PUBLIC POLICY INSTITUTE OF CALIFORNIA

Higher Education and Economic Opportunity in California

Technical Appendices

CONTENTS

Hans Johnson and Marisol Cuellar Mejia

with research support from Sergio Sanchez and Cesar Alesi Perez

Supported with funding from the Bill and Melinda Gates Foundation, the Spencer Foundation, and the Sutton Family Fund

Appendix A. Data and Methods

Wage regressions

For our analysis of wage premiums (Figures 1 and 2 in the report), we rely on Mincer's human-capital wage equations. Specifically, wage premiums were estimated (see Tables B5 and B6) using regressions of log of annual wages on education categorical variables (less than high school, some college, associate degree, bachelor's degree, graduate degree, with high school omitted), age, age squared, categorical variables for race/ethnicity (Latino, African American, Asian, American Indian/Alaska Native, multiple/other, with white omitted) and dummy variables for gender, marital status, birthplace, citizenship status, and ability to speak English well. We obtained average wages for each level of educational attainment and year (for Figure 1) or race/ethnicity (for Figure 2) by calculating dollar estimates at the mean value of each demographic variable for the 2018 sample of interest (see Table B4). The 1990 estimates are adjusted for inflation using the Annual Average Consumer Price Index (CPIURS). As a robustness check, we ran all our specifications for the 2018 sample with metropolitan area fixed effects (2013 Office of Management and Budget metropolitan area delineations are not available for the 1990 decennial census sample) to account for geographic factors that could affect wages. Our estimates of annual wages do not change significantly when including such controls.

The results presented rely on Public Use Microdata Samples for the 1990 decennial census and the 2018 American Community Survey (ACS) one-year file for the U.S. population age 16 and older. We limit our sample to full-time year-round workers employed in the public or private sector with annual wage and salary income above zero dollars. Workers in the military are excluded. Our sample sizes for California are 472,162 and 125,280, respectively.

Regression-adjusted wage premiums take account of the differing age and racial/ethnic distribution of each educational group; consequently, it is a better measure than a "wage premium" computed by simply dividing college-graduate wage by the high-school only wage, for example. Although we controlled for personal characteristics to make comparisons between individuals who are as similar as can be observed, we do not have quasi-experimental variation concerning who goes to college. Thus, caution is necessary in making causal interpretations of the estimated wage premiums, since the potential problem of selection bias from nonrandom sorting on unobservables remains. A potential source of downward bias in the wage estimates for college graduates must also be considered. Specifically, we estimate wages for those who have completed a bachelor's degree separately from wages of those who have gone on to finish a graduate degree. Therefore, this information is only suggestive; it does not directly provide information on whether attendance at college is a worthwhile private or social investment.

A critical question is whether the wage gains enjoyed by college graduates would have occurred for those individuals even if they did not attend college. One argument in the debate over the causality between schooling and earnings is that colleges select individuals who would have succeeded in the labor market even if they did not attend college (known as the selection effect). The other argument is that the skills and knowledge acquired in college lead to better labor market outcomes, including higher wages.

The best research suggests that the college wage premium, as estimated in our standard wage models, is an accurate measure of the causal effect of college. In a thorough review and analysis of the extensive literature on wages and education, David Card (1999) concluded that the selection effect does not exceed 10 percent of the estimated schooling coefficient. That estimate is derived from studies of twins with different educational attainment. Other approaches, including instrumental variable (IV) estimates, are often higher than classic ordinary least squares (OLS) estimates from standard human capital earnings functions. Although it is unclear to what extend this is due to measurement error or inadequate instrumentation, Card notes that one possibility is that OLS approaches actually understate the causal value of a degree (see Trostel et al. 2002).

Wage differences by race and ethnicity

There is an extensive debate in the academic literature about the cause of wage differences between race and ethnic groups. While early research argued that controlling for "ability" eliminated apparent wage gaps, recent research highlights that, conditional on ability, African Americans have higher educational attainment (Rodgers and Spriggs 1996; Lang and Manove 2011). Accordingly, absent discrimination, we should expect equally-abled African Americans to be financially rewarded for their greater education, in effect, having higher wages than their otherwise similar white peers. To this end, Arcidiacono et al. (2011) find no differences in wages between African American and white male college graduates conditional on a host of factors, including "ability" as measured by the AFQT. However, they do find a 6-10 percent wage penalty among those with only a high school degree, concluding that such results "are consistent with the notion that employers use race to statistically discriminate in the high school market but have no need to do so in the college market." While Latino men and women have experienced comparatively larger earnings growth over time, relative gains compared to their white counterparts are significantly reduced when controlling for cost of living. Considering that Latinos live in locations with significantly higher average housing costs, Latino men in particular earn significantly lower hourly wages than white men after controlling for ability, education, and cost of living (McHenry and McInerney 2013).

Even among college completers, wage disparities persist. African Americans and Latinos are more likely to attend colleges with less money to spend on offering a quality education and are significantly underrepresented in highly remunerative majors (Libassi 2018). Work by Raj Chetty and Opportunity Insights finds that most of the income gap between children from low vs. high-income families is explained by differences in the selectivity of college attended. Conditional on college selectivity, they find very little difference in economic outcomes between students who come from low-income families versus those who come from high-income families. See https://opportunityinsights.org/wp-content/uploads/2019/05/Lecture-6-education-1.pdf.

At the same time, improvements in observable skills and credentials have not reduced levels of hiring discrimination (Bertrand and Mullainathan 2004; Quillian, Pager, Hexel, Midtbøen 2017) and have proved unable to overcome unequal access to established social networks (Mora and Davila 2018) and flatter career wage profiles caused by disparities in job switching, job loss, and associated wage growth (Daly, Hobijn, and Pedtke 2020). Furthermore, Witteveen and Attewell (2017) find that "family background casts a long shadow over earnings attainment of offspring in a dataset that followed a large representative sample of baccalaureate graduates for 10 years after graduation. Even when individuals from lower-income families do manage to complete a baccalaureate degree, against the odds, they earn substantially less than graduates from more affluent families, 10 years after graduation. Graduates' pay is related to the selectivity of the college they attended, to their major, and to their academic performance on tests and college GPA. However, those factors do not erase the pay gap associated with disadvantaged family background or even middle-class background relative to higher parental income groups." Using an audit design (a field experiment in which job candidates of different races are created to be equal in qualifications) Gaddis (2015) found that black candidates from elite universities received fewer employer responses and those responses were for jobs with lower starting salaries.

Synthetic cohort pathways

Our synthetic cohort approach to measuring the progress of students from 9th grade to college completion requires the use of multiple data sets and numerous assumptions. The methodological approach is similar to how demographers develop estimates of life expectancies and total fertility rates. As such, the methods are well-established. However, to our knowledge, this is the first time the approach has been used so extensively with California education data. In short, the method relies on recent estimates of probabilities of key events (in this case, graduating from high school conditional on having been in 9th grade, enrolling in college conditional on completing high school, and completing college conditional on enrolling in college) to create a synthetic (or artificial) cohort of students that we can follow from 9th grade to college completion. The end result is the probability that an individual will graduate from college IF recent rates of high school graduation, college enrollment, and college completion persist. The rates are calculated separately for different subpopulations of students, creating multiple cohorts. Additional complexity arises because we consider multiple transitions from high school to college, depending on the type of college. Our model breaks down college enrollment into seven categories: UC, CSU, public universities elsewhere in other states, private non-profit four-year colleges, private for-profit four-year colleges, California Community Colleges, and other two-year colleges.

Consistency in definitions of subpopulations is a challenge. We assume that low-income status and race/ethnicity are identified consistently across educational systems. This assumption seems reasonable enough in the case of race/ethnicity. But the definition of low-income student varies across system. For example, for K-12 students our low-income category is based on the California Department of Education (CDE) data for "socioeconomically disadvantaged" students. CDE defines them as students: (1) who are eligible for the free or reduced-price meal (FRPM) program (also known as the National School Lunch Program, or NSLP), or have a direct certification for FRPMs, or (2) who are migrant, homeless, or foster youth, or (3) where neither of the parents were a high school graduate (source: School Accountability Report Card, Data Element Definitions and Sources, 2017-18, CDE). We define low-income college students as those who receive Pell grants. These two groups are not synonymous. Thus, our estimates of rates of transition and completion among low-income students are accurate only to the extent that the groups we use to define low-income are good proxies for each other (or that the rates between the two groups are similar). This is not as problematic as it might appear. Specifically, because CDE has used National Clearinghouse data to develop college-going rates (by type of college) of recent high school graduates (broken down by demographic and economic group), we are confident in those college-going rates (with the added stipulation that we adjust for "block rates," defined as the share of students enrolled in a college that refuse to have their individual records shared). For low-income students, then, our college-going rate is based on a consistent source for both the numerator and denominator. We then use six-year graduation rates for Pell Students (as reported by UC, CSU, and in IPEDS) as a proxy for the graduation rates of the entire cohort of low-income students who have enrolled in a four-year college. While there might be more low-income students than Pell students (or not), as long as the graduation rate for Pell students is a good measure of the graduation rate for all low-income students, our estimates are accurate. For four-year colleges, this seems to be a very safe assumption as the vast majority of their low-income students receive Pell grants.

The table below shows the data and calculations used to estimate the transition and completion rates (as shown in Figures 3–5 in the report). This process was developed for two income groups and four race/ethnic groups.

TABLE A1

Data and calculations to develop pipeline estimates of 9th grade to baccalaureate completion

Item	Description	Measure used to calculate	Calculation	Source	Notes	Item
a	9th graders	synthetic cohort	set at 1,000			a
b	complete high school	high school graduation rate	1,000 * hs grad rate	CDE	9th grade cohort four-year grad rate	b
c	attend any college	college going rate	d + e		sum of 4-yr and 2-yr college going rates	c
d	attend a four-year college as freshmen	4-year college going rate	b * 4-year college going rate	CDE with PPIC adj.	PPIC adjusted for NSC block rate; any NSC college in US	d
e	attend a community college	2-year college going rate	b * 2-year college going rate	CDE with PPIC adj.	PPIC adjusted for NSC block rate; any NSC college in US	e
f	transfer	transfer rate	e * CCC transfer rate	CCC transfer velocity	assumes transfer rates from non-CCC 2-year colleges are same as those from CCC	f
g	attend a four-year college at any time	4-year plus transfer	d + f		calculated here - sum or attend as freshman plus attend as transfer	g
h	earn a bachelor's degree		I + j			h
i	earn a bachelor's degree as a freshmen	6-year graduation rates	d * six-year grad rate	Multiple	calculated based on institutional data (UC and CSU) and IPEDS data	i
j	earn a bachelor's degree as a transfer	4-year graduation rates	f * four-year grad rate	Multiple	calculated based on institutional data (UC and CSU) and IPEDS data	j

Note that we adjusted the college going rate of recent high school graduates as reported by CDE. CDE works with the National Student Clearinghouse (NSC) to estimate college enrollment rates (within 16 months of high school graduation). CDE does not adjust for students who block NSC from sharing their college enrollment data. However, the NSC does report block rates by state and college type. We used the NSC block rates to adjust upwards the share of California high school students attending college (by our seven types of colleges). Those rates range from about 2 percent for private colleges to 6 percent for the state's public universities to almost 12 percent for community college students in 2016-17.

Appendix B. Additional Tables

TABLES B1–B3

College graduates are more likely to work and have higher quality jobs

All Residents	No HS Diploma	HS Graduate	Some college	Associate degree	Bachelor's degree	Graduate degree
Unemployment rate (%)	7.2	5.7	4.7	4.0	3.3	2.2
Labor force participation rate (%)	66	74	78	80	86	89
Health Insurance through employer (%)	48	66	77	80	86	91
Retirement plan through employer (%)	17	35	40	44	44	52
Full-time employment (%)	53	60	64	65	74	79

Latino Residents	No HS Diploma	HS Graduate	Some college	Associate degree	Bachelor's degree	Graduate degree
Unemployment rate (%)	6.5	5.8	4.2	4.0	3.1	1.6
Labor force participation rate (%)	70	78	82	84	89	91
Health Insurance through employer (%)	46	62	75	78	84	89
Retirement plan through employer (%)	17	29	37	42	48	55
Full-time employment (%)	58	65	68	70	76	81

African-American Residents	No HS Diploma	HS Graduate	Some college	Associate degree	Bachelor's degree	Graduate degree
Unemployment rate (%)	20.4	10.5	8.3	7.6	4.2	2.6
Labor force participation rate (%)	42	63	74	75	88	91
Health Insurance through employer (%)	52	69	77	81	87	92
Retirement plan through employer (%)	17	55	45	50	44	48
Full-time employment (%)	27	48	60	62	76	81

SOURCES: Authors' calculations for California based on 2018 ACS one-year estimates, March Current Population Survey for 25-64 year olds in California.

NOTES: Health insurance and retirement data are restricted to full-time year-round, incorporated workers not living in group quarters, with yearly wage and salary income above $1,000. Retirement plan statistics for African-American residents may not reflect actual distributions due to insufficient data. These tables provide useful clues concerning the relationship between educational attainment and important indicators of individual well-being, but it is worth noting that they do not reliably determine causation or measure the exact size of the effects. They are best interpreted as providing suggestive evidence of the powerful role that higher education plays. Also, there are numerous ways that the observed beneficial effects of college attainment are interdependent.

TABLE B4

Summary statistics of the variables used in the wage premium regressions, ACS 2018

Variable	Obs.	Mean	Std. Dev.	Min	Max
Annual Wage	125,280	77,704	82,648	4	565,000
Age	125,280	43	13	16	94
Age squared	125,280	2054	1176	256	8836
Female	125,280	0.427	0.495	0	1
Married	125,280	0.572	0.495	0	1
Born in California	125,280	0.462	0.499	0	1
Citizen, born in U.S.	125,280	0.670	0.470	0	1
Citizen, naturalized	125,280	0.193	0.395	0	1
Not a citizen	125,280	0.136	0.343	0	1
Years in the U.S.	125,280	9	14	0	83
Speaks English well	125,280	0.928	0.258	0	1
Latino	125,280	0.330	0.470	0	1
White	125,280	0.419	0.493	0	1
African-American	125,280	0.041	0.198	0	1
Asian	125,280	0.179	0.383	0	1
American Indian / Alaska Native	125,280	0.004	0.062	0	1
Multiple	125,280	0.025	0.156	0	1
Less than high school	125,280	0.002	0.047	0	1
High school	125,280	0.101	0.301	0	1
Some college	125,280	0.179	0.383	0	1
Associate degree	125,280	0.207	0.405	0	1
Bachelor's degree	125,280	0.080	0.271	0	1
Graduate degree	125,280	0.262	0.440	0	1

TABLE B5

Wage premium regression estimates for California residents in 1990 & 2018

Variable	1990	2018
Age	0.069 (0.001)***	0.073 (0.001)***
Age squared	-0.001 (0.000)***	-0.001 (0.000)***
Female	-0.290 (0.002)***	-0.232 (0.004)***
Married	0.126 (0.002)***	0.165 (0.004)***
Born in California	0.011 (0.002)***	-0.059 (0.005)***
Naturalized citizen	-0.046 (0.003)***	-0.100 (0.007)***
Not a citizen	-0.165 (0.003)***	-0.150 (0.008)***
Speaks English well	0.238 (0.004)***	0.264 (0.008)***
Latino	-0.119 (0.002)***	-0.194 (0.005)***
African American	-0.127 (0.003)***	-0.217 (0.010)***
Asian	-0.093 (0.003)***	-0.073 (0.006)***
AI / AN	-0.169 (0.010)***	-0.246 (0.028)***
Multiple/other	-0.139 (0.022)***	-0.076 (0.012)***
Less than high school	**-0.137** (0.003)***	**-0.163** (0.008)***
Some college	**0.125** (0.002)***	**0.146** (0.006)***
Associate degree	**0.187** (0.003)***	**0.228** (0.007)***
Bachelor's degree	**0.392** (0.003)***	**0.565** (0.006)***
Graduate degree	**0.581** (0.004)***	**0.832** (0.007)***
Intercept	8.944 (0.010)***	8.784 (0.024)***
Observations R-Squared	472,162 0.359	125,280 0.363

SOURCES: Authors' calculations based on 1990 decennial census and 2018 American Community Survey one-year estimates.

NOTES: Omitted groups include individuals that are citizens by birth, white, and high school graduates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.2.

TABLE B6

Wage premium regression estimates for California residents by race and ethnicity in 2018

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Age	0.073 (0.001)***	educ1#race1	-0.000 (0.025)	educ5#race5	-0.146 (0.097)
Age squared	-0.001 (0.000)***	educ1#race3	-0.018 (0.081)	educ5#race6	-0.040 (0.039)
Female	-0.229 (0.004)***	educ1#race4	0.028 (0.033)	educ6#race1	-0.080 (0.018)***
Married	0.164 (0.004)***	educ1#race5	0.128 (0.138)	educ6#race3	-0.131 (0.034)***
Born in California	-0.060 (0.005)***	educ1#race6	-0.173 (0.093)*	educ6#race4	0.268 (0.019)***
Naturalized citizen	-0.101 (0.007)***	educ3#race1	-0.031 (0.013)**	educ6#race5	0.003 (0.110)
Not a citizen	-0.168 (0.008)***	educ3#race3	-0.110 (0.028)***	educ6#race6	-0.046 (0.044)
Speaks English well	0.248 (0.008)***	educ3#race4	0.028 (0.020)	Intercept	8.783 (0.024)***
Latino (race1)	-0.161 (0.009)***	educ3#race5	-0.107 (0.072)	Observations R-squared	125,280 0.366
African American (race3)	-0.137 (0.020)***	educ3#race6	-0.082 (0.041)**		
Asian (race4)	-0.185 (0.015)***	educ4#race1	0.003 (0.017)		
AI / AN (race5)	-0.185 (0.053)***	educ4#race3	-0.065 (0.037)*		
Multiple/other (race6)	-0.028 (0.033)	educ4#race4	0.081 (0.023)***		
Less than high school (educ1)	-0.171 (0.023)***	educ4#race5	-0.129 (0.089)		
Some college (educ3)	0.169 (0.010)***	educ4#race6	-0.028 (0.049)		
Associate degree (educ4)	0.231 (0.012)***	educ5#race1	-0.126 (0.014)***		
Bachelor's degree (educ5)	0.597 (0.009)***	educ5#race3	-0.102 (0.029)***		
Graduate degree (educ6)	0.807 (0.010)***	educ5#race4	0.089 (0.018)***		

SOURCES: Authors' calculations based on 2018 American Community Survey one-year estimates.

NOTES: Omitted groups include individuals that are citizens by birth, white, and high school graduates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.2.

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.

Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
T: 415.291.4400
F: 415.291.4401
PPIC.ORG

PPIC Sacramento Center
Senator Office Building
1121 L Street, Suite 801
Sacramento, CA 95814
T: 916.440.1120
F: 916.440.1121