

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

September 2010

How Does Your Kindergarten Class Affect Your Earnings?

- What are the long-term impacts of early childhood education?
- Limited evidence to date because few datasets link information on early childhood test scores with data on adult outcomes
- We link data from the STAR experiment to US tax records to analyze how KG class assignment affects long-term outcomes

Project STAR: Background

- Student/Teacher Achievement Ratio (STAR) experiment is one of the most widely studied education interventions
 - Conducted from 1985 to 1989 in Tennessee
 - One cohort of 11,571 children in grades K-3 at 79 schools
 - Most children in the experiment 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
- 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)
 - Students assigned to more experienced teachers have higher scores
 - 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

- Access to selected variables in anonymous U.S. tax records to conduct research on behavioral responses to economic policies
- 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
 - Match rate orthogonal to treatments

Table 1: Summary Statistics

	Mean (1)	St. Dev. (2)
Average Wage Earnings (2005-07)	\$15,910	\$15,558
Fraction With Zero Earnings ('05-'07)	13.9%	34.5%
Attended College in 2000 (age 20)	26.4%	44.1%
Ever Attended College by age 28	45.5%	49.8%
Average Parental Income ('96-'98)	\$48,010	\$41,622
Fraction Black	35.9%	48.0%

Outline

1. Test scores and adult outcomes in the cross-section
2. Re-evaluate validity of STAR experimental design
3. Class size impacts on adult outcomes
4. Teacher/peer impacts on adult outcomes
5. Fade-out, Re-Emergence, and Non-Cognitive Skills
6. Conclusion: Cost-Benefit Analysis

Part 1: Cross-Sectional Correlations

- Begin with OLS estimates of return to higher early childhood test scores by correlating KG test scores with adult outcomes
- Useful to benchmark estimates obtained from randomized interventions
- Test score: percentile score on Stanford Achievement Test (math + reading), constructed as in Krueger (1999)
- Estimate both raw correlations and 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

Figure 1: Wage Earnings in 2007 vs. KG Test Score

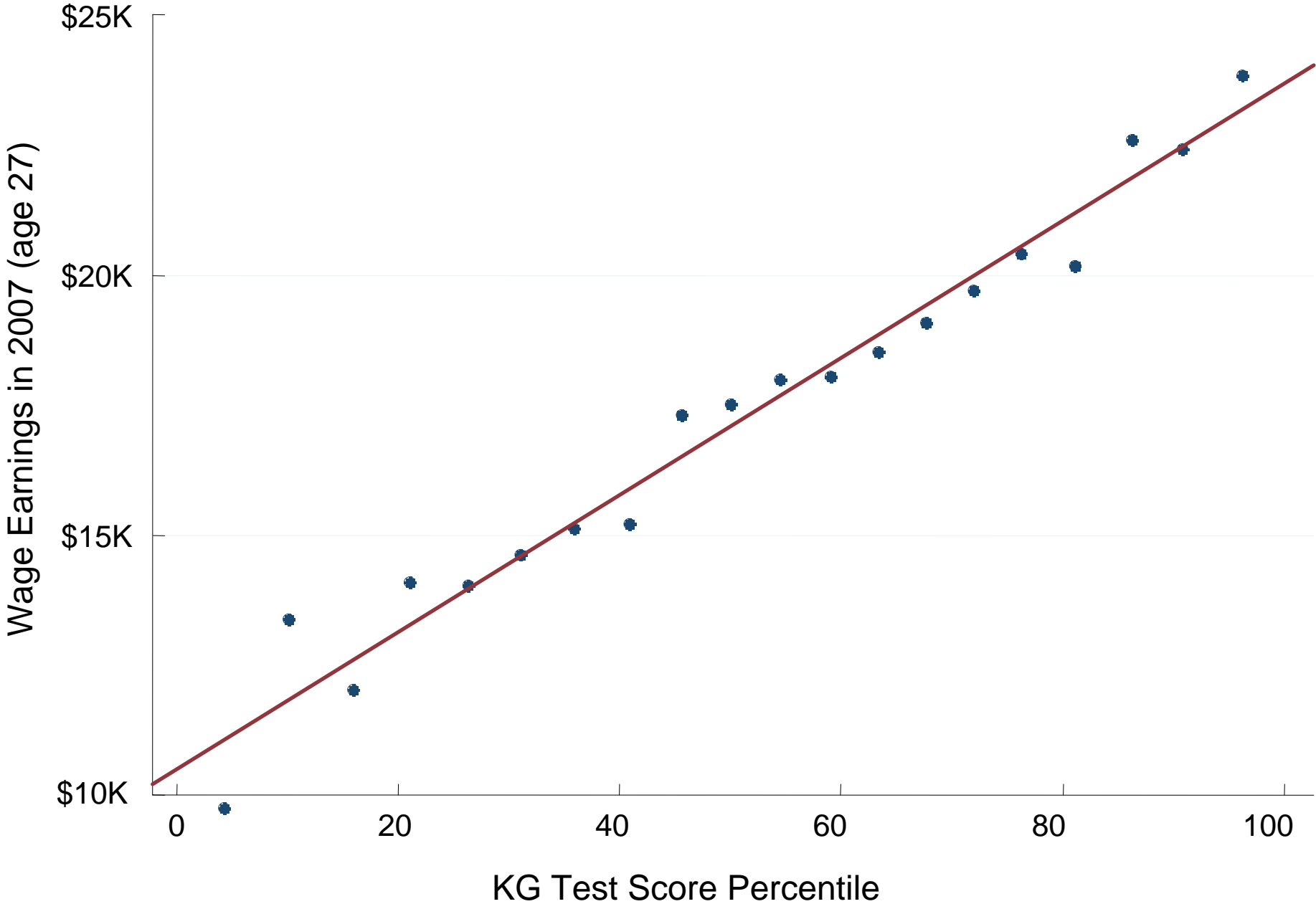
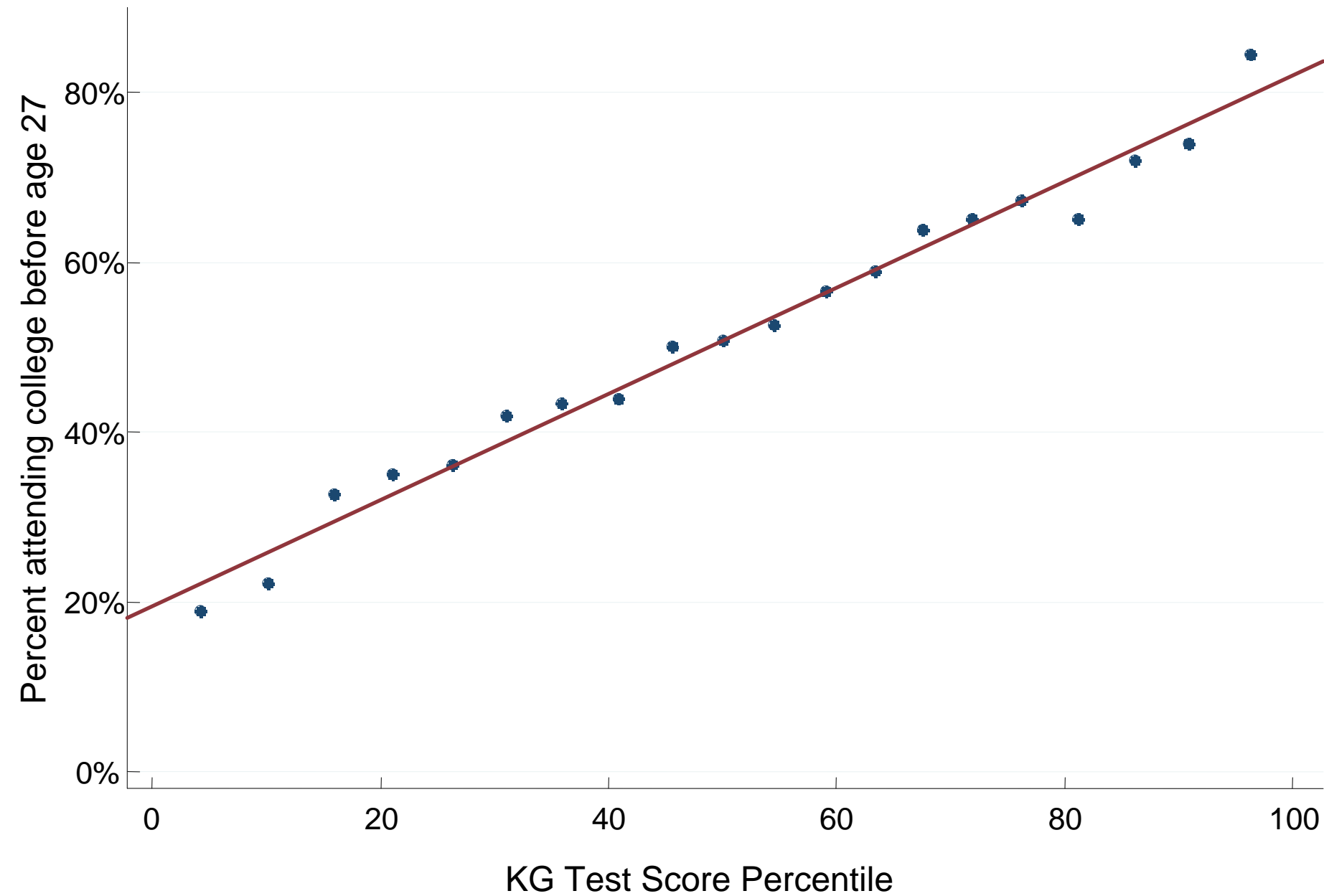


Table 2: Test Scores and Earnings in the Cross-Section

Dependent Var.:	Wage Earnings			Log Wage Earnings	Wage Earnings
	(1)	(2)	(3)	(4)	(5)
KG Test percentile	\$131.74 (\$7.63)	\$142.98 (\$9.32)	\$93.79 (\$9.56)		\$105.54 (\$9.46)
KG Test z score				0.180 (0.026)	
Parental Income Percentile					\$157.68 (\$8.54)
Black			-\$46.77 (\$878.70)	0.086 (0.085)	
Controls			x	x	
Class Fixed Effects		x	x	x	x
Adjusted R ²	0.05	0.09	0.17	0.11	0.16
Observations	5,621	5,621	5,621	5,154	5,621

Figure 2a: 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 automatically generates a 1098-T form linking student and institution
 - Form filed even if student did not pay out-of-pocket and received a full scholarship
- Find everyone age 20 enrolled in college in 1999
- Calculate average wage earnings in 2007 (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

Table 3: An Earnings-Based Index of College Quality

US News Ranking	College	Mean Earnings at age 28
1	Harvard Princeton Yale Cal Tech MIT	\$79,643
2		
3		
4		
5		
6	Stanford U Penn Columbia U Chicago Duke	\$75,570
7		
8		
9		
10		
....		
121	Arizona St. Catholic U MI Tech U Buffalo U San Fran	\$46,390
122		
123		
124		
125		

College Mean Wage Earnings by US News Ranking

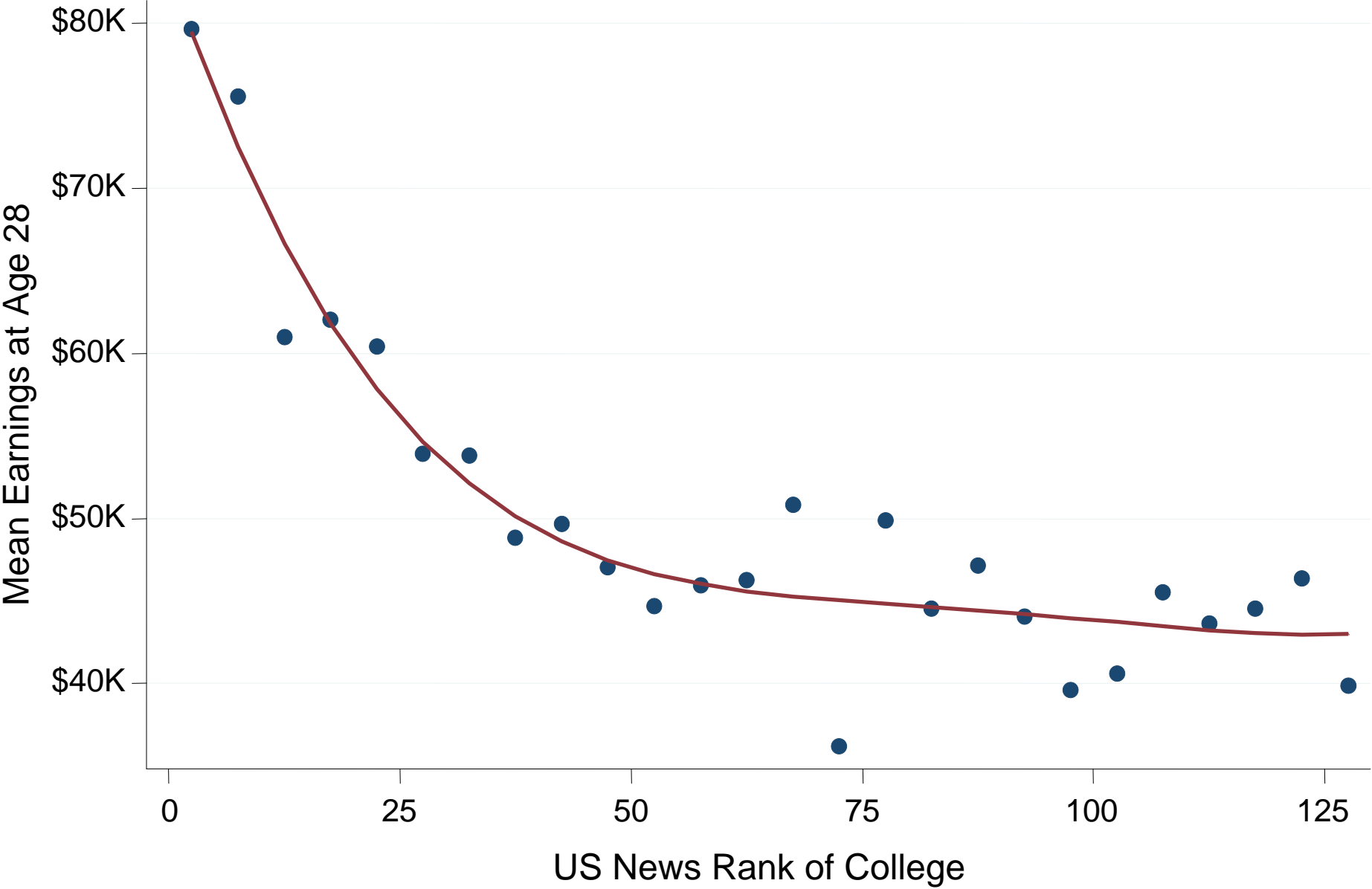


Figure 2b: College Quality vs. KG Test Score

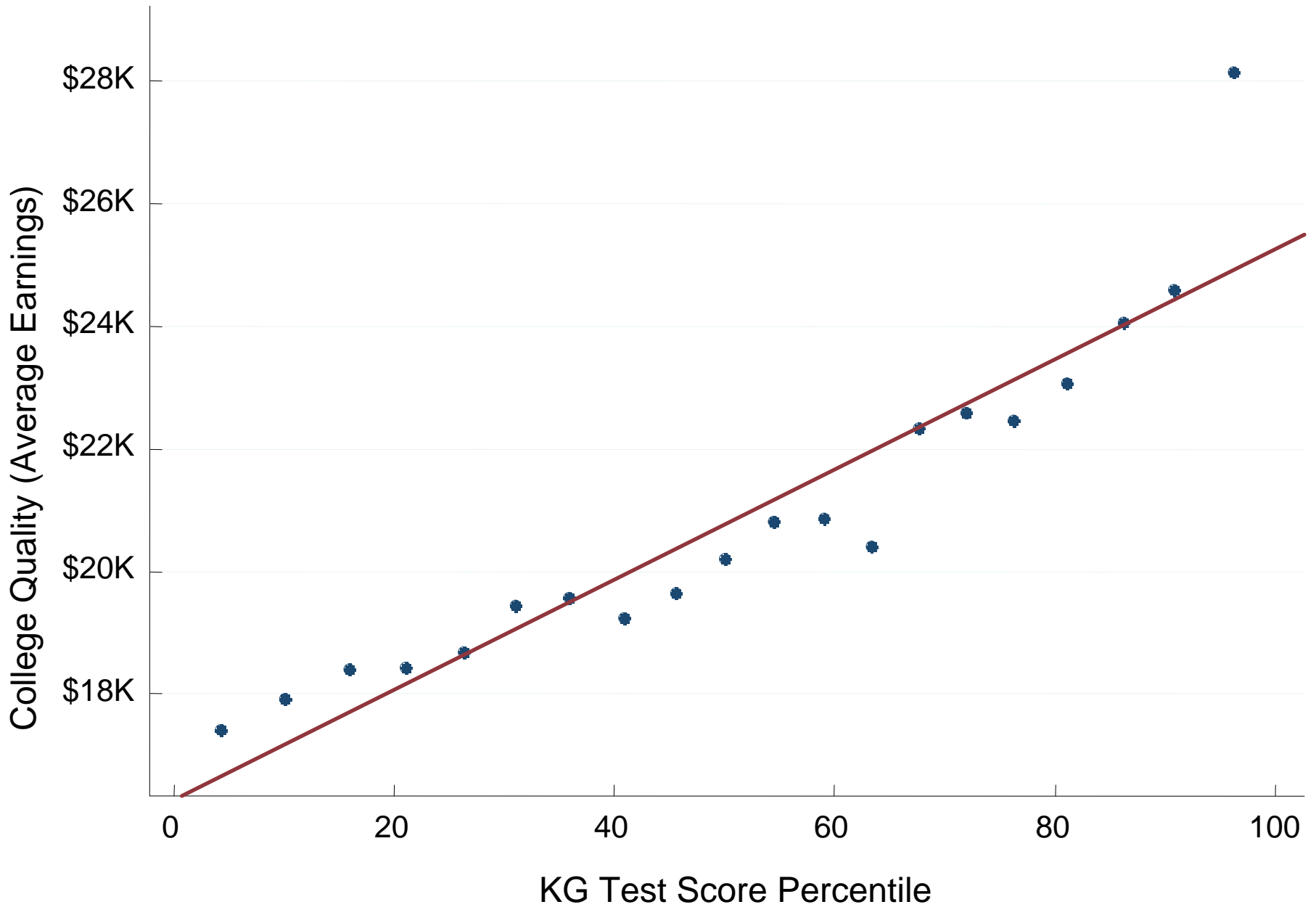


Figure 3a: Home Ownership vs. KG Test Score

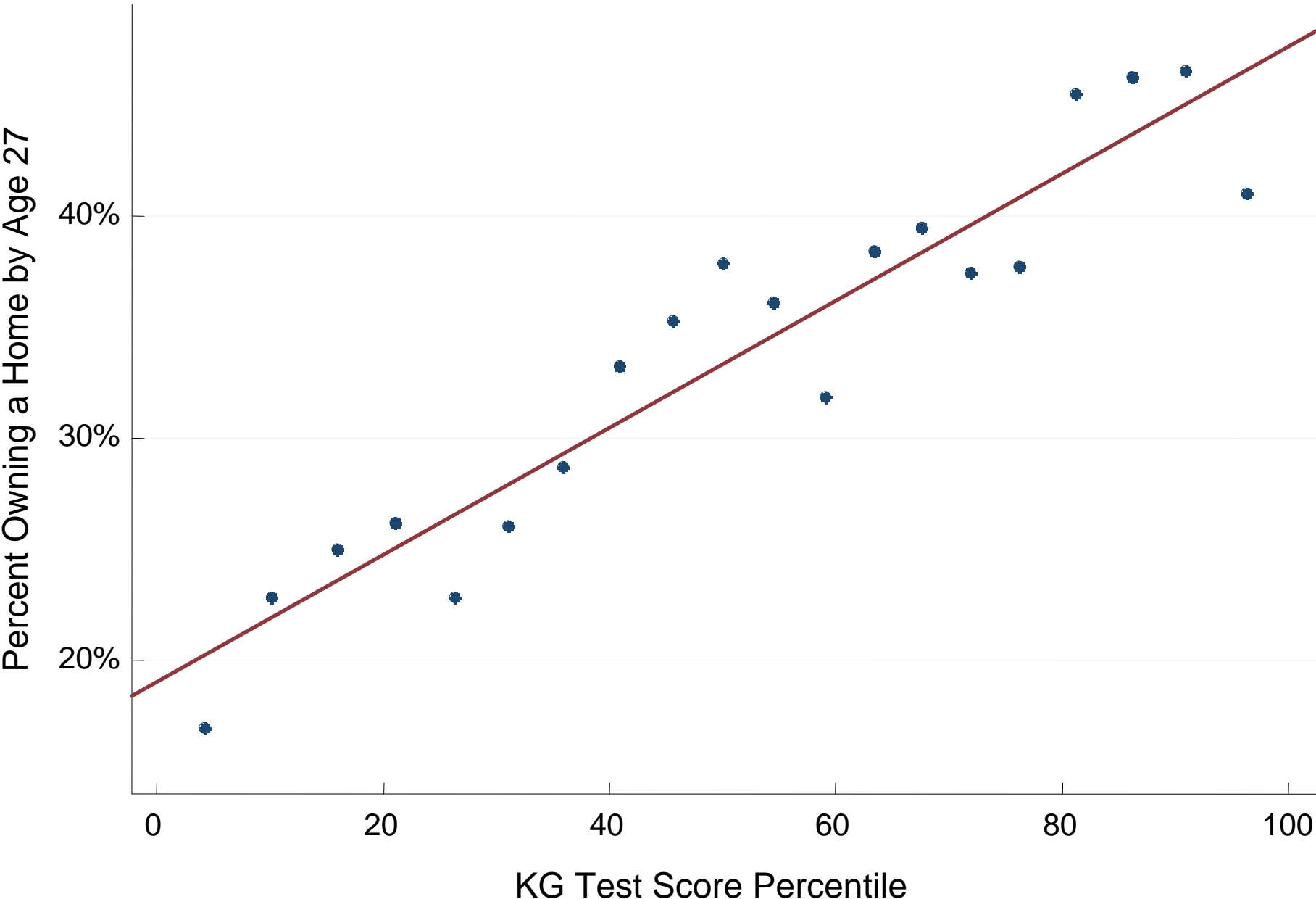


Figure 3b: Retirement Savings vs. KG Test Score

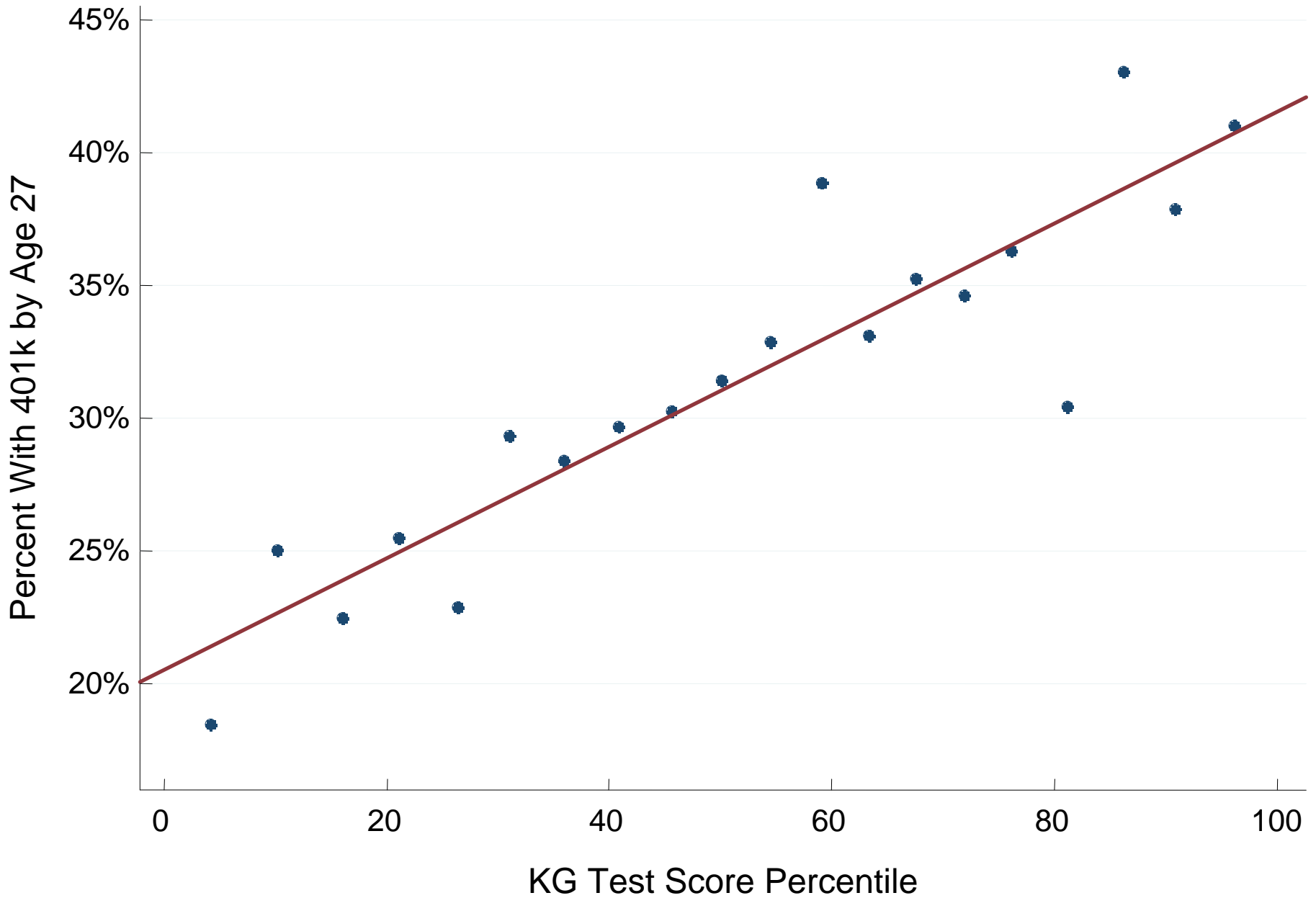


Figure 3c: Cross-State Mobility vs. KG Test Score

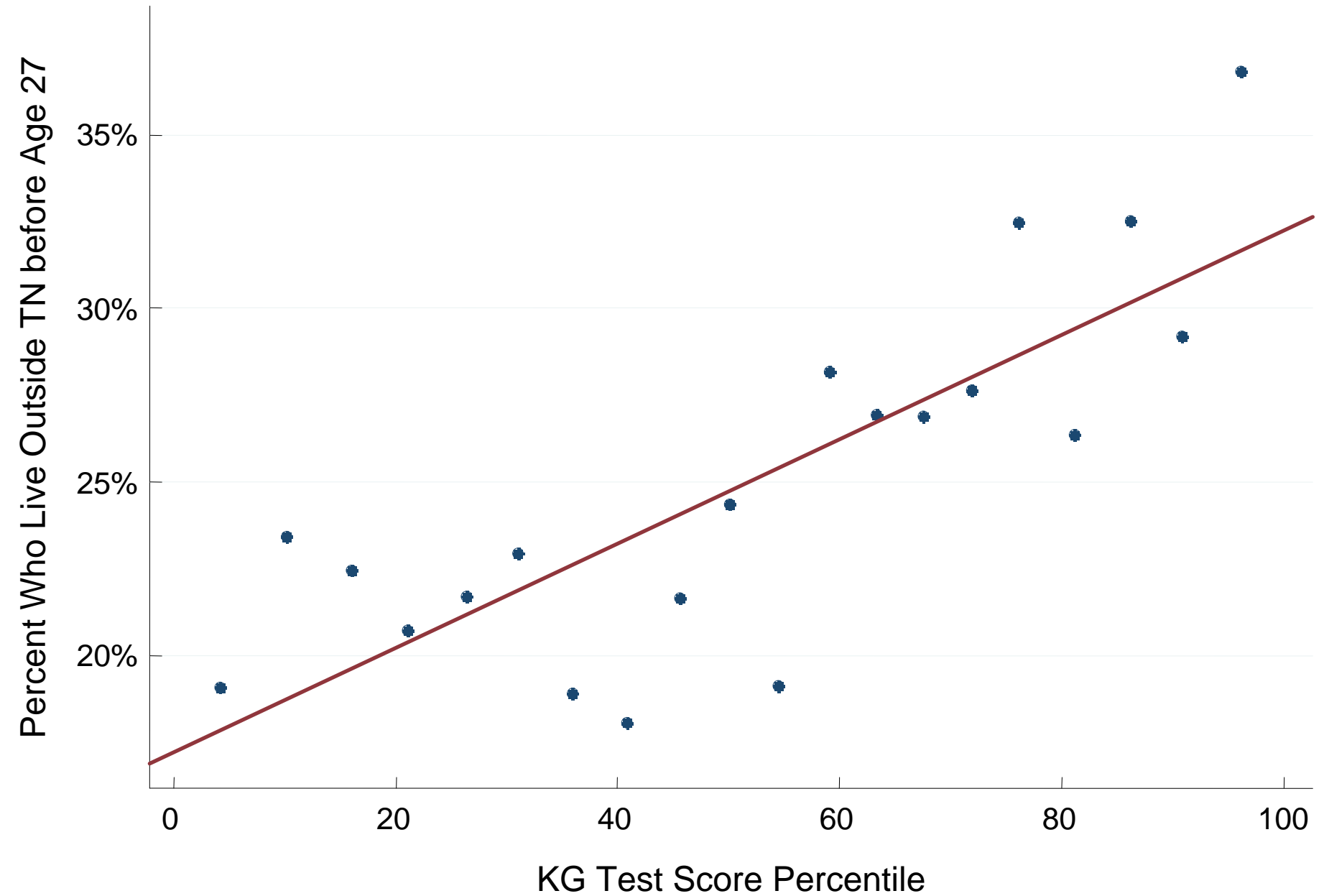


Figure 3d: College Graduate in ZIP code vs. KG Test Score

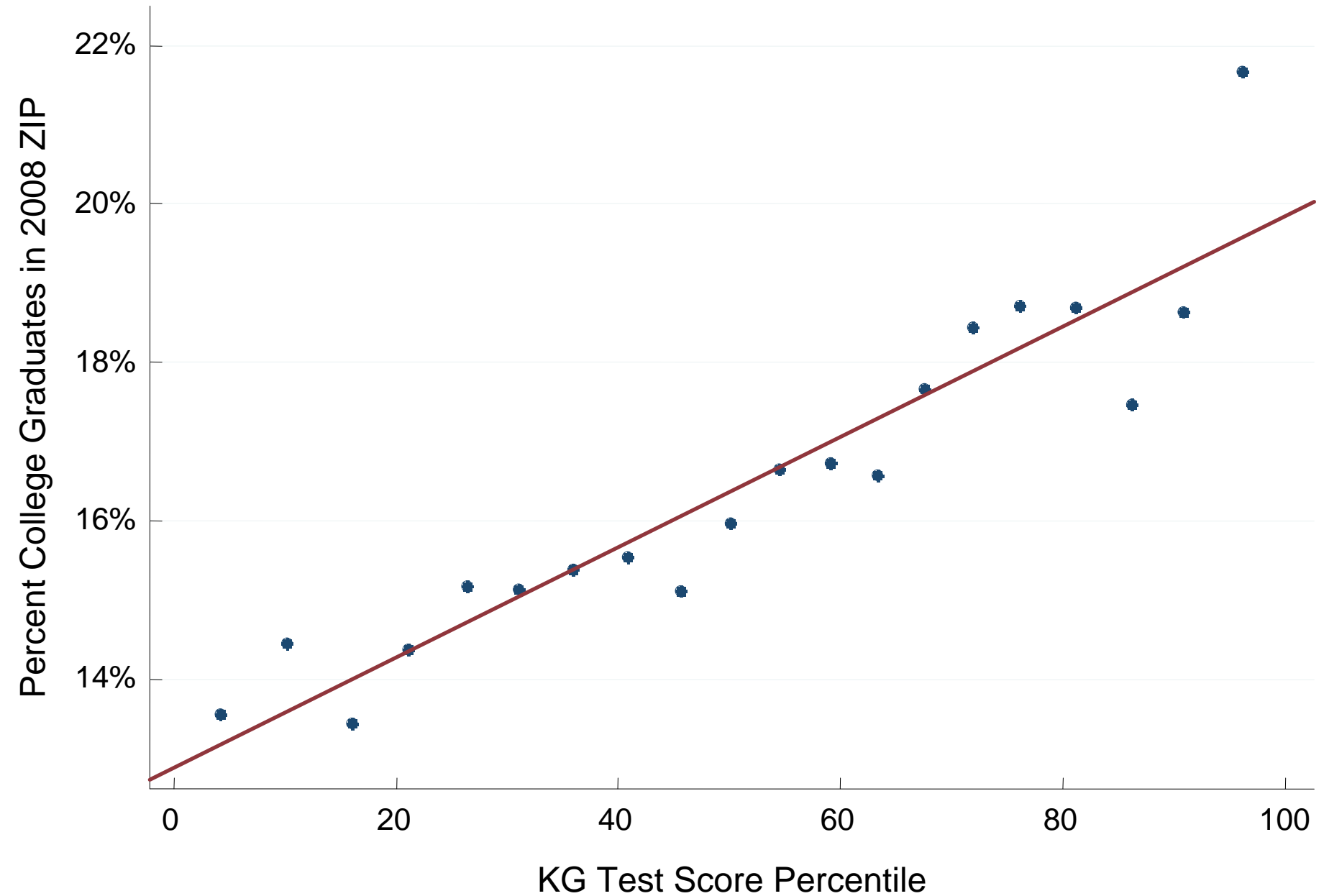


Figure 3e: Marriage by Age 27 vs. KG Test Score

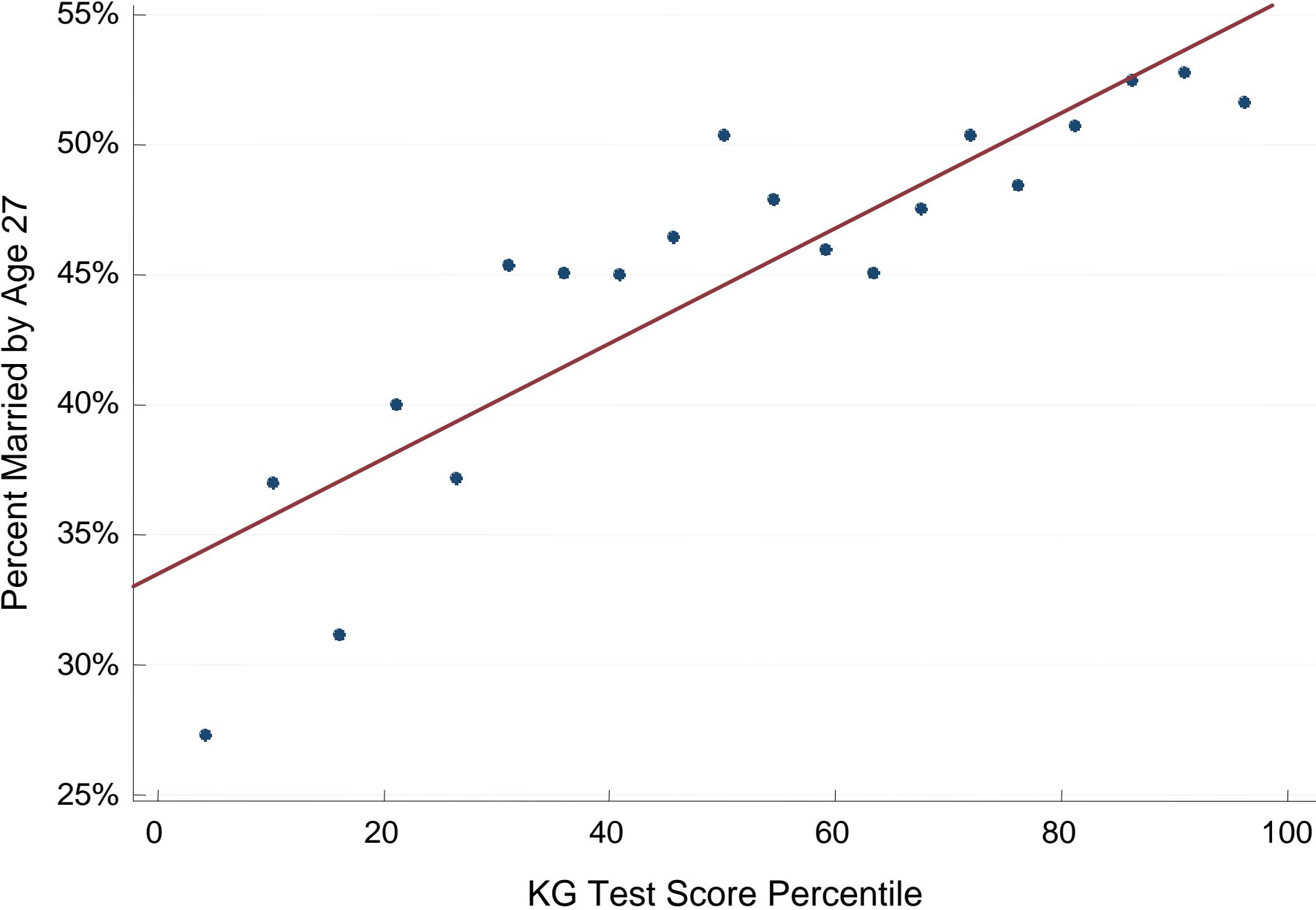
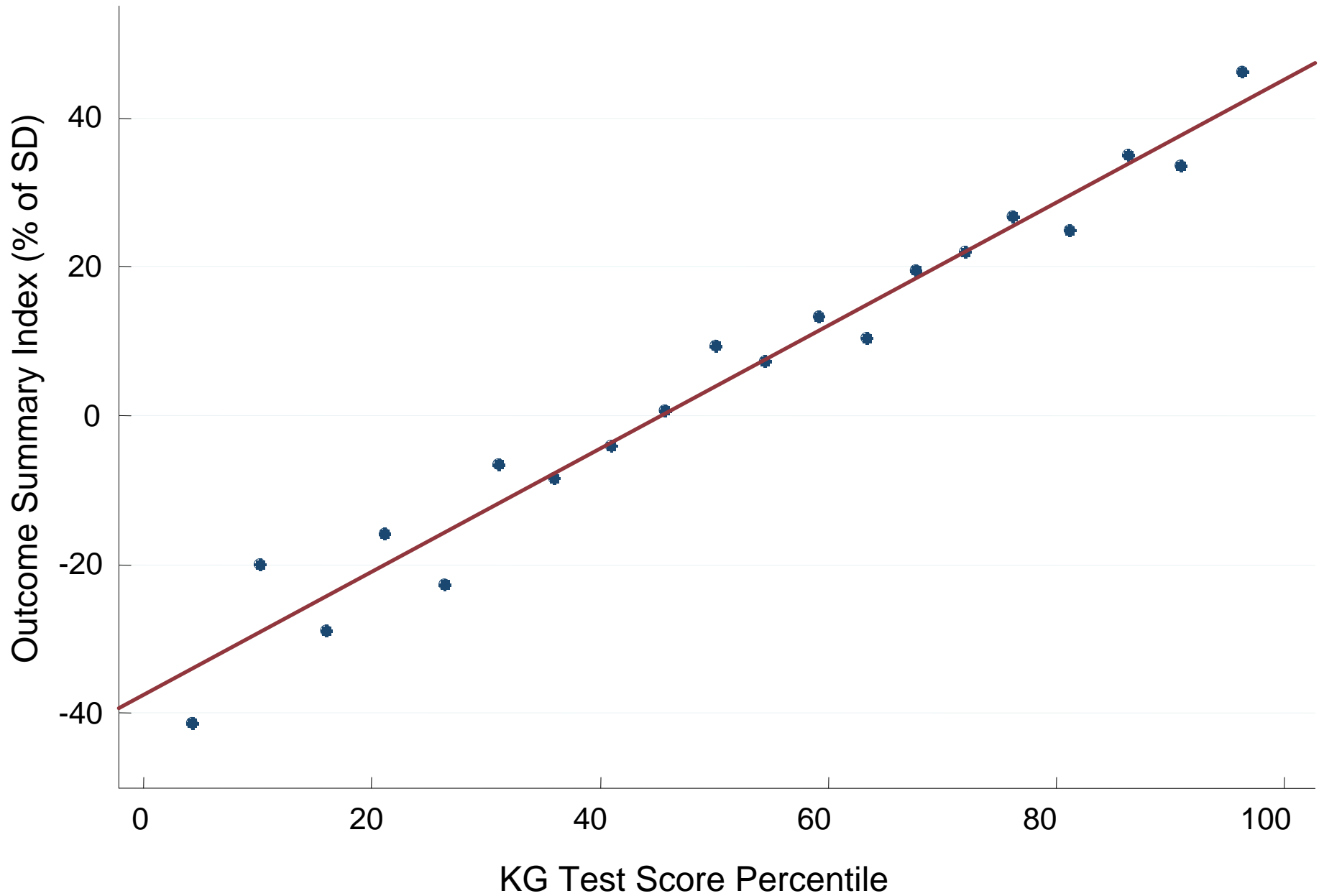


Figure 3f: Summary Outcome Index vs. KG Test Score



Part 2: Validity of the STAR Experimental Design

- Experimental analysis rests on two assumptions: randomization and no differential attrition across treatment groups
 - All pre-determined variables (e.g. parent characteristics) are balanced across classrooms
 - No evidence of differences in match rates across classrooms

Part 2: Validity of the STAR Experimental 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 4: Randomization Tests

Dependent Variable:	First Obs. Test Score	Small Class
	(1)	(2)
Household Income (\$1,000s)	0.085 (0.008) [11.20]	-0.003% (0.014%) [-0.248]
Parents Age at STAR Birth	0.178 (0.039) [4.520]	0.029% (0.072%) [0.411]
Parents Ever Have 401k	1.863 (0.518) [3.598]	1.455% (0.942%) [1.545]
Female	3.510 (0.460) [7.632]	-.224% (0.832%) [-0.270]
Black	-8.779 (0.920) [-9.546]	0.205% (1.656%) [0.124]
p-Value on F-Statistic	0.000	0.259
Observations	9,393	10,992

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

Validity of the STAR Experimental 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 5: Tests for Selective Attrition

Specification:	Match Rates		Death Rates	
	(%)	(%)	(%)	(%)
	(1)	(2)	(3)	(4)
Difference	-0.019 (0.476)	0.079 (0.418)	-0.011 (0.293)	-0.006 (0.293)
F-Test P-Value on Class F.E.'s	0.951	0.888	0.388	0.382
Controls		x		x
Observations	11,571	11,571	10,988	10,988

Part 3: Class Size Impacts

- Replicate specifications in previous studies
- Independent variable: dummy for small class *assignment* (ITT)
- Focus on four outcomes:
 1. College attendance in 2000
 2. College quality index
 3. Mean earnings (ages 25-27)
 4. Standardized (SD = 1) summary index of other outcomes:
$$\text{Index} = 401K + \text{Home Owner} + \text{Married} + \text{Moved (Leave TN)} + \text{Pct. College Grads. in Zip}$$

Figure 4a: Effect of Class Size on College Attendance by Year

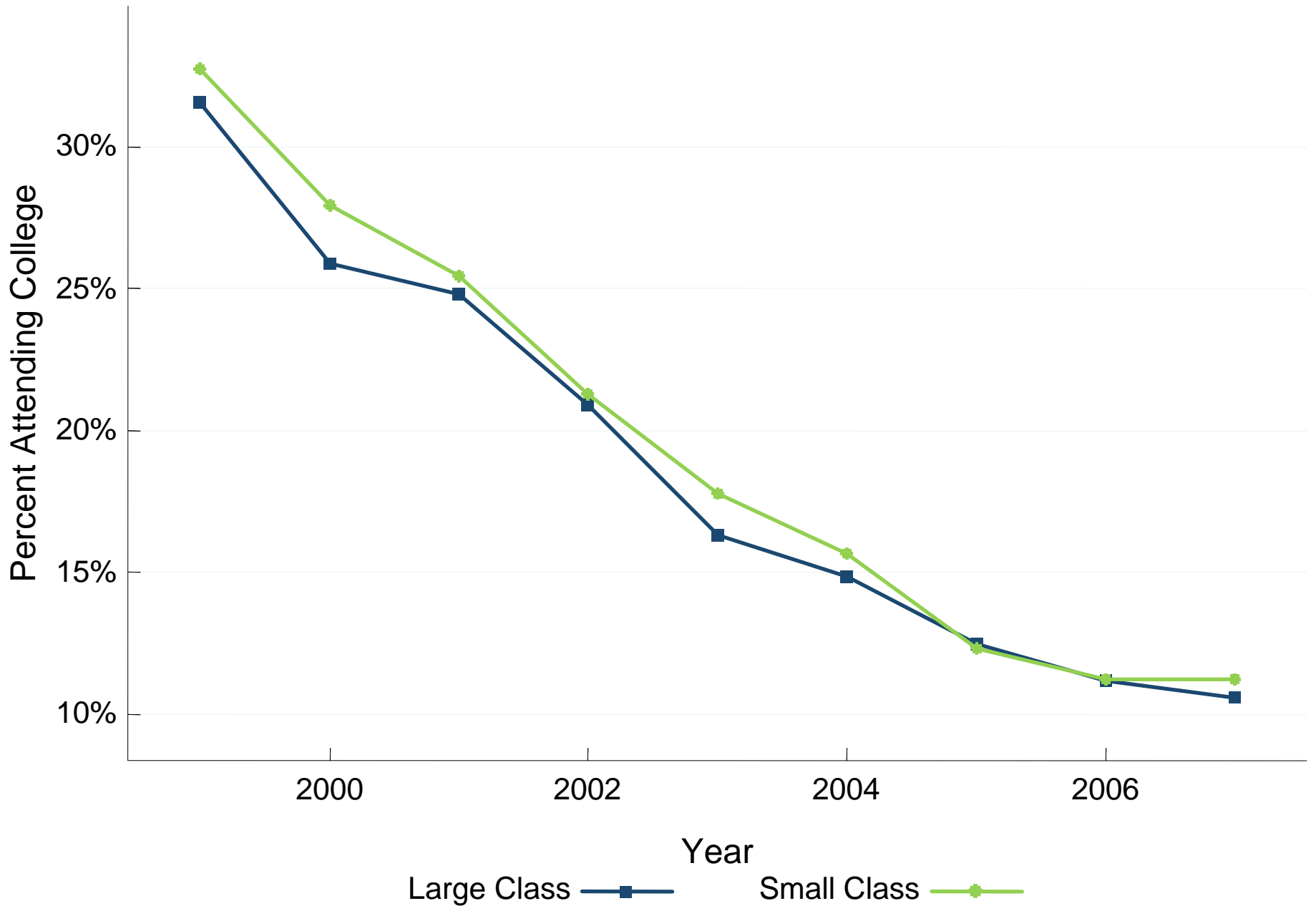


Figure 4b: College Earnings Quality by Class Size

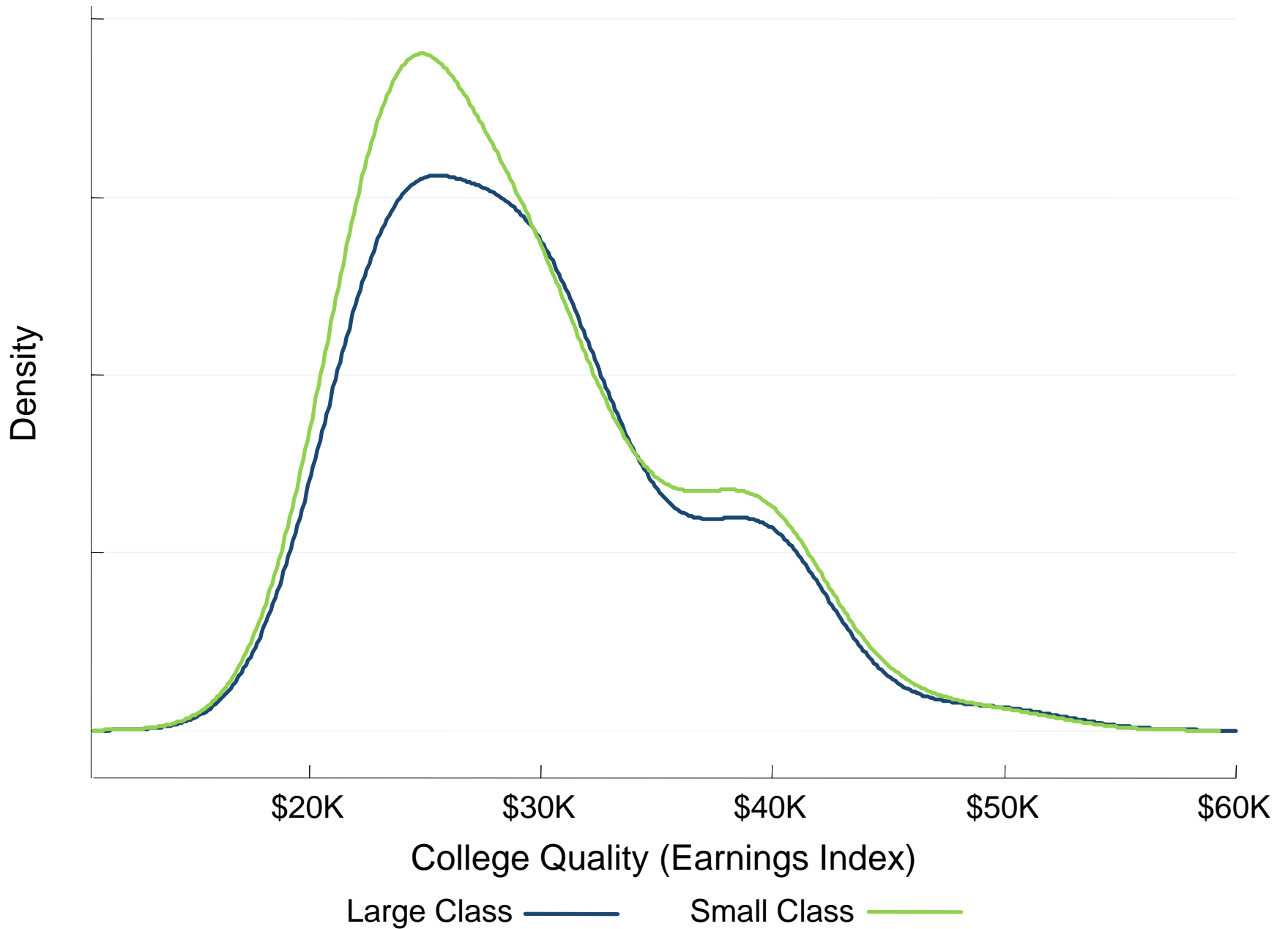


Figure 4c: Effect of Class Size on Wage Earnings by Year

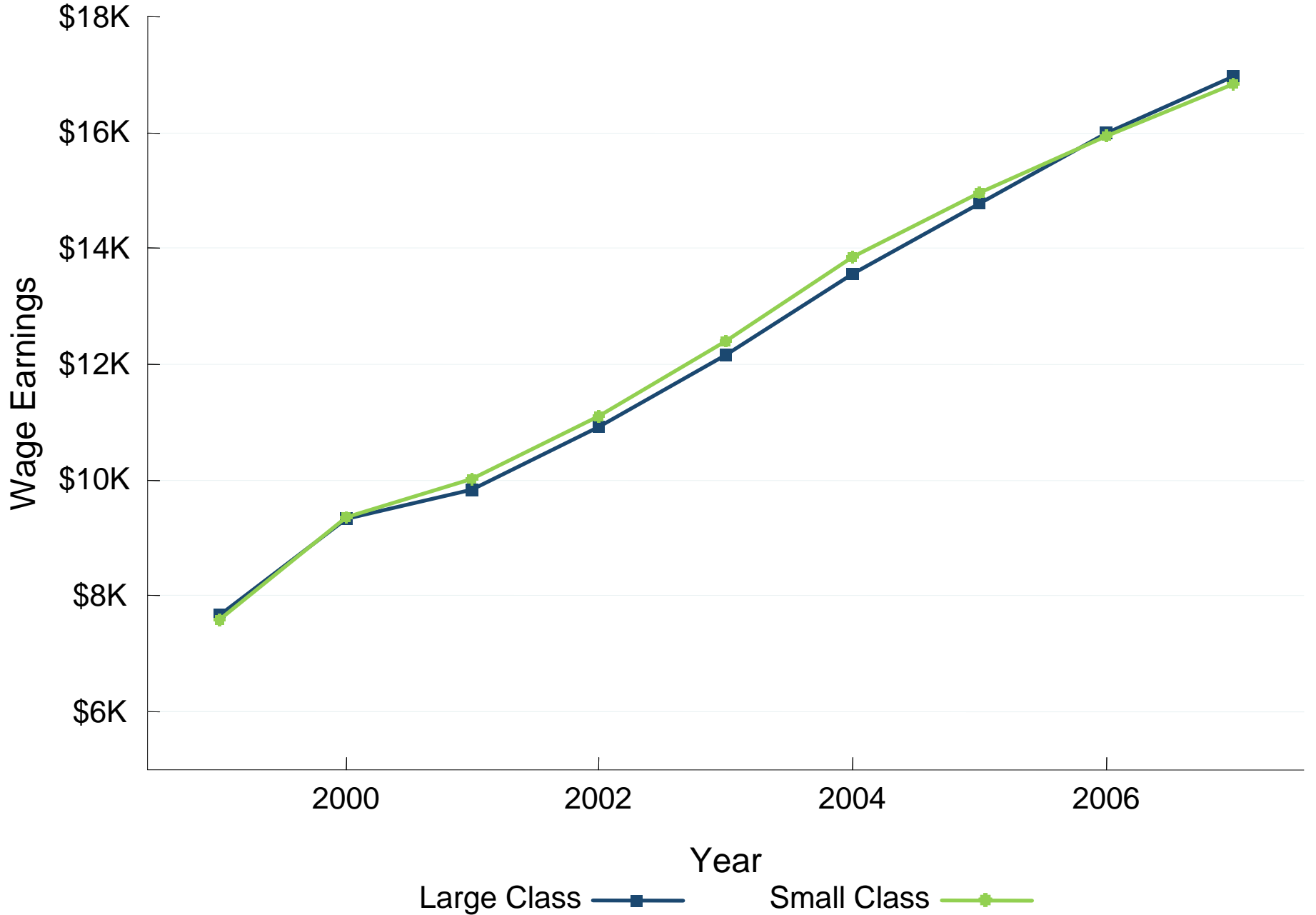


Table 6: Impacts of Class Size on Adult Outcomes

Dependent Var.:	Attended College In 2000 (1)	College Quality (3)	Wage Earnings (4)	Index of Other Outcomes (6)
Small Class	1.78% (0.87%)	\$212.9 (\$145.6)	-\$124.3 (\$324.5)	4.74% (2.05%)
Observations	10,992	10,992	10,992	10,605
Mean of Dep. Var.	26.4%	\$20,147	\$15,912	0.00

Note: All specifications control for school by entry wave effects and class size and the following other controls: a quartic in parental income interacted with marital status, mother's age at child's birth, student gender, free-lunch status, age, and race. Summary index includes a dummy for owning a home, having a 401(k), being married, having moved outside TN, and percent college graduates in 2008 zip.

Part 4: Teacher/Peer Effects

- Students randomly assigned to classes that differ in teacher and peer quality
- Do teachers/peers affect adult outcomes?
 - First examine impacts of observable characteristics of teachers and peers (e.g. teacher experience)
- Throughout remainder of talk, isolate variation across classes within schools *and* class type (small vs. large)
 - Control for school fixed effects and indicator for class size

Figure 5a: Causal Effect of Teacher Experience on Earnings

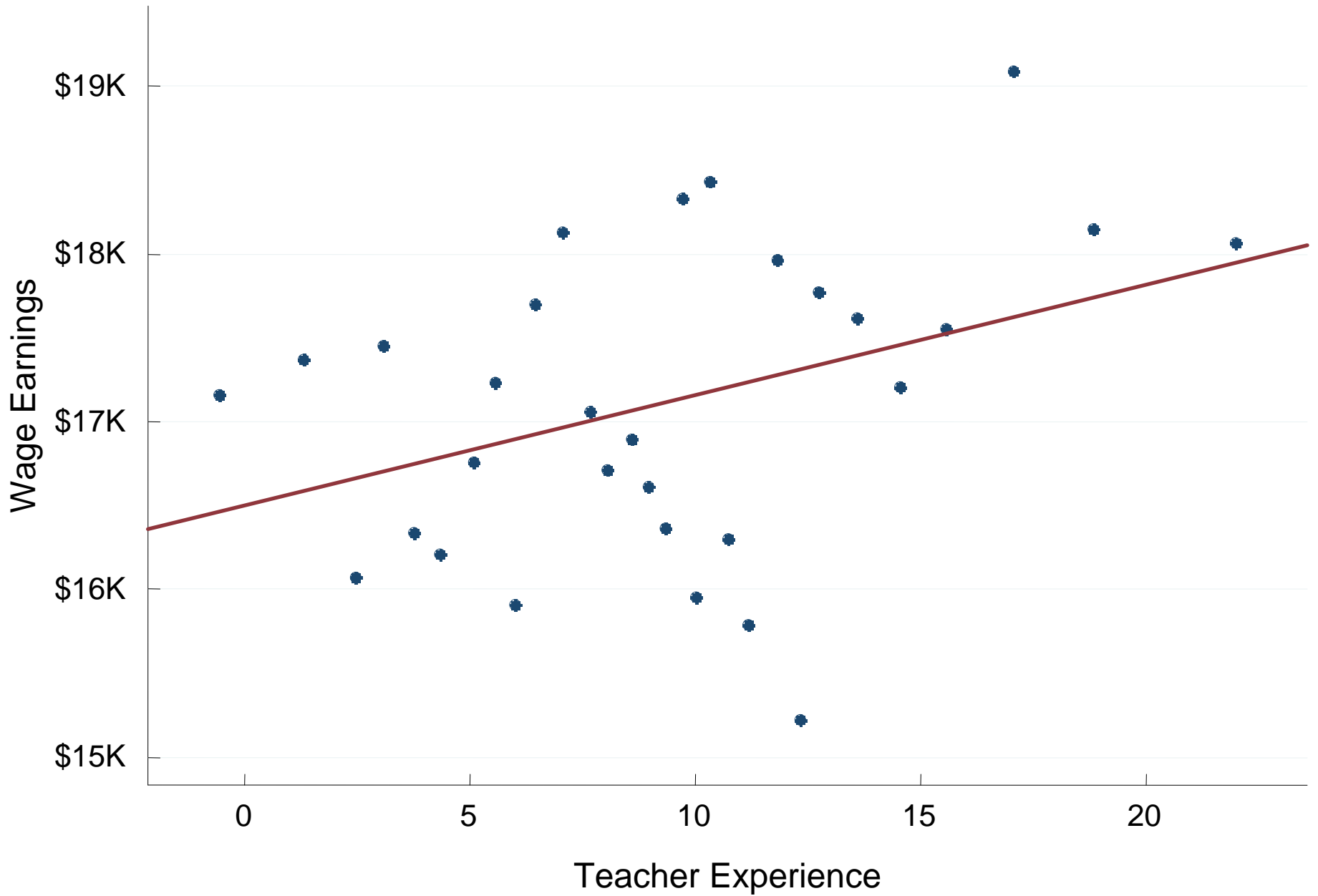


Figure 5b: Effect of Teacher Experience on Earnings by Year

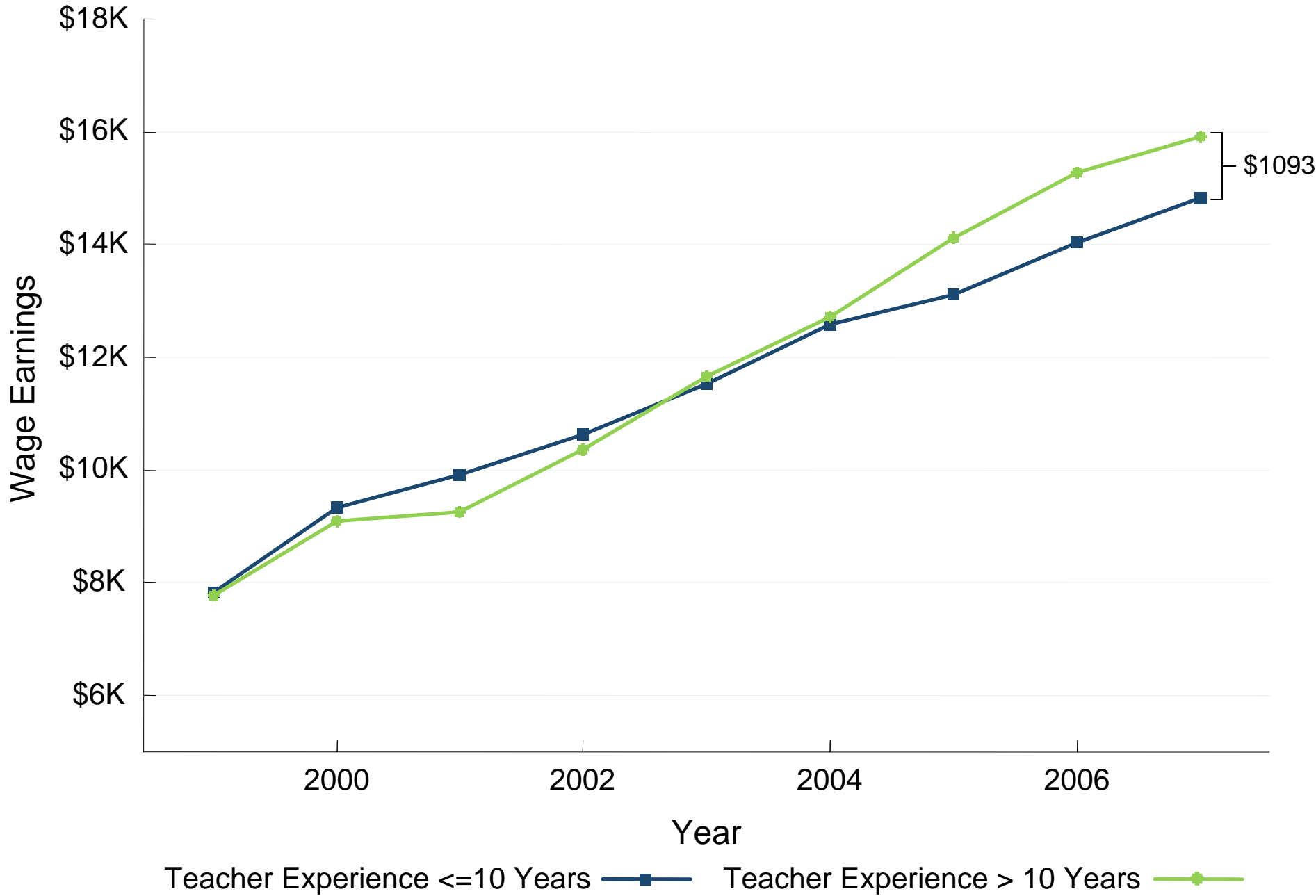


Table 7: Observable Teacher vs. Peer Effects

Dependent Var.:	Wage Earnings			
	(1)	(2)	(3)	(4)
Teacher with >10 Years Experience	\$1093 (\$453.7)			
Teacher Experience (years)		\$57.13 (\$37.70)		
Teacher Degree Higher than a BA	-\$261.1 (\$489.7)	-\$204.7 (\$493.5)		
% Black Peers			-\$1,758 (\$3,063)	
% Female Peers			-\$67.56 (\$1,535)	
% Free-Lunch Peers			-\$284.1 (\$1,593)	
Average Age of Peers			-\$26.67 (\$1,440)	
Predicted Peer Scores				-\$15.91 (\$90.62)
Observations	6,005	6,005	10,992	10,992

Note: All specifications control for school fixed effects and class size, as well as family characteristics listed above.

Class Effects

- Most of teacher and peer quality (e.g. clarity of instruction, enthusiasm) not captured by observable measures
- Well known problem in literature on teacher effects (see e.g. Rockoff and Staiger 2010)
- Modern literature captures unobserved teacher characteristics by estimating distribution of teacher effects on scores
- We use this approach and test for class-level effects on adult outcomes
- This “class effect” includes effect of teachers, peers, and any class-level shocks such as noise outside classroom

Class Effects: ANOVA

- Test for class effects using analysis of variance
- Do earnings vary across classes by more than what would be predicted by random variation in student abilities?
 - F test for significance of class fixed effects
 - Random effects estimate of class-level SD for outcomes

Table 8: F-Tests for Kindergarten Class Effects

Dependent Var.:	Grade	Grade	Wage Earnings		
	K Scores	8 Scores	(3)	(4)	(5)
	(1)	(2)			
P-value of F-Test on KG Class Fixed Effects	0.000	0.411	0.043	0.023	0.020
SD of Class Effects (RE estimate)	8.367	0.001	\$1,514	\$1,536	\$1,703
Demographic Controls				x	x
School Fixed Effects	x	x	x	x	x
Indicator for Small Class	x	x	x	x	
Large Classes Only					x
Observations	5,621	4,448	6,025	6,025	4,208

Note: All specifications control for school fixed effects and class size as well as family background controls listed above.

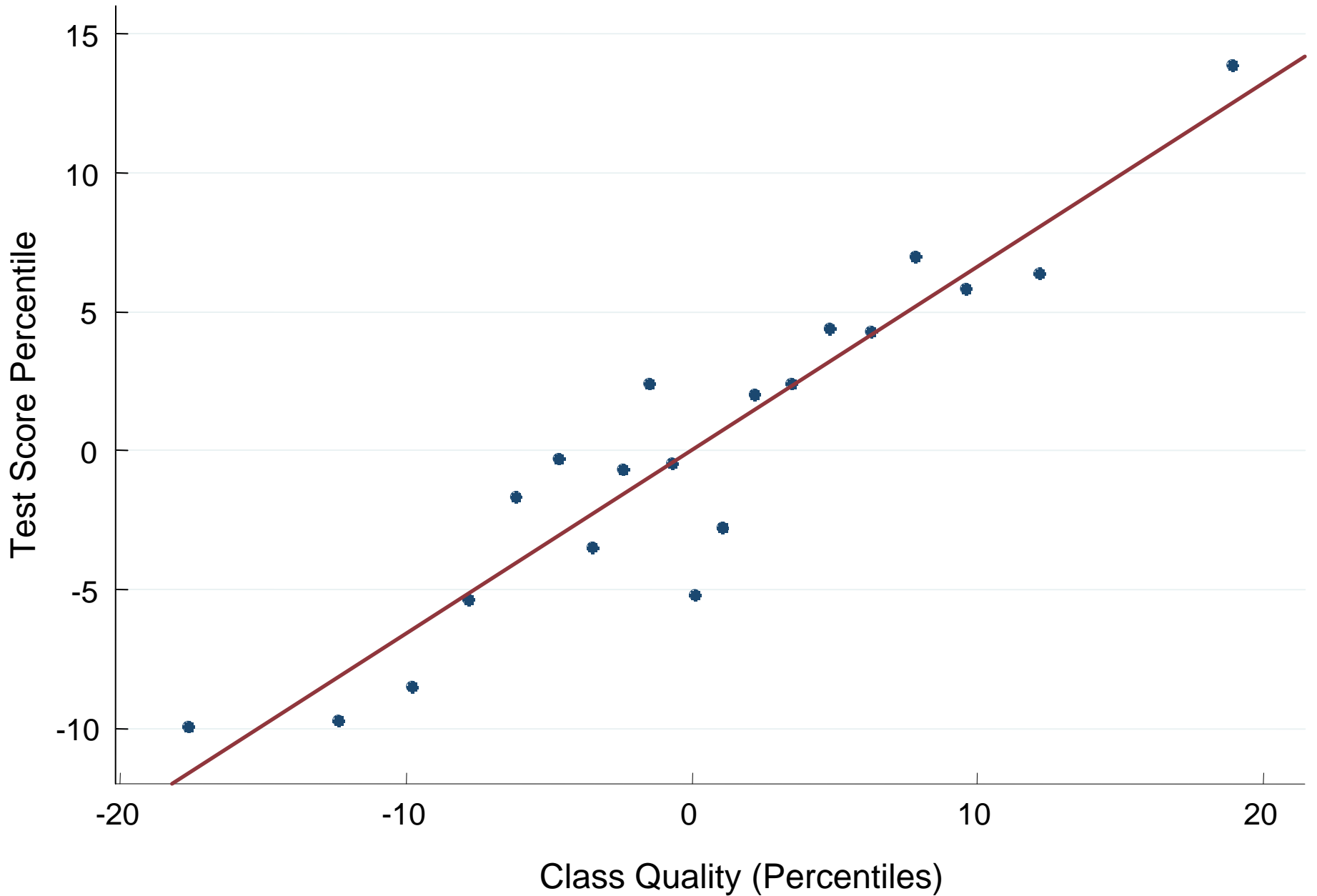
Class Effects on Scores and Earnings

- Key question: Are class effects on KG scores correlated with class effects on earnings?
 - Do (unobserved) interventions that raise test scores also improve adult outcomes?
- Cannot simply regress score class f.e.'s on earnings class f.e.'s
 - Class fixed effect includes a student's own test score → bias toward OLS regression of earnings on scores
 - Equivalent to a weak-instruments problem because class size is finite
 - We address this using a leave-out mean (i.e. jackknife)

Peer-Score Measure of Class Quality

- Measure unobserved “class quality” using leave-out mean:
 - $\text{Class Quality} = \text{Mean Peer Scores in Class} - \text{Mean Peer Scores in School}$
 - How good are your classmates’ scores, compared with the classmates you could have had?
- Because we use peers’ test scores at the *end* of KG, class quality reflects teacher + peer effects + random class-level shocks
- Good teachers and class environments raise peer test scores
- Because students were randomly assigned to classes, class quality varies randomly within schools

Figure 6a: Causal Effect of Class Quality on Own Score



Peer-Score Measure of Class Quality: Placebo Test

- To confirm that leave-out mean captures causal effect of class quality on own score, run a placebo test
 - Predict each student's score from regression of actual score on pre-determined demographics and parental background
 - Test whether class quality is correlated with predicted test score
 - Effectively a test for balance in student characteristics across classes of different quality

Figure 6b: Actual Test Score vs. Predicted Test Score

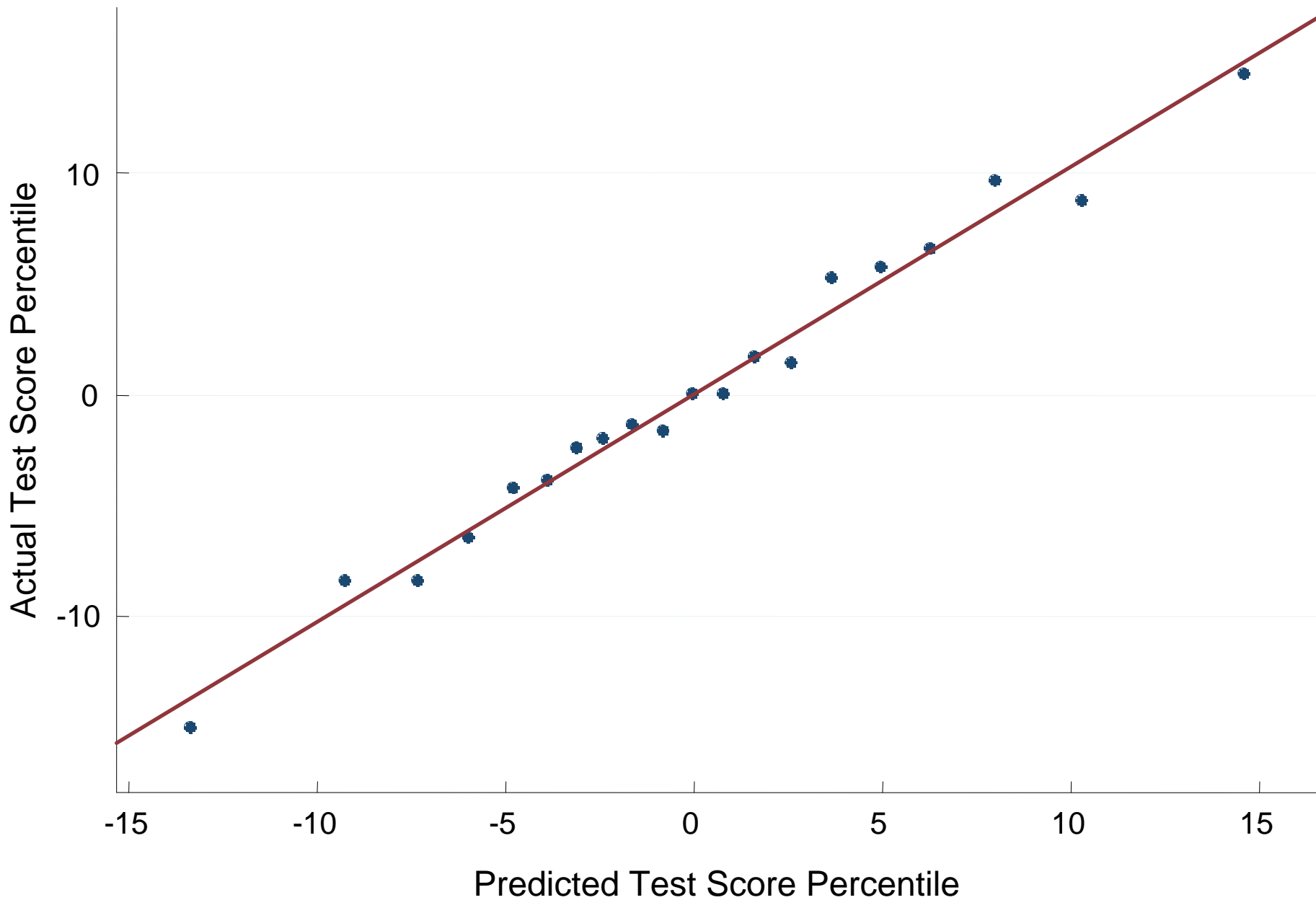


Figure 6c: Placebo Test: Class Quality and Predicted Own Score

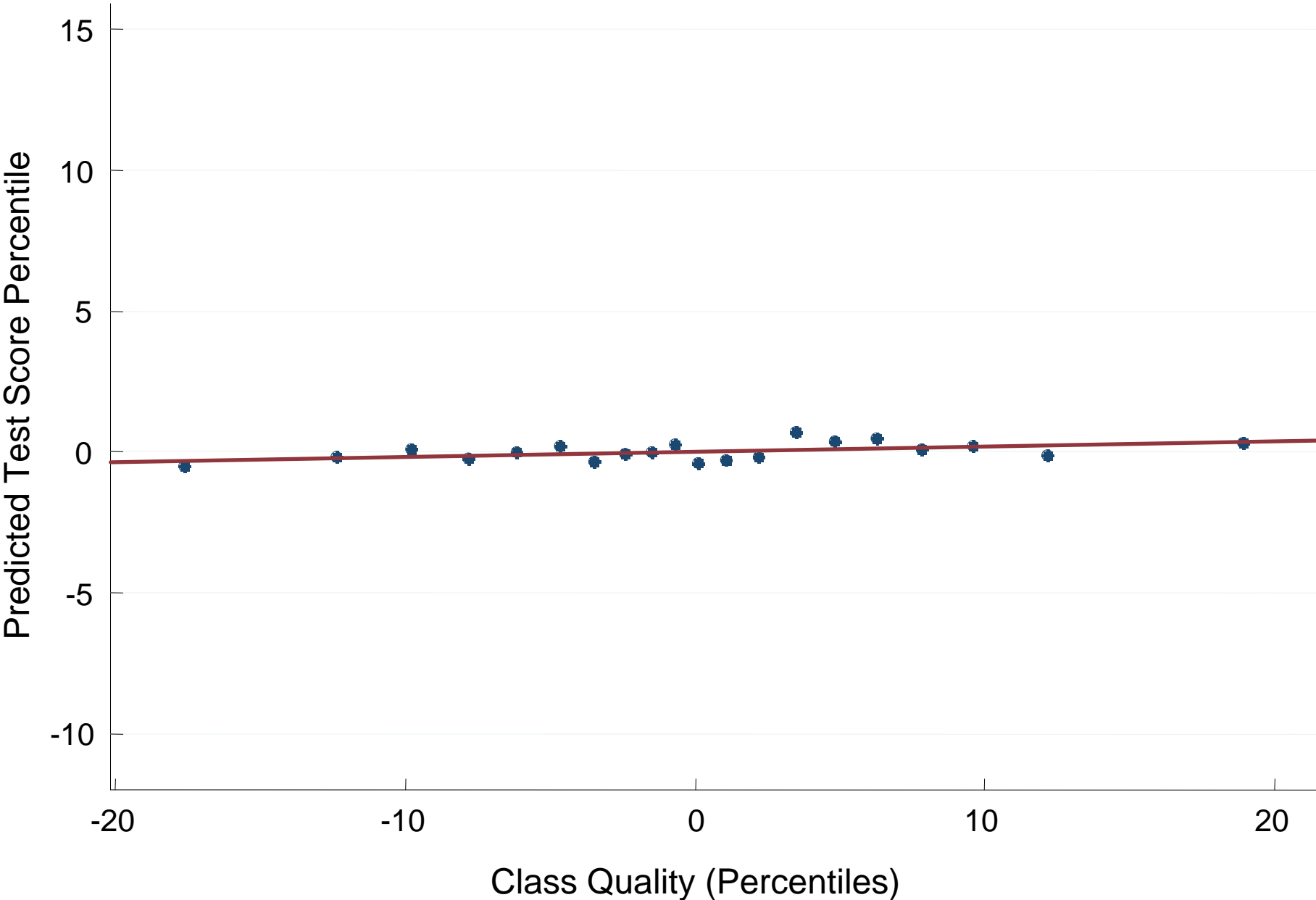


Figure 6d: The Causal Effect of Kindergarten Classrooms on Earnings

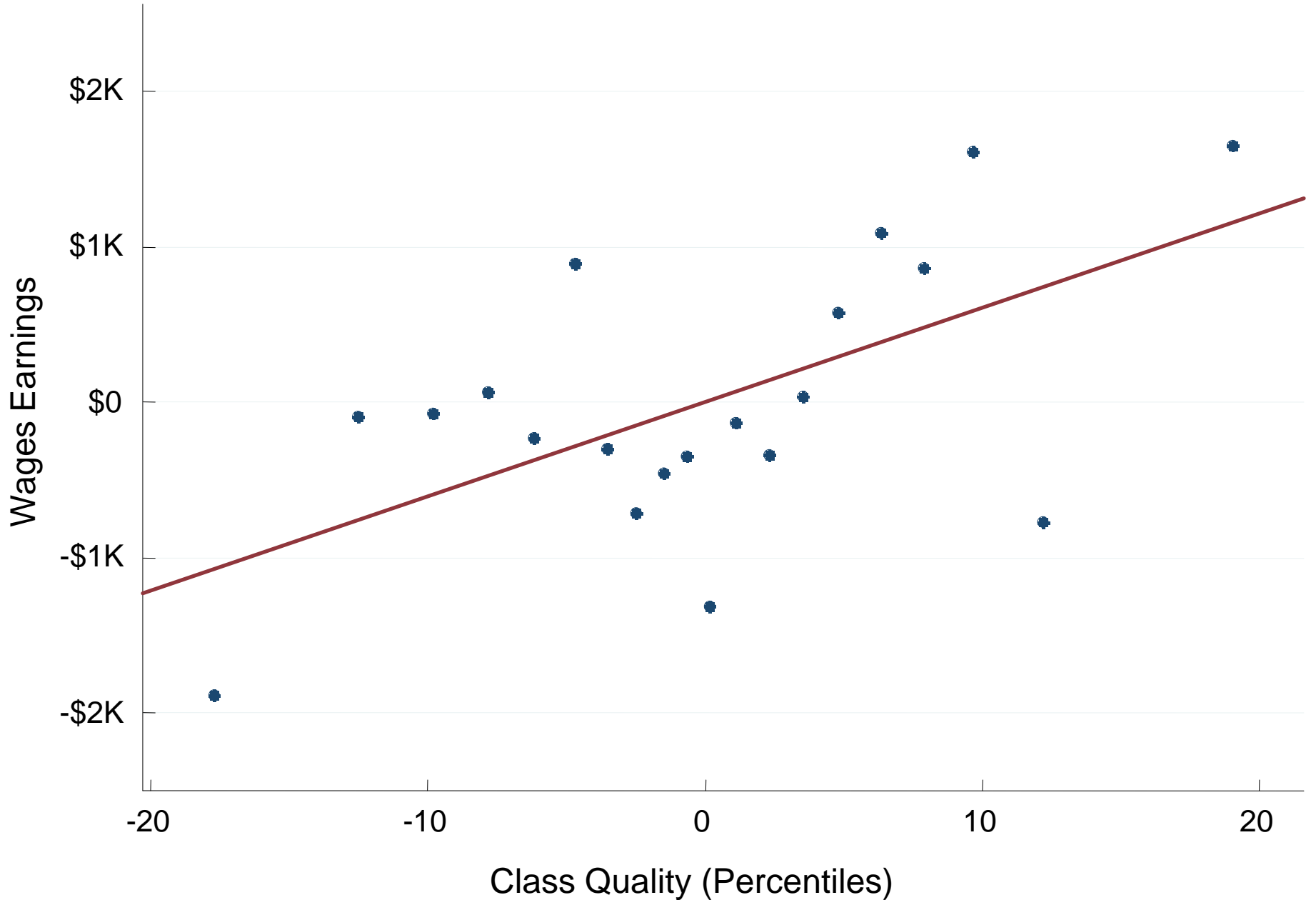


Figure 6e: Placebo Test: Class Quality and Predicted Earnings

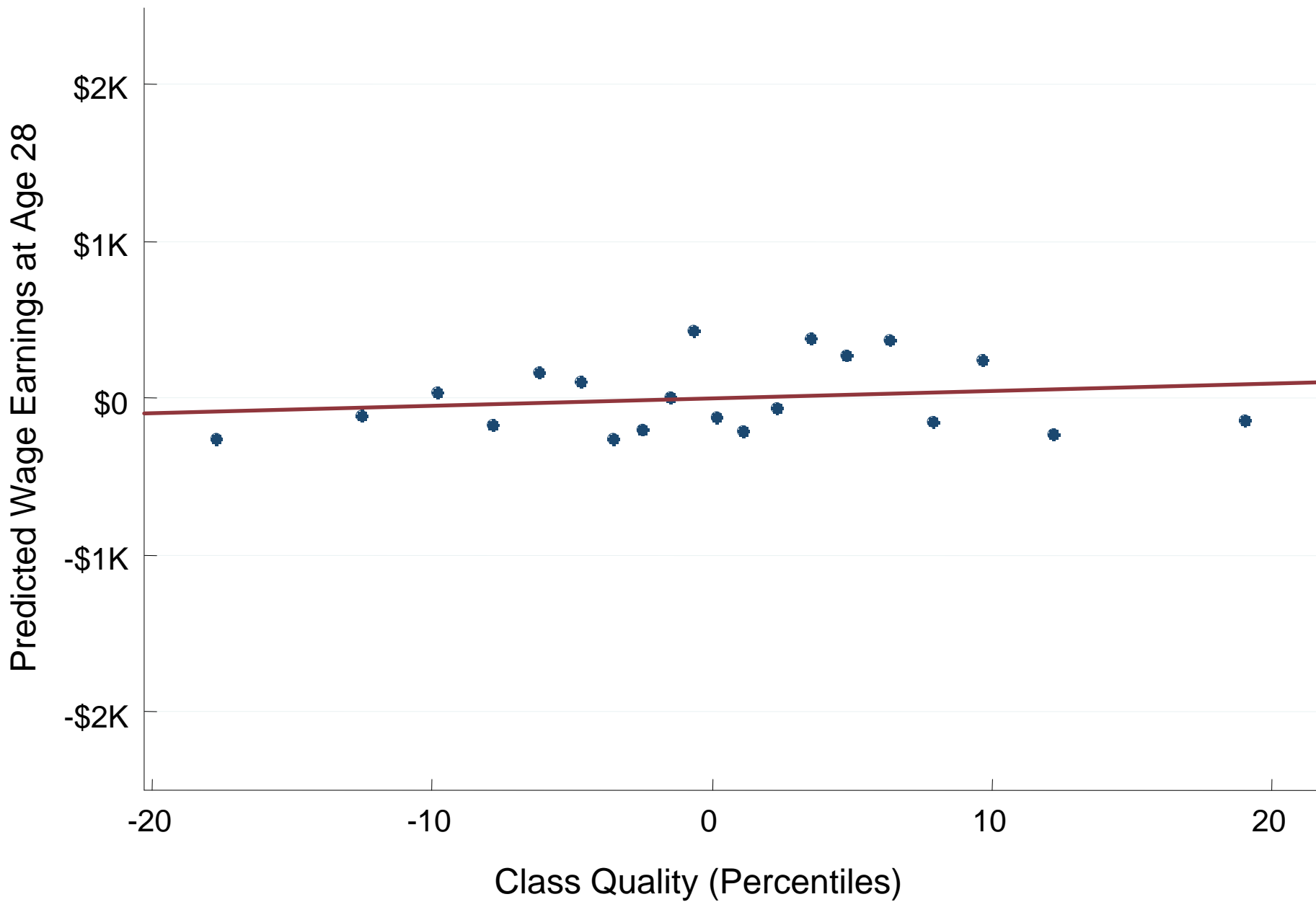


Figure 6f: Effect of Class Quality on Earnings by Year

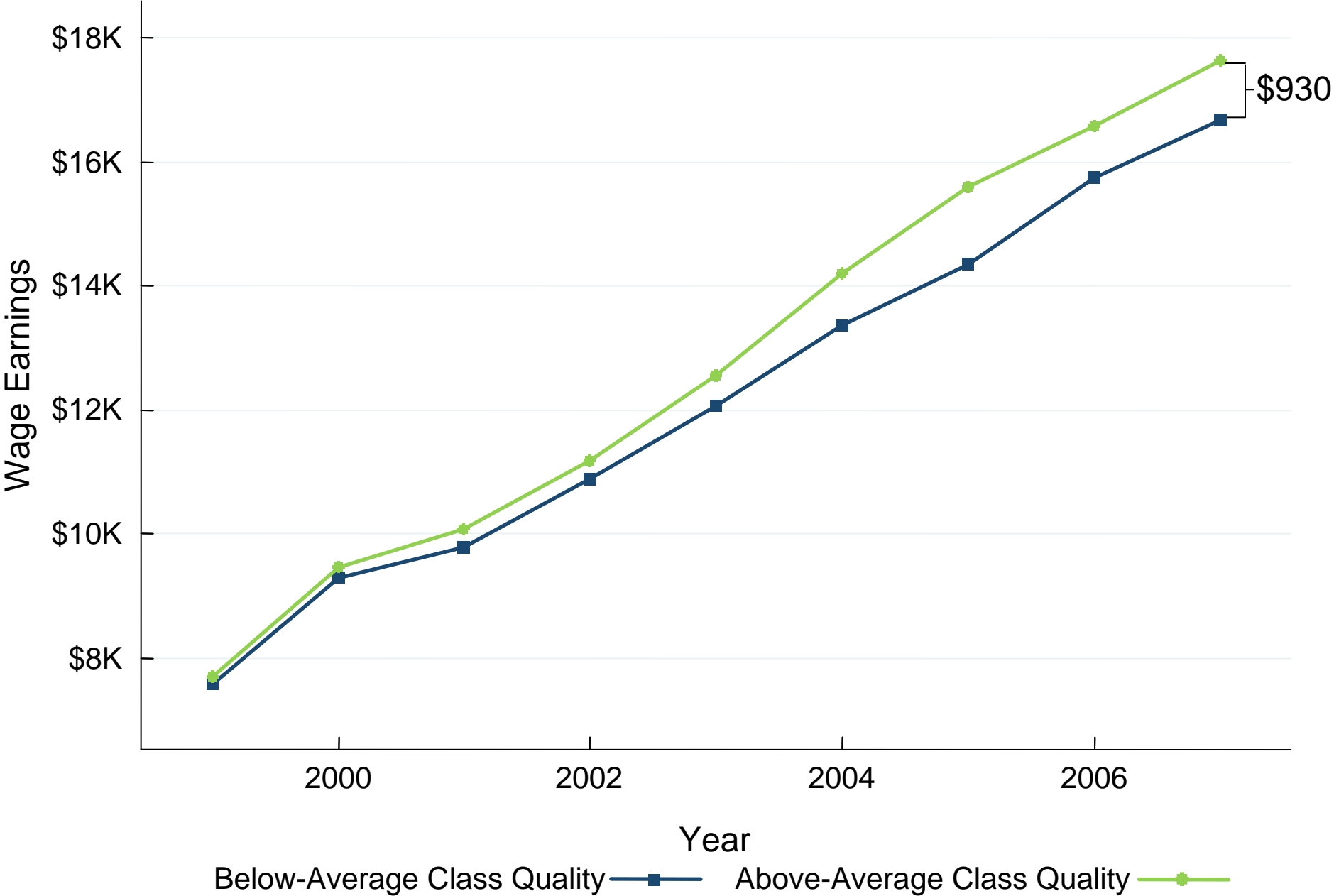


Table 8: Impacts of Class Quality on Adult Outcomes

Dependent Var.:	Test Percentile	Wage Earnings	College in 2000	Ever College	College Quality	Summary Index
	(%)	(\$)	(%)	(%)	(\$)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)
Class Quality	0.66	50.60	0.10	0.11	17.20	0.21
	(0.02)	(15.35)	(0.04)	(0.05)	(6.90)	(0.10)

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, mother's age at child's birth, parent savings, student gender, free-lunch status, and race.

Figure 7: Fadeout of Class Effects
Effect of 1 SD of Class Quality on Test Scores by Grade

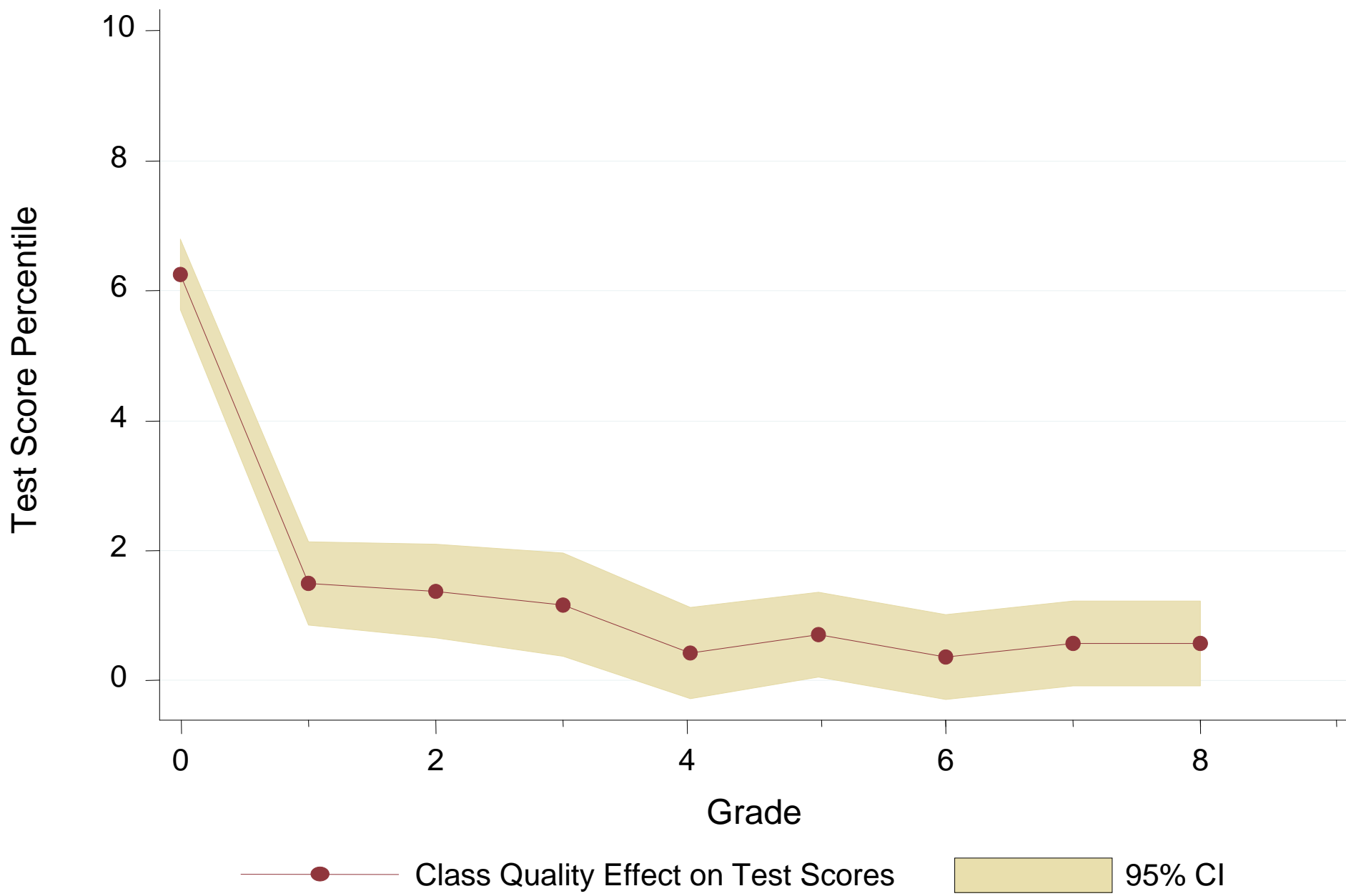
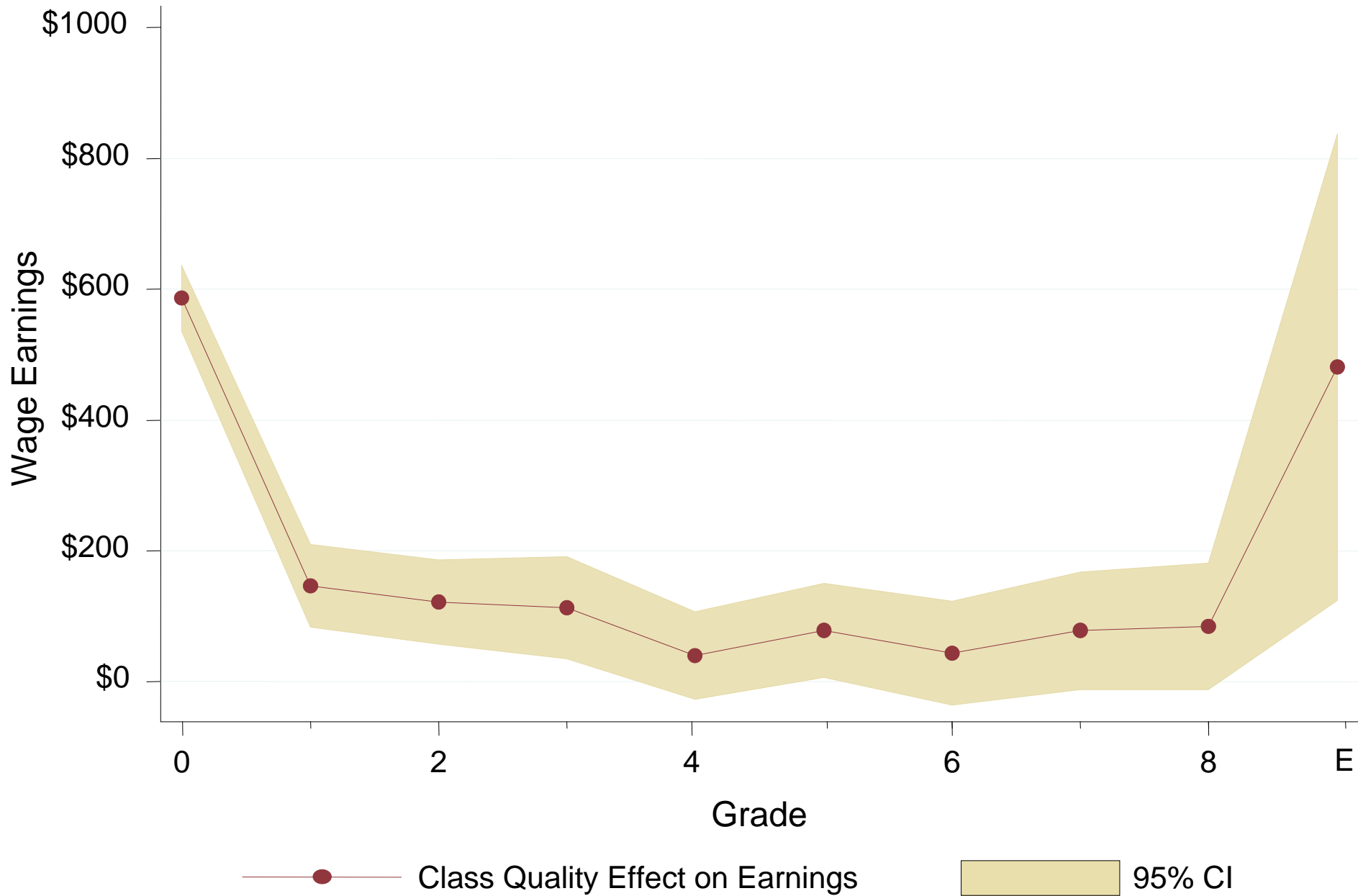


Figure 7: Fadeout of Class Effects
Effect of 1 SD of Class Quality on Earnings



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

- Why do effects of kindergarten class fade out on test scores and then reappear in adulthood?
- One explanation: non-cognitive skills (Heckman 2000)
- Data on non-cognitive measures was collected for random subset of STAR students in 8th grade
 - Effort, initiative disruptive behavior rated by English and Math teachers
 - Convert mean score to percentile scale
- Do effects of KG class quality persist on non-cognitive skills?

Table 9: Impacts of Class Quality on Non-Cognitive Skills

Dependent Variable:	Wage Earnings	Grade 4 Scores		Grade 8 Scores	
		Test	Non-Cognitive	Test	Non-Cognitive
	(1)	(2)	(3)	(4)	(5)
Grade 4 Non-Cognitive	\$87.68 (\$19.55)				
Grade 4 Test Score	\$36.42 (\$23.32)				
Class Quality		0.047 (0.040)	0.153 (0.070)	0.064 (0.037)	0.129 (0.061)
Observations	1,360	4,023	1,671	1,780	1,772

Cost-Benefit Analysis: Class Quality

- Assume: 3% real discount rate, constant percent income gains, income follows average US income profile
- One SD increase in KG class quality leads to \$1,536 (9.7%) increase in wage earnings at age 27
 - Lifetime earnings gain of **\$39K** for average individual in US
 - Total present value earnings gain for class of 20 students of **\$784,000** from a 1 SD improvement in class in a *single year*
- This calculation highlights stakes at play in early childhood education
- But not directly useful for policy because we cannot manipulate class quality directly

Cost-Benefit Analysis: Class Size Reduction of 33%

- Method 1: use estimated earnings impacts at age 27
 - NPV earnings gain per student of \$1K, 95% CI = **(-\$9K,17K)**
- Problem: estimates imprecise and earnings impacts may not appear by age 27
- Method 2: project earnings gains based on estimated impacts of class quality on scores and earnings
 - 1 pctile improvement in class quality → 0.52 percentile gain in scores and \$50 increase in earnings (\$95 per pctile gain)
 - Class size reduction → 5 pctile gain in scores = \$475
 - NPV earnings gain per student of **\$11,500** per student
- Benefits may be larger than per-student cost of \$9,355

Cost-Benefit Analysis: Teacher Quality

- Method 1: use estimated impacts of teacher experience
 - 1 SD increase in teacher experience raises NPV earnings by **\$7,500** per student
- Method 2: project earnings gains based on estimated impacts of class quality on scores and earnings
 - Rockoff (2004) and Kane and Staiger (2008) estimate that a 1 SD improvement in teacher quality raises scores by 0.2 SD
 - We estimate \$95 per pctile score gain → 1 SD improvement in teacher quality raises earnings by \$515 (3%) at age 27
 - 1 SD increase in teacher quality yields NPV earnings gains of **\$13,000** per student
 - **\$260K** for a class of 20 students