

How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR

Raj Chetty, Harvard
John N. Friedman, Harvard
Nathaniel Hilger, Harvard
Emmanuel Saez, UC Berkeley
Diane Schanzenbach, Northwestern
Danny Yagan, Harvard

February 2011

Introduction

- What are the long-term impacts of early childhood education?
- Limited evidence because few datasets link information on childhood education with adult outcomes
- We link data from the STAR experiment to U.S. tax records to analyze how class assignment in grades K-3 affects adult outcomes

Project STAR Background

- Student/Teacher Achievement Ratio (STAR) experiment:
 - Conducted from 1985 to 1989 in Tennessee
 - One cohort of 11,571 children in grades K-3 at 79 schools
 - Most children born in 1979-80 → graduate high school in 1998
- Students and teachers randomized into classrooms within schools
 - Class size differs: small (15 students) or large (22 students)
 - Classes also differ in teachers and peers
 - Randomized into classroom upon entry into participating school and kept in small/large track from grades K-3
- Only one cohort treated → no repeat teacher observations

Project STAR Background

- Large literature on STAR shows that class size, teacher quality, and peer quality have causal impacts on scores
 - Students in small classes have 5 percentile point (0.2 sd) higher test scores in K-3 (Krueger 1999)
 - But test score gains fade out to 1-2 percentiles by grade 8
 - Similar fade out effects observed in other early childhood interventions (e.g. Currie and Thomas 1995, Deming 2009)
- Do early test score gains translate into impacts on adult outcomes?

United States Tax Data

- Dataset covers full U.S. population from 1996-2008
- Approximately 90% of working age adults file tax returns
- Third-party reports yield data on many outcomes even for non-filers
 - Employer and wage earnings from W-2 forms
 - College attendance from 1098-T forms
- 95% of STAR records were linked to tax data

Table 1: Summary Statistics

	STAR Sample (1)	U.S. 1979-80 cohort (2)
Mean Wage Earnings (2005-07)	\$15,912	\$20,500
Zero Wage Earnings (2005-07)	13.9%	15.6%
Attended College in 2000 (age 20)	26.4%	34.7%
Attended College by age 27	45.5%	57.1%
Mean Parents' Income (1996-98)	\$48,010	\$65,660

Outline

1. Test scores and adult outcomes in the cross-section
2. Impacts of observable classroom characteristics
3. Impacts of unobservable classroom characteristics
4. Fade-out, re-emergence, and non-cognitive skills
5. Cost-benefit analysis

Part 1: Cross-Sectional Correlations

- Begin by correlating KG test scores with adult outcomes
 - Useful to benchmark estimates from randomized interventions
- Estimate both raw correlations and OLS regressions with controls:
 - quartic in parental household income interacted with marital status
 - mother age at child's birth
 - parent's 401K contributions, home ownership
 - child's gender, free lunch status, race, and age
- Test score: percentile score on Kindergarten Stanford Achievement Test (math + reading)

What is a kindergarten test?

- Instructions:
 - I'll say a word to you. Listen for the *ending* sound.
 - You circle the picture that *starts* with the same sound.

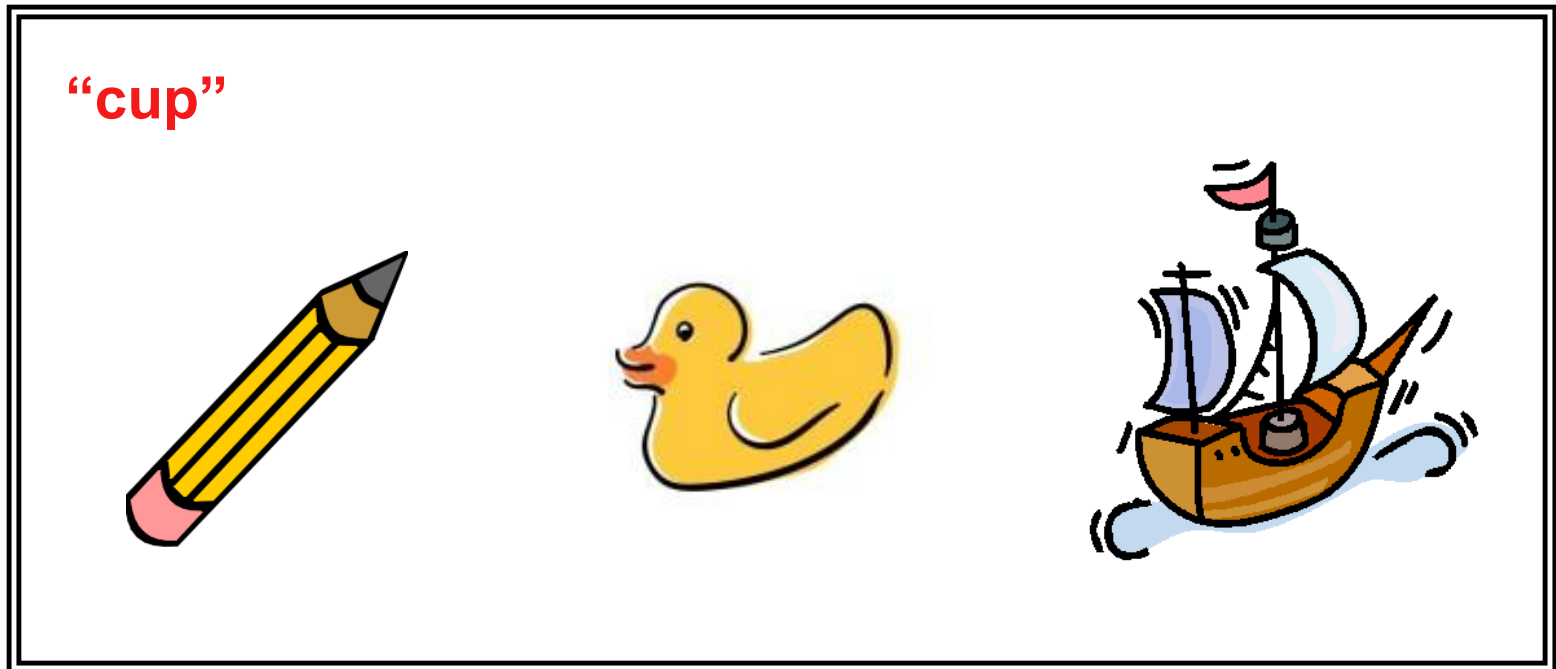
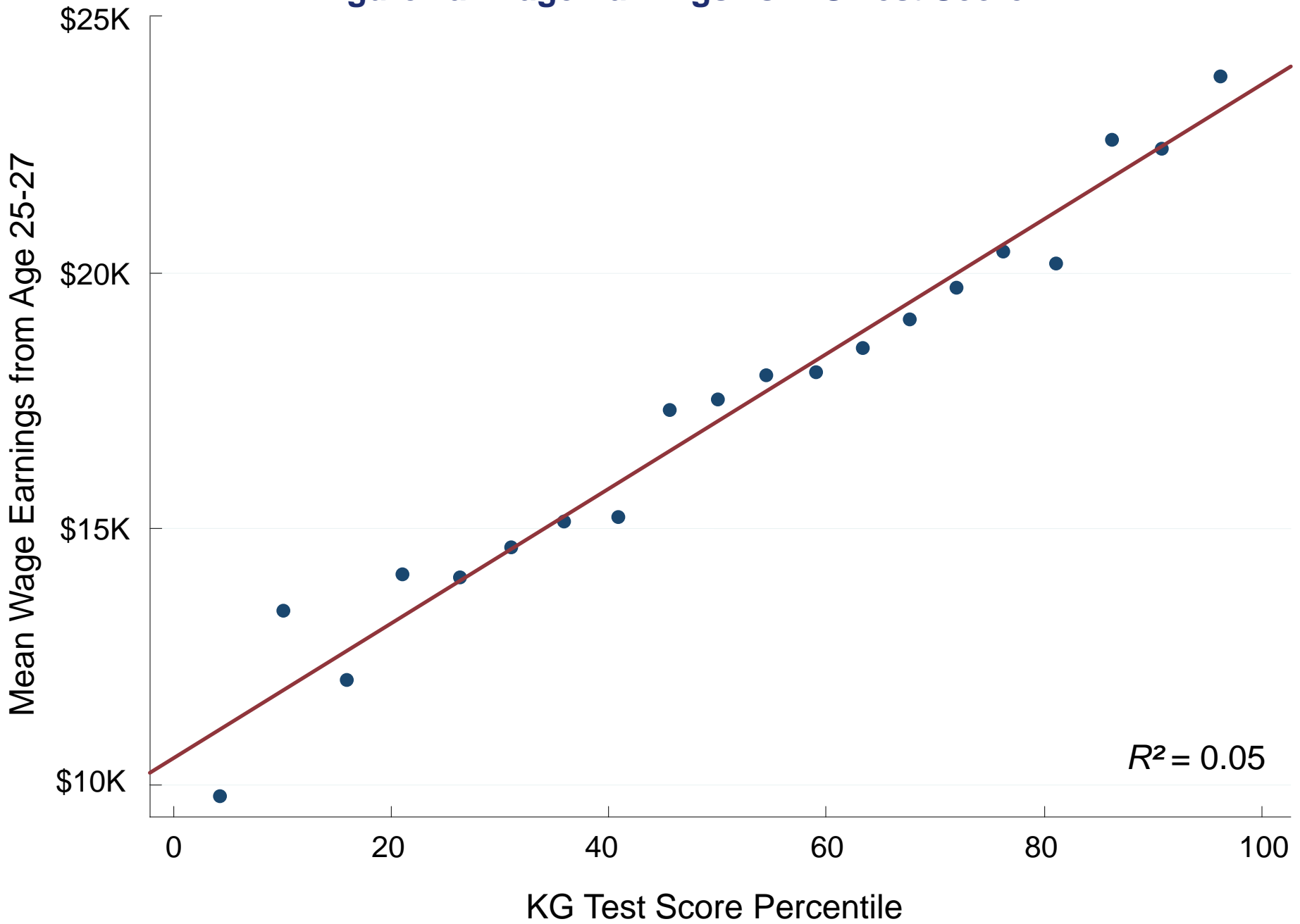


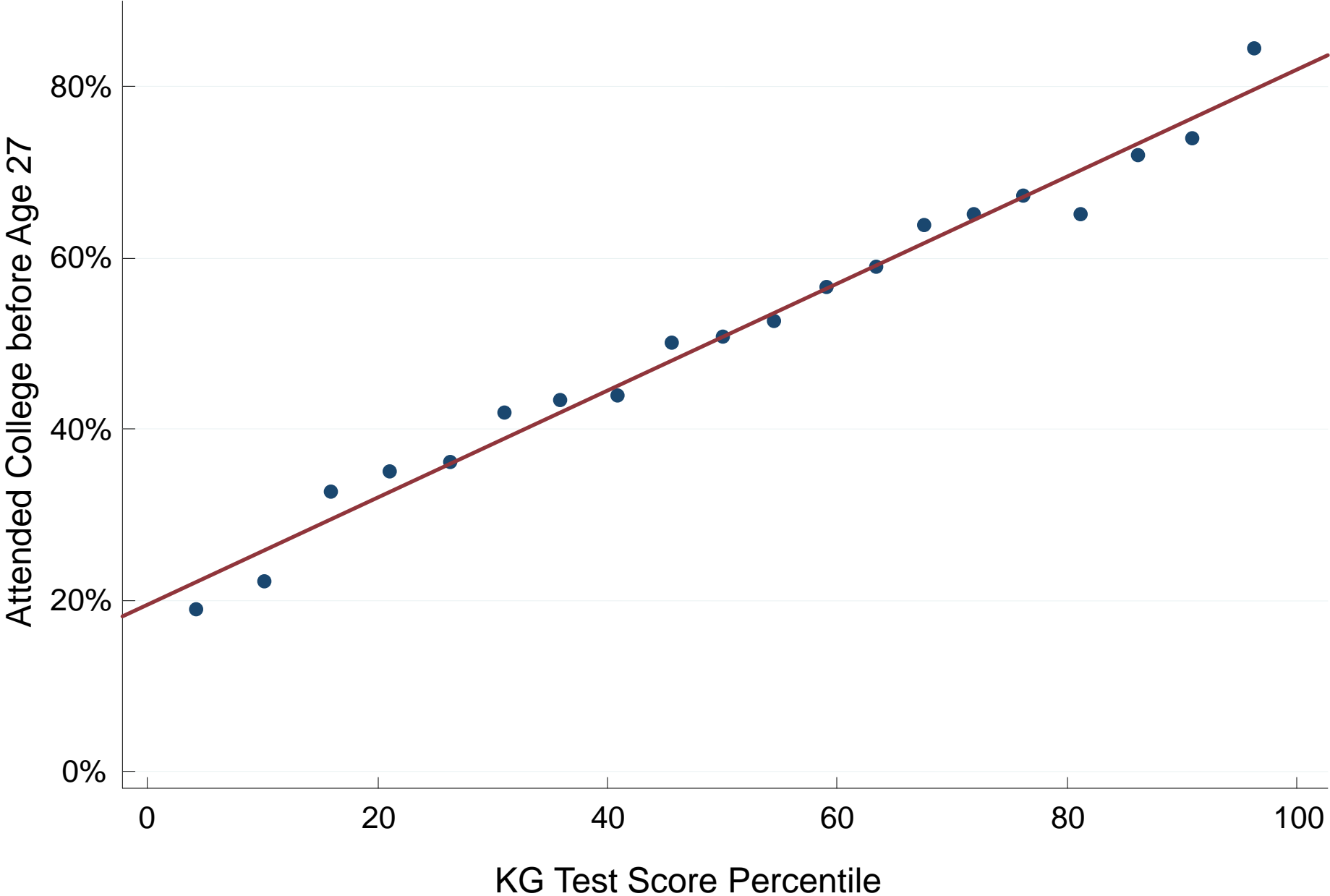
Figure 1a: Wage Earnings vs. KG Test Score



Test Scores and Earnings in the Cross-Section

Dependent Var.:	Wage Earnings			
	(1)	(2)	(3)	(4)
KG Test Percentile	\$132 (\$12.2)	\$93.8 (\$11.6)	\$90.0 (\$8.65)	\$97.7 (\$8.47)
Parental Income Percentile				\$146 (\$8.15)
Entry Grade	KG	KG	All	All
Class Fixed Effects		x	x	x
Student Controls		x	x	x
Parent Controls		x	x	
Adjusted R ²	0.05	0.17	0.17	0.16
Observations	5,621	5,621	9,939	9,939

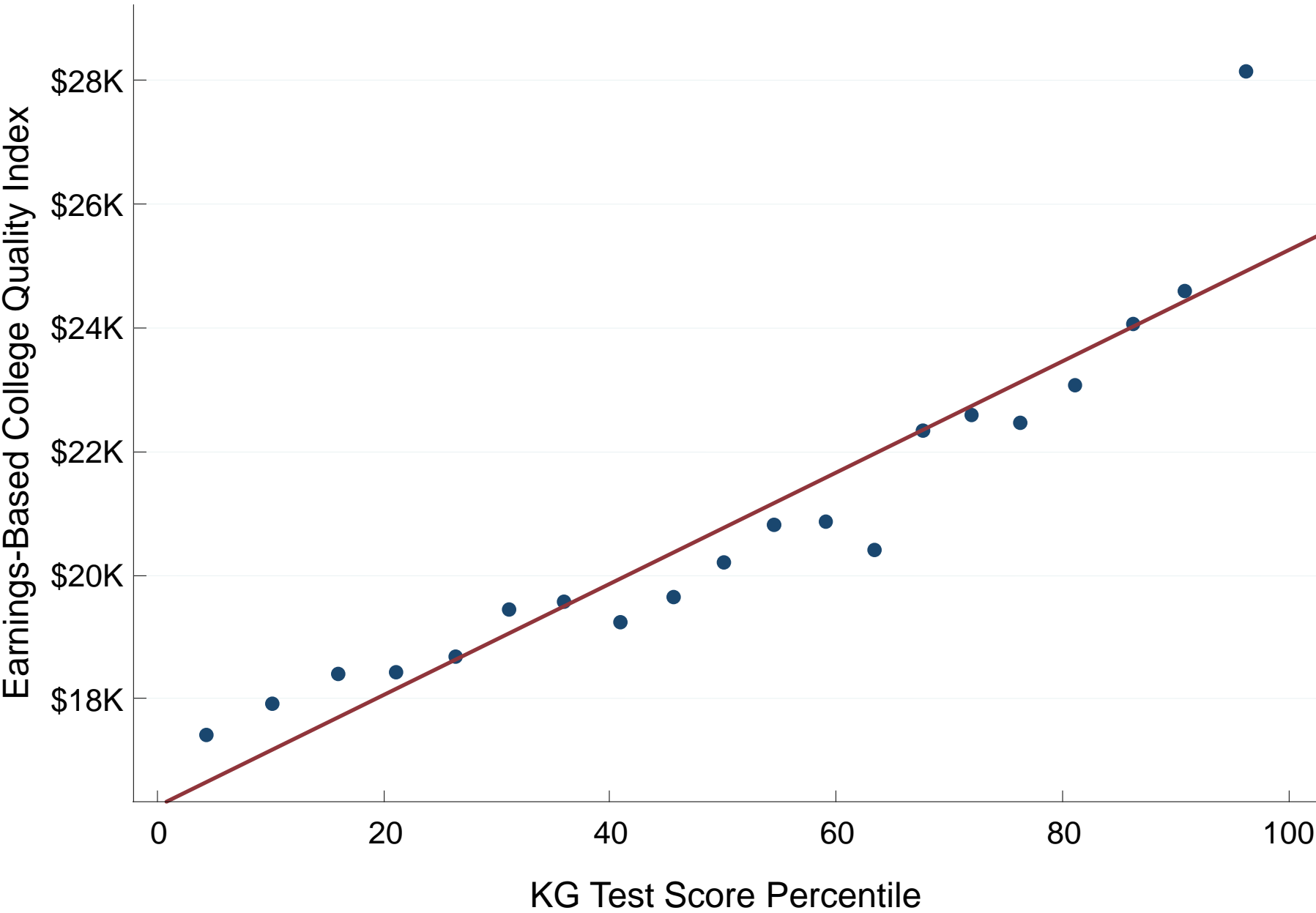
Figure 1b: College Attendance Rates vs. KG Test Score



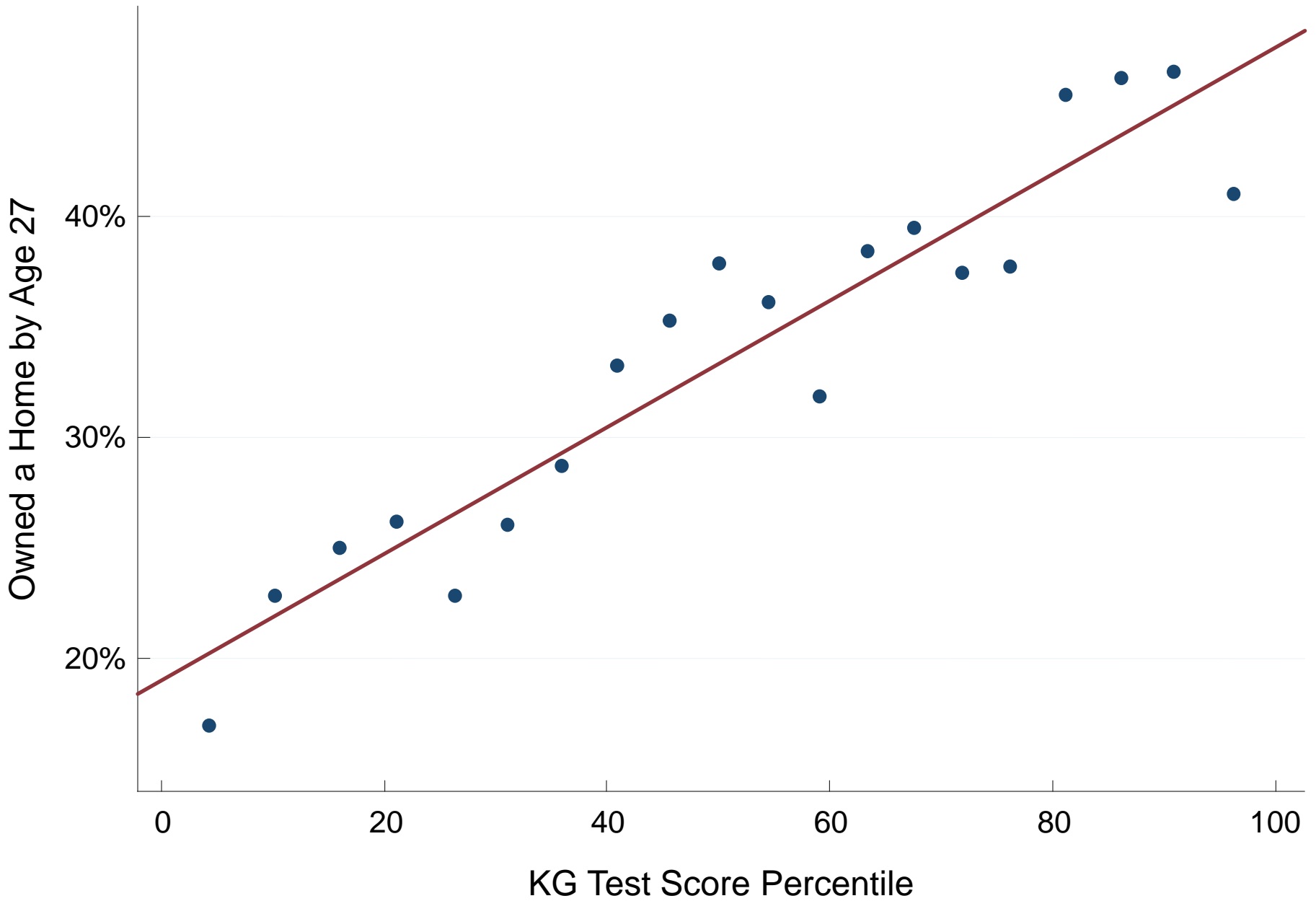
An Earnings-Based Index of College Quality

- We construct an index of college quality using tax data
- Tuition paid to any higher ed. institution (Title IV) automatically generates a 1098-T form linking student and institution
- Calculate average wage earnings (from W-2s) by college
- For those who do not attend college, define college quality index as mean earnings for those not in college in 1999

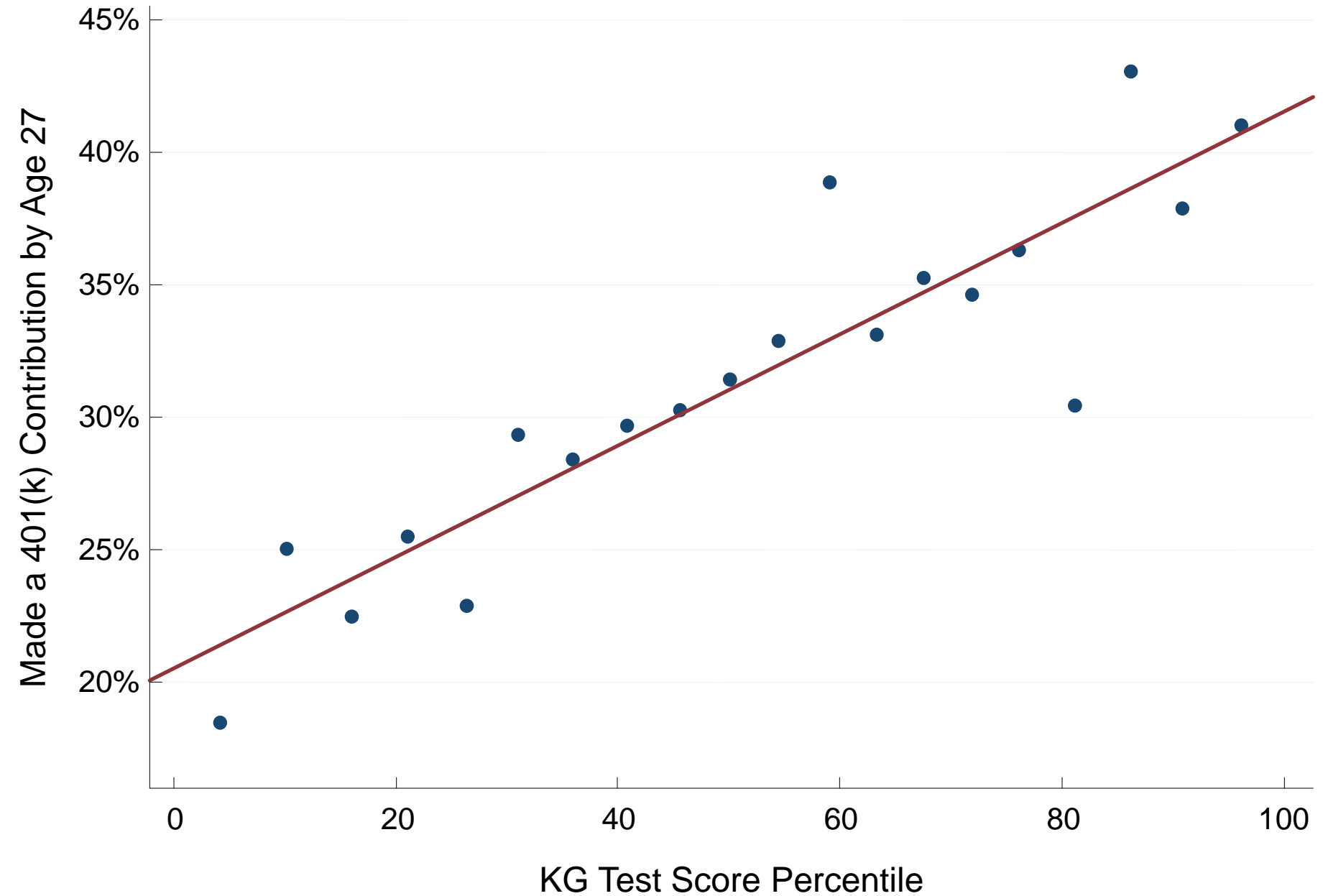
College Quality vs. KG Test Score



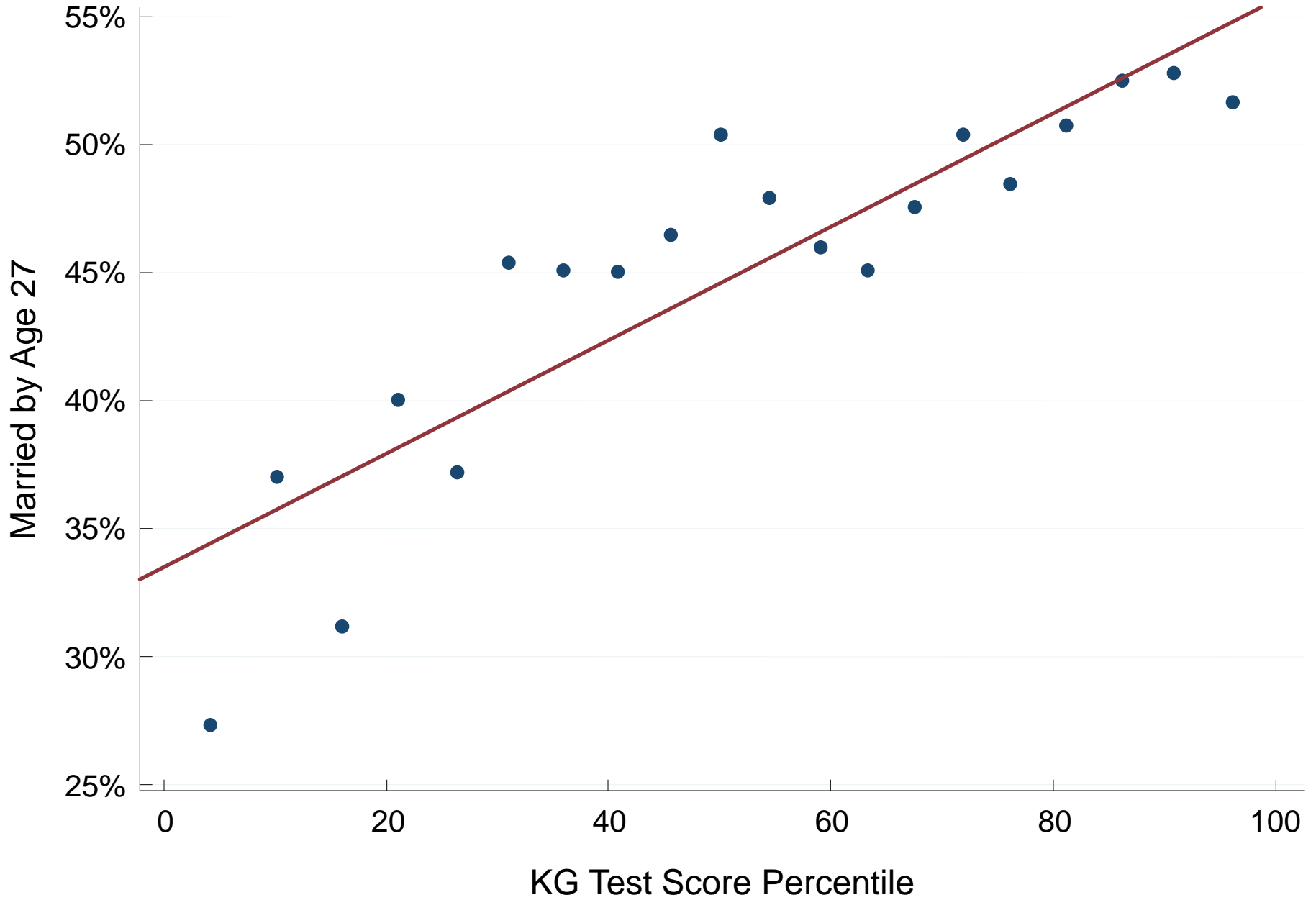
Home Ownership vs. KG Test Score



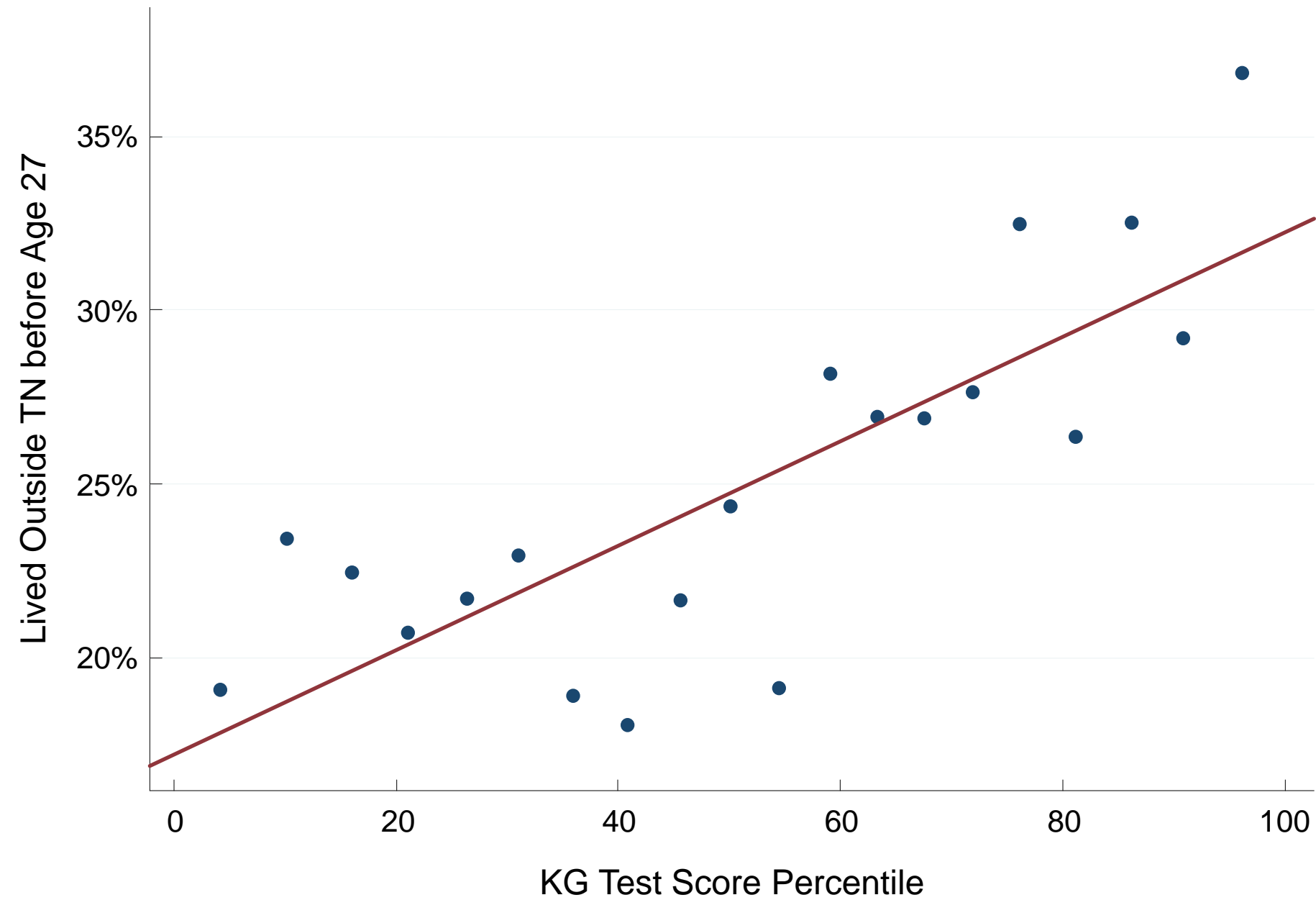
Retirement Savings vs. KG Test Score



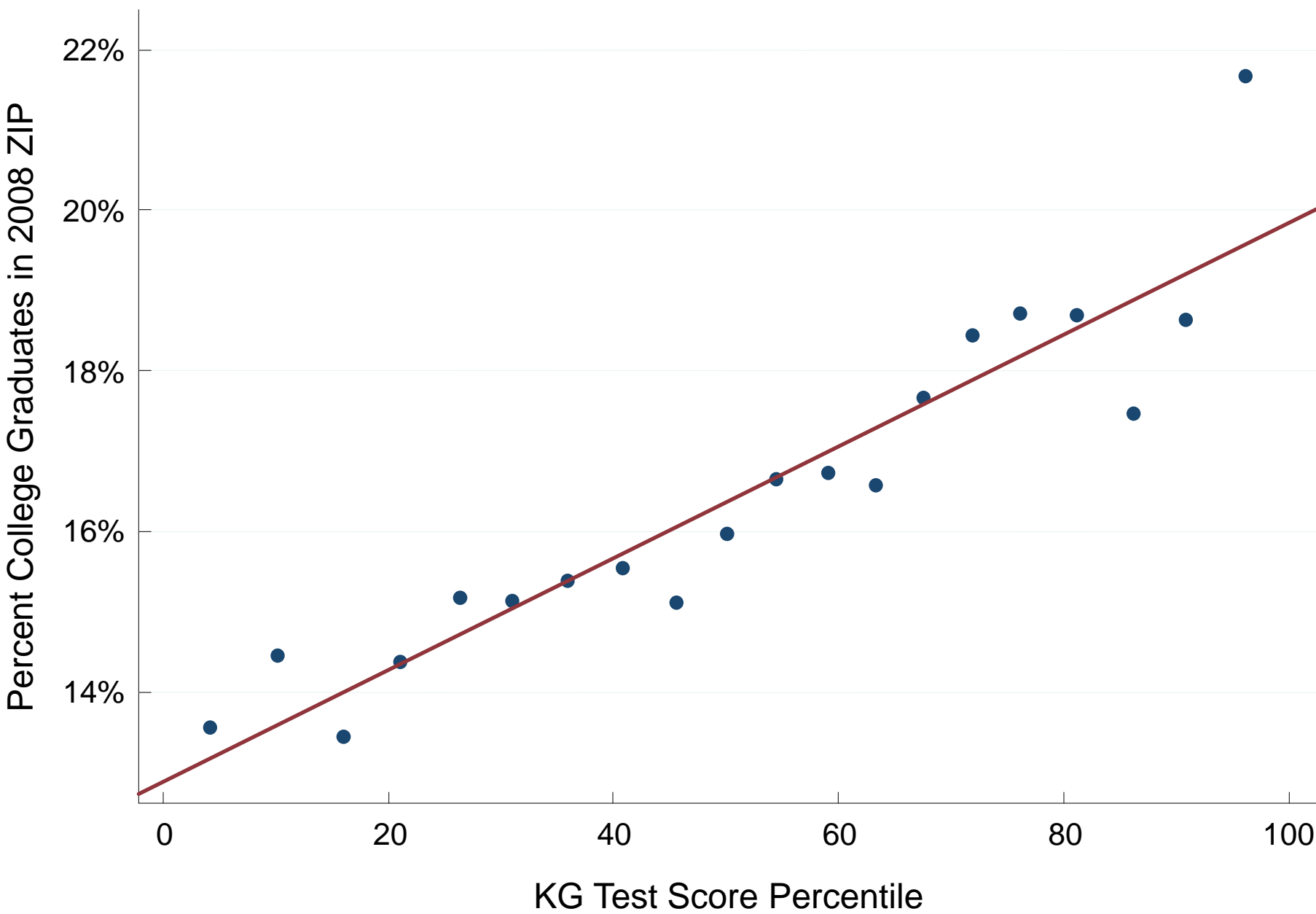
Marriage by Age 27 vs. KG Test Score



Cross-State Mobility vs. KG Test Score



Percent College Graduates in ZIP code vs. KG Test Score



Part 2: Validity of the STAR Experimental Design

- Validity of experimental analysis rests on two assumptions:
- Assumption 1: *Randomization*
 - All pre-determined variables (e.g. parent characteristics) are balanced across classrooms
- Assumption 2: *No Differential Attrition*
 - 95% match rate → little attrition here
 - No evidence of differences in match rates across classrooms
 - No evidence of differences in death rates across classrooms

Part 2: Validity of the STAR Experiment Design

- Threat #1: *Failure of Randomization*
 - Prior studies had few baseline measures, limiting ability to evaluate randomization protocol (Schanzenbach 2006)
- We test for balance across class types with an expanded set of parent/sibling characteristics in two ways:
 1. Do characteristics vary across small vs. large class types?
 2. Do characteristics vary across classrooms within schools?

Table 2: Randomization Tests

Dependent Variable: Wage Earnings (%) (1)	Small Class (%) (2)	Teacher Exp. (%) (3)	Class Effects p-value (4)	
Parent's Income (\$1,000s)	65.47 (6.634) [9.87]	-0.003 (0.015) [-0.231]	-0.001 (0.002) [-0.509]	0.848
Mother's Age at STAR Birth	53.96 (24.95) [2.162]	0.029 (0.076) [0.384]	0.022 (0.012) [1.863]	0.654
Parents Have 401(k)	2273 (348.3) [6.526]	1.455 (1.063) [1.368]	0.111 (0.146) [0.761]	0.501
Student Female	-2317 (425.0) [-5.451]	-0.226 (0.864) [-0.261]	0.236 (0.111) [2.129]	0.502
Student Black	-620.8 (492.0) [-1.262]	0.204 (1.449) [0.141]	0.432 (0.207) [2.089]	0.995
p-Value of F Test	0.000	0.261	0.190	
Observations	10,992	10,992	10,914	

Note: Regressions include school-by-entry-grade fixed effects.

Validity of the STAR Experiment Design

- Threat #2: *Selective Attrition*
- Much less attrition than in prior studies of STAR because we follow 95% of the sample
- Test for selective attrition through two channels:
 1. Does match rate vary across treatment groups?
 2. Does death rate vary across treatment groups (Muennig et al. 2010)?

Table 3: Tests for Selective Attrition

Dependent Variable:	Matched		Deceased	
	(%)	(%)	(%)	(%)
	(1)	(2)	(3)	(4)
Small Class	-0.019 (0.467)	0.079 (0.407)	-0.010 (0.286)	-0.006 (0.286)
p Value on F test on Class Effects	0.951	0.888	0.388	0.382
Demographic Controls		x		x
Mean of dep. Var.	95.0	95.0	1.70	1.70
Observations	11,571	11,571	10,992	10,992

Part 3: Class Size Impacts

- Regress outcomes on dummy for small class assignment (intent to treat) with school fixed effects
- Analyze impacts on four outcomes:
 1. College attendance
 2. College quality index
 3. Mean earnings (ages 25-27)
 4. Standardized (SD = 1) summary index of other outcomes:

$$\text{Index} = 401K + \text{Home Owner} + \text{Married} + \\ \text{Moved out of TN} + \text{Pct. College Grads. in Zip}$$

Figure 2a: Effect of Class Size on College Attendance by Year

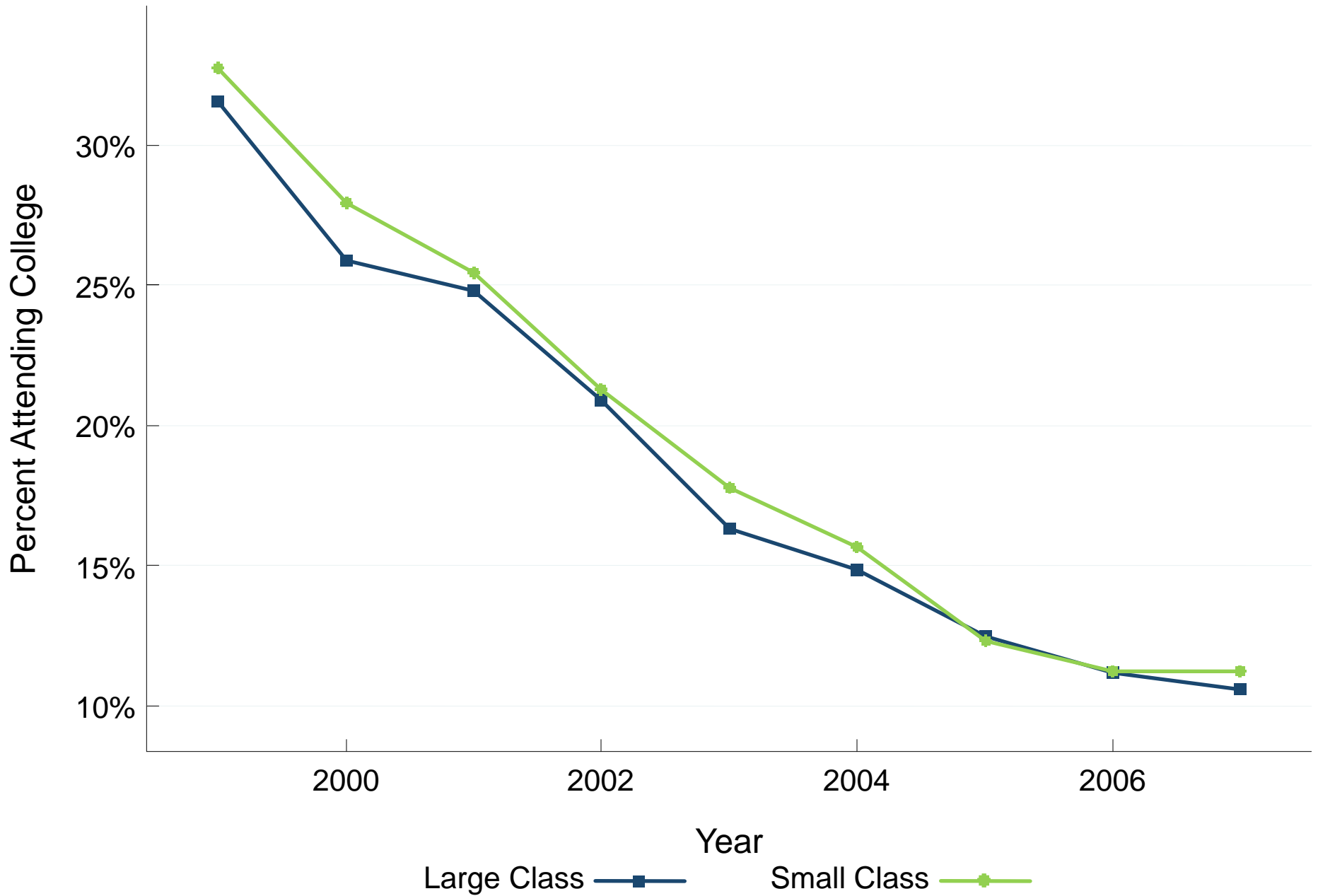


Figure 2b: College Earnings Quality by Class Size

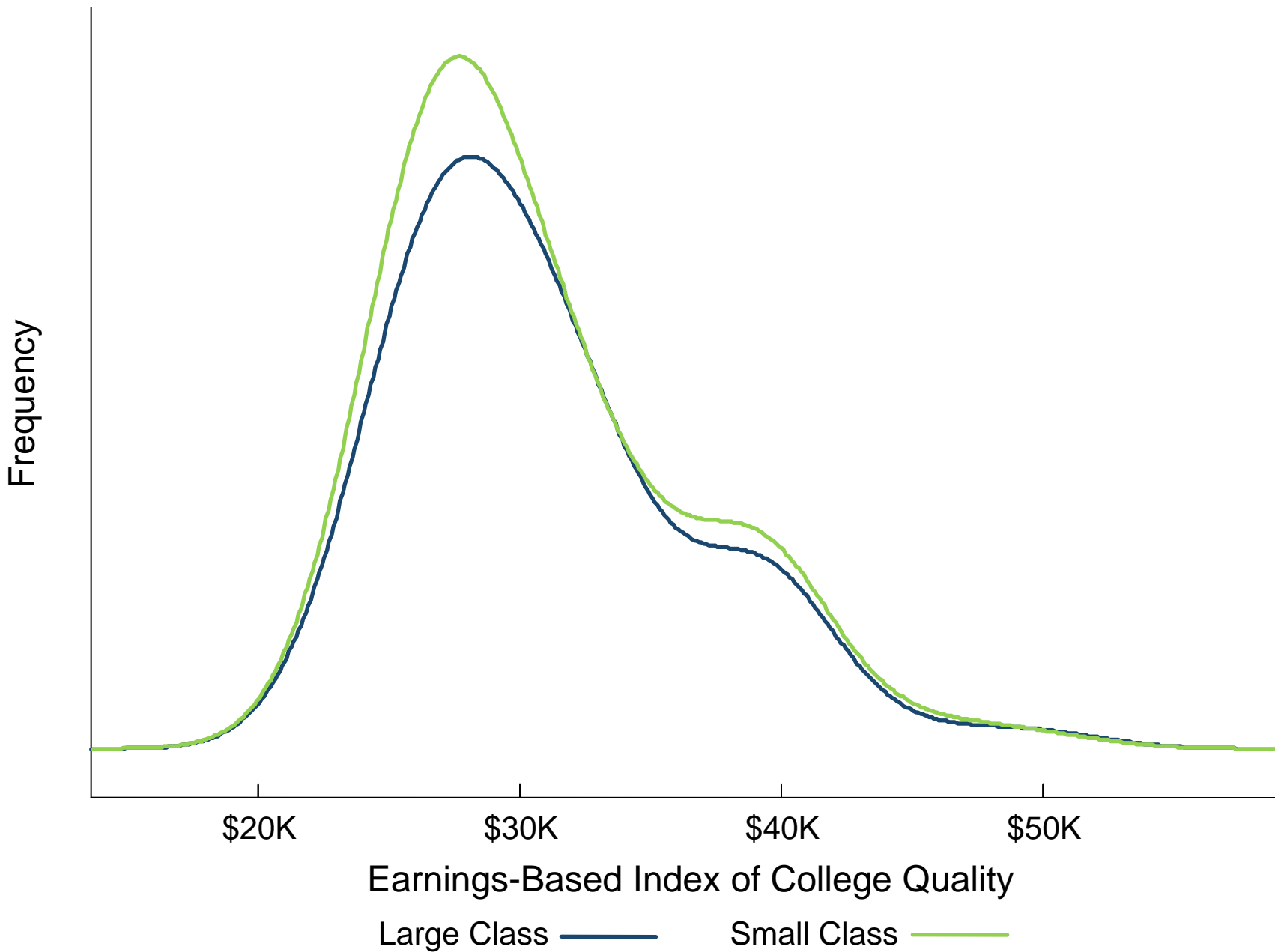


Figure 2c: Effect of Class Size on Wage Earnings by Year

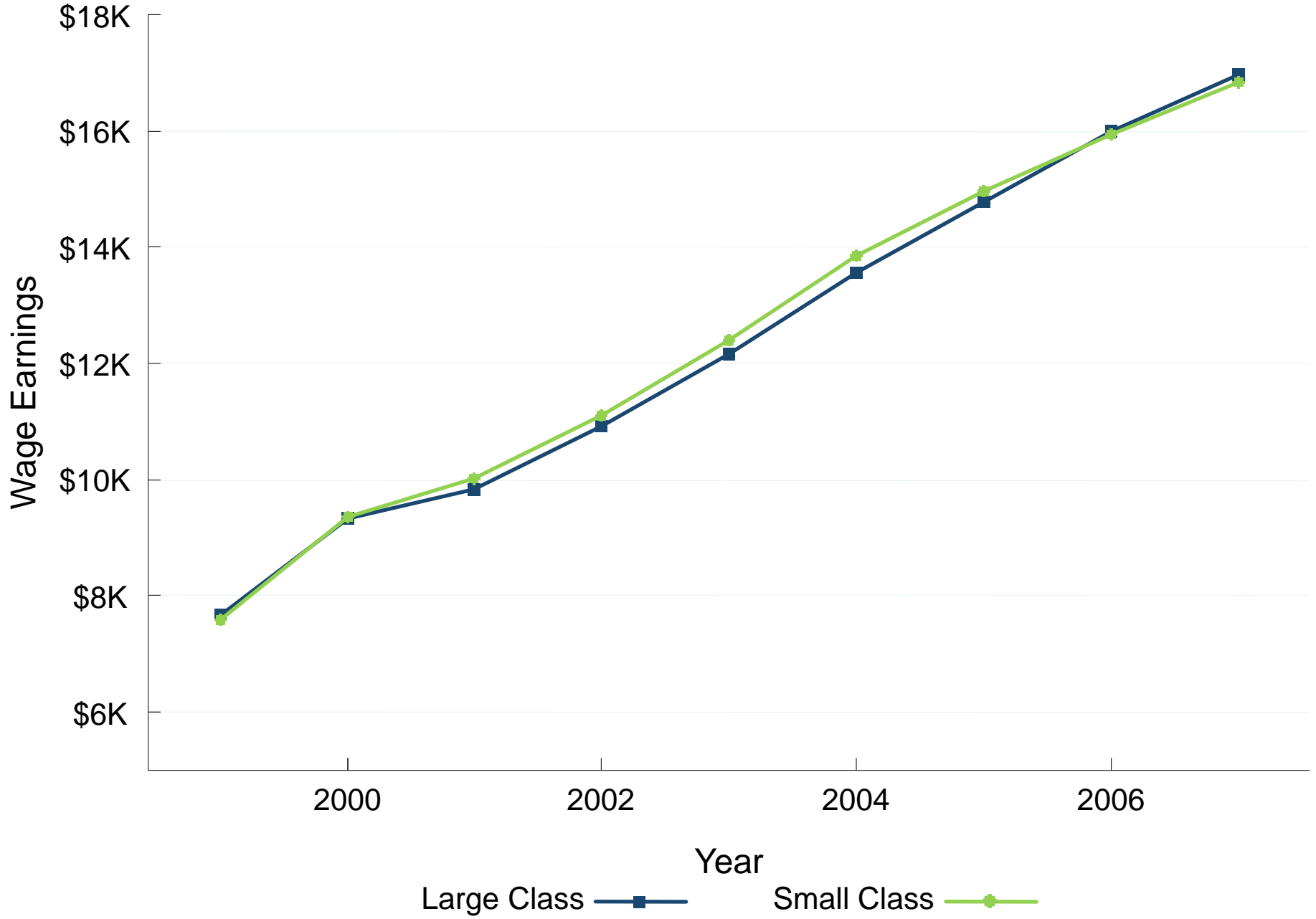


Table 5: Impacts of Class Size on Adult Outcomes

Dependent Var.:	College In 2000 (1)	College Quality (2)	Wage Earnings (3)	Summary Index (4)
Small Class	2.02% (1.10%)	\$119 (\$97)	\$4 (\$327)	5.06% (2.16%)
Observations	10,992	10,992	10,992	10,992
Mean of Dep. Var.	26.4%	\$27,115	\$15,912	0.00

Note: All specifications control for school-by-entry-grade effects.

Part 3: Teacher/Peer Effects

- Students randomly assigned to classes that differ in teacher and peer quality
 - Do teachers/peers affect adult outcomes?
- First test: does random assignment to a more experienced KG teacher improve adult outcomes?
 - Not necessarily causal effect of raising teacher experience *per se*
 - Experienced teachers may also differ on other dimensions such as dedication to teaching

Figure 3a: Effect of Teacher Experience on Test Scores

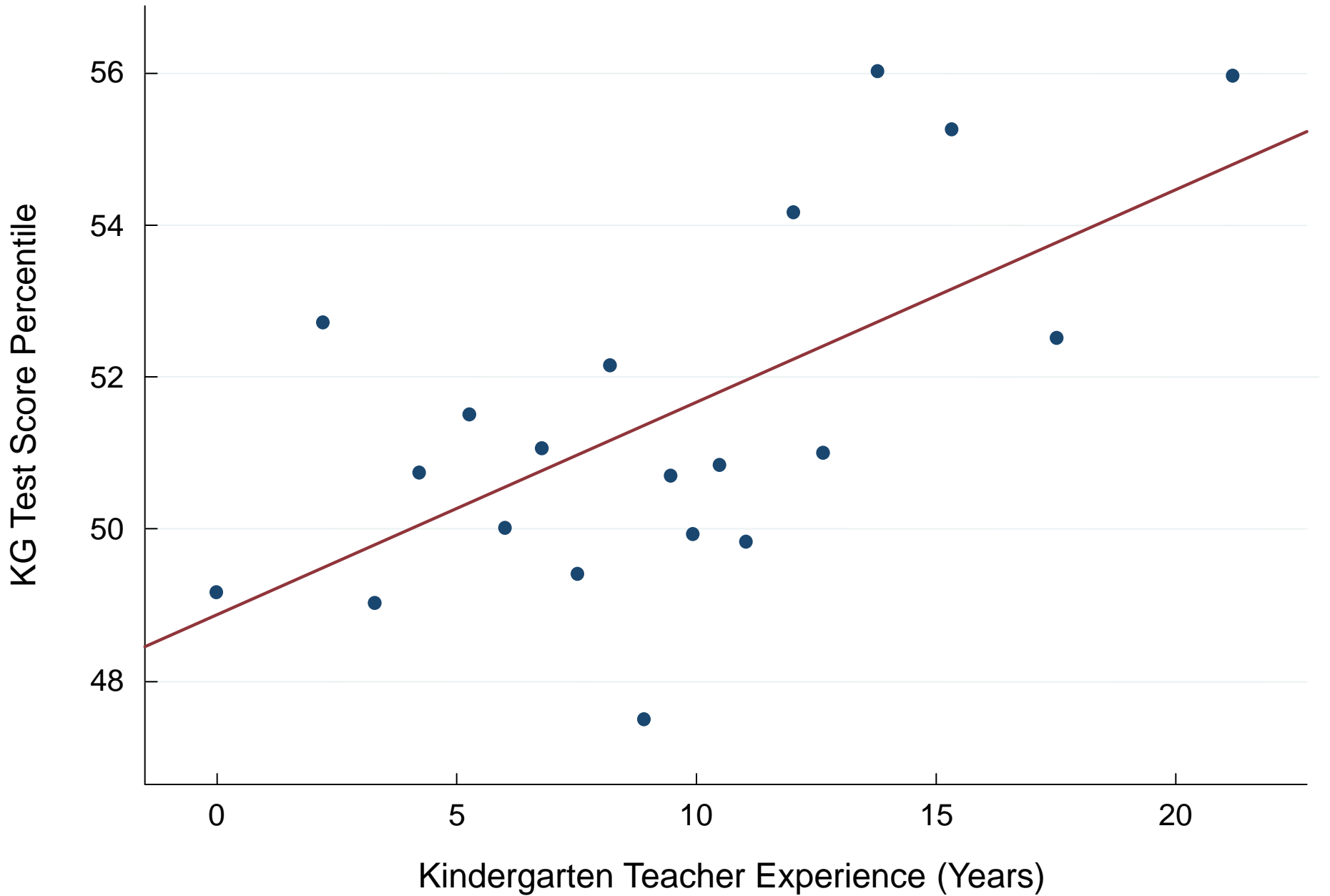


Figure 3b: Effect of Teacher Experience on Earnings

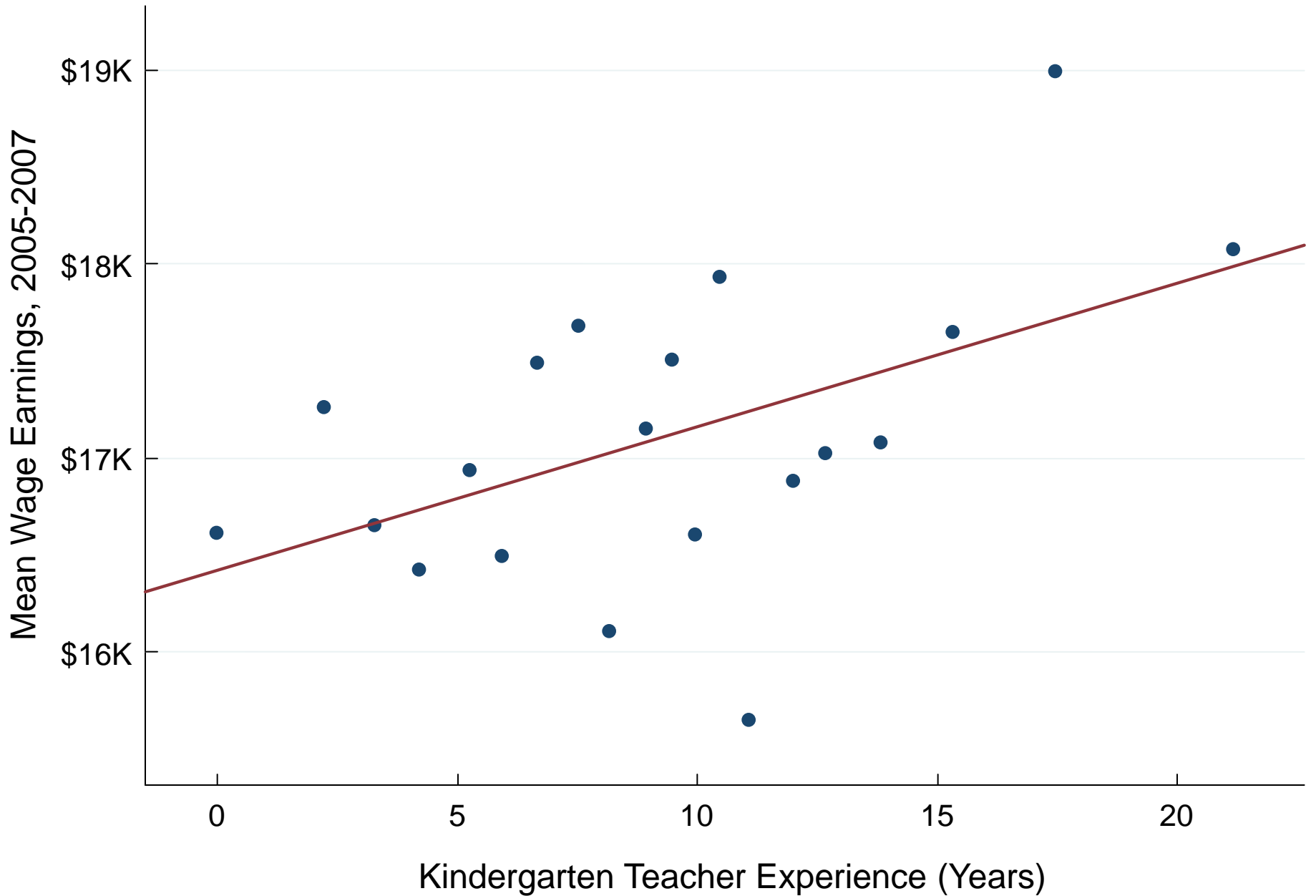


Figure 3c: Effect of Teacher Experience on Earnings by Year

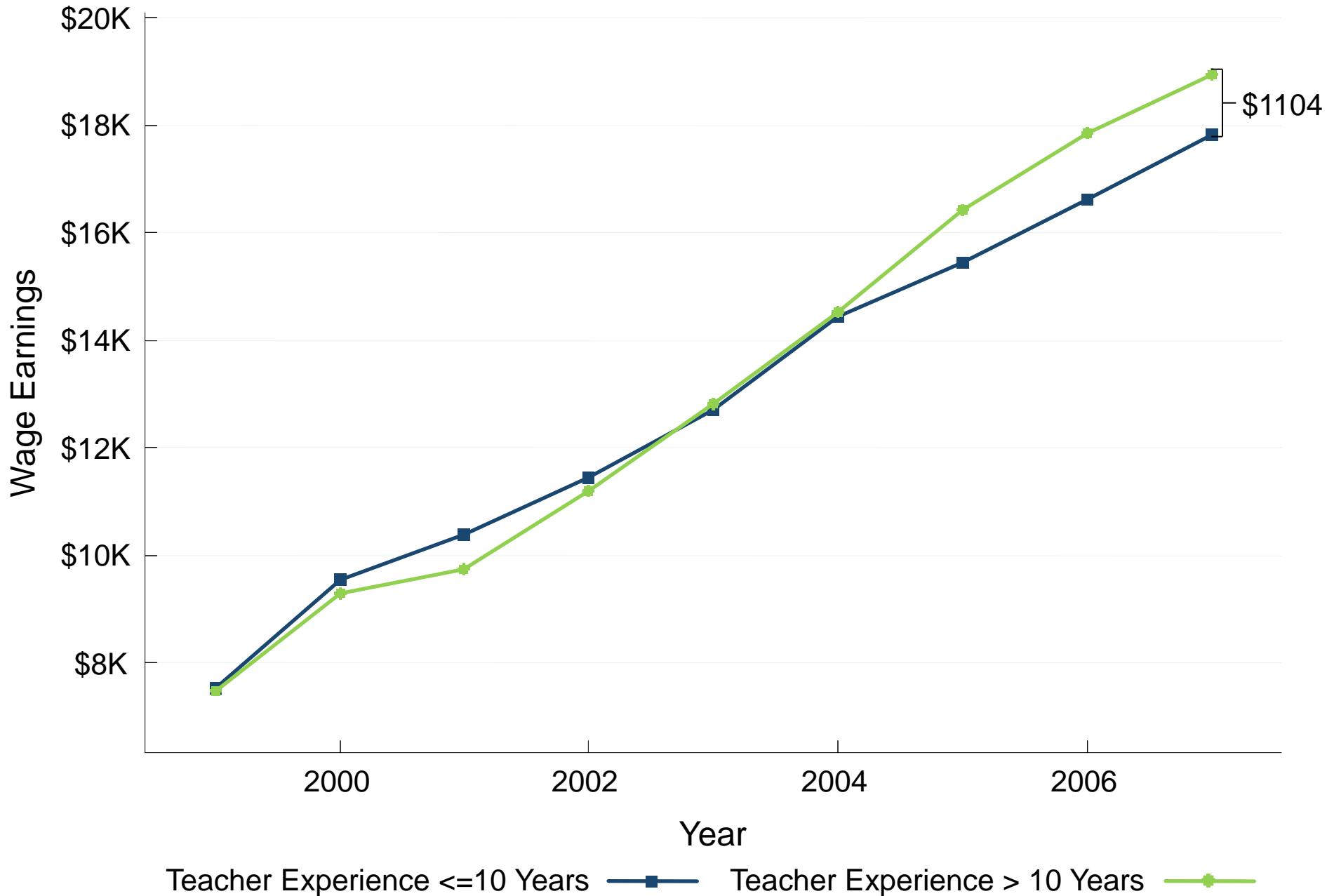


Table 6: Observable Teacher vs. Peer Effects

	Dependent Var.: Test Score		Wage Earnings	
	(1)	(2)	(3)	(4)
Teacher with >10 Years Experience	3.18% (1.26%)	\$1093 (\$546)		
Teacher has post-BA deg.	-0.85% (1.15%)	-\$261 (\$449)		
% Black Classmates			-\$1,757 (\$2,692)	
% Female Classmates			-\$67.5 (\$1,539)	
% Free-Lunch Classmates			-\$285 (\$1,731)	
Classmates' Mean Age			-\$25.8 (\$1,359)	
Classmates' Mean Pred. Score				-\$23.3 (\$93.7)
Entry Grade	KG	KG	All	All
Observations	5,601	6,005	10,992	10,992

Note: All specifications control for school fixed effects and class size, as well as student and parent demographics.

Part 4: Unobservable Class Effects

- Many elements of teacher and peer quality (e.g. clarity of instruction, enthusiasm) are not observable
 - Well known problem in literature on teacher effects
- Test for “class effects” on adult outcomes using analysis of variance
 - Is there significant intra-class correlation in student’s outcomes?
 - This class effect includes effect of teachers, peers, and any class-level shocks such as noise outside classroom
- Formally, we are testing for clustering of outcomes by (randomly assigned) classroom

A Model of Class Effects

- Test scores and earnings for individual i in class c :

$$s_{ic} = z_c + a_{ic}$$

$$y_{ic} = \beta z_c + \gamma z_c^Y + \rho a_{ic} + v_{ic}$$

- z_c = class-level intervention (e.g. better teaching) that affects scores and earnings
- z_c^Y = intervention that affects earnings but not scores

A Model of Class Effects

- Test scores and earnings for individual i in class c :

$$s_{ic} = z_c + a_{ic}$$

$$y_{ic} = \beta z_c + \gamma z_c^Y + \rho a_{ic} + v_{ic}$$

- z_c = class-level intervention (e.g. better teaching) that affects scores and earnings
- z_c^Y = intervention that affects earnings but not scores
- a_{ic} = academic ability
- v_{ic} = earnings ability orthogonal to academic ability

A Model of Class Effects

- Test scores and earnings for individual i in class c :

$$s_{ic} = z_c + a_{ic}$$

$$y_{ic} = \beta z_c + \gamma z_c^Y + \rho a_{ic} + v_{ic}$$

- z_c = class-level intervention (e.g. better teaching) that affects scores and earnings
- z_c^Y = intervention that affects earnings but not scores
- a_{ic} = academic ability
- v_{ic} = earnings ability orthogonal to academic ability
- $\beta + \gamma$ = impacts of interventions on earnings
- β = covariance of class effects on scores and earnings

A Model of Class Effects

- Test scores and earnings for individual i in class c :

$$s_{ic} = z_c + a_{ic}$$

$$y_{ic} = \beta z_c + \gamma z_c^Y + \rho a_{ic} + v_{ic}$$

- Thus far, we have estimated β directly by using observable z 's that affect test scores (e.g. teacher experience)
- How can we estimate β and γ when class-level interventions are unobserved?

Strategy 1: Analysis of Variance

- Test for class effects on earnings ($\beta + \gamma > 0$) using ANOVA
- Do earnings vary across classes by more than what would be predicted by random variation in student abilities?
- Two steps:
 - [Fixed effects] Test for significance of class fixed effects
 - [Random effects] Estimate class-level SD of outcomes assuming normally distributed class effects

Table 7: Analysis of Variance: Kindergarten Class Effects

Dependent Var.:	Grade	Grade	Wage Earnings			
	K Scores	8 Scores	(3)	(4)	(5)	(6)
	(1)	(2)				
P-value of F-Test on KG Class Fixed Effects	0.000	0.419	0.047	0.026	0.020	0.042
SD of Class Effects (RE estimate)	8.77%	0.000%	\$1,497	\$1,520	\$1,703	\$1,454
Demographic Controls	x	x		x	x	x
Large Classes Only					x	
Observable Class Chars.						x
Observations	5,621	4,448	6,025	6,025	4,208	5,983

Note: All specifications control for school fixed effects and class size.

Strategy 2: Covariance of Class Effects on Scores and Earnings

- ANOVA does not tell us whether class effects on scores are correlated with class effects on earnings ($\beta > 0$)
 - Do class-level interventions that raise test scores also improve adult outcomes?
- Turn to a second strategy to measure *covariance* between class effects on scores and earnings (β)
 - What is the correlation of class effects on scores and class effects on earnings?
 - Derive estimator for β and prove it is unbiased in paper; give a heuristic explanation here

Peer-Score Measure of Class Quality

- Average *end-of-year* test scores in class relative to school s_c is a (noisy) measure of class effect on scores:

$$s_c = z_c + \frac{1}{I} \sum_{j=1}^I a_{jc}$$

- Motivates regression of the form:

$$y_{ic} = a + b^M s_c + \varepsilon_{ic}$$

- Own-observation bias: with finite class size, $Eb^M > 0$ even if $b^M = 0$
 - Smart kid raises average class score and has high earnings
 - Analogous to bias in 2SLS estimate with weak instruments

→ Use jackknife (leave-out mean) to measure class effect on scores: s_c^{-i}

Peer-Score Measure of Class Quality

- Regression specification:

$$y_{ic} = a + b^{LM} s_c^{-i} + \varepsilon_{ic}$$

- This regression does not estimate peer effects because we are using *end-of-year* test scores
 - Class quality s_c^{-i} captures teacher quality + class-level shocks
 - Good teachers raise peers' end of year scores
- Class quality s_c^{-i} varies randomly within schools
 - Can test whether classes that generate test score gains also generate earnings gains

Peer-Score Measure of Class Quality

- Regression specification:

$$y_{ic} = a + b^{LM} s_c^{-i} + \varepsilon_{ic}$$

- Three remaining sources of bias in b^{LM}
 1. Mechanical: Peers below-avg. \rightarrow you are above avg. (Guryan, Kroft, Notowidigdo 2009). Solution: define intercept using leave-out mean
 2. Attenuation: s_c^{-i} is a noisy measure of class quality
 3. Reflection: with peer effects, smart kid raises peers' scores and earns a lot, driving up b^{LM}
- After presenting results, we bound reflection bias and show it is smaller than attenuation bias

Figure 4a: Effect of Early Childhood Class Quality on Own Score

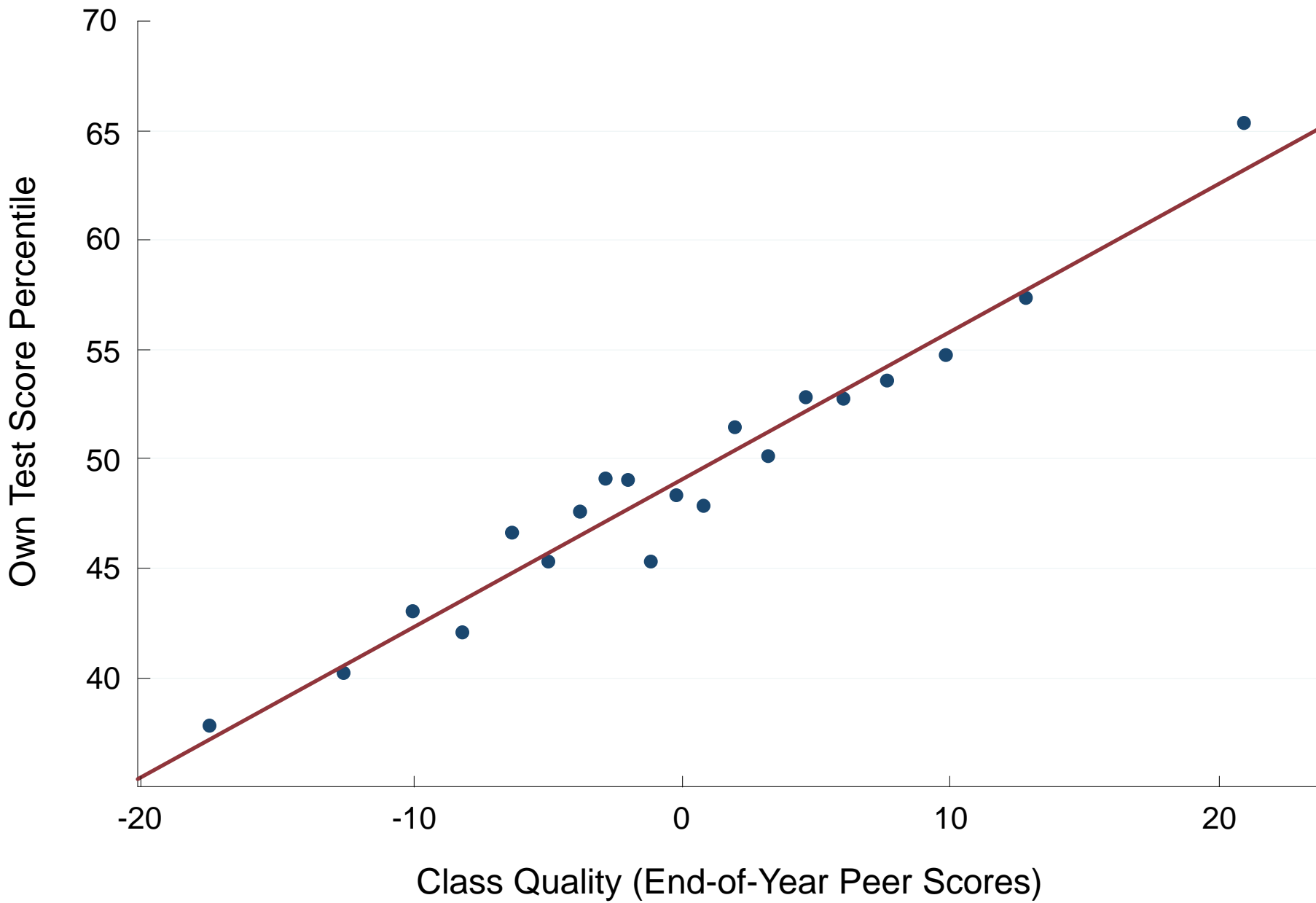


Figure 4c: Effect of Early Childhood Class Quality on Earnings

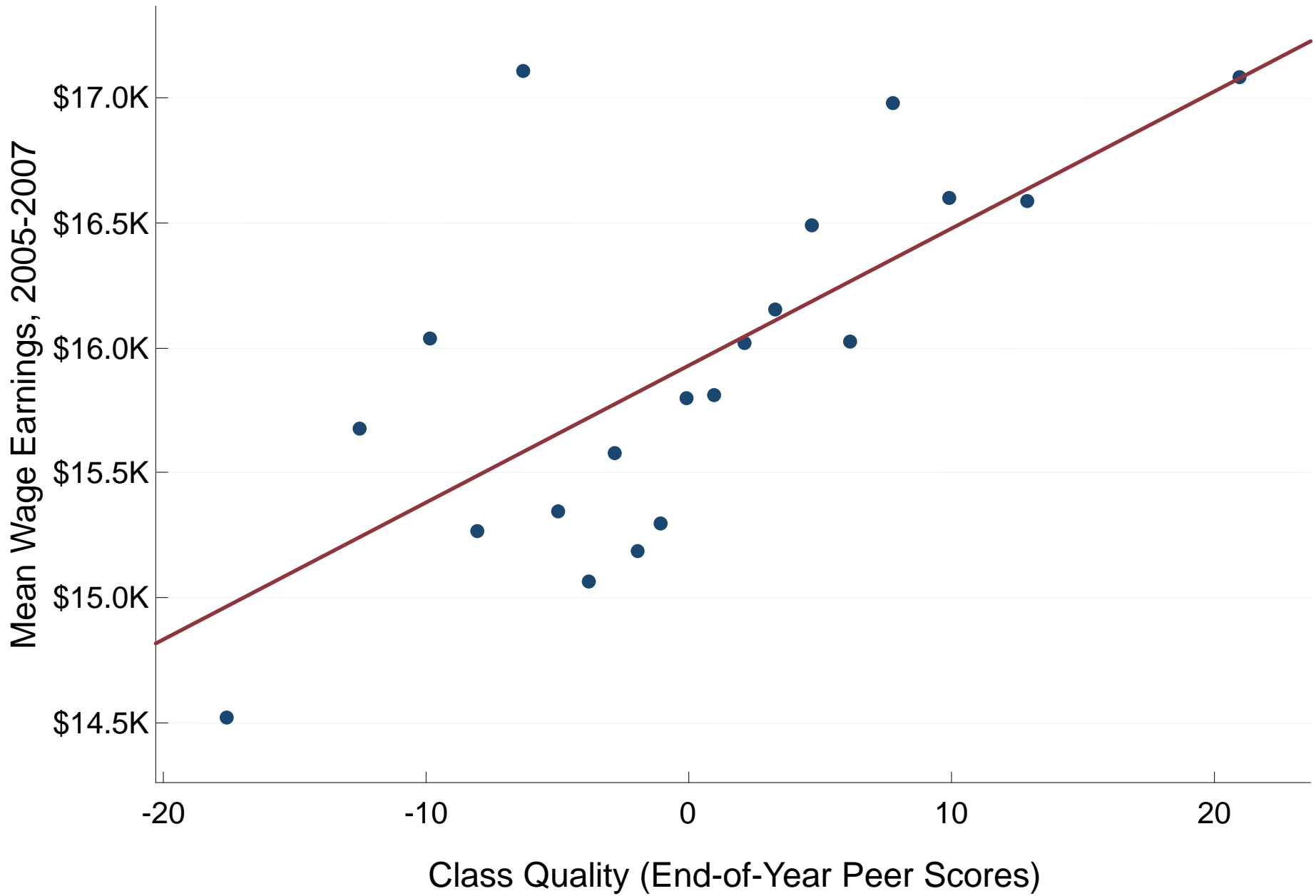


Figure 5a: Effect of Class Quality on Earnings by Year

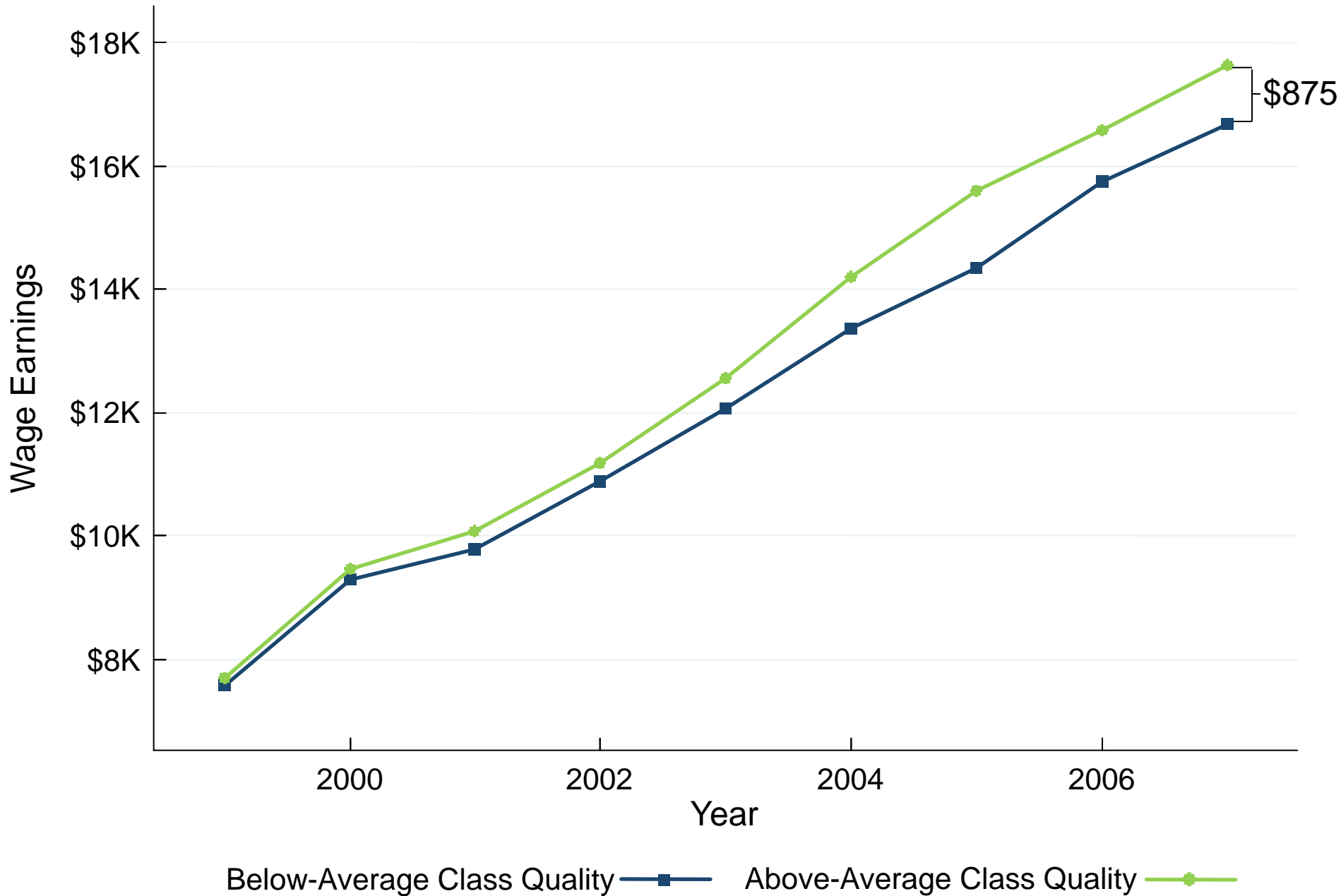


Table 8a: Impacts of Class Quality on Earnings

Dependent Variable:	Wage Earnings (\$)			
	(1)	(2)	(3)	(4)
Class Quality (peer scores)	50.61 (17.45)	61.31 (20.21)	53.44 (24.84)	47.70 (18.63)
Entry Grade	All	All	KG	Grade ≥1
Observable class chars.		x		
Observations	10,959	10,859	6,025	4,934

NOTE--All regressions control for student and parent demographics and school-by-entry-grade fixed effects.

Table 8b: Impacts of Class Quality on Other Adult Outcomes

Dependent Variable:	College in 2000	College by Age 27	College Quality	Summary Index
	(%)	(%)	(\$)	(%)
	(1)	(2)	(3)	(4)
Class Quality (peer scores)	0.096 (0.046)	0.108 (0.053)	9.328 (4.573)	0.250 (0.098)
Observations	10,959	10,959	10,959	10,959

NOTE--All regressions control for student and parent demographics and school-by-entry-grade fixed effects.

Figure 6a: Fadeout of Class Effects
Effect of 1 SD of Class Quality on Test Scores by Grade

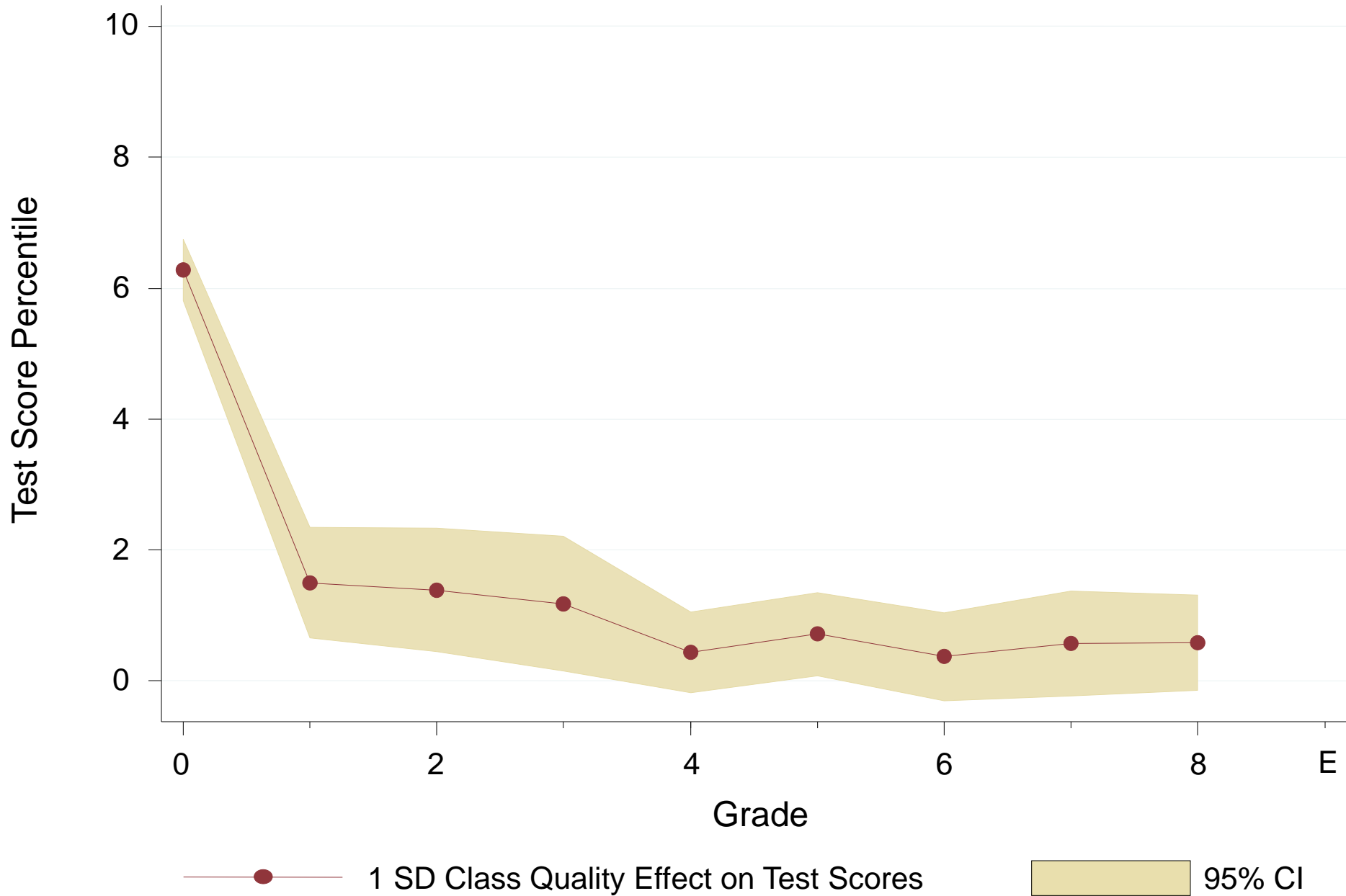
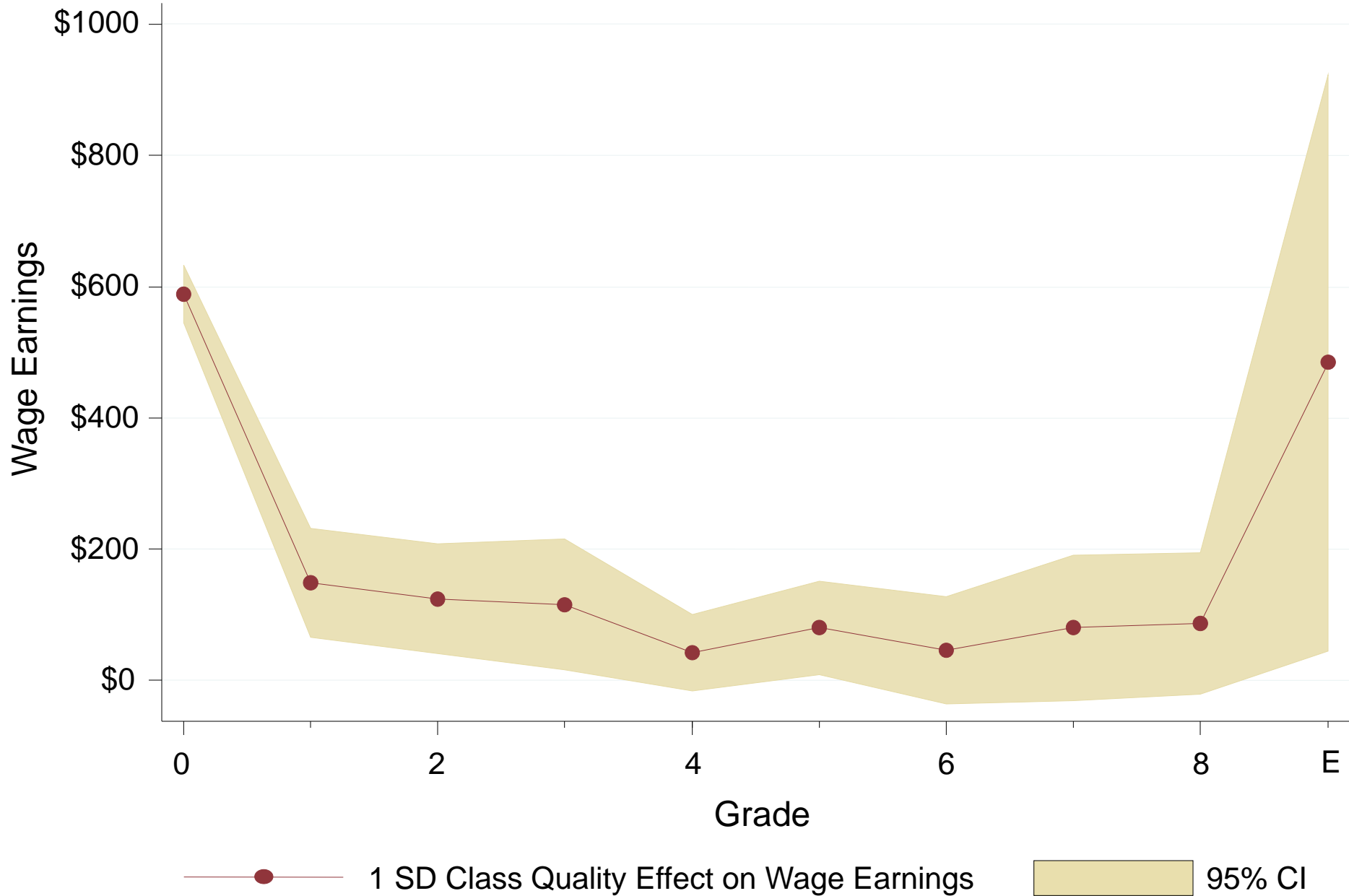


Figure 6b: Fadeout of Class Effects
Effect of 1 SD of Class Quality on Earnings



Bounding Reflection Bias

- Small impact of KG class quality on subsequent test scores places a tight upper bound on reflection bias
- Smart kids score high on all tests (test scores highly autocorrelated)
- To have large reflection bias, smart kid must raise peer scores a lot
 - Large correlation between peer scores and own score in later grades
- We formalize this intuition in a linear-in-means model and derive a bound on the degree of reflection bias
 - Observed correlation between KG peer scores and 8th grade score places an upper bound on reflection bias of 20%
 - Variance in scores implies attenuation bias of 20% as well, implying preceding estimates are downward biased on net

Fade-out and Re-emergence: The Role of Non-Cognitive Skills

- Why do effects of kindergarten class fade out and re-emerge?
- One explanation: non-cognitive skills (Heckman 2000)
- Data on non-cognitive measures (effort, initiative, disruption) collected for random subset of STAR students in 4th and 8th grade

Please consider the behavior of Jim Smith over the last 2-3 months. Circle the number that indicates how often the child exhibits the behavior.

	<u>Never</u>		<u>Some-</u> <u>times</u>		<u>Always</u>
#1. Acts restless, is often unable to sit still	1	2	3	4	5
#2. Annoys or interferes with peers' work	1	2	3	4	5

- Convert mean non-cog score to percentile scale as above

Mean Wage Earnings vs. Grade 4 Non-Cognitive Percentile

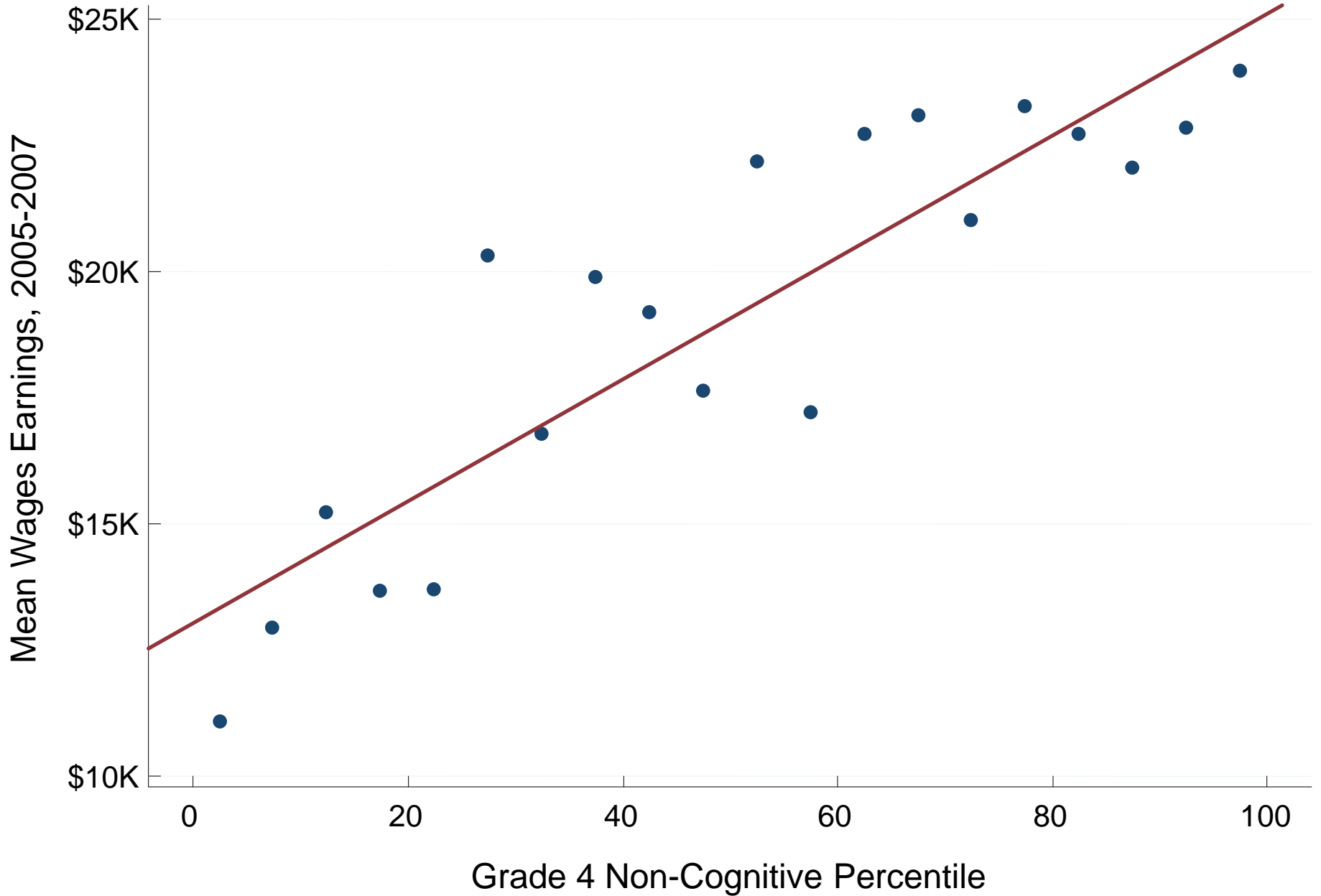


Table 9: Wages Earnings, Class Quality and Non-Cognitive Skills

Dependent Variable:	Wage Earnings	Grade 8	Grade 4 Scores		Grade 8 Scores	
		Math+ Read	Math+ Read	Non-Cognitive	Math+ Read	Non-Cognitive
		(1)	(2)	(3)	(4)	(5)
Grade 4 Non-Cog. Score	\$87.7 (\$20.4)	0.059 (0.017)				
Grade 4 Math + Reading Score	\$36.4 (\$24.7)	0.671 (0.023)				
Class Quality (peer scores)			0.047 (0.035)	0.153 (0.065)	0.064 (0.041)	0.128 (0.054)
Observations	1,360	1,254	4,023	1,671	4,448	1,780

Fade-out and Re-emergence: The Role of Non-Cognitive Skills

- Non-cognitive skills provide a simple explanation of our findings
- High quality KG teachers raise KG test scores partly through good classroom management
 - Good classroom management instills social skills
 - Social skills not directly measured in standardized tests but have returns in the labor market
- Rapid fadeout in math and reading tests after KG
- But significant earnings gains from better KG class

Part 6: Cost-Benefit Analysis

- Assume: 3% real discount rate, constant percent income gains, income follows average US income profile, constant effects of class quality
1. One SD increase in KG class quality for a single year
 - Total NPV earnings gain for class of 20 students of \$782K
 2. 33% reduction in class size
 - \$4K-\$189K per class (very imprecisely estimated)
 3. One SD improvement in teacher quality
 - \$170-\$214K per class
 - Moving from below-avg (25th pctile) to above-avg (75th pctile) teacher generates NPV of \$320K for a class of 20 students

School Quality and Income Inequality

- Intergenerational income correlation of around 0.3 (Solon 1999)
 - How much of this can be explained by the fact that higher income families have access to better public schools?
- In STAR data, each \$10K of parents' income increases class quality in each grade by 0.7% of a SD
- Use our estimates of effect of class quality on child's earnings and assume constant class-quality effects across grades
- Roughly 1/3 of intergenerational income transmission runs through differences in school quality in K-12