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

July 2010

Motivation

- Large literature on how interventions such as better teacher quality or smaller classes affect achievement as measured by test scores

- Much less evidence on whether interventions that increase test scores improve outcomes such as earnings

- Problem: 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 evaluate the long-term impacts of education interventions

→ Question: Are higher test scores a good proxy for improvements in adult outcomes?

    - Do small classes and better teachers/peers improve adult outcomes to the extent they improve test scores?

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 has investigated impacts of STAR on test scores

    - Students in small classes have 5 percentile point (0.2 sd) higher test scores in K-3 (Krueger 1999)

- 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)

→ Concern that early test score gains may not 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 information from W-2 forms

  - College attendance information for all kids from 1098-T forms

- 93.4% 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)	$14,790	$13,808
Fraction With Zero Earnings ('05-'07)	13.6%	34.3%
Attended College in 2000 (age 20)	26.5%	44.1%
Ever Attended College by age 28	46.7%	49.9%
Parental Household Income ('00-'07)	$39,030	$37,953
Fraction Black	36.4%	48.1%

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. 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
    - parent age at child's birth
    - parent's IRA + 401K contributions
    - filing parent's gender, child's gender, free lunch status, race

Figure 1: Wage Earnings in 2007 vs. KG Test Score
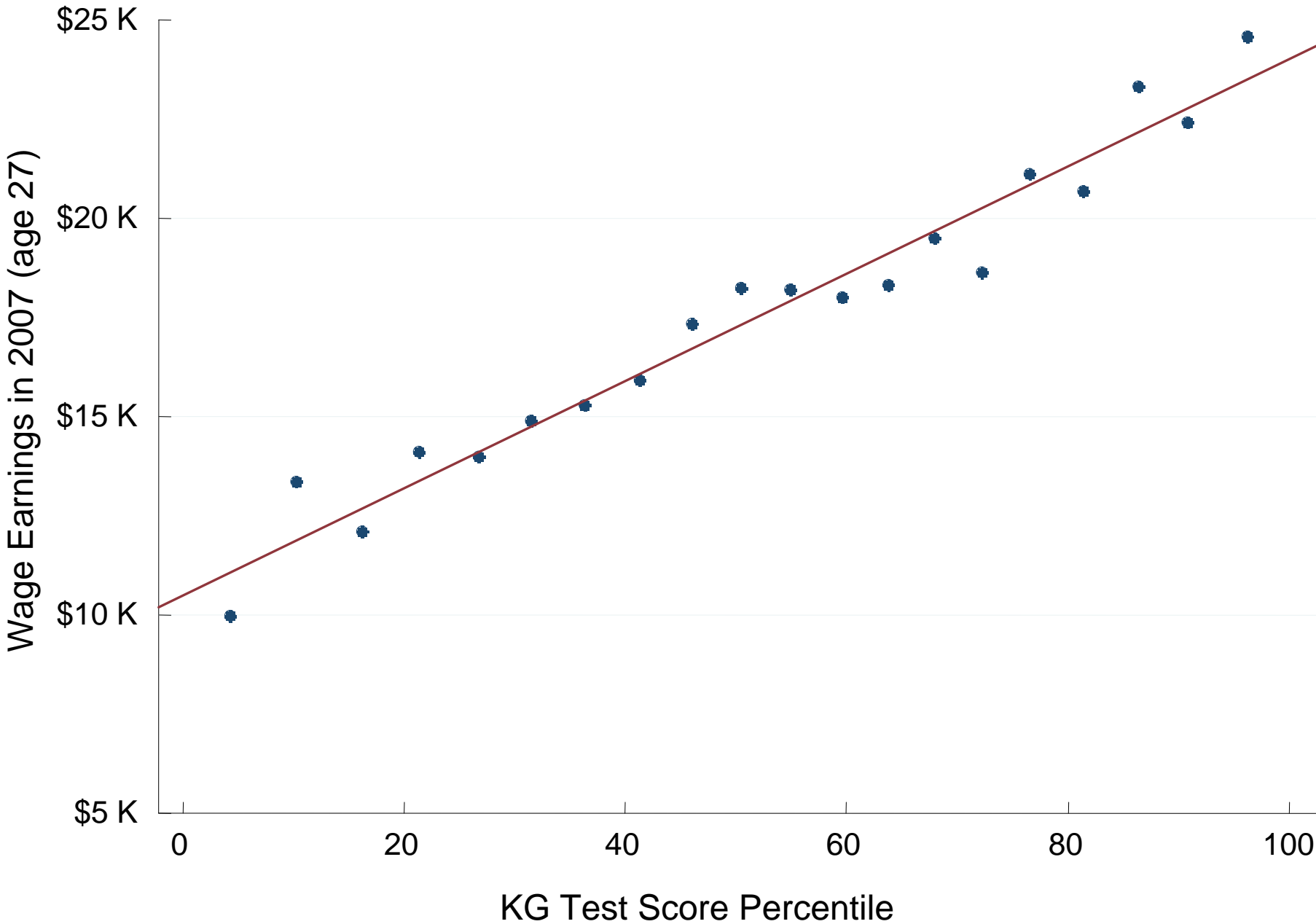
Table 2: Test Scores and Earnings in the Cross-Section

Dependent Var.:	Wage Earnings			Log Wage Earnings	Wage Earnings
	(1)	(2)	(3)	(4)	(5)
KG Test percentile	$119.01	$128.44	$81.21		$97.80
	($6.75)	($8.26)	($8.42)		($8.31)
KG Test z score				0.174	
				(0.027)	
Parental Income Percentile					$119.74
					($6.79)
Black			-$326.70	0.070	
			($779.30)	(0.086)	
Parental + Demog. Controls			x	x	
Class Fixed Effects		x	x	x	x
Adjusted R^2	0.05	0.09	0.17		0.15
Observations	5,609	5,609	5,609		5,609

Note: Parental controls are a quartic in parental household income interacted with marital status, parent age at child's birth, filing parent's gender, child's gender, and free lunch

**Figure 2a: Percentage Attending College in 2000 vs KG Test Score**
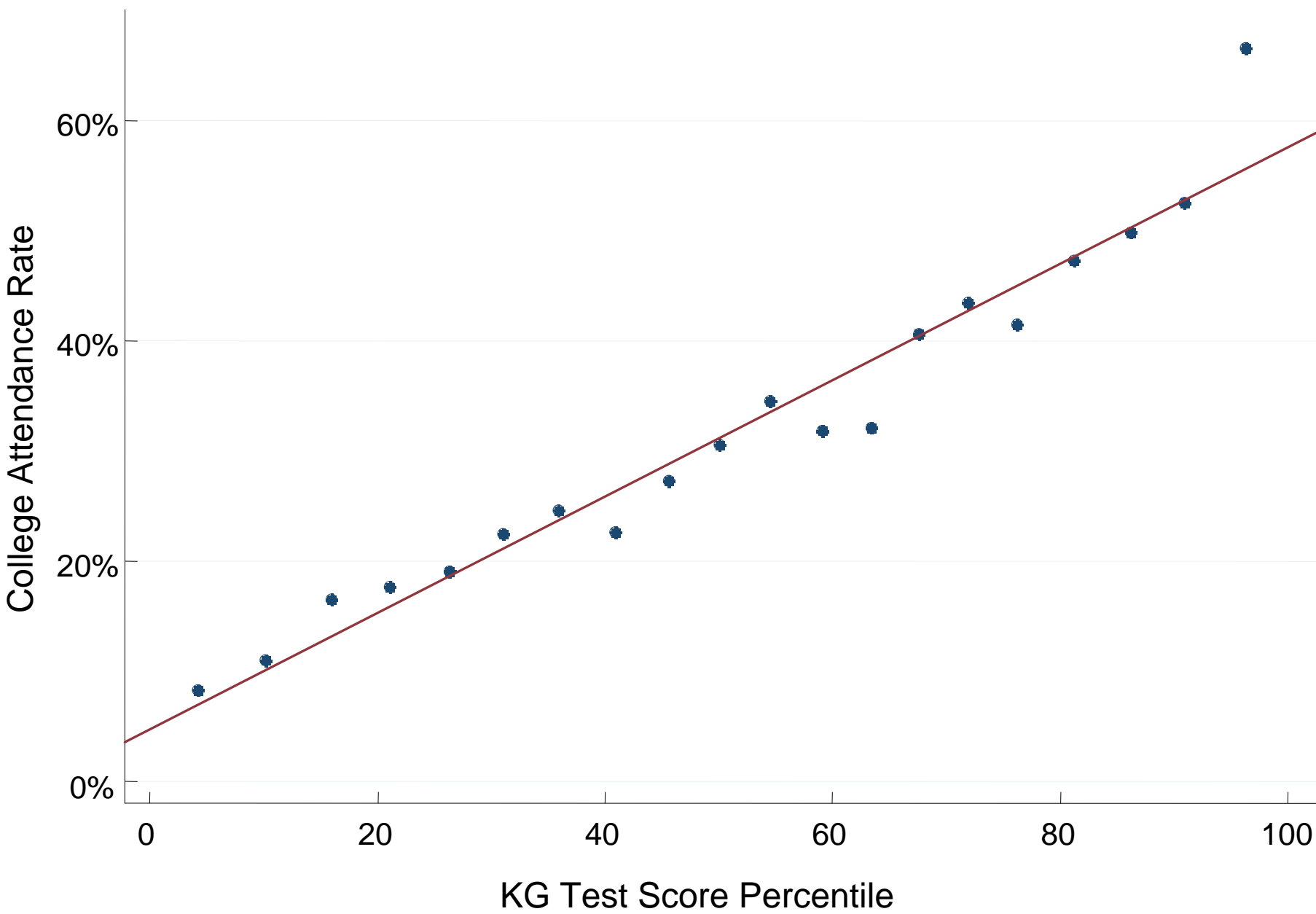
Figure 2a.1: Percentage Ever Attending College vs KG Test Score
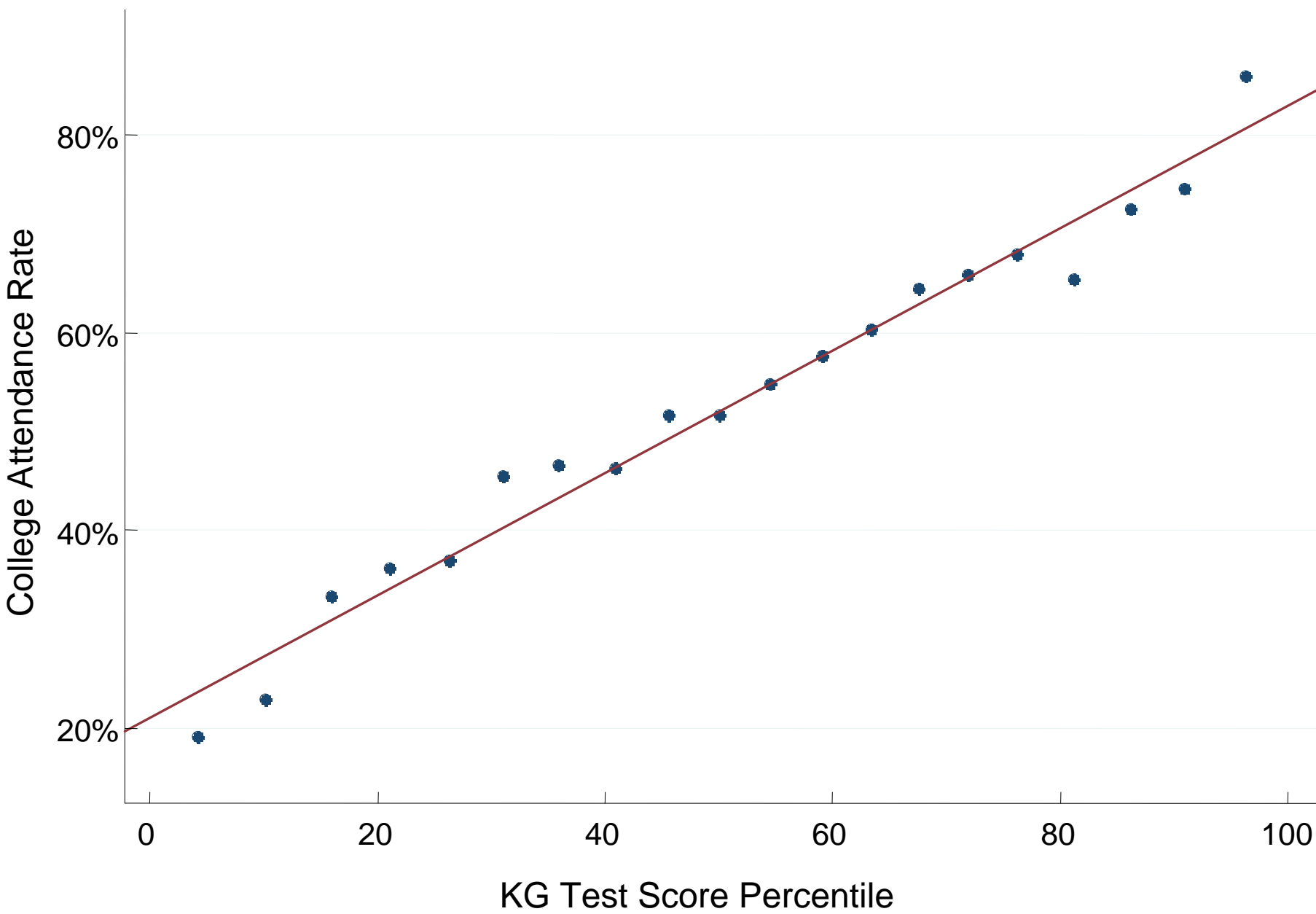
An Earnings-Based Index of College Quality

- 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
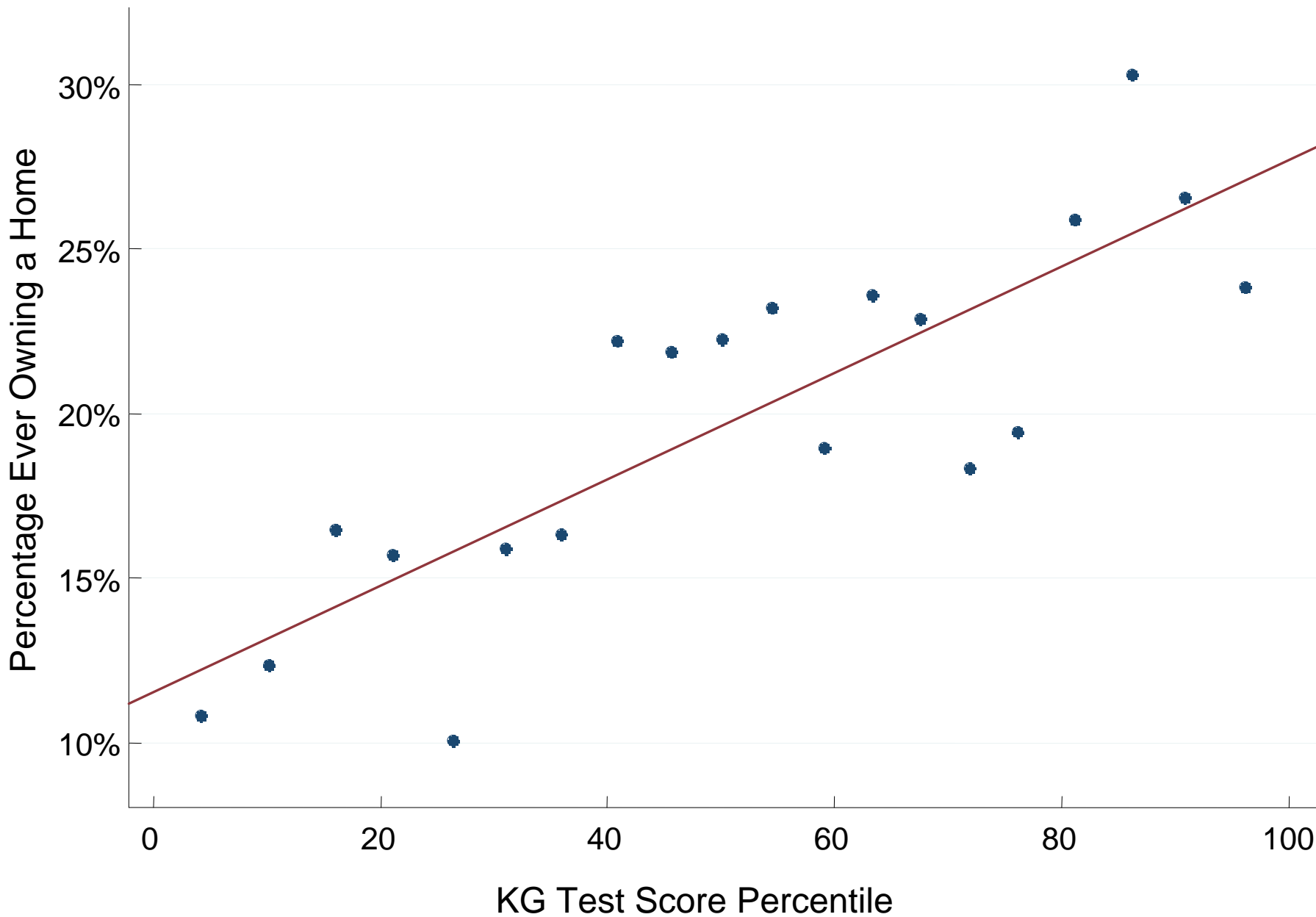
College Mean Wage Earnings by US News Ranking
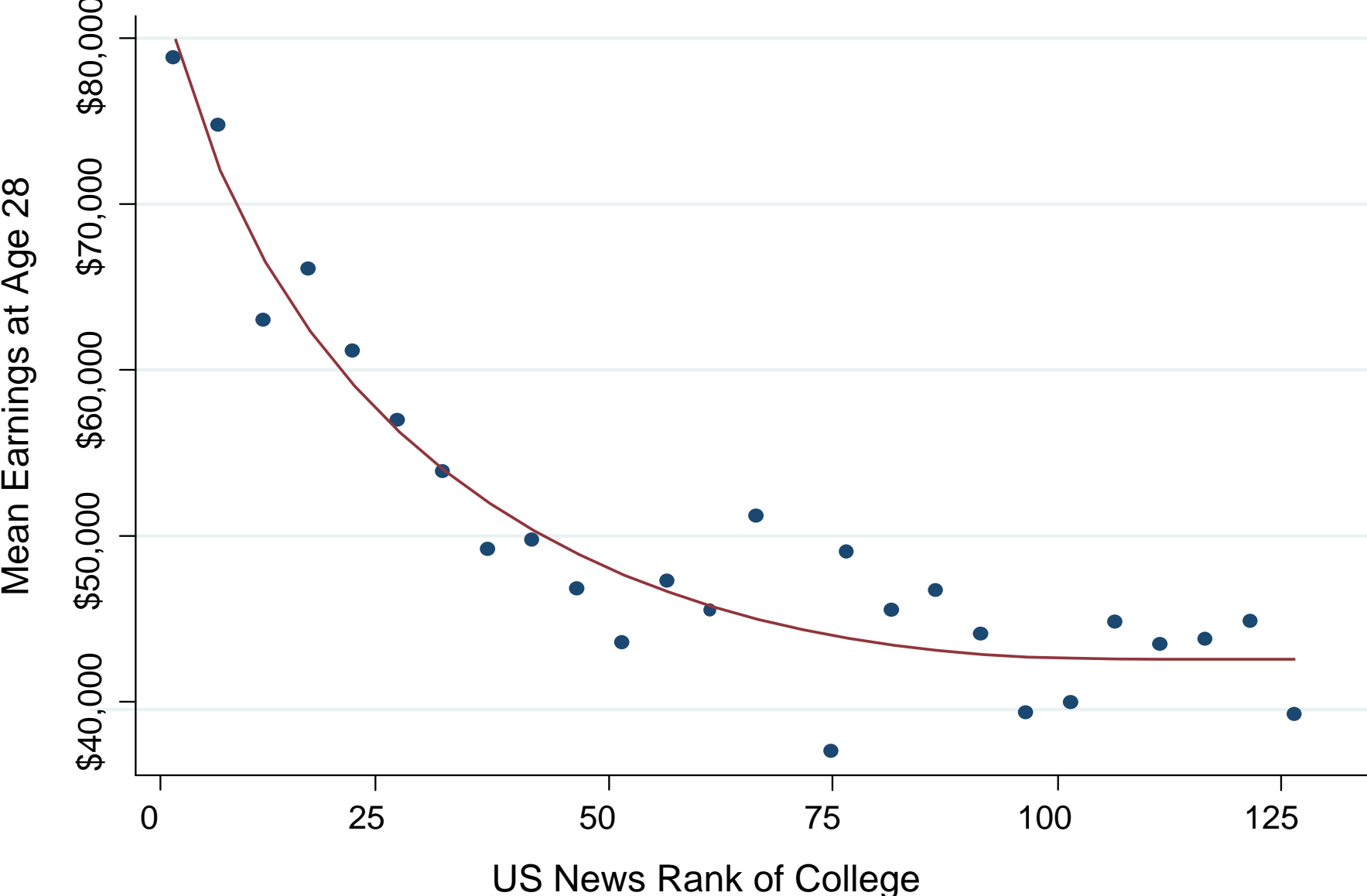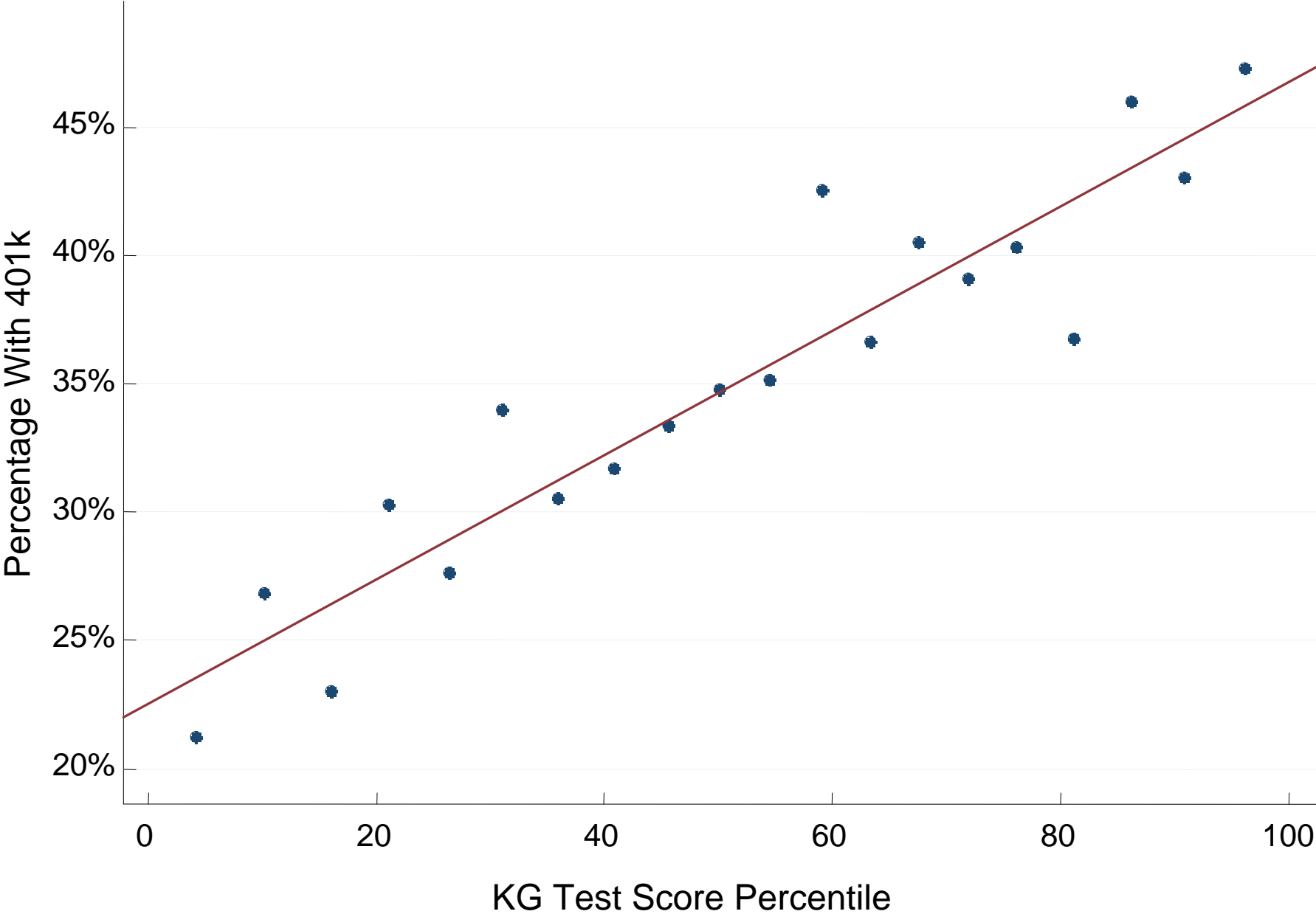
Figure 2b: College Quality in 2000 vs. KG Test Score
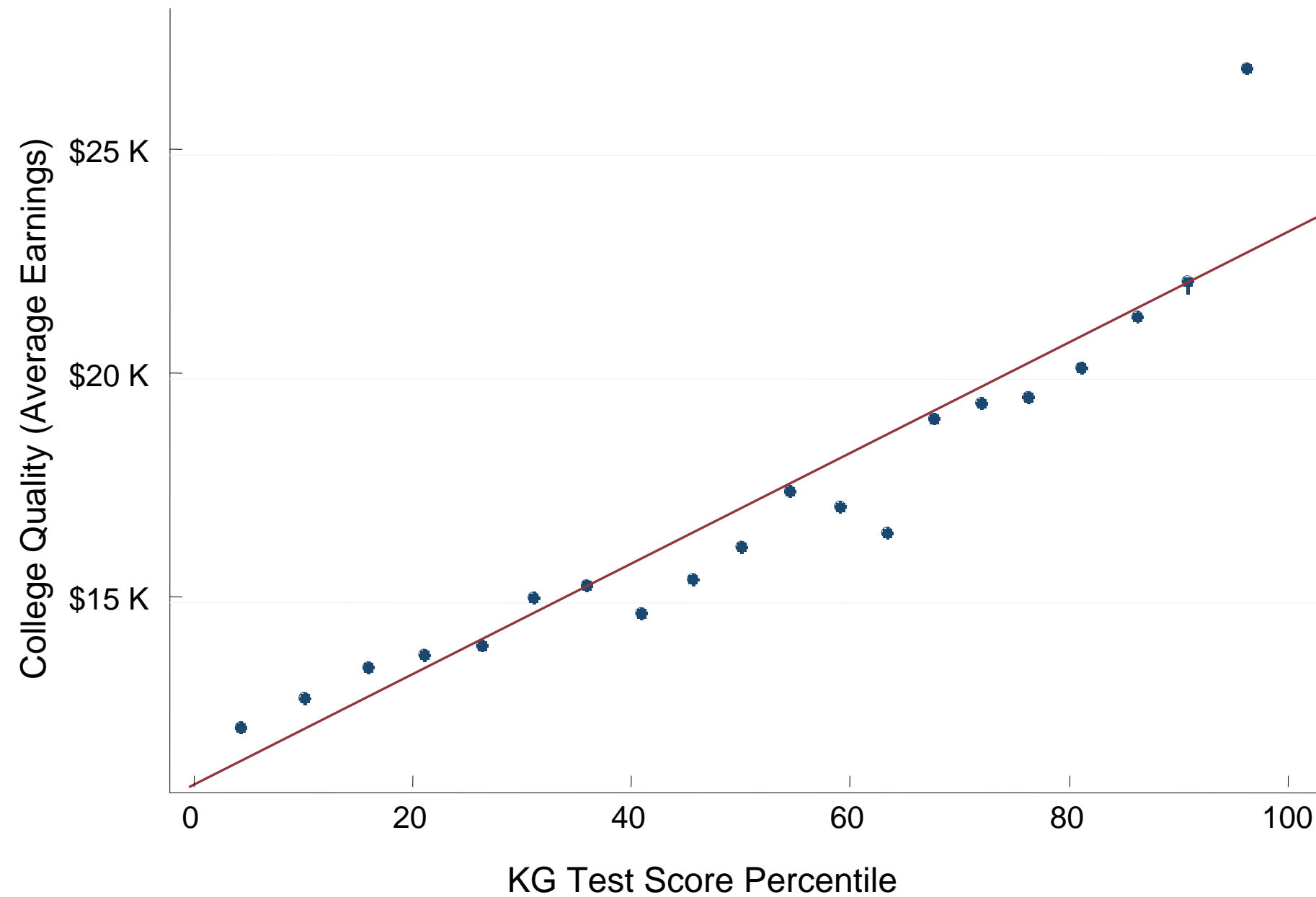
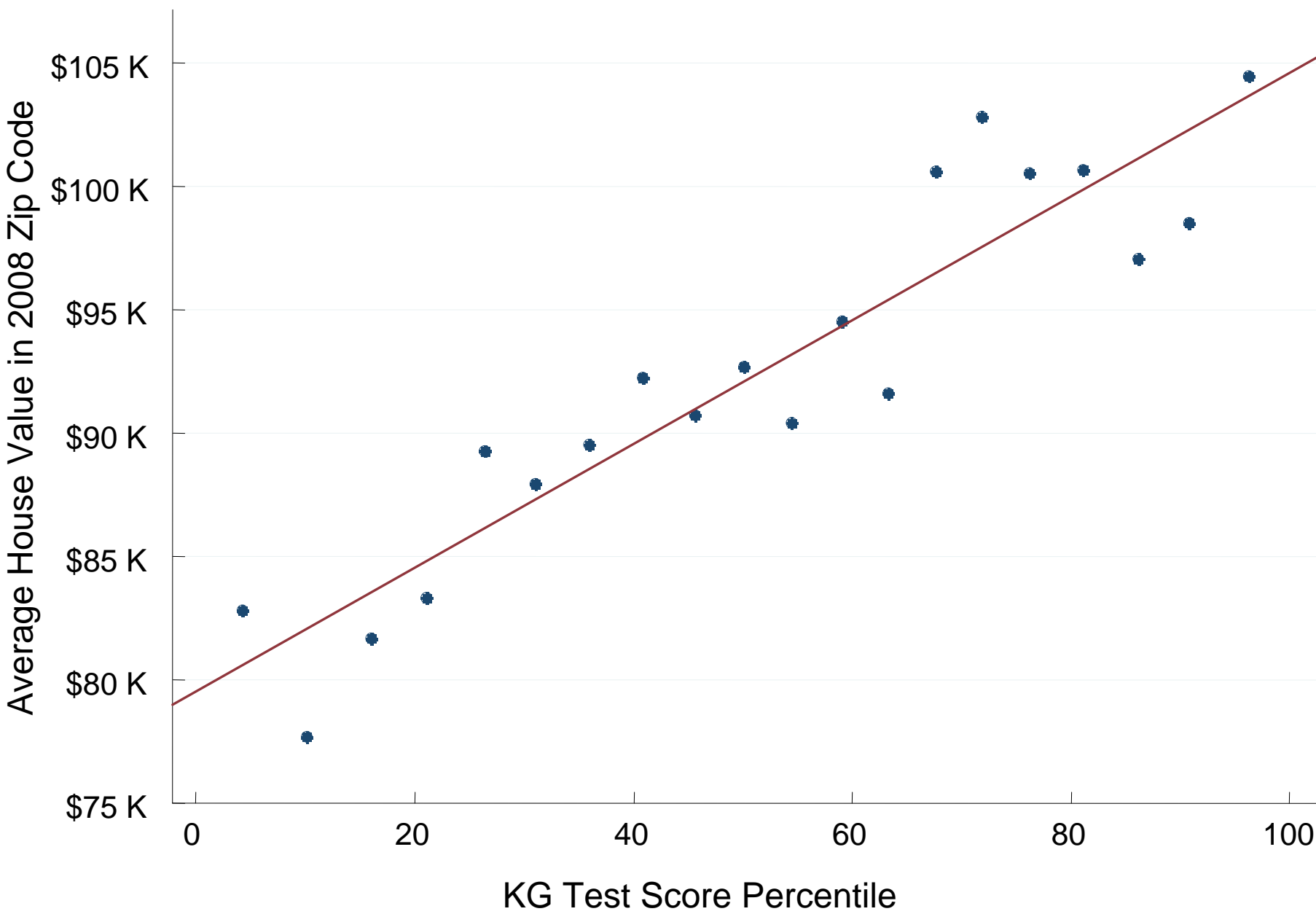**Figure 3a: Percentage Owning a Home by Age 28 vs. KG Test Score**

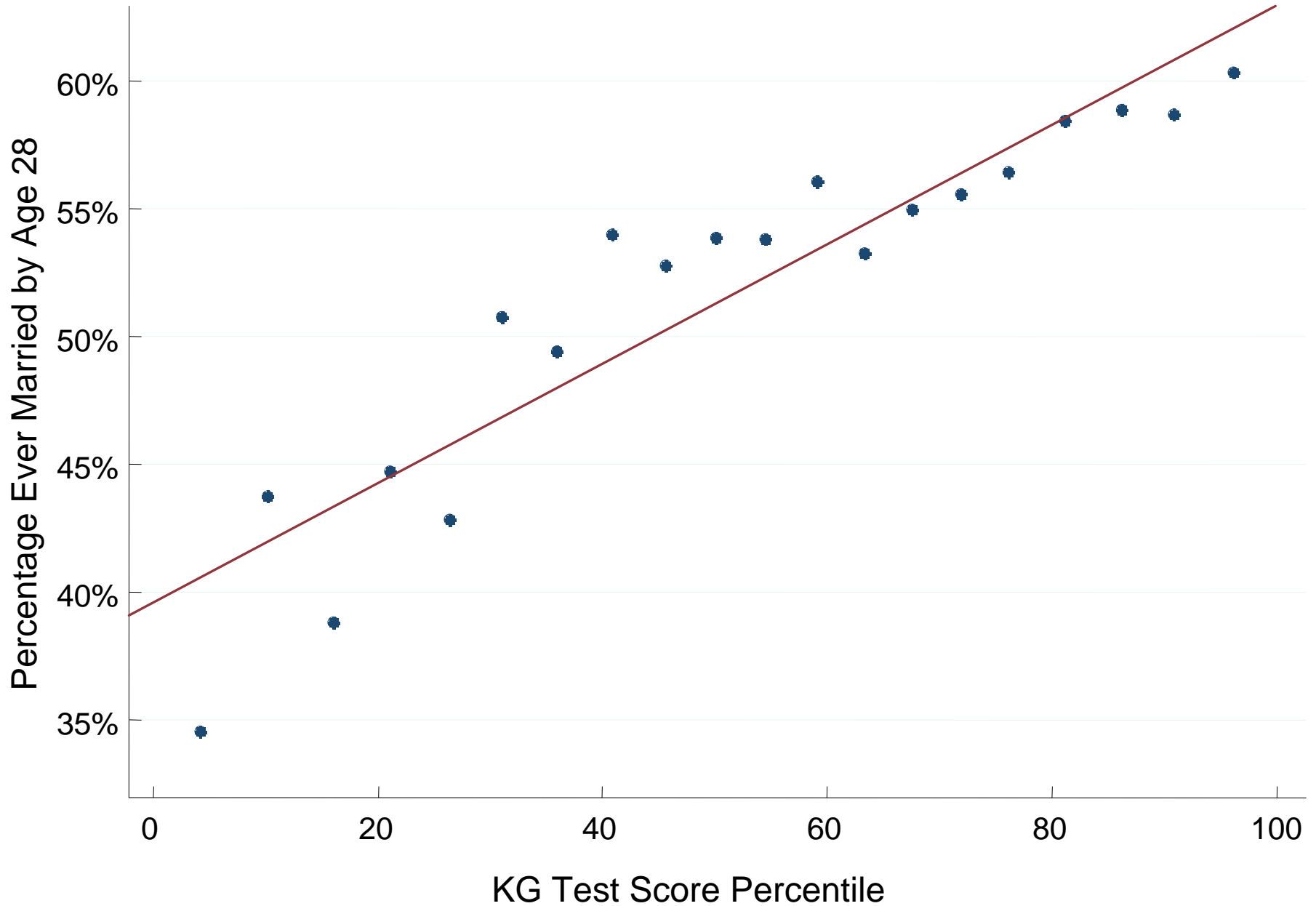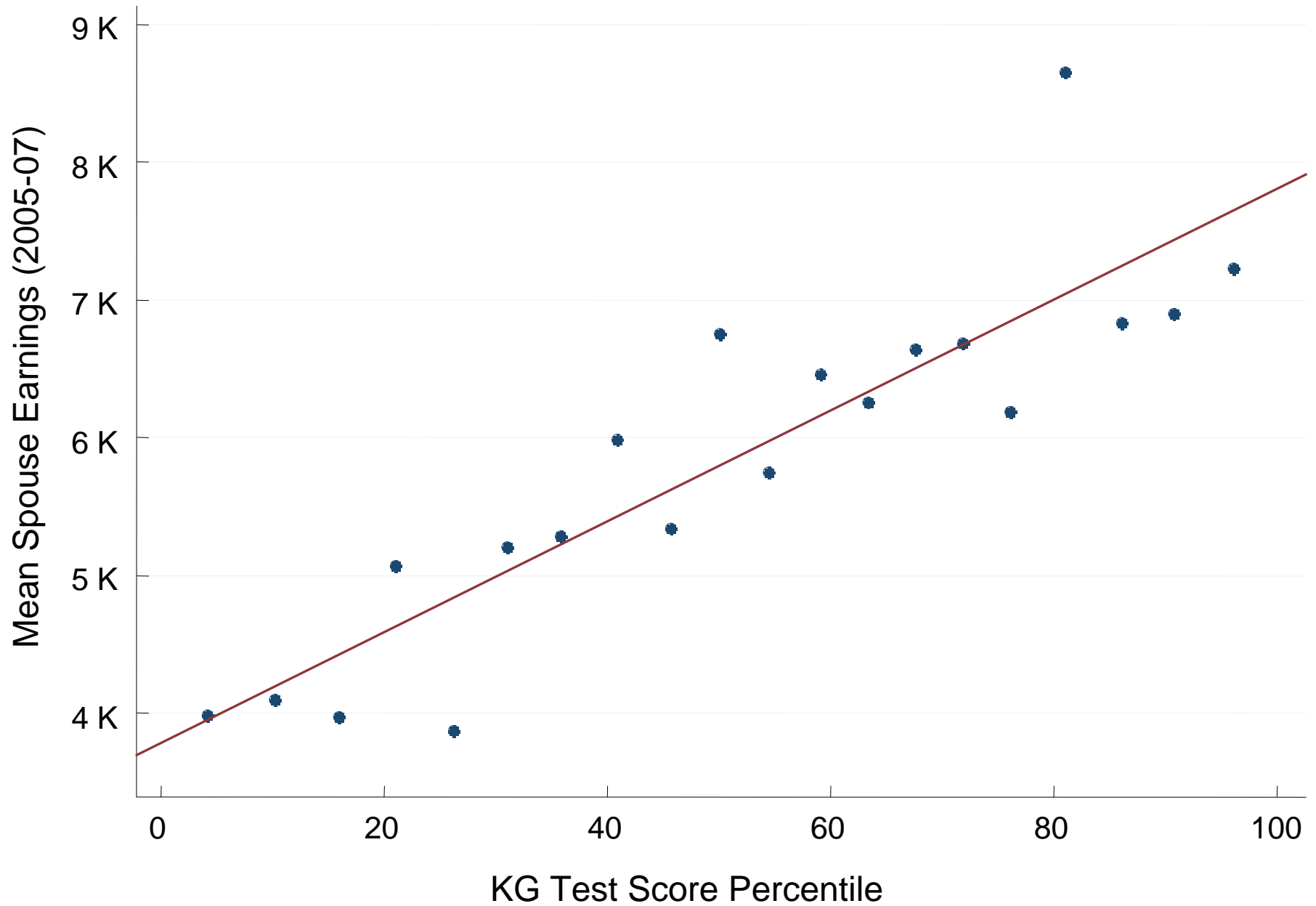Figure 3b: Percentage With 401k by Age 28 vs. KG Test Score

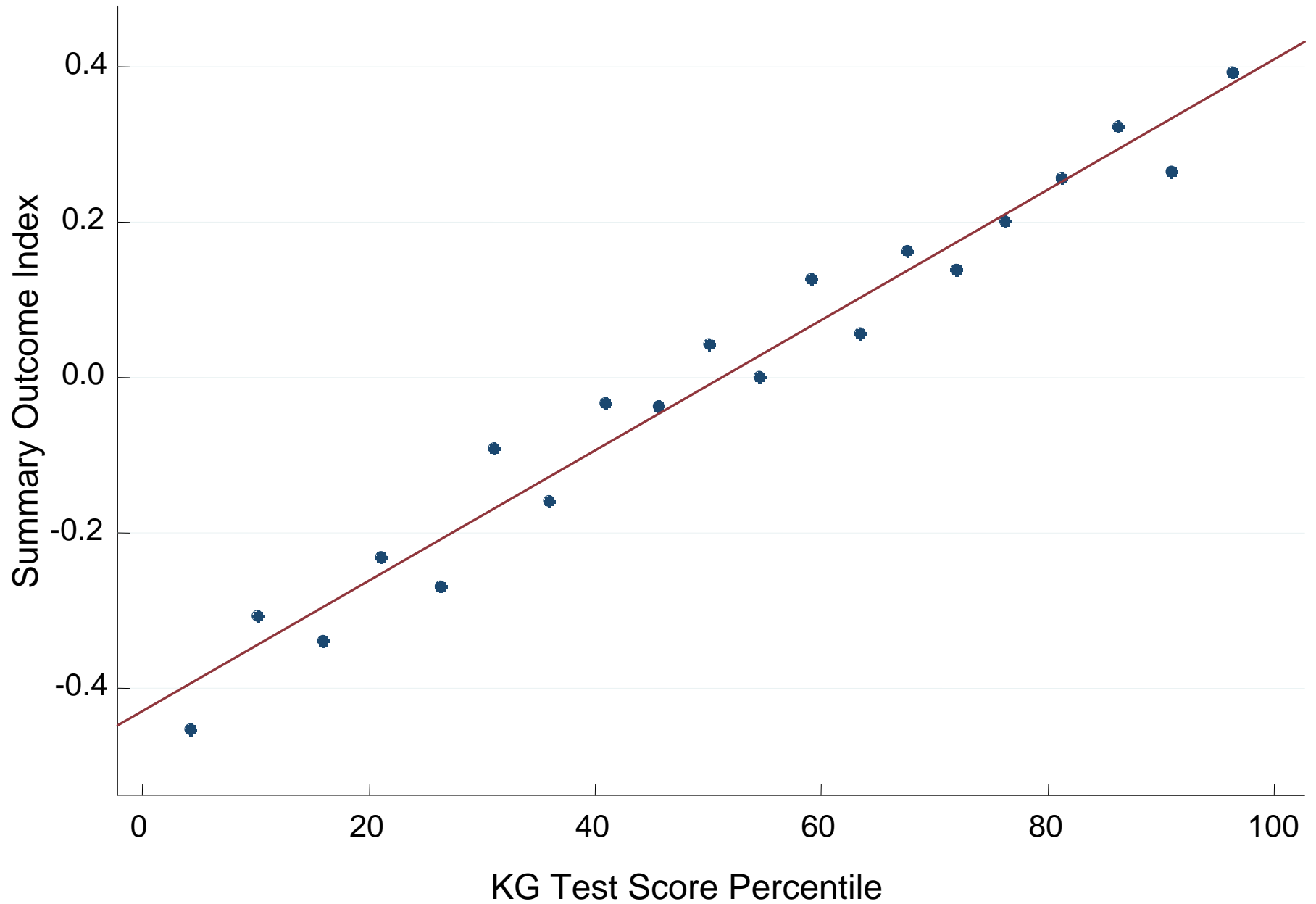**Figure 3c: Average House Value in 2008 Zip Code vs. KG Test Score**

Figure 3d: Percentage Married by Age 28 vs. KG Test Score

Figure 3e: Spouse Earnings vs. KG Test Score

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

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 per	0.10	0.00%
Claiming Parent ($1,000s)	(0.01)	(0.01%)
	[12.52]	[0.16]
Parent's Age at STAR Birth	0.19	0.05%
	(0.03)	(0.06%)
	[5.57]	[0.78]
Parent's 401k+IRA Savings	0.86	-0.15%
($1,000s)	(0.19)	(0.35%)
	[4.51]	[-0.44]
Female	3.52	0.01%
	(0.46)	(0.83%)
	[7.71]	[0.02]
Black	-10.77	0.69%
	(0.90)	(1.62%)
	[-11.98]	[0.43]
p-Value on F-Statistic	0.000	0.979
Observations	10,306	10,919

Note: Regressions include school fixed effects.

Validity of the STAR Experimental Design

- Threat #2: *Selective Attrition*

  - 50% of children starting in KG are missing scores in grade 3; 28% missing scores in grade 8

- Much less attrition here because we follow 93% of the sample

- Test for selective attrition in our data in three ways:

  1. Does match rate vary across treatment groups?

  2. Does death rate vary across treatment groups (Muennig et al. 2010)?

  3. Do characteristics of attriters vary across treatment groups?

Table 5: Match Rates and Death Rates

Dependent Var.:	Matched		Dead	
	(1)	(2)	(3)	(4)
Small Class Mean	93.68%	93.73%	1.69%	1.69%
Large Class Mean	94.61%	94.59%	1.59%	1.59%
Difference	-0.93%	-0.86%	0.10%	0.10%
	(0.55%)	(0.51%)	(0.27%)	(0.27%)
Class Fixed Effects	x	x	x	x
Controls		x		x

F-Test of Match Rate on Class Fixed Effects: p value = 0.95

Part 3: Class Size Impacts

- Replicate specifications in previous studies to estimate effect of class size on long-term outcomes

- 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:

    Index = 401K + Home Owner + Married + Spousal Earnings +
                    Moved (Leave TN) + Average Home Value in Zip

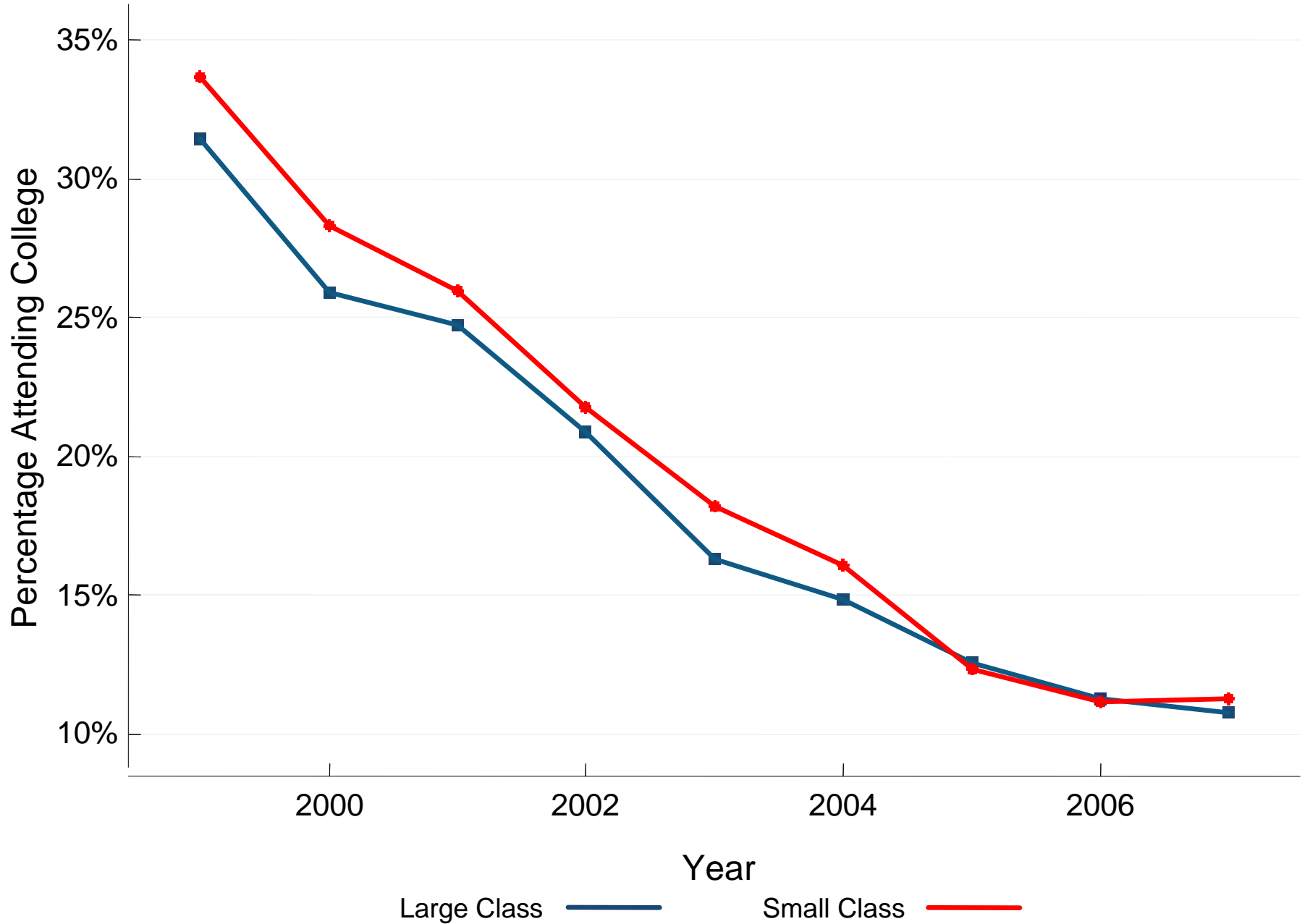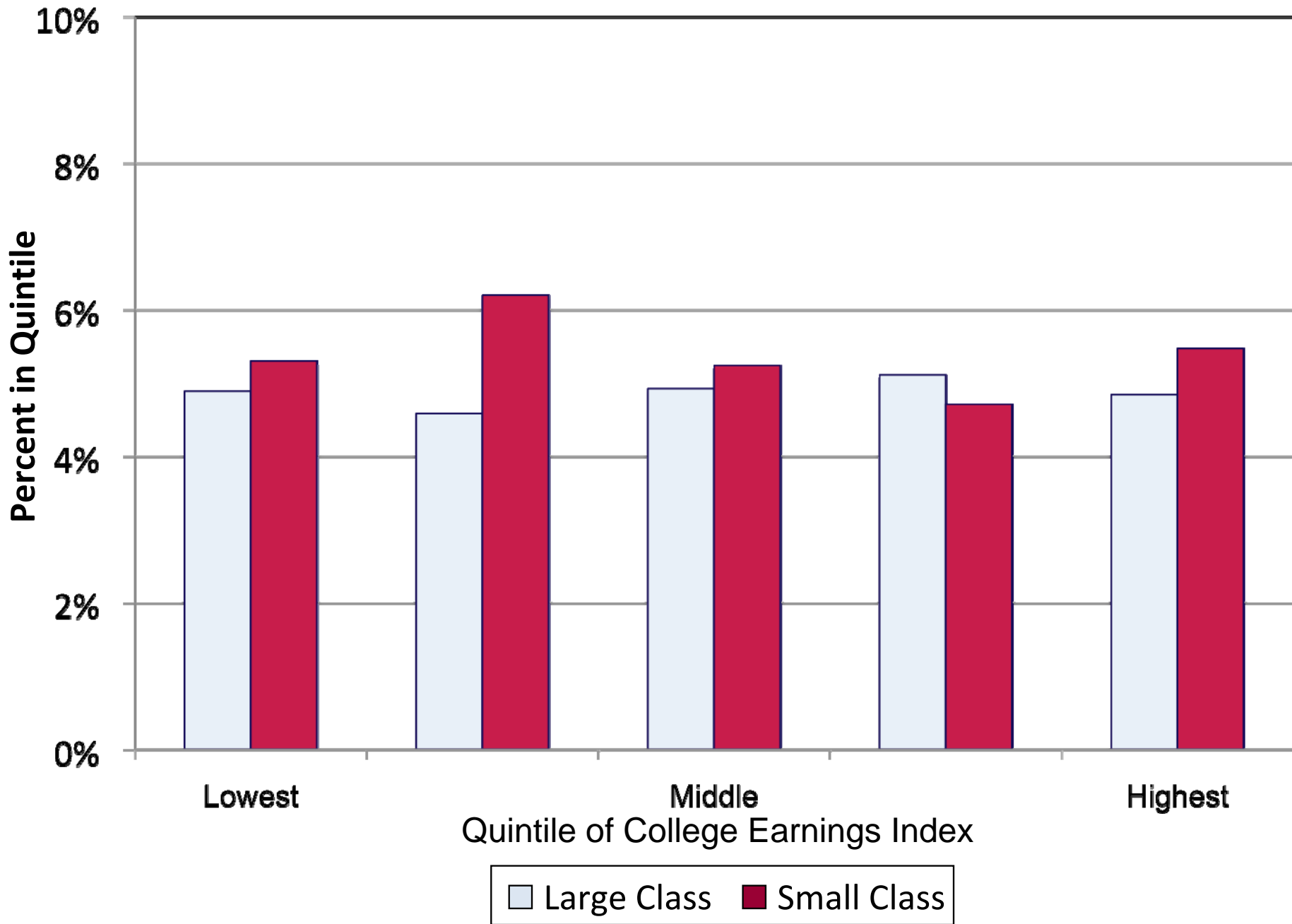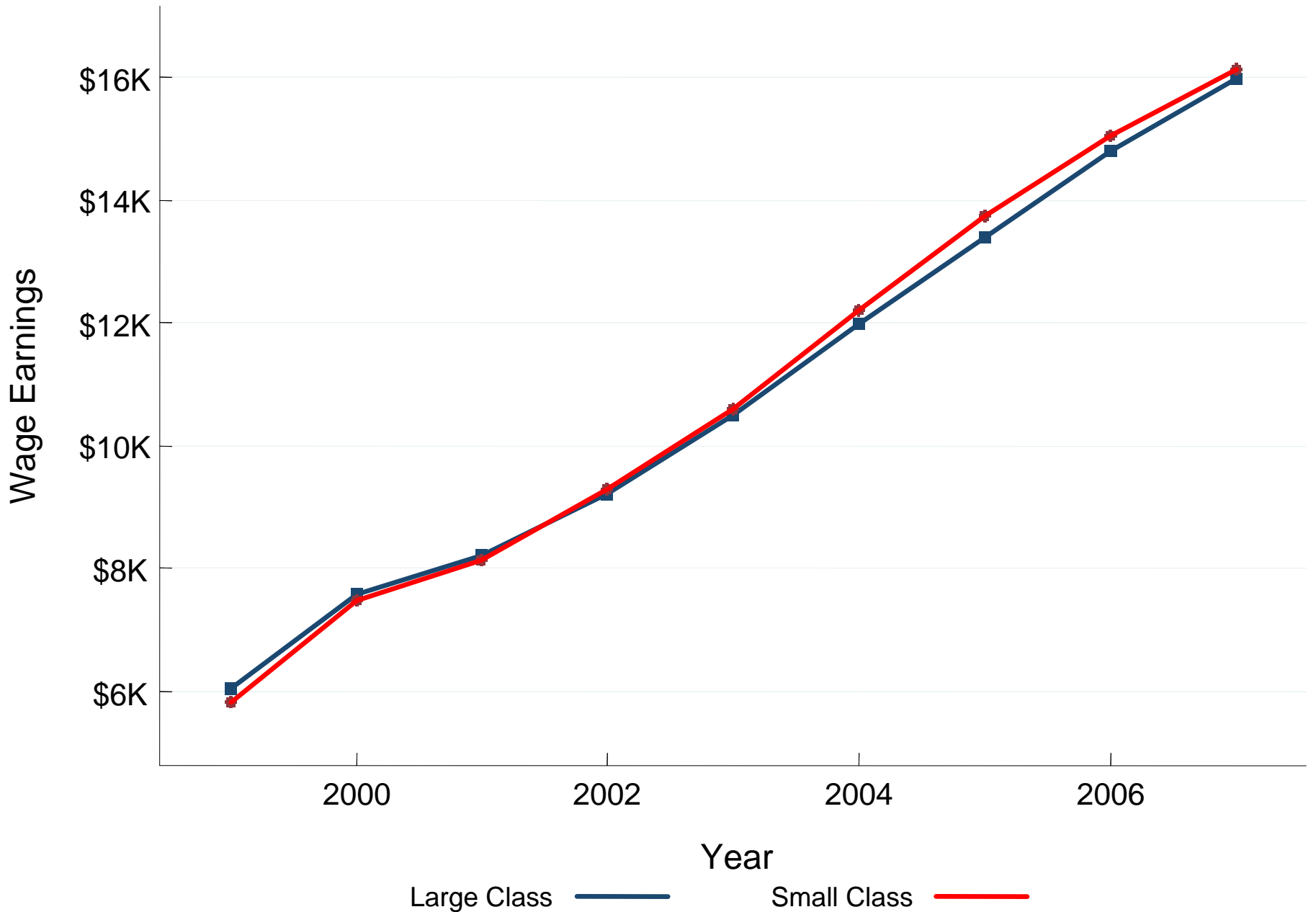**Figure 4a: Effect of Class Size on College Attendance by Year**

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 of Adult Outcomes

Dependent Var.:	Attended College In 2000		College Mean Earnings	Wage Income		Index of Other Outcomes
	(1)	(2)	(3)	(4)	(5)	(6)
Small Class	1.91%	3.38%	$327.5	$79.62	$294.6	4.38%
	(0.88%)	(1.48%)	($197.6)	($289.7)	($453.2)	(2.09%)
Controls	x	x	x	x	x	x
Blacks Only		x			x	
Observations	10,919	3,914	10,919	10,919	3,914	10,911
Mean of Dep. Var.	26.5%	20.7%	$16,098	$14,792	$11,592	0.00

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, parent age at child's birth, parent gender, student gender, free-lunch status, and race. Summary index includes a dummy for owning a home, having a 401(k), having moved outside TN, average house value in 2008 zip, a dummy for being married, and spousal earnings (imputing 0 when single). Index is scaled such that a 1 pp. increase in test scores increases the index by 1 unit.

Part 4: Teacher/Peer Effects

- Students randomly assigned to classes that differ in teacher and peer quality

- Do teachers/peers affect adult outcomes?

- Approach 1: Examine impacts of observable characteristics of teachers and peers

  - Begin by examining teacher experience, which predicts test scores

- 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

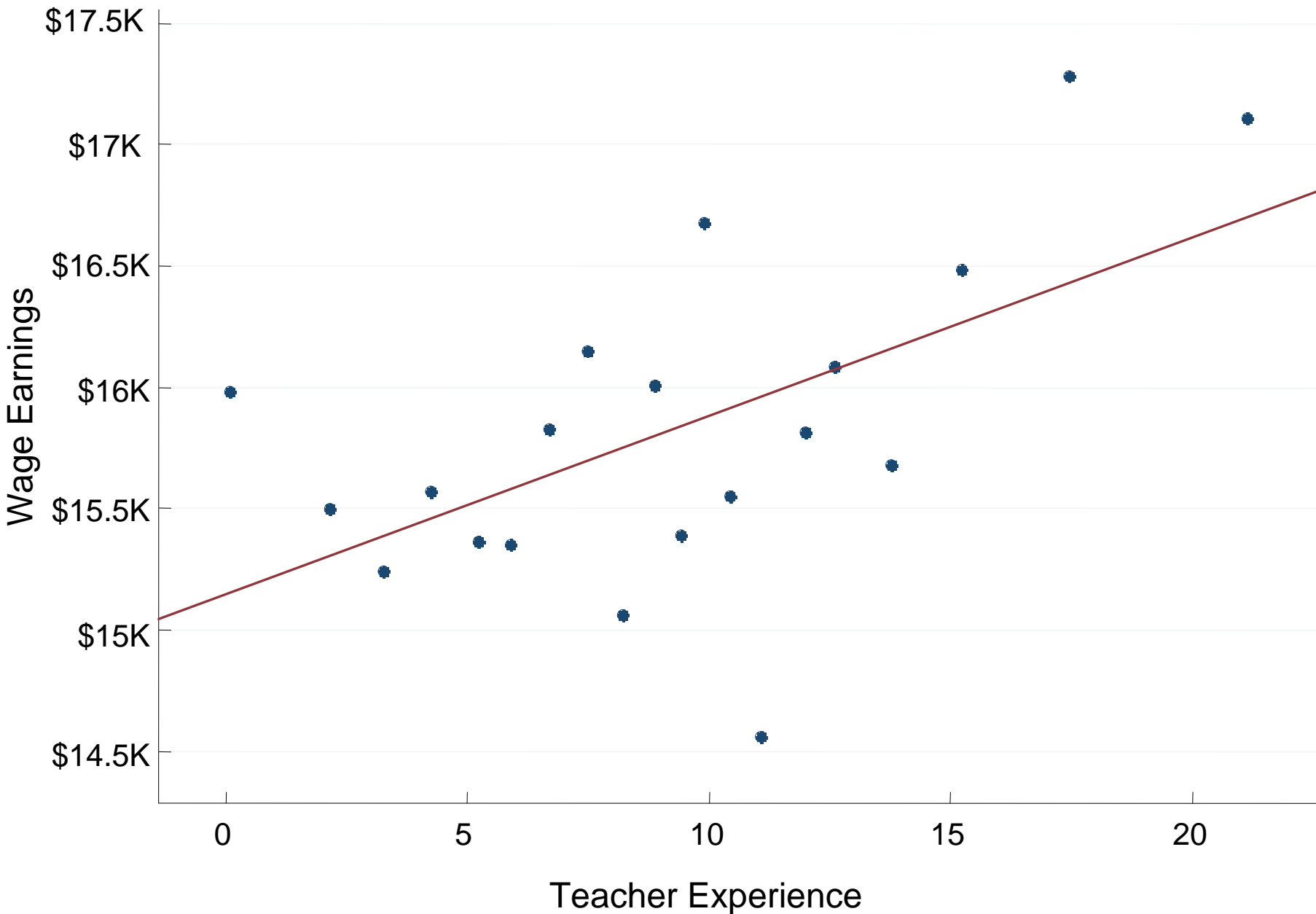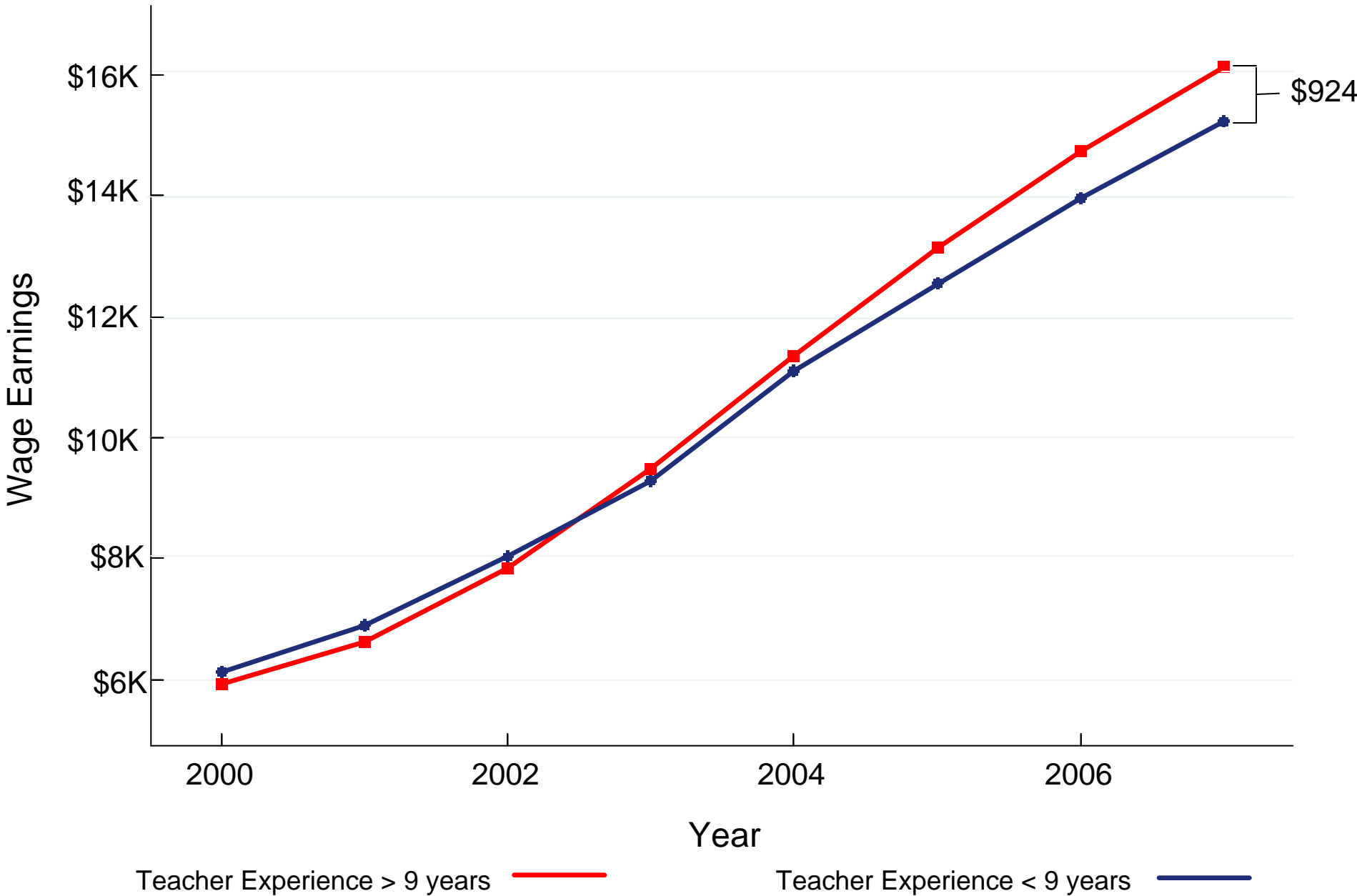Figure 5a: Causal Effect of Teacher Experience on Earnings

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

$924

Wage Earnings

Year

Teacher Experience > 9 years   Teacher Experience < 9 years

Table 7: Observable Teacher vs. Peer Effects

Dependent Var.:	Wage Income			
	(1)	(2)	(3)	(4)
Above Median Teacher Experience	**$818.0** ($397.4)			
Teacher Experience (years)		**$61.93** ($33.59)		
Teacher Degree Higher than a BA	$37.47 ($434.0)	$21.39 ($437.8)		
% Black Peers			$3,335 ($3,213)	
% Female Peers			-$838.6 ($1,504)	
% Free-Lunch Peers			-$2,094 ($1,759)	
Predicted Peer Scores				$95.73 ($114.2)
Observations	5,993	5,993	5,876	5,879

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

<u>Class Effects</u>

- We find significant impacts of teacher experience on wages

- Effects of observed peer characteristics are imprecisely estimated

- But most of teacher and peer quality is not captured by these observable measures

  - Similar problem arises in all studies of teacher effects (e.g. Rockoff 2004, Kane and Staiger 2008)

  - To capture these unobservable aspects of class quality, we look directly for class-level effects on wages

  - Class effect includes effect of teachers, peers, and any class-level shocks

Class Effects: ANOVA

- Begin by testing 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 K Scores (1)	Grade 8 Scores (2)	Mean Wage		
			(3)	(4)	(5)
P-value of F-Test on KG Class Fixed Effects	0.000	0.650	0.036	0.032	0.011
SD of Class Effects (RE estimate)	8.765	0.000	$1,373	$1,338	$1,329
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	5,869	4,470	6,014	6,014	4,202

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

Class Effects on Scores and Earnings

- Key question: Are class effects on KG scores correlated with class effects on earnings?

  - Are improvements in test scores a good proxy for impacts of classes on 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 $\rightarrow$ 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)

Jackknife Estimate of Class Effects

- Define leave-out mean measure of "class quality" as follows:

  - "How good are your classmates' scores, compared with the classmates you could have had?"

  - Class Quality =

    Mean Peer Scores in Class – Mean Peer Scores in School

- Note that because we are measuring peers' test scores at the *end* of KG, class quality reflects teacher + peer effects

- Because students were randomly assigned to classes, class quality varies randomly within schools

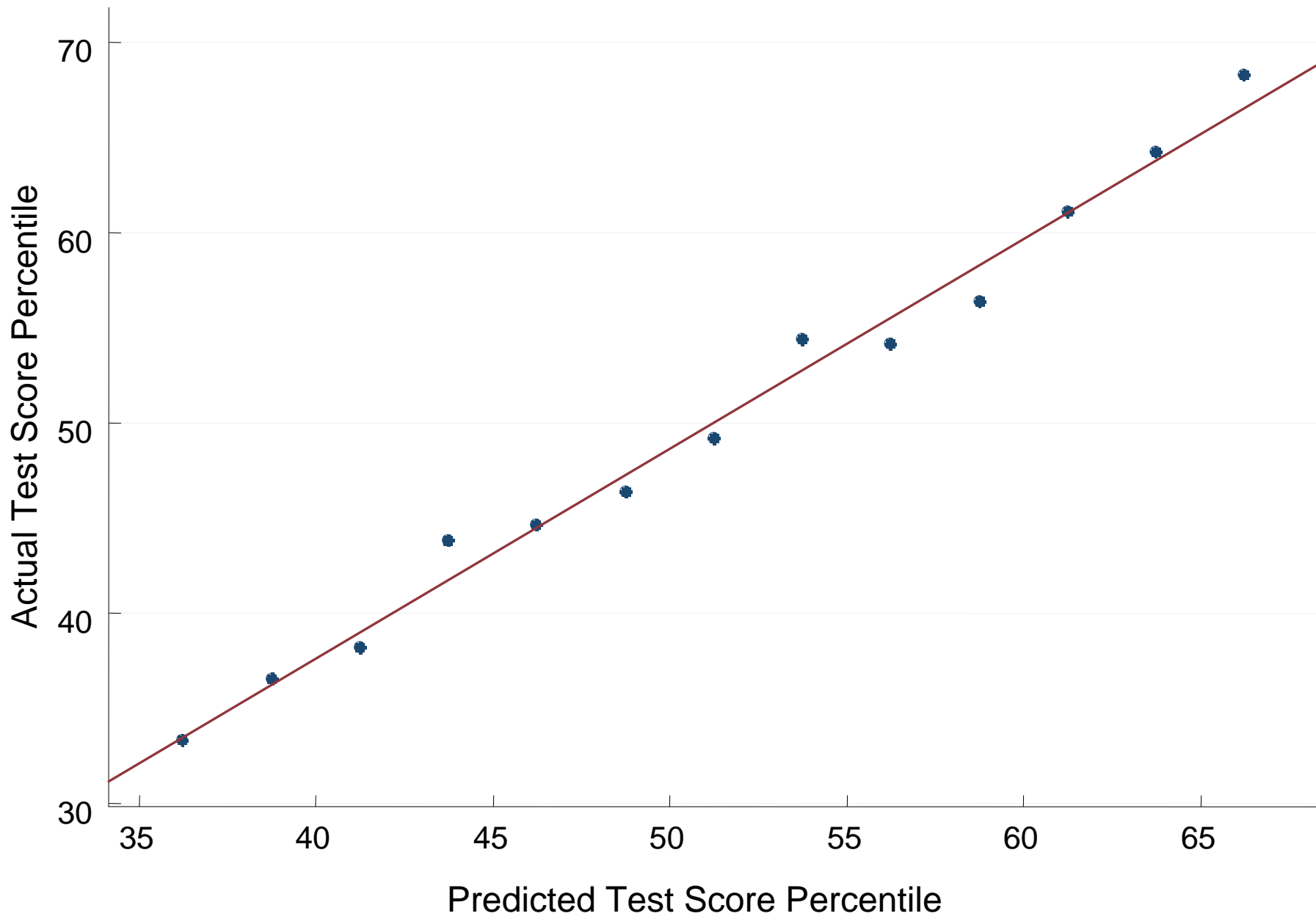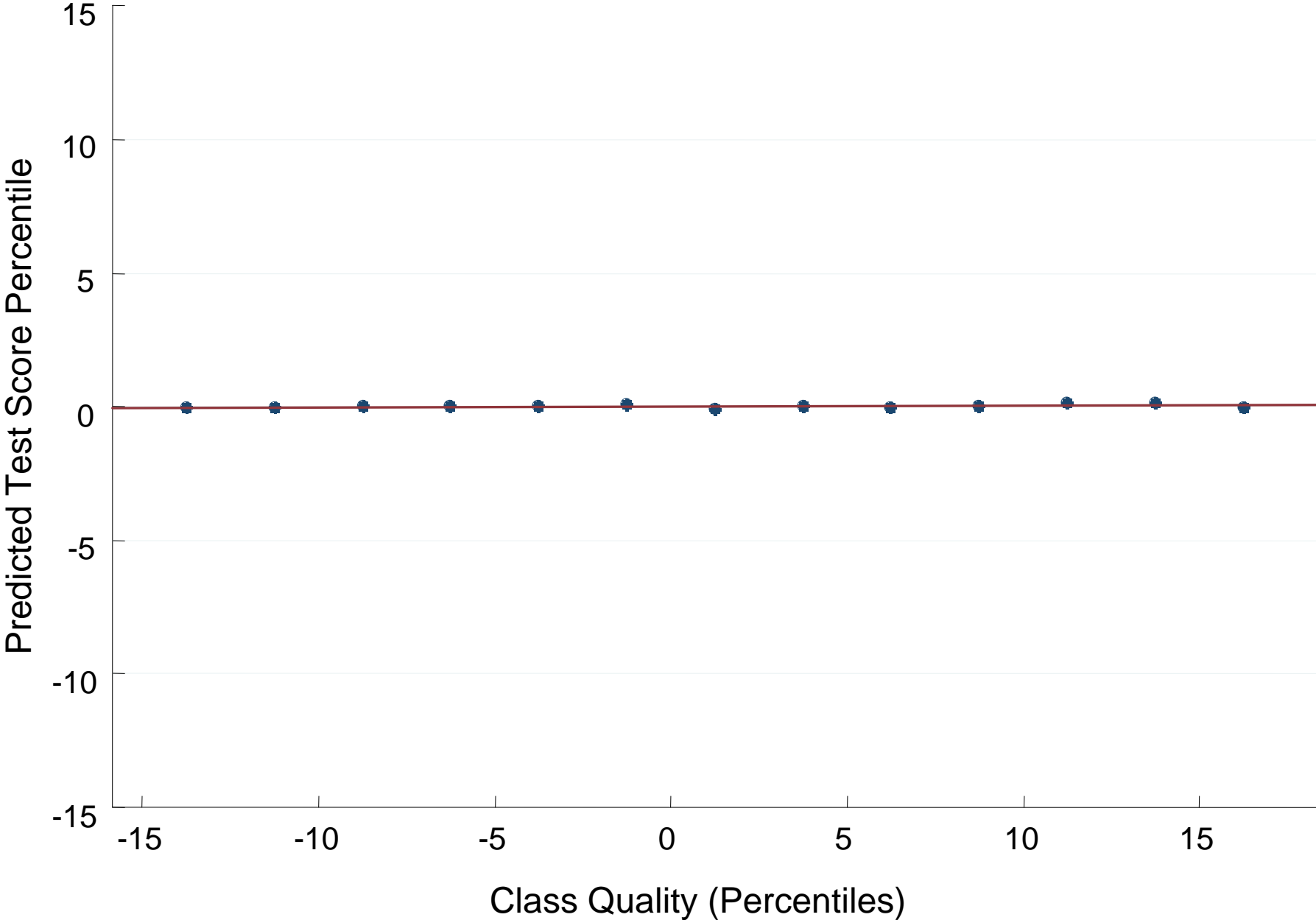# Figure 6a: Causal Effect of Class Quality on Own Score
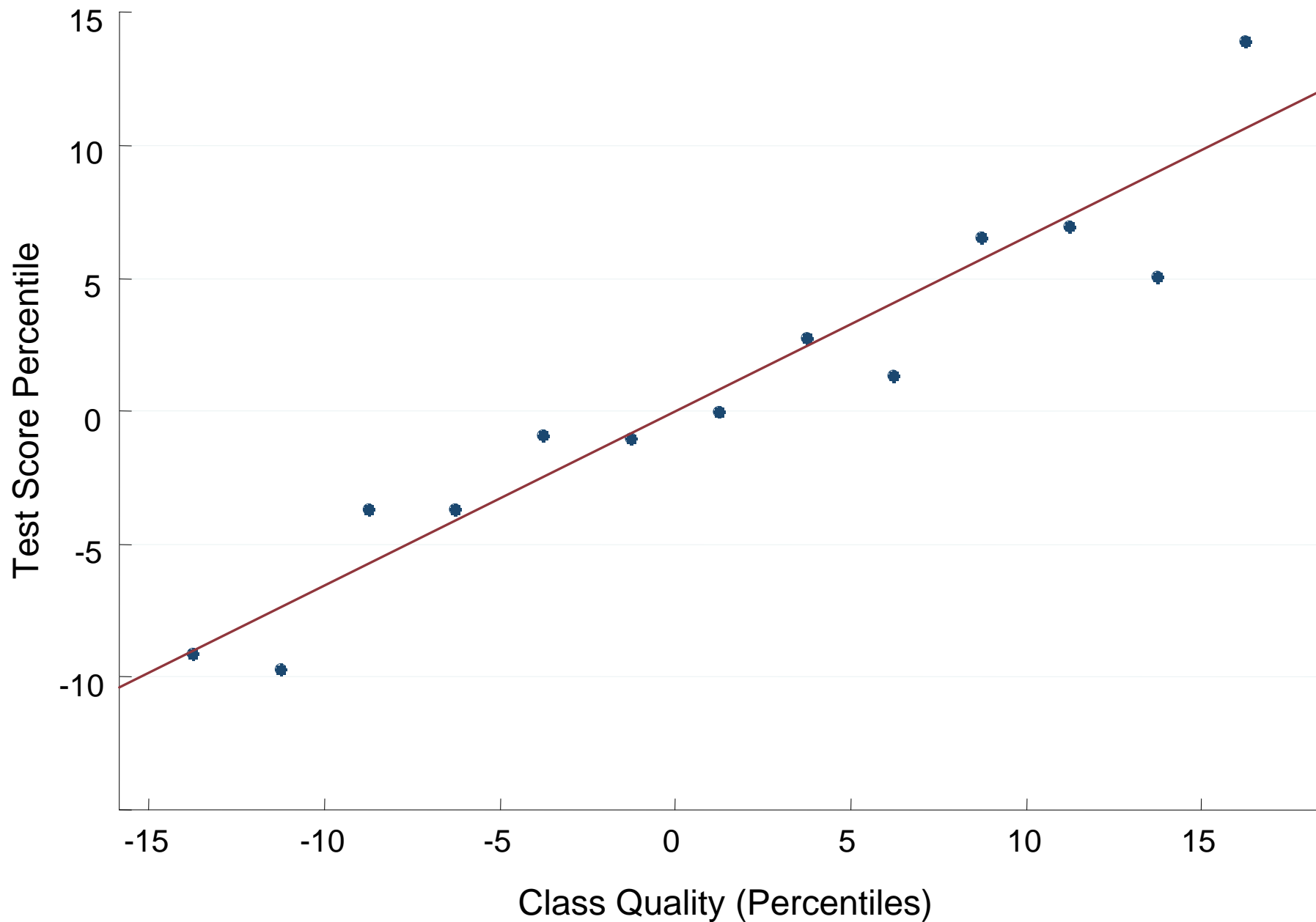
Jackknife Estimate of Class Effects: Placebo Test

- To confirm that jackknife estimate reflects 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
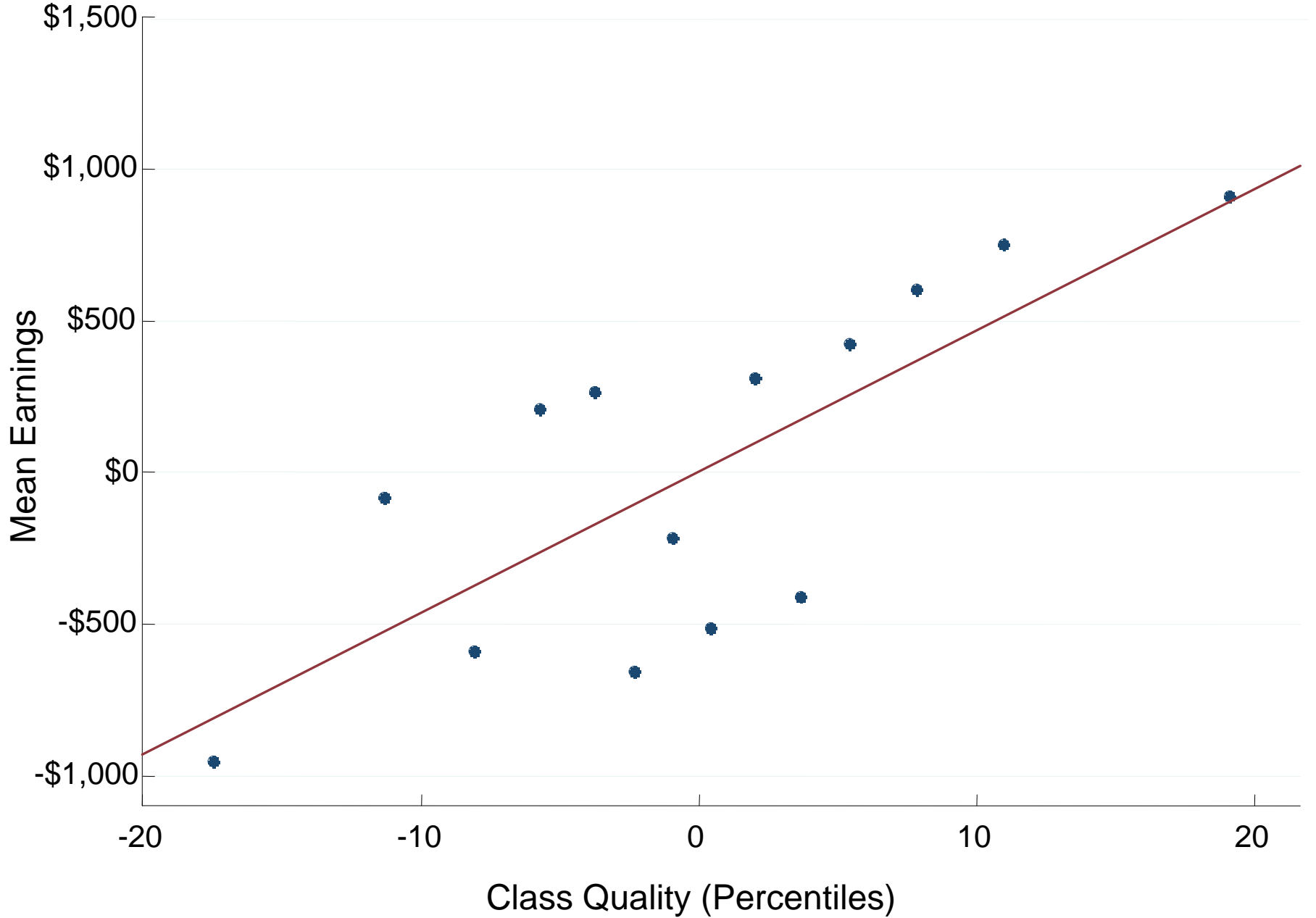
Figure 6b: Actual Test Score vs. Predicted Test Score

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

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

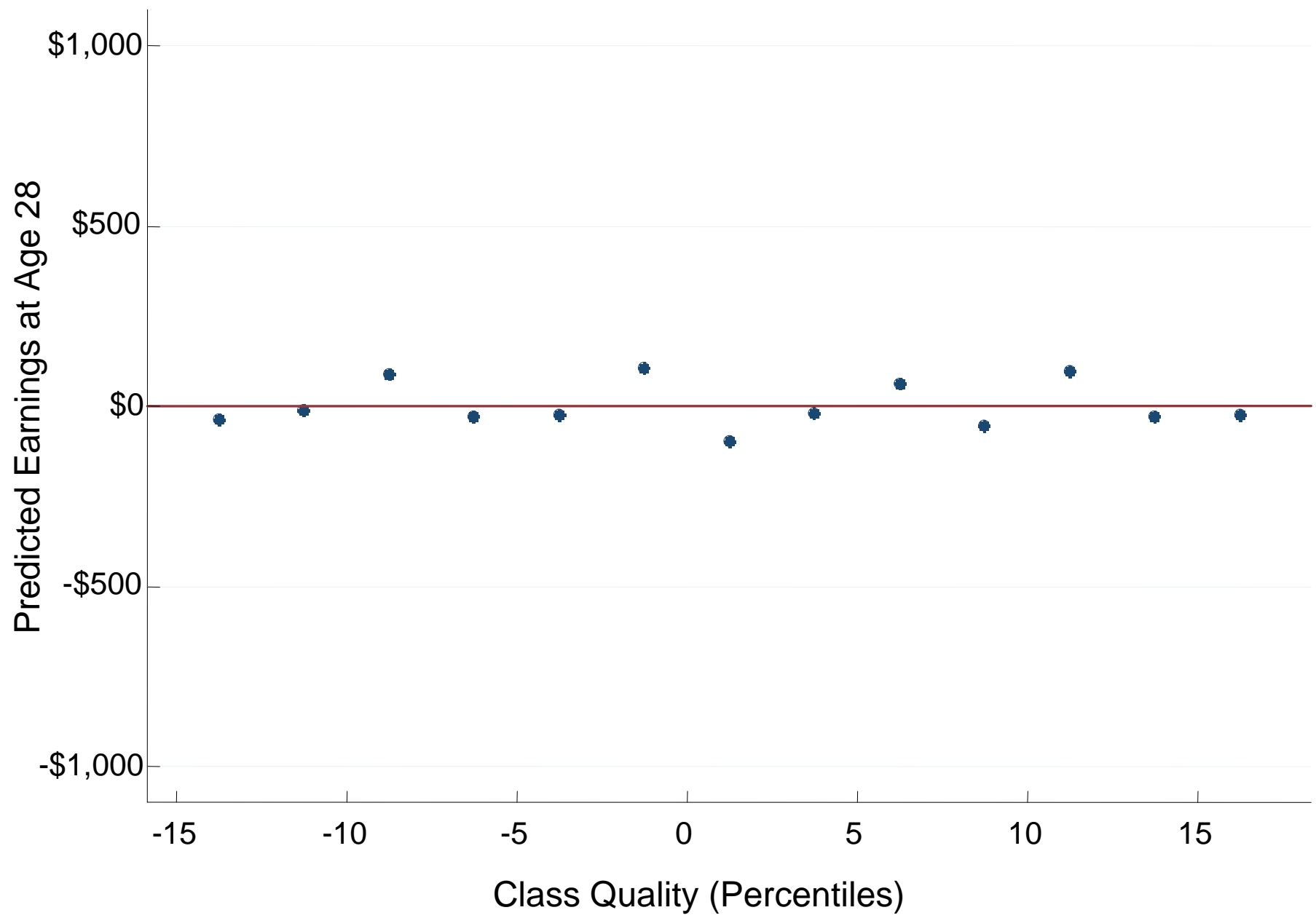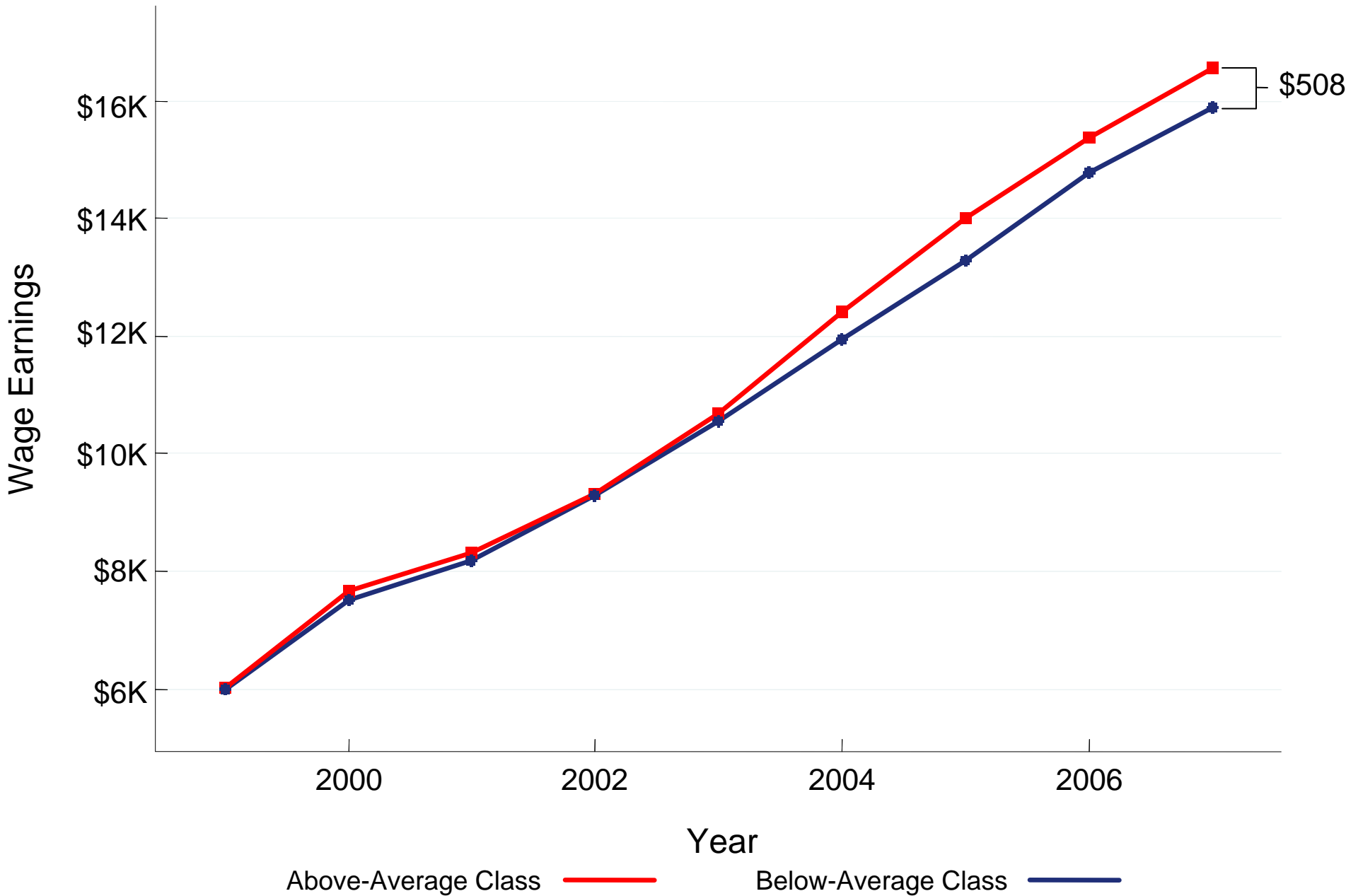**Figure 6e: Placebo Test: Class Quality and Predicted Earnings**

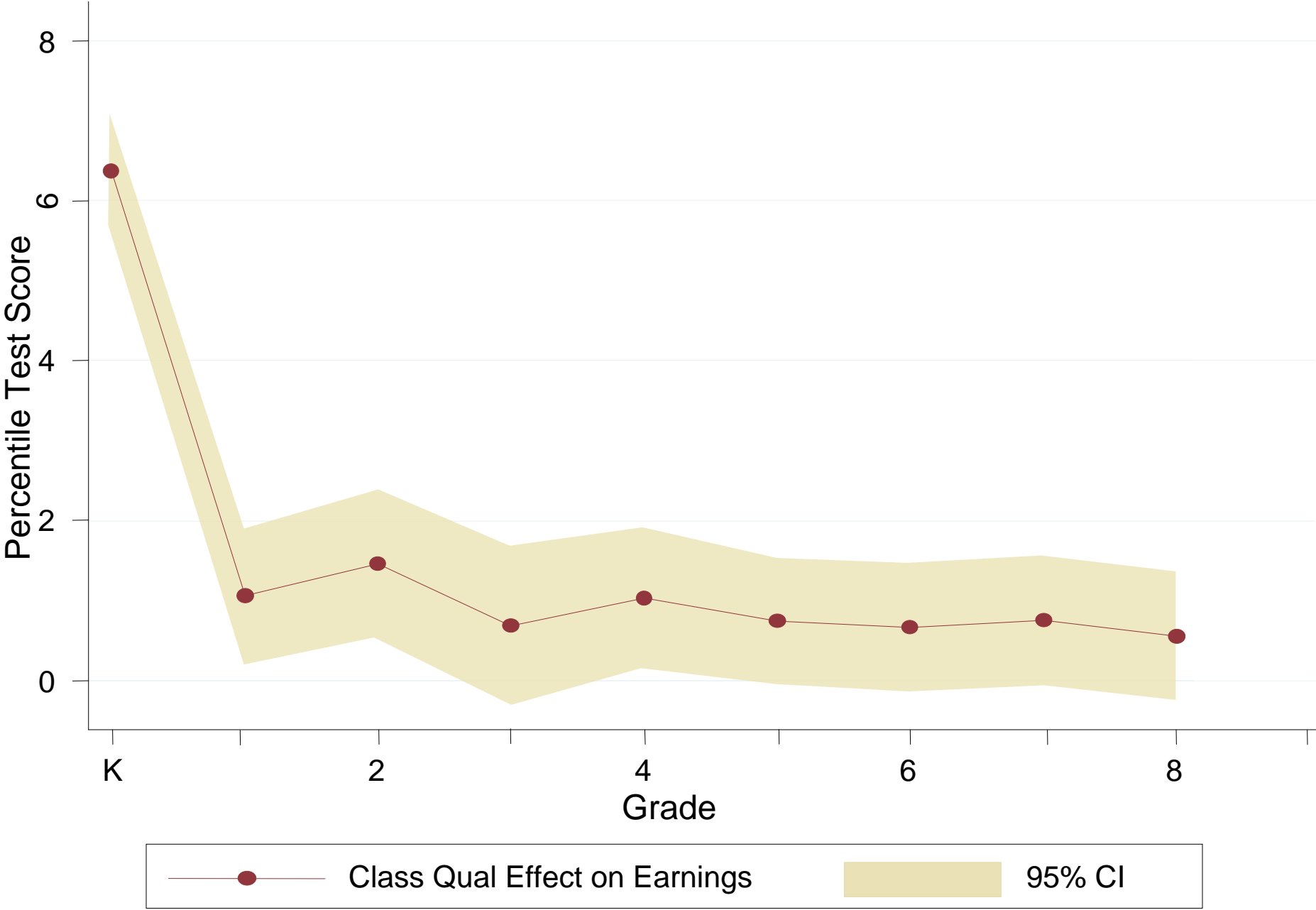**Figure 6f: Effect of Class Quality by Year**

Above-Average Class ———— Below-Average Class ————

Table 9: Class Effects on Test Scores and Earnings

| Dependent Var.:	Test Percentile	Wage Earnings					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test Score			$80.61 ($28.05)	$76.75 ($30.82)	$78.97 ($20.21)	$78.62 ($24.13)	$82.26 ($5.65)
Class Quality	0.578 (0.023)	$46.55 ($13.44)					
Controls	x	x	x	x	x	x	x
Estimation Method	First Stage	Reduced Form	Jackknife 2SLS	SSIV 2SLS	2SLS with Class FX	LIML with Class FX	OLS
Observations	10,306	10,306	10,306	5,553	5,553	5,553	10,306

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

Class Qual Effect on Earnings
95% CI
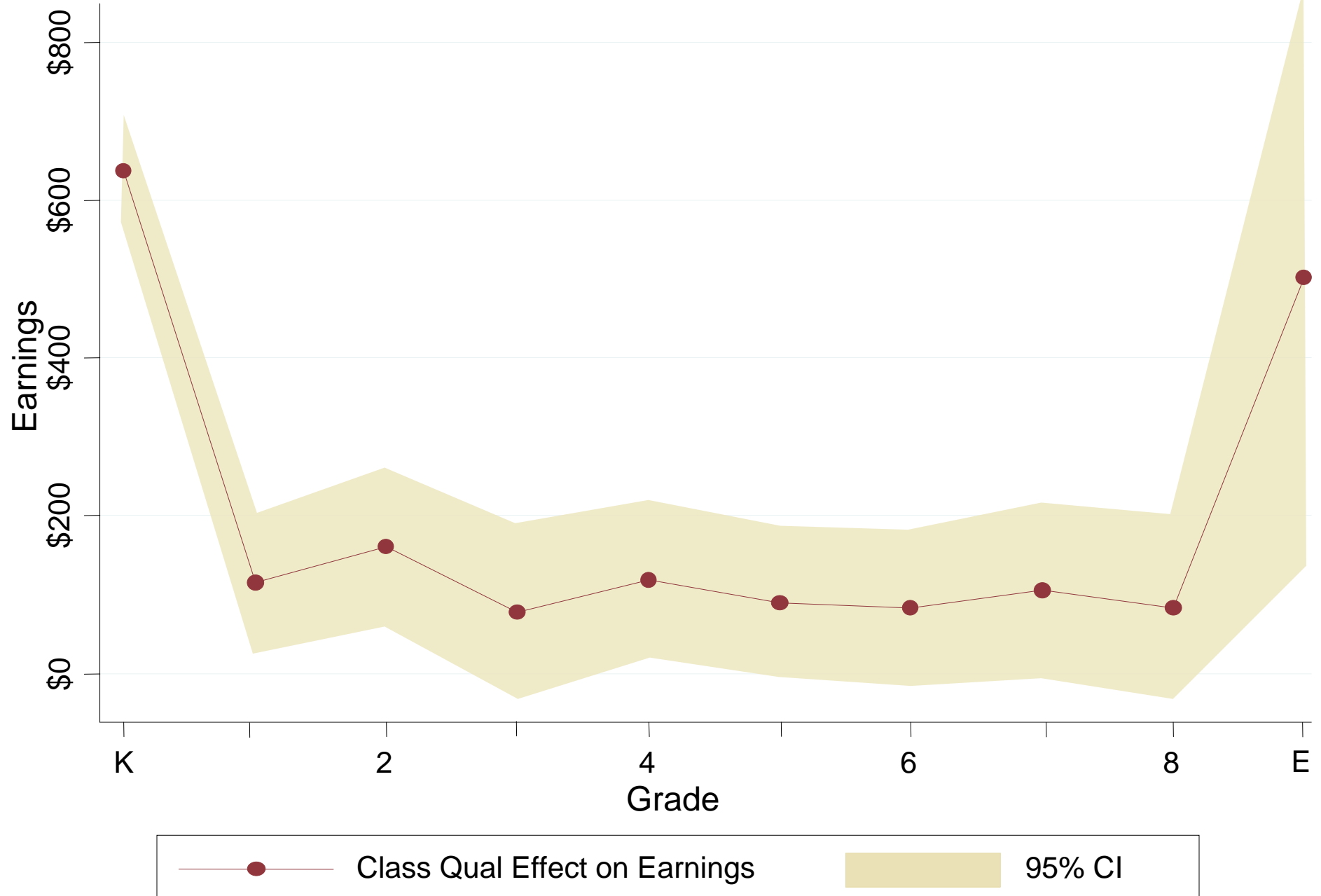
Table 10: Class Effects: Impacts on Adult Outcomes

Dependent Var.:	Attended College in 2000	College Quality (Mean Earnings)	Index of Other Outcomes
	(1)	(2)	(3)
First Obs. Test Score	0.20%	$38.33	0.27%
	(0.07%)	($15.38)	(0.16%)
Individual Controls	x	x	x
Observations	10,306	10,306	10,299
Mean of Dep. Var.	26.50%	$16,098	0.00

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

Conclusions

1. Early childhood class effects fade out in test scores but reappear in adult outcomes

2. Contemporaneous test scores are a valid proxy for the benefits of early childhood interventions

   - 1 SD higher in test scores increases earnings by 14.8%

   - Intervention-based estimates similar to OLS with controls

Cost-Benefit Analysis: Class Size

- Small classes raise test scores by 0.23 SD = $431

    - Assume: 3% real discount rate, constant percent income gains, income follows average US income profile

    → NPV earnings gains of $11,842

- STAR intervention lowered class size by one-third

    - Average school costs = $8,848 per student per year

    - Average treatment = 2.14 years

    → Total Cost per Student = $9,355

- Costs of class size reduction are of the same order of magnitude as earnings gains

Cost-Benefit Analysis: Teacher Quality

- Rockoff (2004) and Kane and Staiger (2008) estimate that a 1 SD improvement in teacher quality raises test scores by 0.2 SD

- Based on this estimate, our results imply that 1 SD improvement in teacher quality raises students' annual earnings by 2.9%

- Assume: 3% real discount rate, constant percent income gains, income follows average US income profile

  → 1 SD increase in teacher quality yields NPV earnings gains of $12,000 per student, or $240K per class of 20 students

- Implies that moving from a 25th percentile (below average) teacher to a 75th percentile (above average) teacher → $320,000

- One year of teacher experience raises annual earnings by $62

  → NPV gain of $34K per class of 20 students