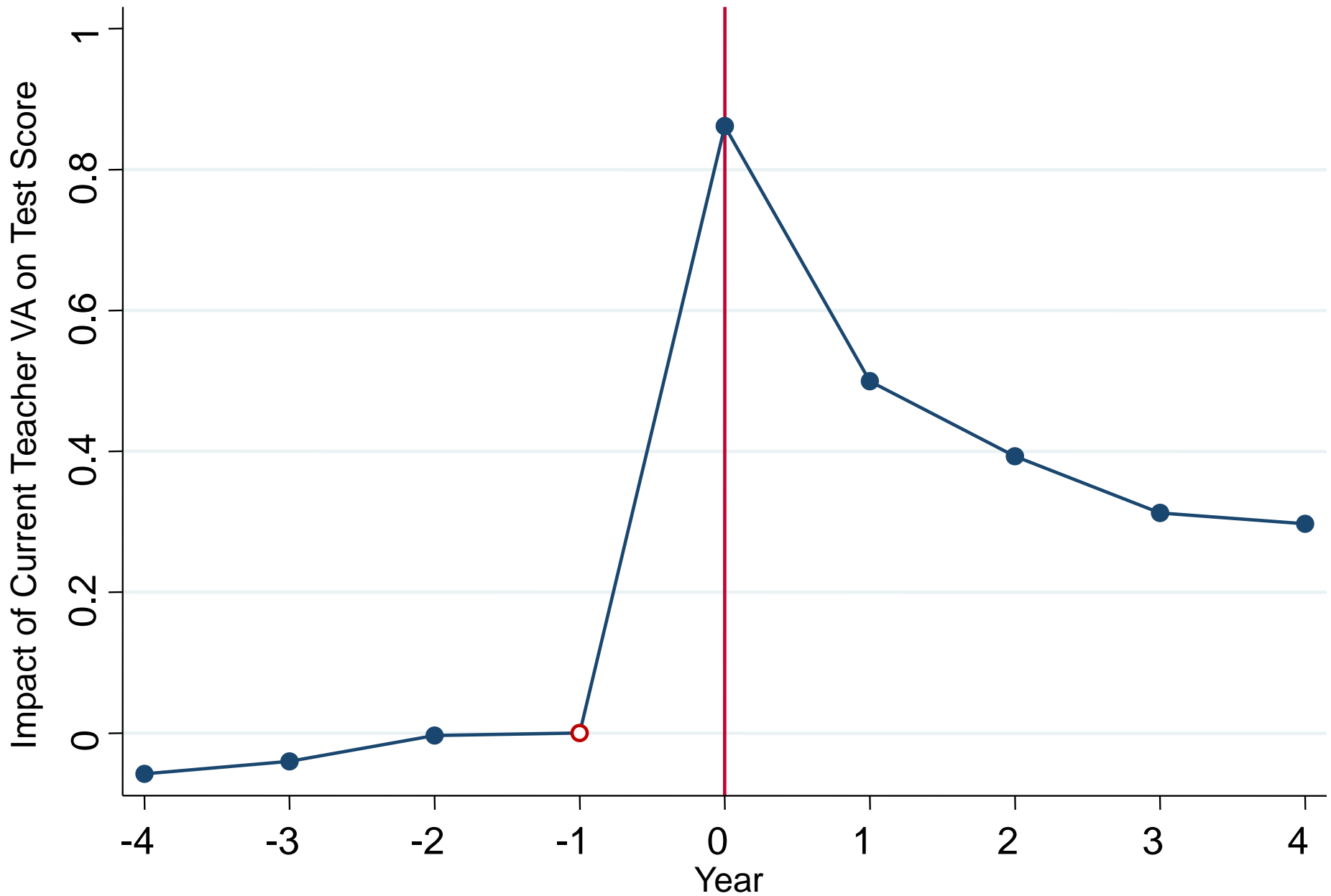
Predicted Scores based on Parent Chars. vs. Teacher Value-Added



Twice-Lagged Score vs. Current Teacher VA



Impacts of Teacher Value-Added on Lagged, Current, and Future Test Scores



Tests for Balance Using Parent Characteristics and Lagged Scores

Dep. Var.:	Score	Score in year t	Predicted Score	Score in year t	Score in year t-2
	(SD)	(SD)	(SD)	(SD)	(SD)
	(1)	(2)	(3)	(4)	(5)
Value-Added	0.861 (0.010) [82.68]	0.864 (0.011) [75.85]	0.006 (0.004) [1.49]	0.804 (0.011) [70.63]	-0.002 (0.011) [-0.21]
Pred score (parent chars.)		0.175 (0.012) [62.70]			
t-2 Score				0.521 (0.001) [363.3]	
Observations	3,721,120	2,877,502	2,877,502	2,771,865	2,771,865

Fade-Out of Test Score Impacts

Dep. Var.:	Score at End of Year t	Score in t+1	Score in t+2	Score in t+3	Score in t+4
	(SD)	(SD)	(SD)	(SD)	(SD)
	(1)	(2)	(3)	(3)	(4)
Value-Added	0.861 (0.010)	0.499 (0.011)	0.393 (0.012)	0.312 (0.013)	0.297 (0.018)
Observations	3,721,120	2,911,042	2,247,141	1,578,551	790,173

Sensitivity of Teacher Value-Added Measures to Controls

	(1)	(2)	(3)	(4)	(5)	(6)
	baseline	add parent chars.	add t-2 scores	t-1 scores only	no controls	Quasi- Experimental Estimate of Bias
Baseline	1.000					3.1% (7.6)
add Parent	0.999	1.000				2.6% (7.6)
add t-2 Scores	0.975	0.974	1.000			1.7% (7.4)
t-1 Scores only	0.945	0.943	0.921	1.000		14.3% (6.9)
No Controls	0.296	0.292	0.279	0.323	1.000	87.8% (1.4)

Quasi-Experimental Validation: Teacher Switchers

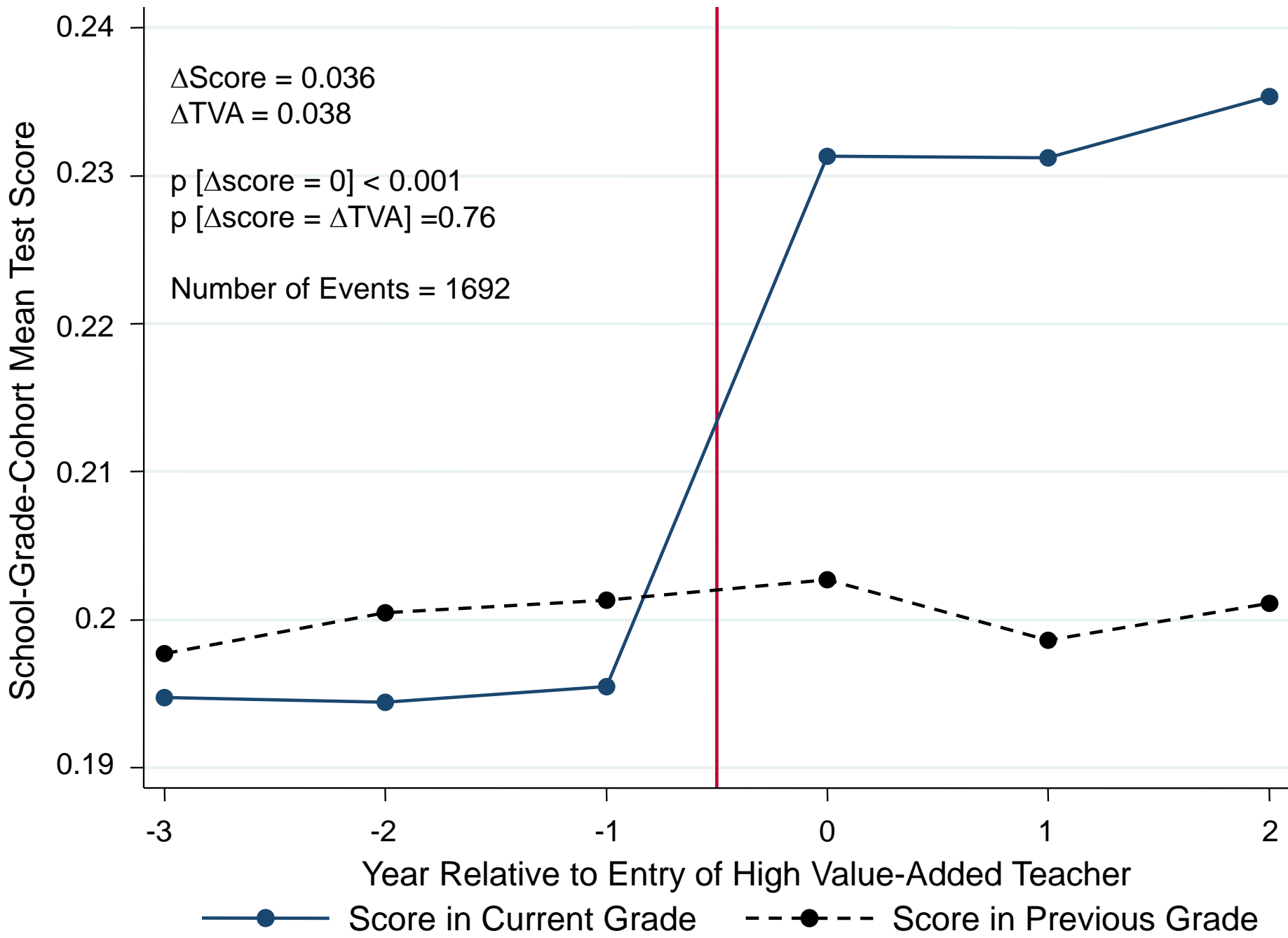
- Leave-out mean VA measures orthogonal to observables in our data
- But selection on unobservables could still be a problem (Rothstein 2010)
- Ideal test: out-of-sample forecasts in experiments (Kane and Staiger 2008)
 - Does a student who is randomly assigned to a teacher previously estimated to be high VA have higher test score gains?
- We use teacher switching as quasi-experiments

Teacher Switchers in School-Grade-Subject-Year Level Data

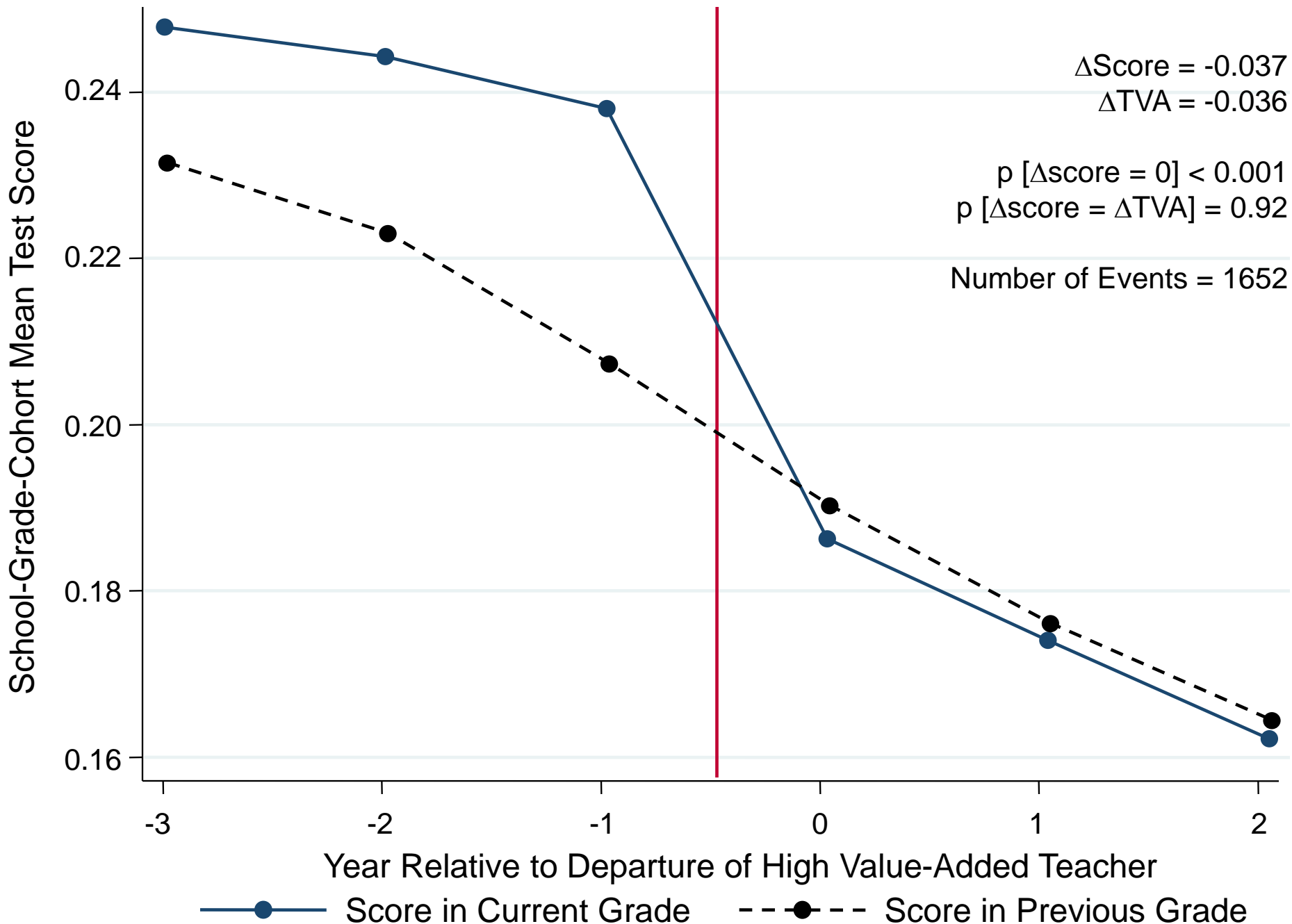
School	Grade	Subject	Year	Teachers	Mean Score	Mean Age 28 Earnings
1	5	math	1992	Smith, Solow, ...	-.09	\$15K
1	5	math	1993	Smith, Solow, ...	-.04	\$17K
1	5	math	1994	Smith, Solow, ...	-.05	\$16K
1	5	math	1995	Samuelson, Solow, ...	0.01	\$18K
1	5	math	1996	Samuelson, Solow, ...	0.04	\$17K
1	5	math	1997	Samuelson, Solow, ...	0.02	\$18K

- Smith switches to grade 4 in 1995; Samuelson switches into grade 5

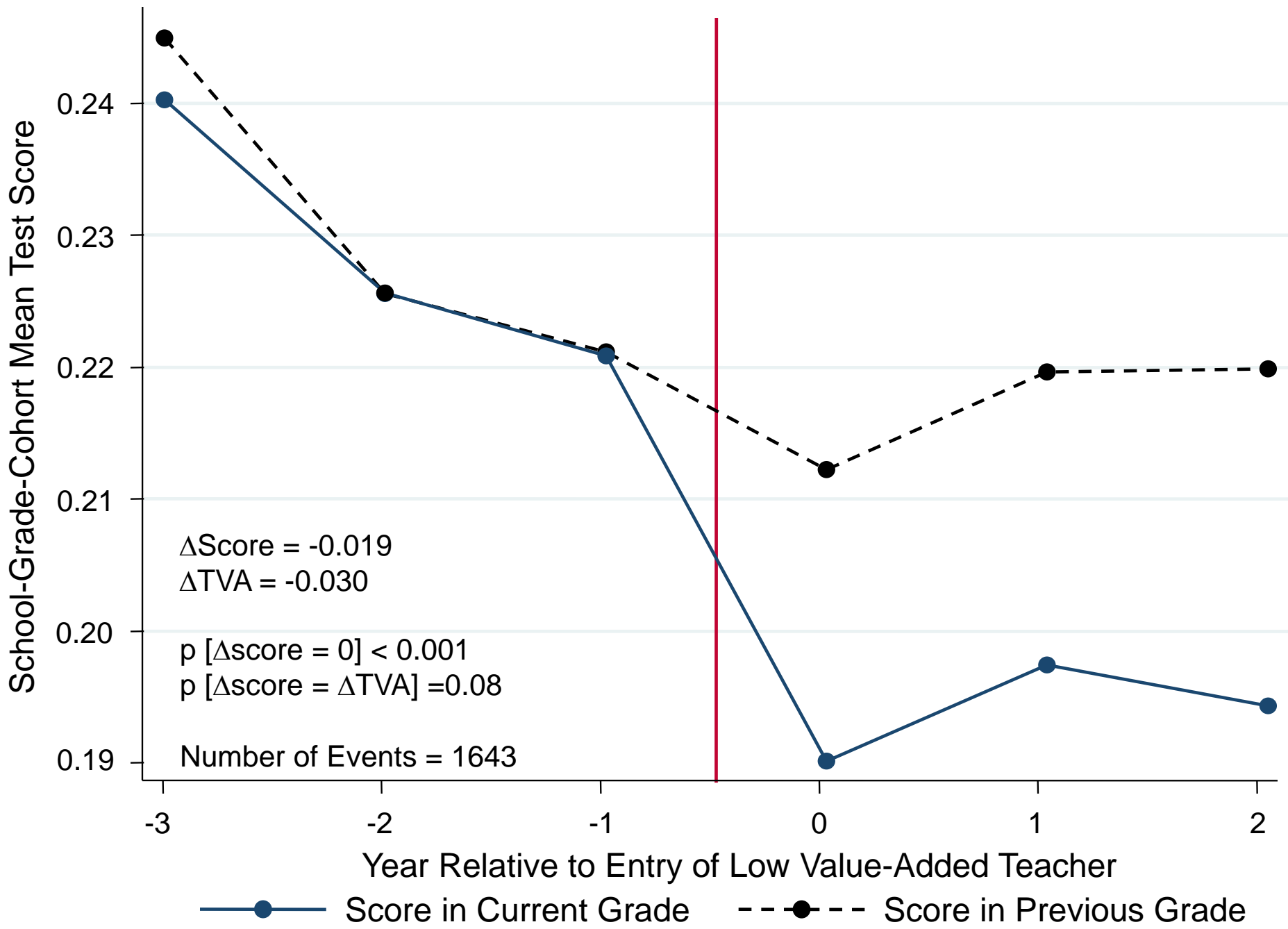
Impact of High Value-Added Teacher Entry on Cohort Test Scores



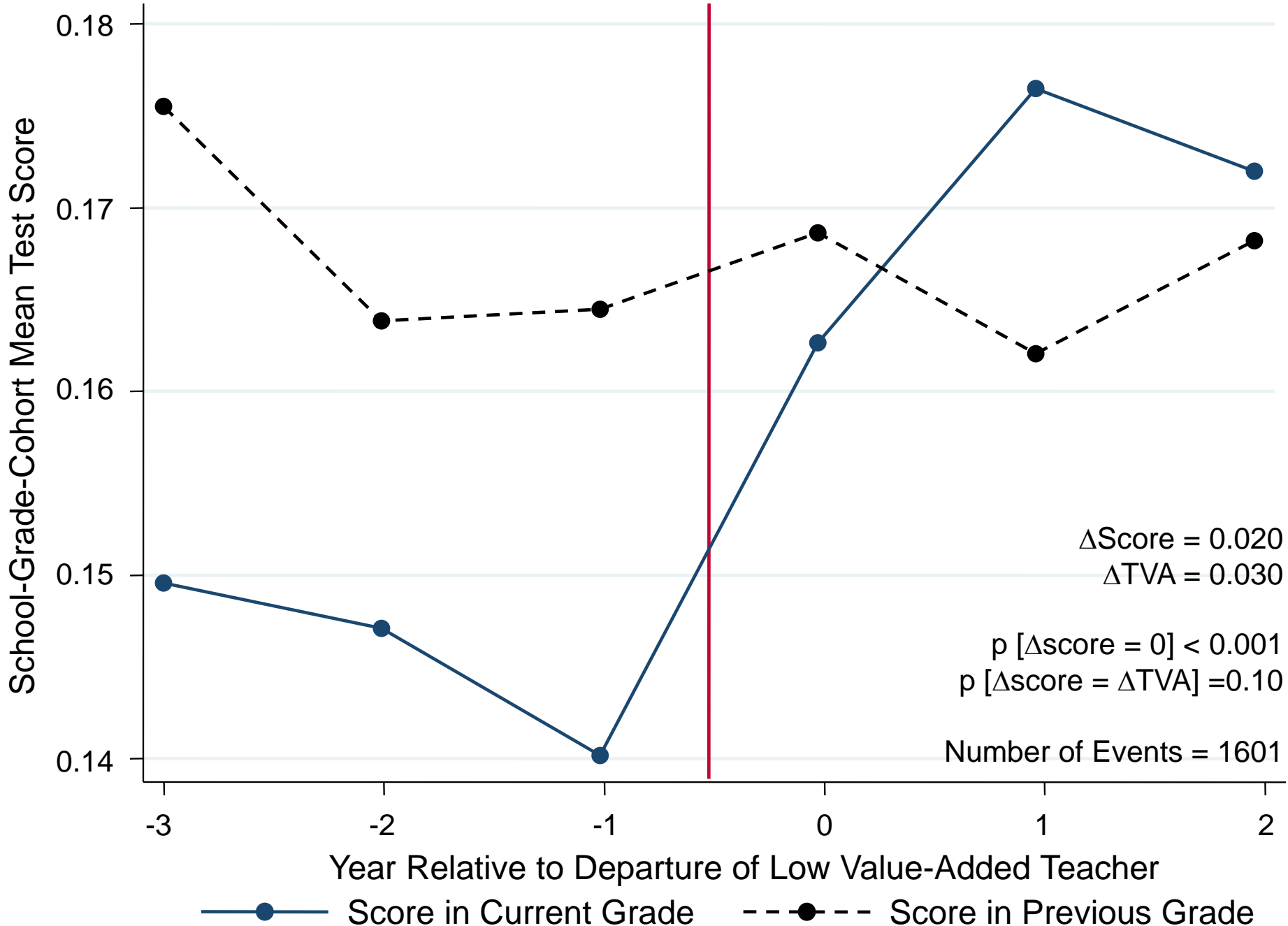
Impact of High Value-Added Teacher Exit on Cohort Test Scores



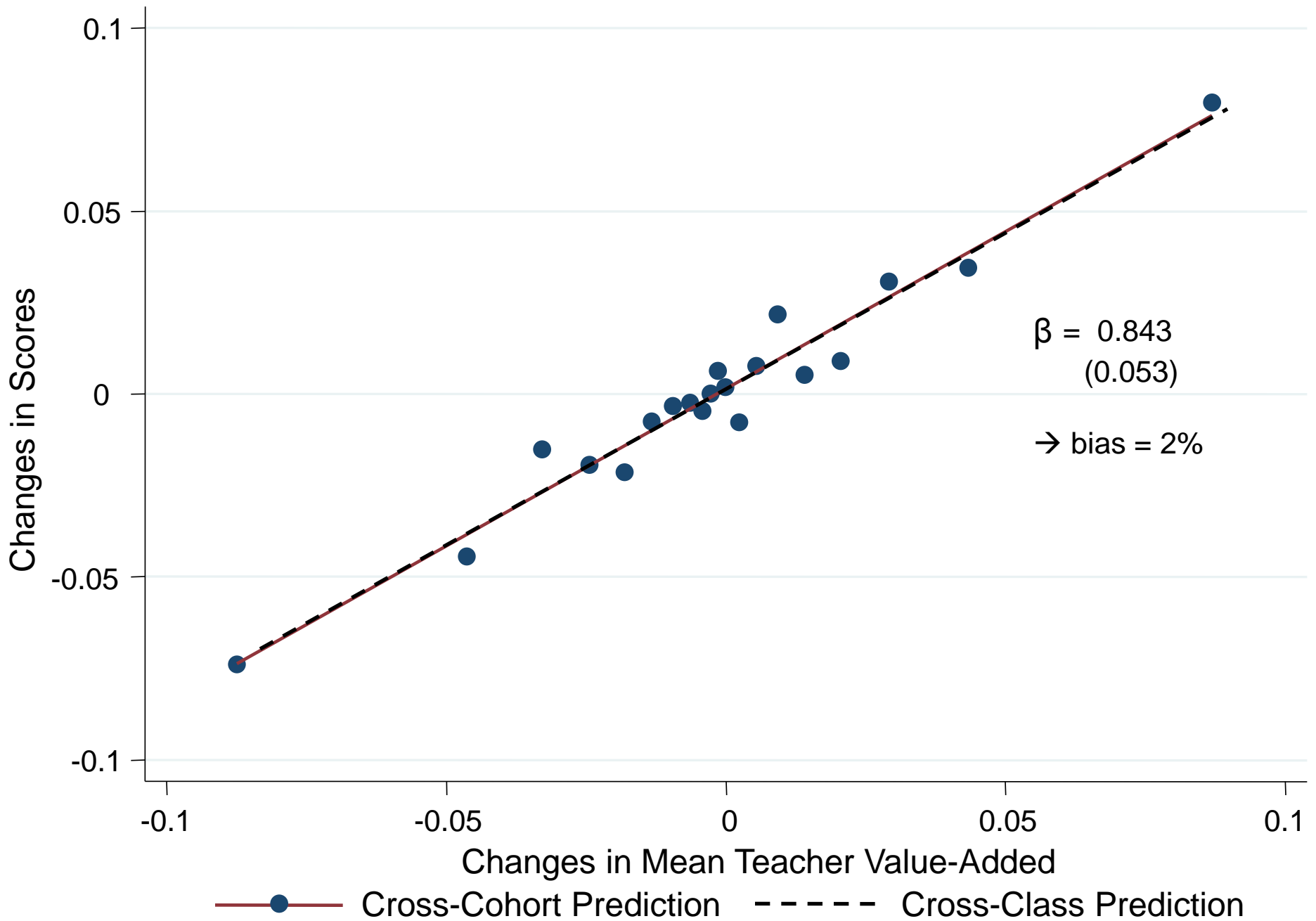
Impact of Low Value-Added Teacher Entry on Cohort Test Scores



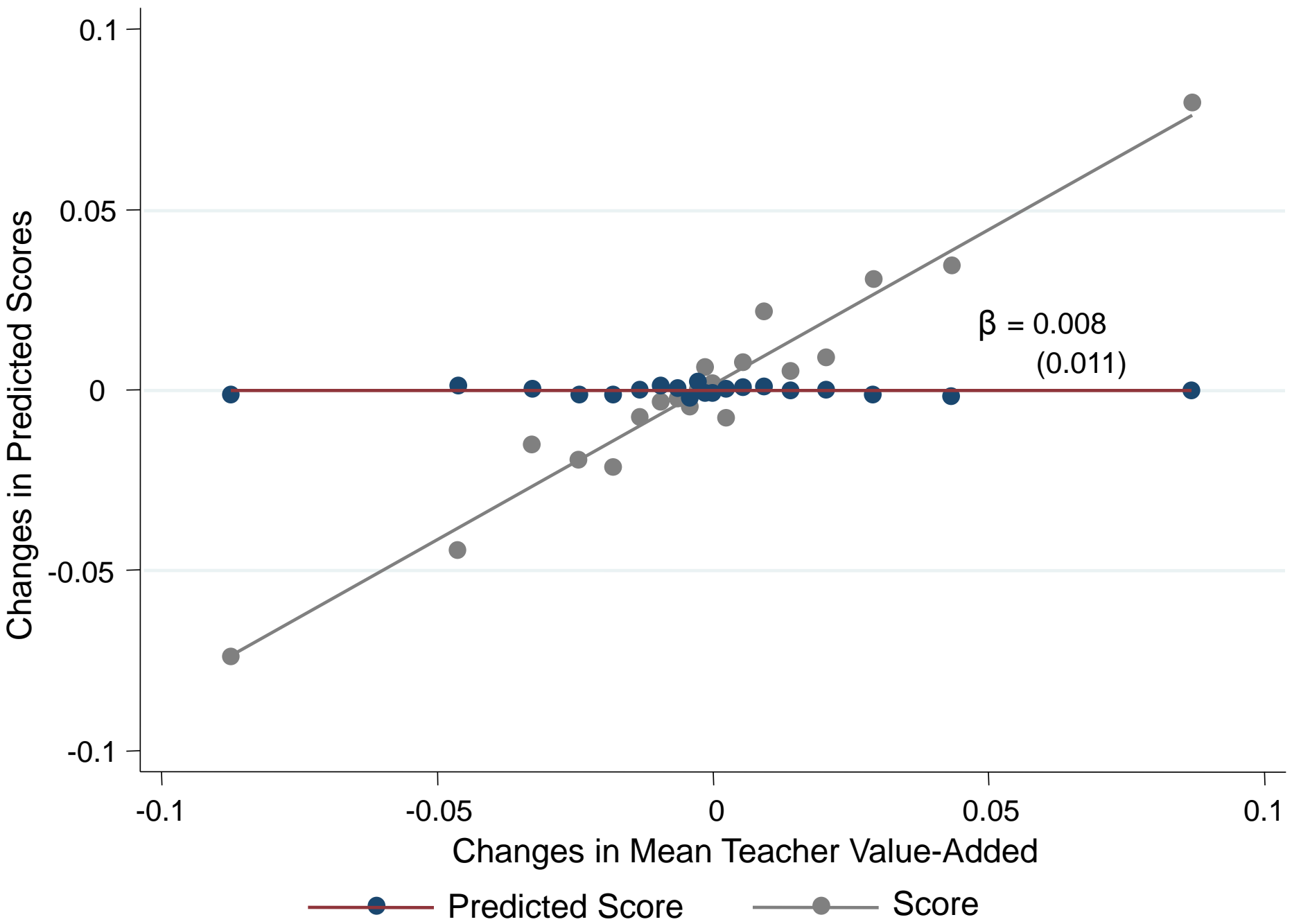
Impact of Low Value-Added Teacher Exit on Cohort Test Scores



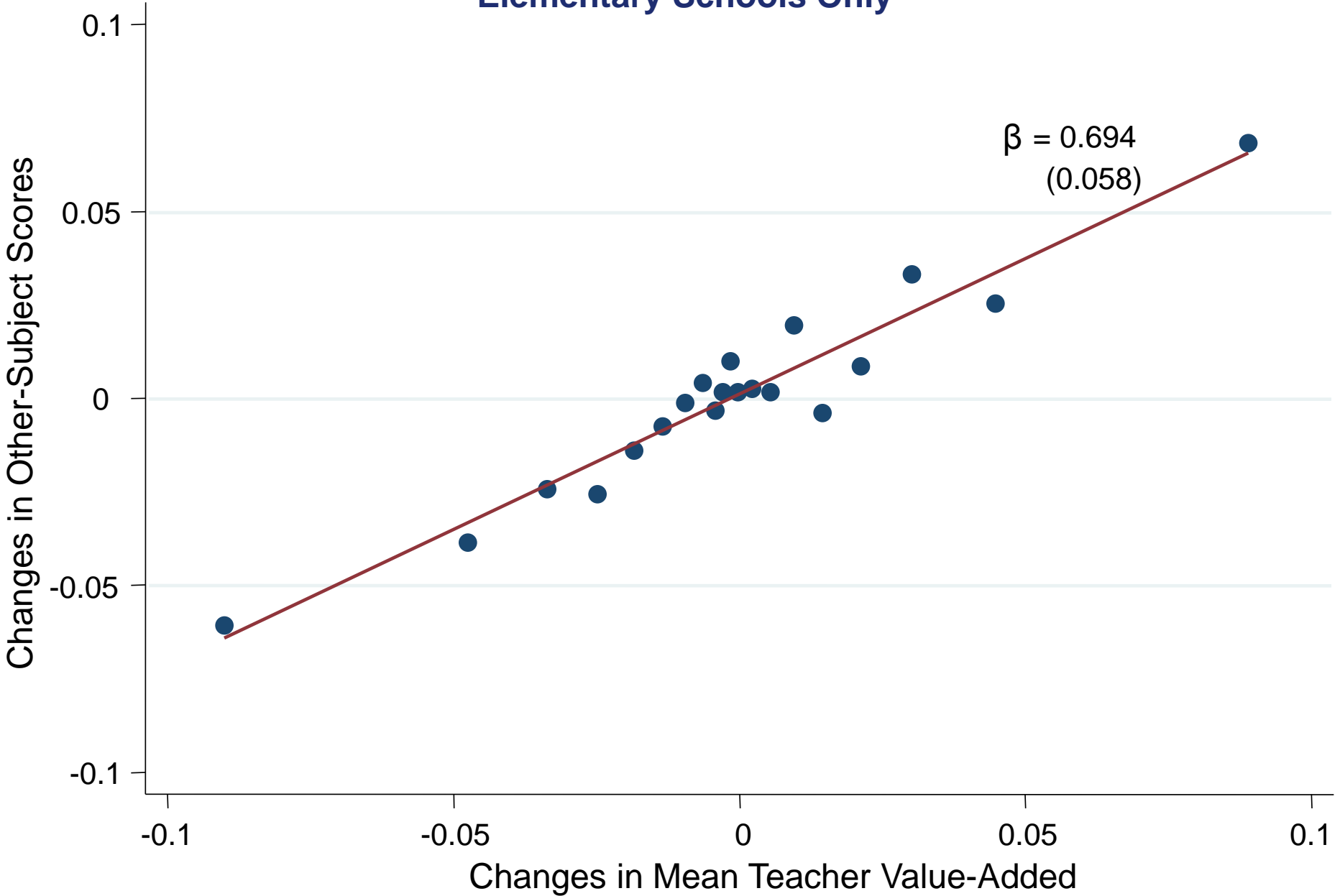
Teacher Switchers Design: Changes in Scores vs. Changes in Mean Teacher VA



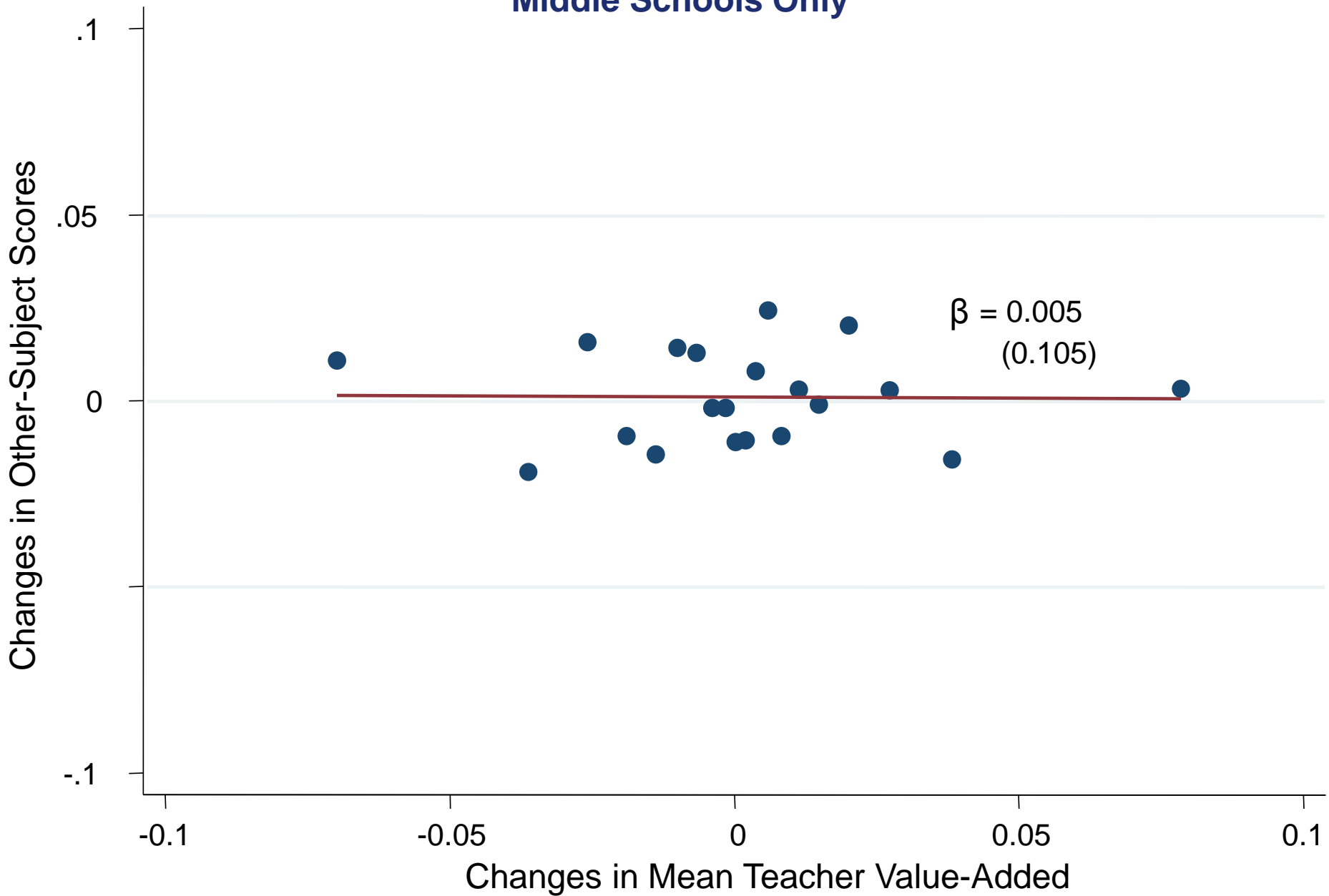
Changes in Predicted Scores vs. Changes in Mean Teacher VA



Changes in Other-Subject Scores vs. Changes in Mean Teacher VA Elementary Schools Only



Changes in Other-Subject Scores vs. Changes in Mean Teacher VA Middle Schools Only



Teacher Switchers: Quasi-Experimental Validation

- Teacher switcher (“cross-cohort”) research design:
 - Does *not* rely on classroom assignment for variation (as in Rivkin, Hanushek, Kain 2005)
 - Does not rely on any controls
- Can be used to validate VA estimates in any dataset
- Provides direct evidence on impacts of hiring/dismissing teachers

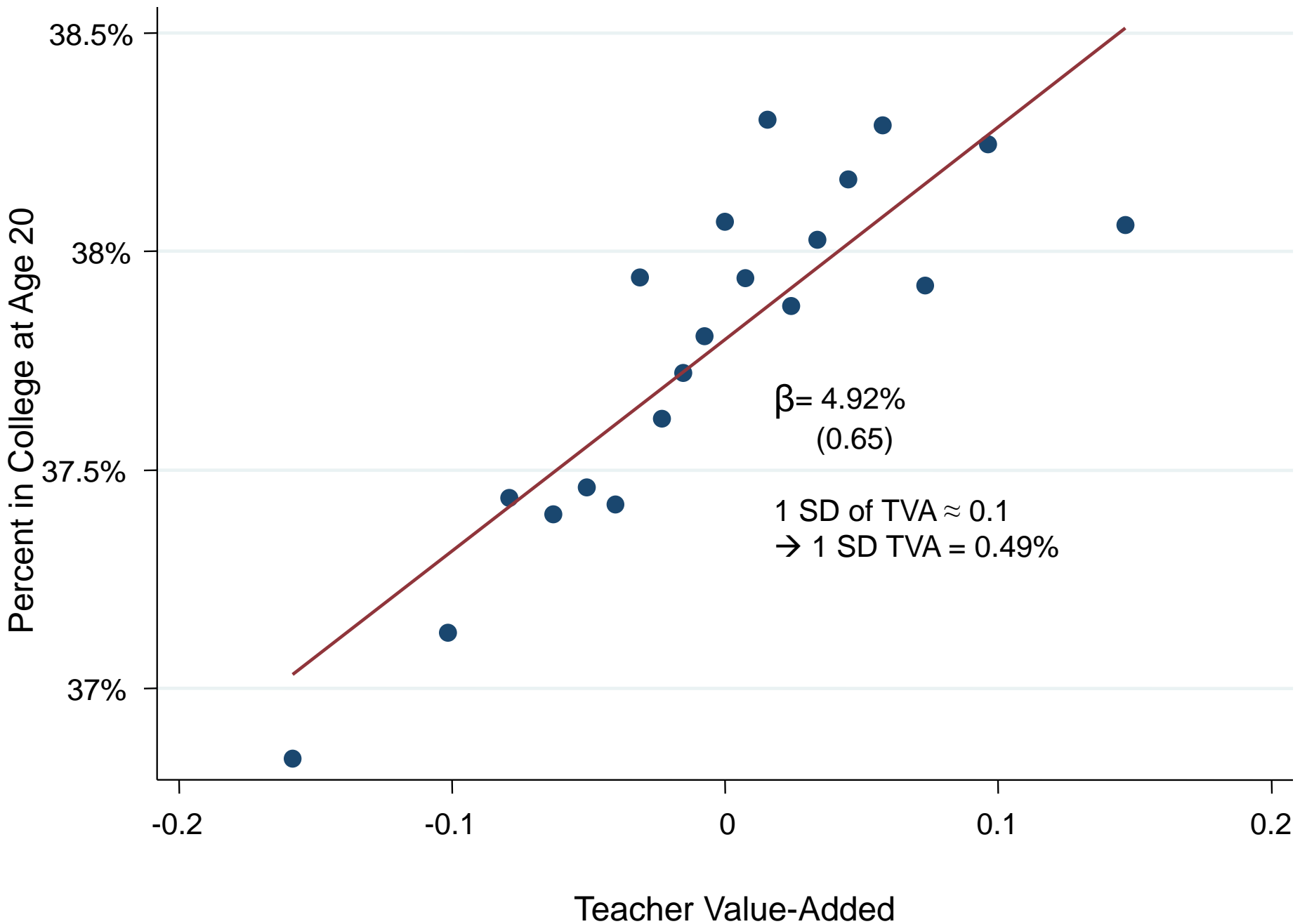
Relation to Rothstein (2010) Findings on Sorting

- Rothstein result 1: Students are sorted into classrooms based on pre-determined variables such as grade $g-2$ test scores
 - We confirm this result in our data
- Rothstein result 2: Selection on observables is minimal conditional on grade $g-1$ controls
 - Controlling for grade $g-2$ score does not affect VA estimates
 - Consistent with our findings that VA does not predict $g-2$ score
- Rothstein notes that his findings do not imply bias in VA estimates
 - But they raise concerns about potential selection on unobservables
 - Our quasi-experimental teacher switcher tests indicate that selection on unobservables turns out to be modest in practice

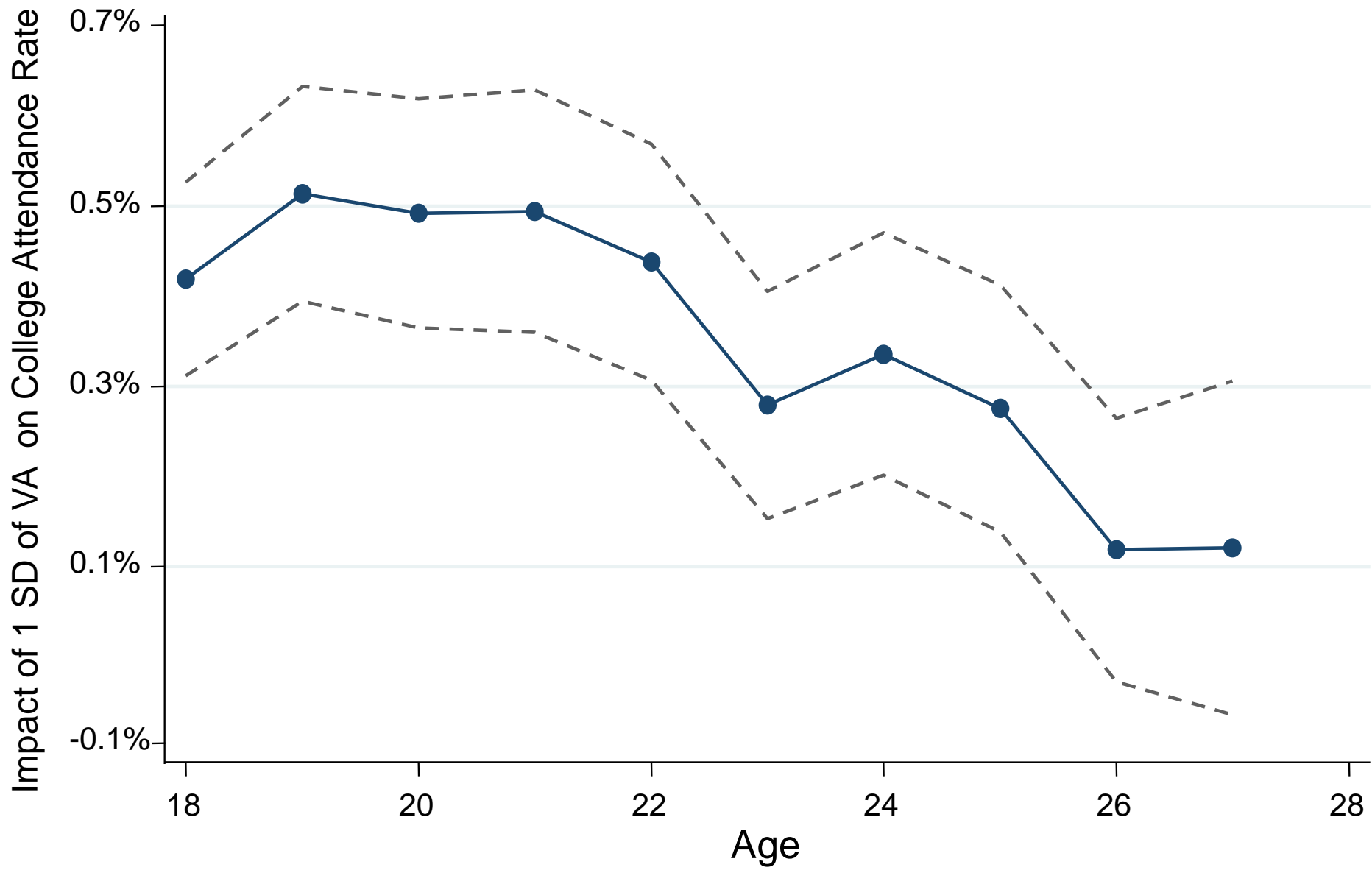
Question 2: Impacts on Outcomes in Adulthood

- Now test whether teachers who raise test scores also increase students' long-run outcomes
- Regress adult outcomes on value-added measures using all classes
 - Then validate using cross-cohort design that does not rely on class assignment
- Interpretation:
 - β is *reduced-form* impact of having better teacher, as measured by VA, for a **single year** during grades 4-8 on earnings
 - Includes benefit of better teachers, peers, etc. in later grades via tracking, as well as any complementarity with future teacher quality

College Attendance at Age 20 vs. Teacher Value-Added

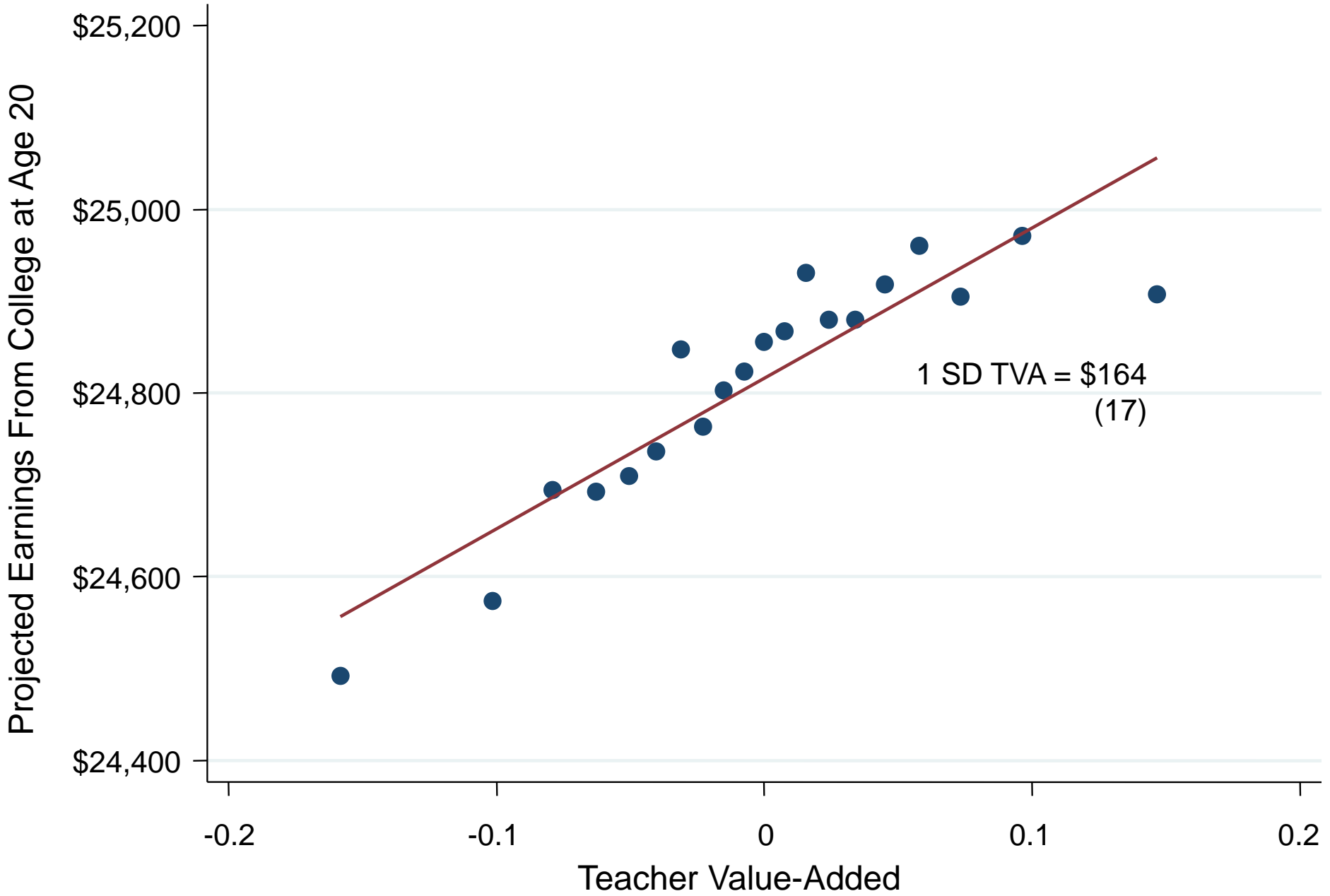


Impact of Teacher Value-Added on College Attendance by Age



● Impact of 1 SD Increase in Teacher VA on College Attendance - - - - - 95% CI

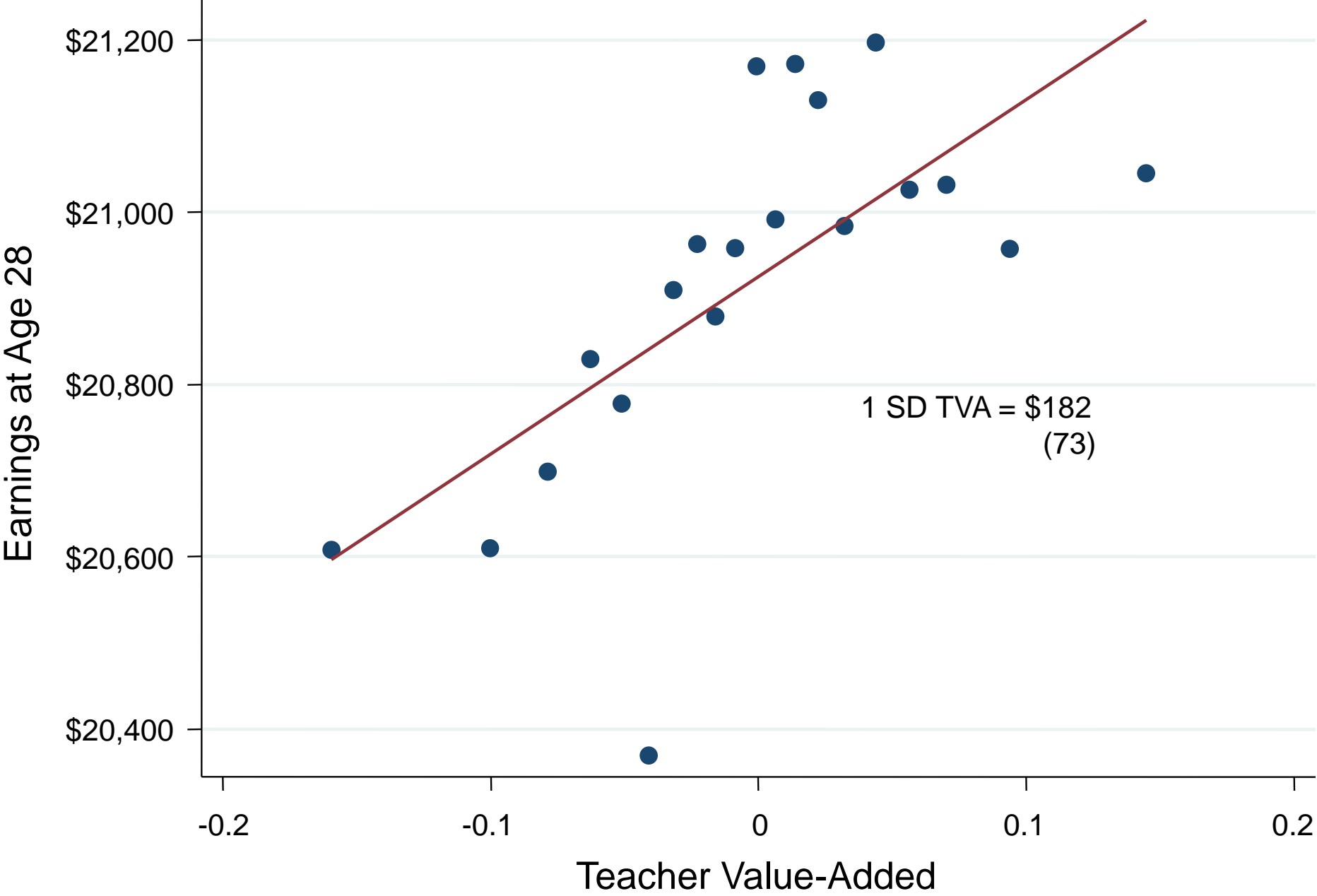
College Quality (Projected Earnings) at Age 20 vs. Teacher Value-Added



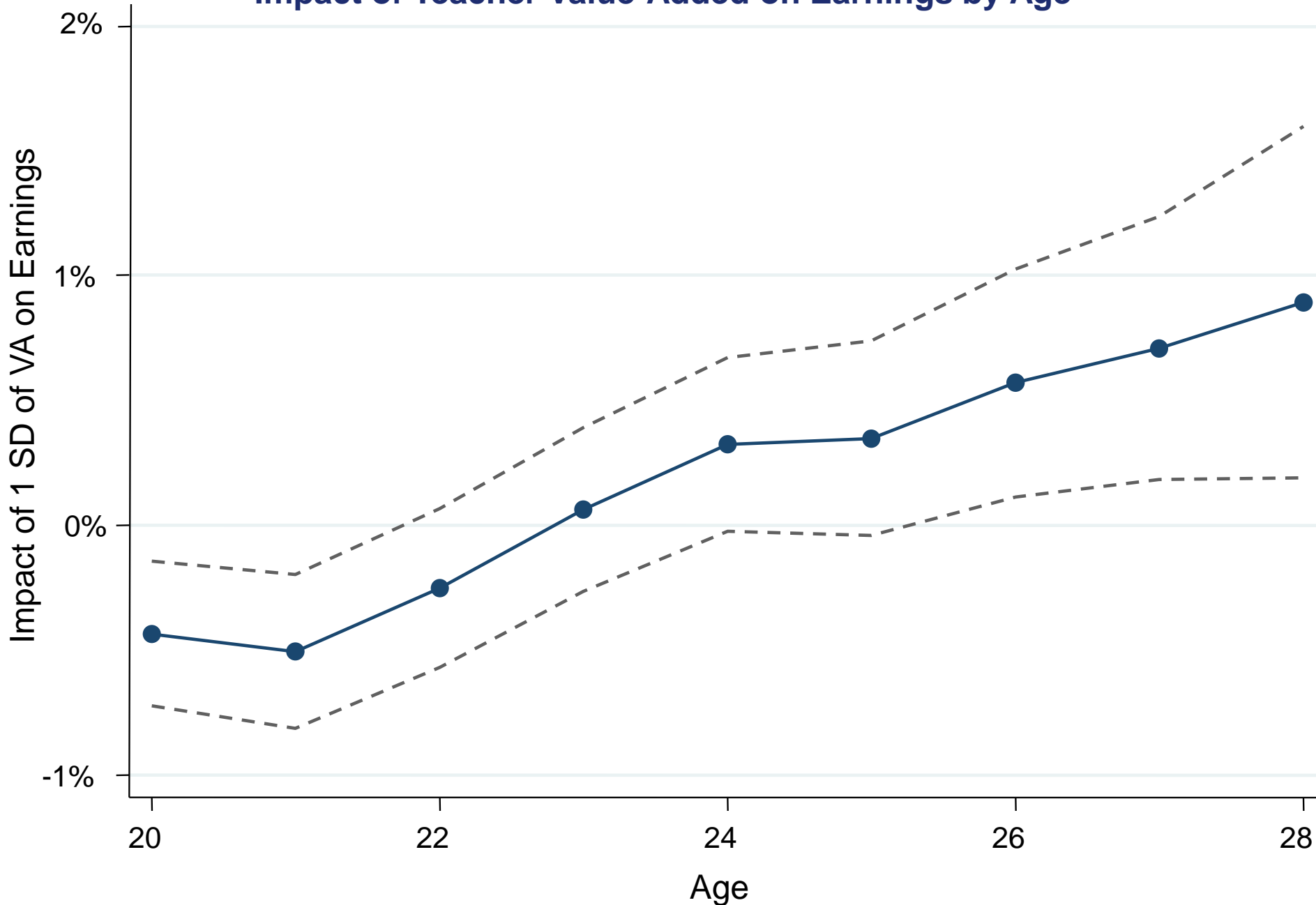
Impacts of Teacher Value-Added on College Attendance

Dependent Var:	College at Age 20	Predicted College at Age 20	Changes in College Attendance	College Quality at Age 20	Changes in College Quality	High Quality College	College at Age 25
	(%)	(%)	(%)	(\$)	(\$)	(%)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Value-Added	4.917 (0.646)	0.463 (0.261)		1,644 (173)		3.588 (0.612)	2.752 (0.697)
Changes in Mean TVA			6.101 (2.094)		1,319 (539)		
Controls	x	x		x		x	x
Source of Variation	X-Class	X-Class	X-Cohort	X-Class	X-Cohort	X-Class	X-Class
Observations	3,095,822	3,097,322	25,073	3,095,822	24,296	3,095,822	985,500
Mean of Dep. Variable	37.8	37.8	35.9	24,815	24,293	19.8	18.1

Earnings at Age 28 vs. Teacher Value-Added

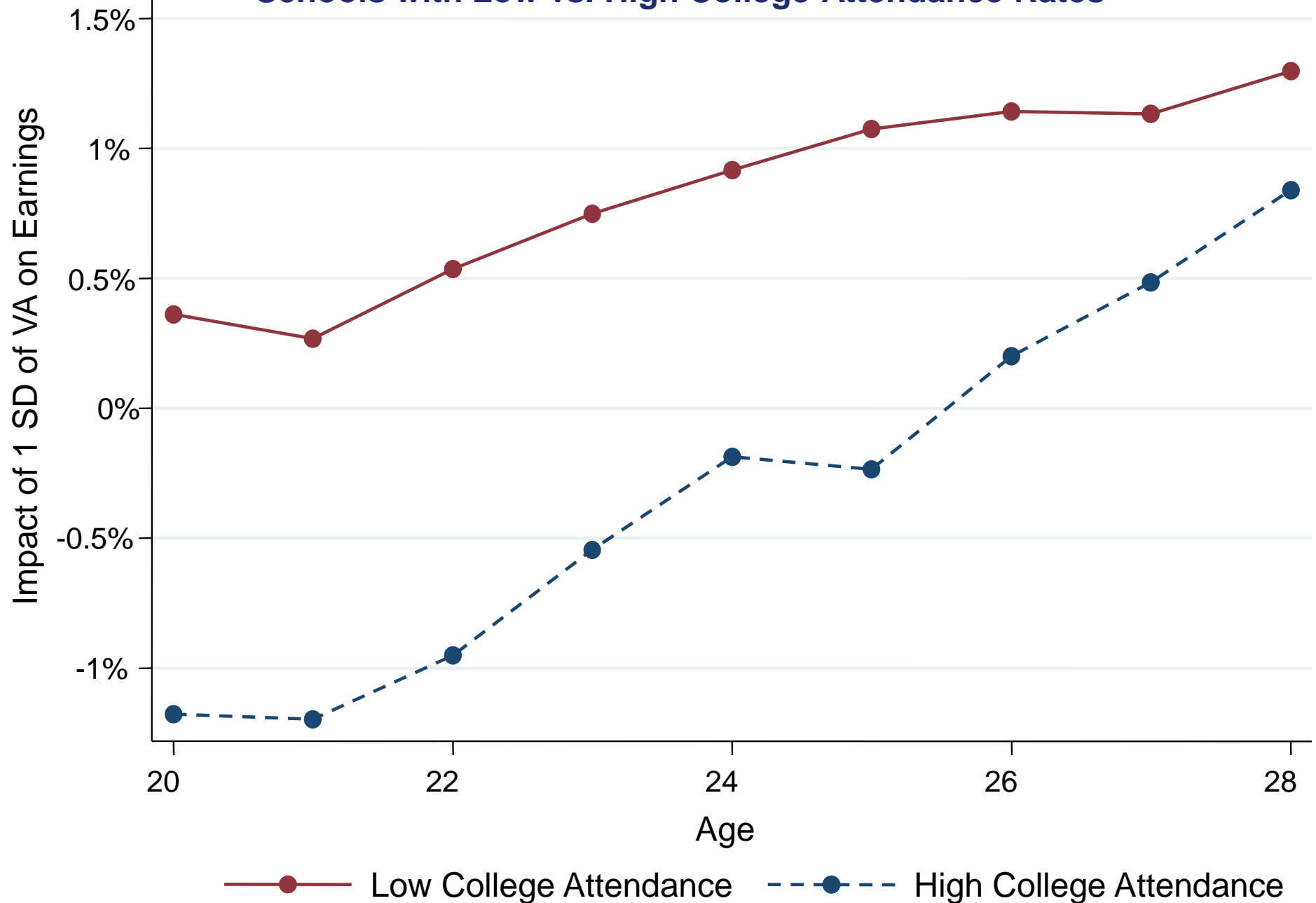


Impact of Teacher Value-Added on Earnings by Age



—●— Impact of 1 SD Increase in Teacher VA on Earnings - - - - - 95% CI

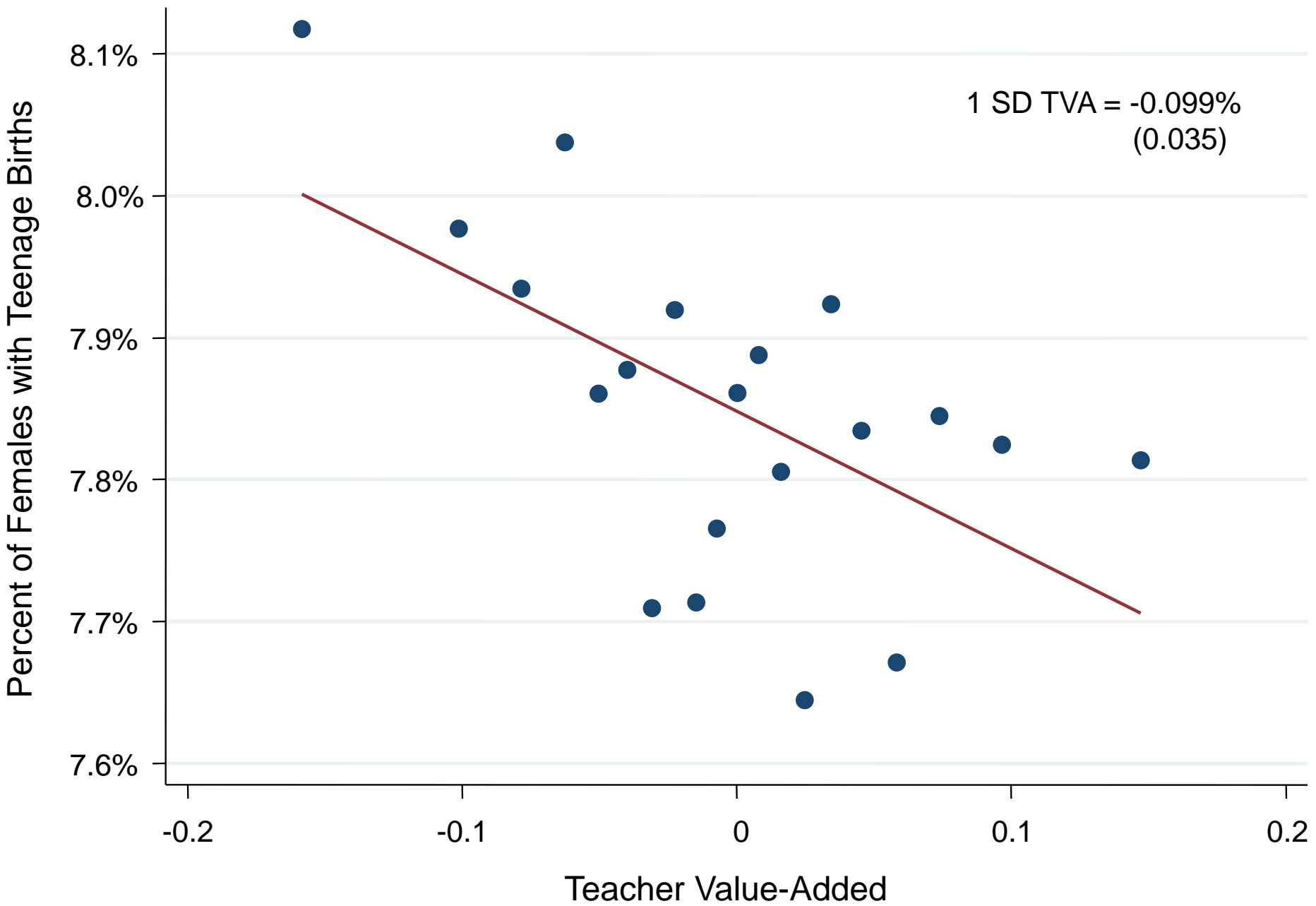
Impact of Teacher Value-Added on Earnings by Age Schools with Low vs. High College-Attendance Rates



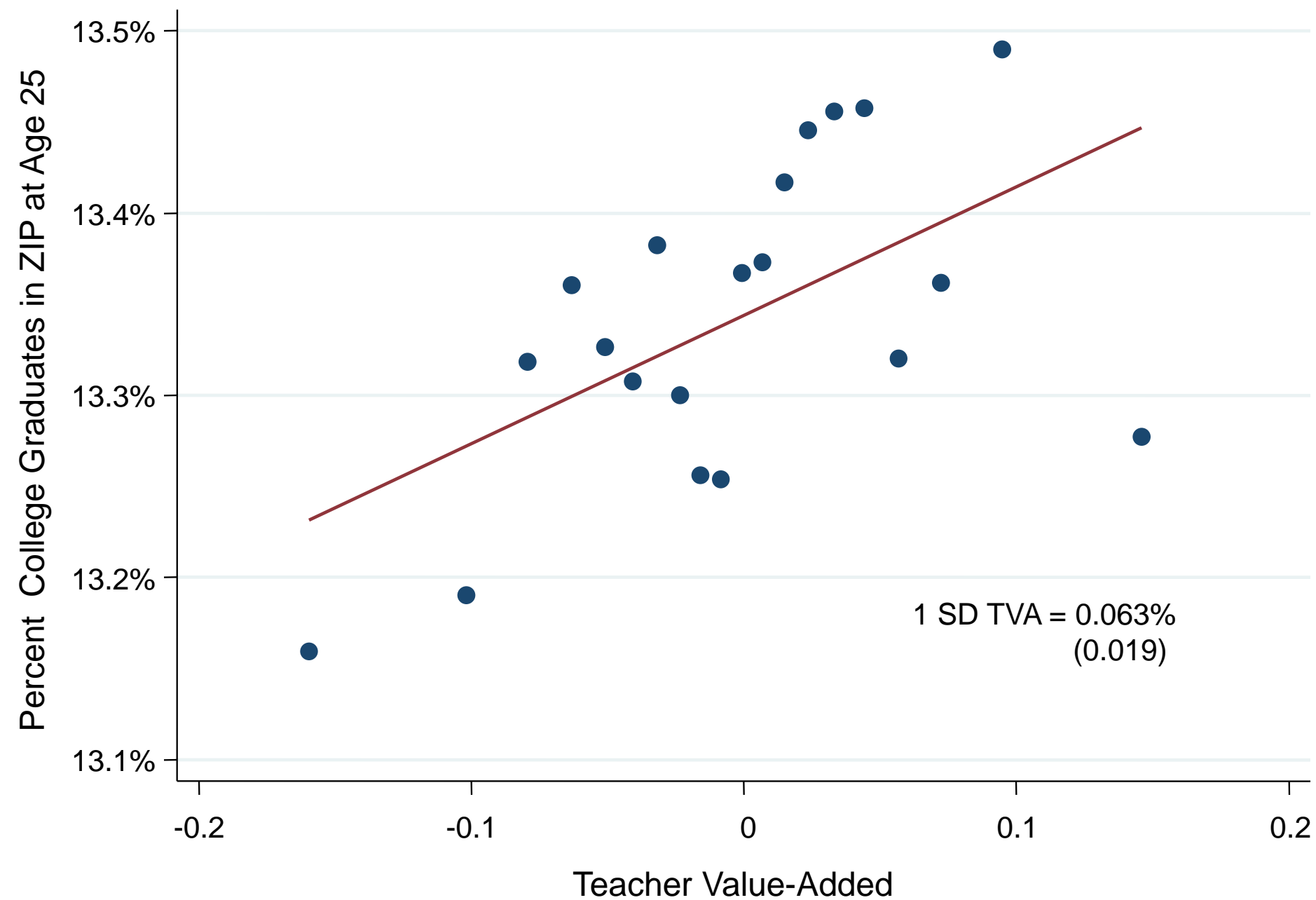
Impacts of Teacher Value-Added on Earnings

Dep. Variable:	Wage Growth Ages 22-28 (\$) (3)	College at Age 25 (%) (4)	College at Age 25 (%) (5)	Wage Growth Ages 22-28 (\$) (6)	Wage Growth Ages 22-28 (\$) (7)
Teacher VA	1,802 (636)	0.526 (0.789)	4.728 (1.152)	1,403 (661)	2,838 (1,118)
Observations	368,405	528,065	457,435	201,933	166,472
Schools	All	Low Coll.	High Coll.	Low Coll.	High Coll.
Mean of Dep. Var.	14,039	14.30	22.43	10,159	18,744

Women with Teenage Births vs. Teacher Value-Added



Neighborhood Quality at Age 25 vs. Teacher Value-Added



Heterogeneity in Impacts of 1 SD of Teacher VA by Demographic Group

Dependent Variable:	College Quality at Age 20 (\$)					
	Girls (1)	Boys (2)	Low Income (3)	High Income (4)	Minority (5)	Non-Minority (6)
Value-Added	1,903 (211)	1,386 (203)	1,227 (174)	2,087 (245)	1,302 (154)	2,421 (375)
Mean College Quality	25,509	24,106	21,950	27,926	21,925	31,628

Heterogeneity in Impacts of 1 SD of Teacher VA by Subject

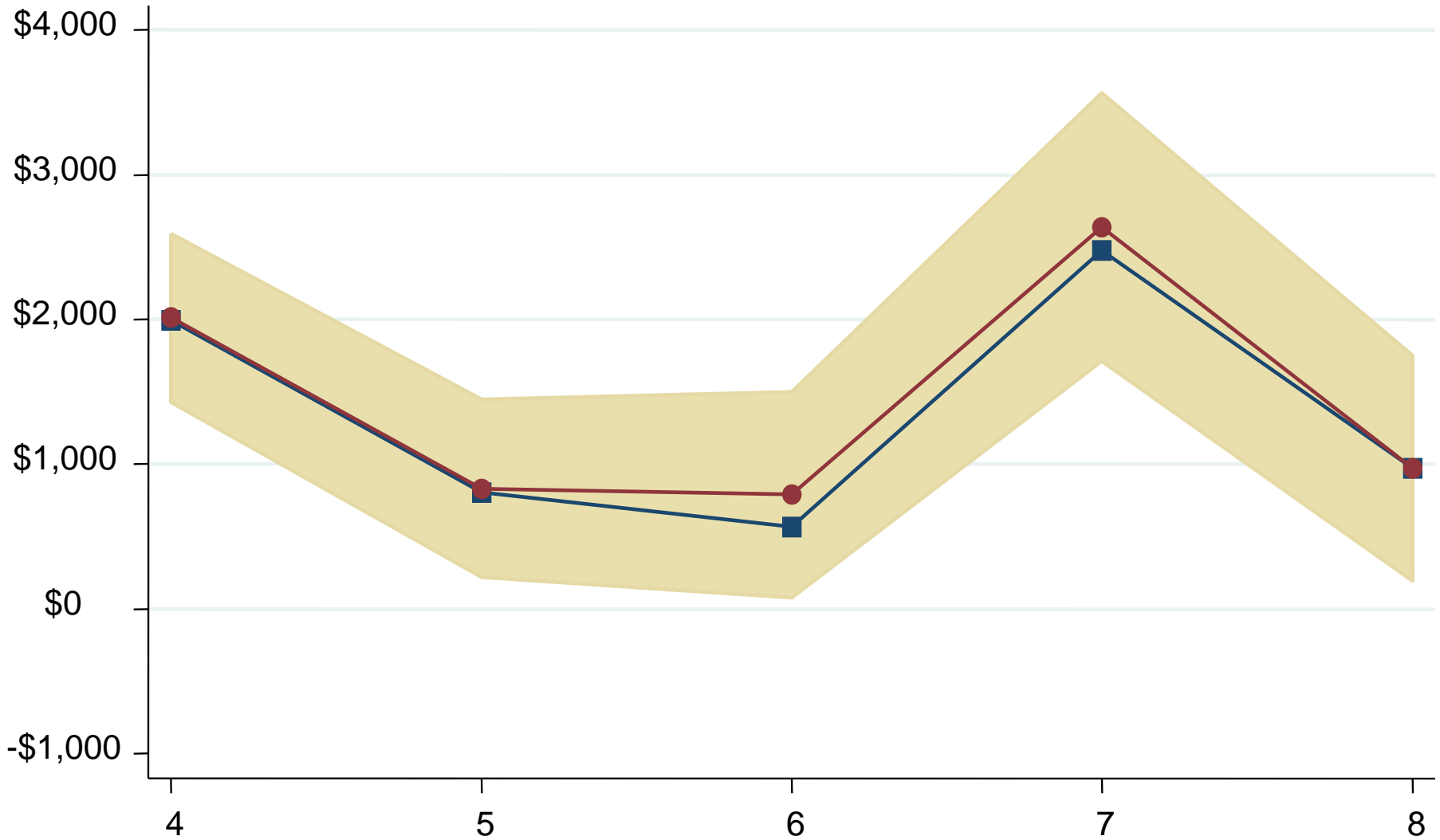
Dependent Variable:	College Quality at Age 20 (\$)					
	Elementary School			Middle School		
	(1)	(2)	(3)	(4)	(5)	(6)
Math Teacher Value-Added	1,095 (176)		638 (219)		1,648 (357)	1,374 (347)
English Teacher Value-Added		1,901 (303)	1,281 (376)		2,896 (586)	2,543 (574)

Teacher Impacts by Grade

- Reduced-form impacts of having better teachers in each grade include tracking to better teachers in future grades
- We can net-out the impact of tracking from the reduced-form coefficients by estimating tracking process
 - Estimate impact of current teacher VA on VA of future teachers
 - Subtract out impacts of future teachers

Effect of Value-Added on Earnings by Grade

Impact of 1 SD of VA on College Quality at Age 20



Grade



Reduced Form Coefficients



Net of Teacher Tracking



95% CI

Policy Implications

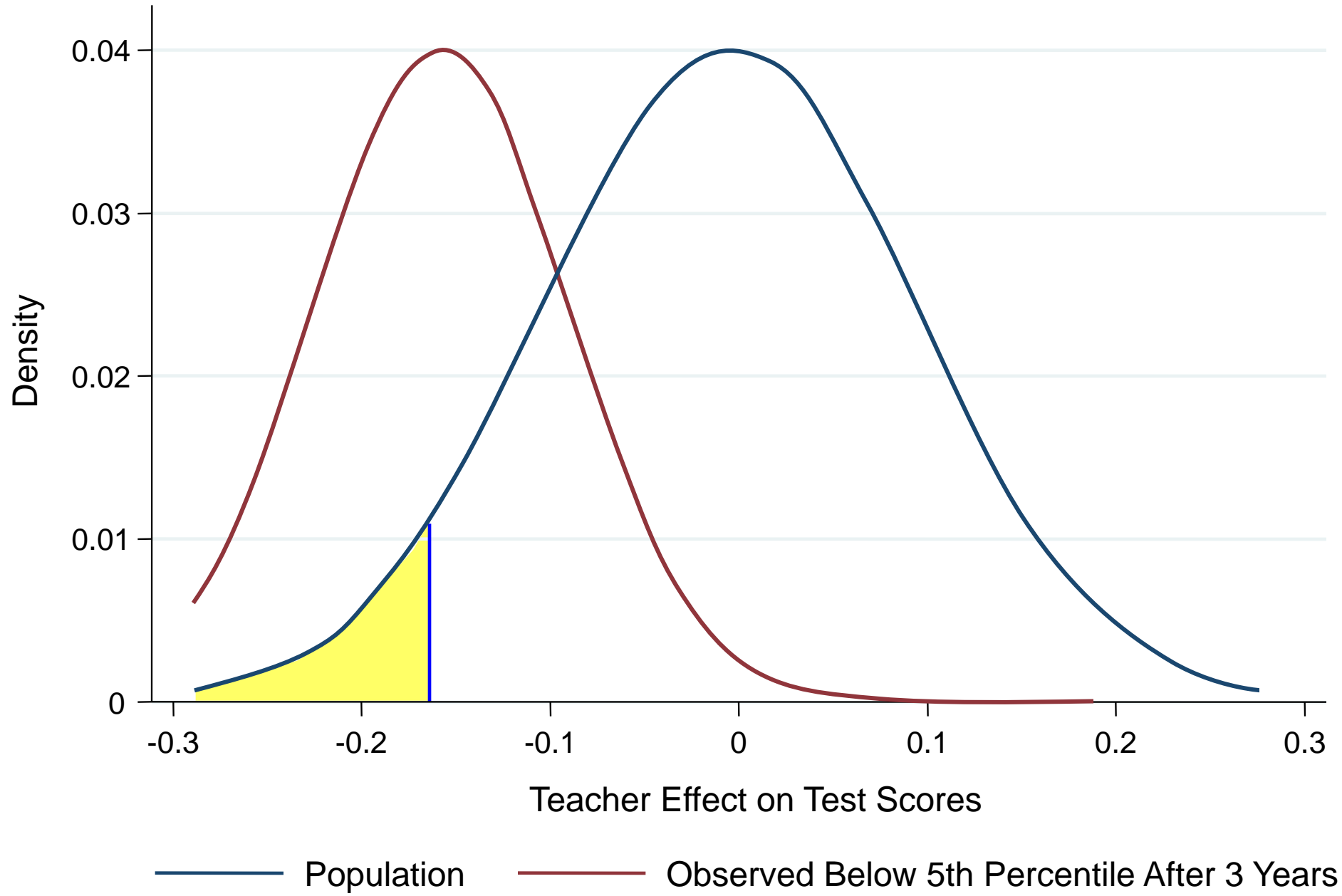
Policy Proposal 1: Deselection of Low VA Teachers

What are the gains from replacing teachers with VA in bottom 5% with teachers of median quality (Hanushek 2009)?

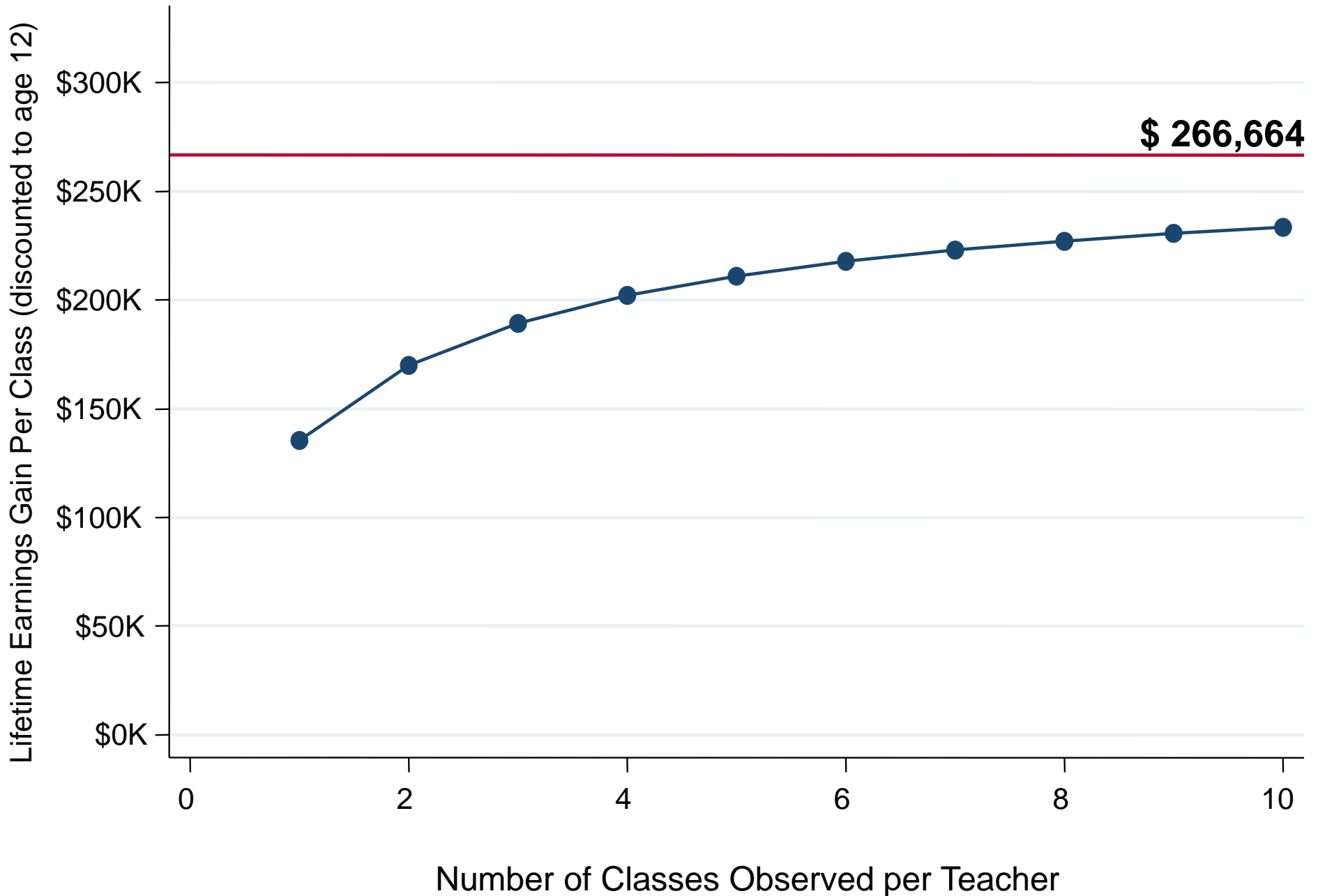
Policy Implications

- Use estimates to evaluate gains from improving teacher quality
- Measure impact of teacher quality on present value of lifetime earnings for a class of average size (28 students)
- Assumptions
 - Ignores general equilibrium effects and non-monetary gains [Oreopoulos and Salvanes 2011, Heckman 2000]
 - Constant percentage impact on earnings over life
 - Life-cycle earnings follows cross-sectional life-cycle path in 2010
 - 2% wage growth with 5% discount rate back to age 12
 - Undiscounted lifetime earnings gains are roughly 5 times larger

Deselecting Teachers on the Basis of Value-Added



Present Value Earnings Gain from Deselecting Teachers Below 5th Percentile



Policy Implications

Policy Proposal 2: Retention of High VA Teachers

What are the gains from increasing retention of high value-added teachers by paying salary bonuses?

Gains from Retaining High VA Teachers

- Retaining a teacher whose VA is at the 95th percentile (based on 3 years of data) for an extra year would generate PV earnings gain of **\$192K**
- Clotfelter et al. (2008) analyze impacts of bonus payments to teachers
 - \$2,000 bonus would raise teacher retention by 1.5 percentage points → earnings gain of **\$3,200**
- Net return relatively small because most of the bonus payments go to teachers who would not have left anyway
 - Have to pay bonuses to 60 teachers to keep retain 1 teacher on average

Policy Calculations

- Further work needed to assess value-added as a policy tool
 - Using VA measures in high-stakes evaluation could induce negative behavioral responses such as teaching to the test or cheating
 - Errors in personnel decisions must be weighed against mean benefits
- Results highlight large potential returns from developing policies to improve teacher quality
- From a purely financial perspective, high income (~\$100K) parents should be willing to pay about **\$6,500/year** to get a 1 SD better teacher