When Wages Rise, Who Goes to College? Impact of Labor Market Shocks on College Attendance*

Steven Lee

October 16, 2025
PRELIMINARY

Abstract

This paper examines how labor market shocks influence college attendance decisions and subsequent earnings among high school graduates. Using comprehensive administrative data from Texas spanning 2004–2015, I study students' responses to changes in local and industry-level wage conditions. I first implement a shift-share identification strategy that combines industry-specific wage shifts with students' predicted industry propensities, and I validate the results using the mid-2000s fracking boom as a natural experiment. I find that positive labor market conditions reduce the likelihood of college matriculation: a one standard deviation increase in labor market strength leads to a 3.2 percentage point decline in college attendance. The students induced out of college are drawn broadly across the achievement distribution, even up to the second quintile of achievement. Despite entering a temporarily strong labor market, these students earn less within five years after high school graduation, suggesting limited long-term gains from substituting work for college.

1 Introduction

College enrollment in the United States has stagnated over the last decade and a half, in sharp contrast to the steady growth in college attendance throughout the late twentieth and early twenty-first centuries (National Center for Education Statistics, 2025). This slowdown contrasts with evidence of a growing earnings gap between college graduates and non-college graduates, which has nearly doubled since 1980 (Autor, 2014). In recent years, U.S. policymakers have increasingly focused on improving labor demand for lower skilled workers through industrial and labor policy. In

*I appreciate valuable comments from Anna Aizer, John Friedman, Peter Hull, and Bryce Millett Steinberg, as well as feedback from seminar participants at Brown University. I am grateful to the Population Studies and Training Center at Brown University, which receives funding from the NIH, for training support (T32 HD007338) and general support (P2C HD041020). I additionally acknowledge funding support from the Bravo Center for Economic Research at Brown University, the Annenberg Institute for School Reform at Brown University, and the American Institute for Boys and Men. The findings of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas.

particular, several federal initiatives incentivize firms to use apprenticeships, providing alternative pathways to employment without earning a college degree. State policies have mirrored these efforts, including recent efforts by several states to remove college degree requirements from government jobs.

Little is known about the types of students who might be pulled away from higher education due to changes in labor market prospects for lower skilled workers. Existing research shows students, particularly men, have reduced educational attainment due to resource booms (Black et al., 2005; Cascio and Narayan, 2022), single-sector shocks (Morissette et al., 2015; Atkin, 2016), or general economic conditions (Charles et al., 2018). However, much less is known about who these marginal college attendees are. Whether the marginal student has low or high potential gains from college attendance affects the degree to which these "blue-collar" policies help low-skilled workers at the expense of discouraging students from pursuing higher education.

Assessment of the marginal students affected by changes in labor market conditions requires three ingredients. First, students' realized gains from college attendance or some proxies for the potential gains need to be known. Without an individual link between students and future outcomes—such as earnings—it is difficult to assess the return to education and/or labor force participation. Barring this, data must contain information such as high school school achievement which can serve as a proxy for potential gains from college. Second, the studied treatment or shock must be unrelated to these attributes or outcomes. Often studies that include rich student achievement information or linkages to later-life earnings exploit quasi-random assignment directly across the variables of interest, such as high school grade point average (Zimmerman, 2014) and college admissions tests (Goodman et al., 2017; Smith et al., 2025). Finally, the shocks have to be directly related to labor market conditions to be relevant for future policy which raises or lowers labor market opportunities for young workers. Most research studying the returns to college largely focus on college "pull" factors, such as admissions thresholds (Zimmerman, 2014; Goodman et al., 2017; Smith et al., 2025), tuition support (Denning, 2017; Gurantz, 2020; Acton, 2021; Carruthers et al., 2023), and proximity to college (Card, 1995; Mountjoy, 2022), which may feature different marginal students than those influenced by labor market conditions.

This paper seeks to directly examine the set of students who opt out of college and join the labor force when labor market conditions change. Using the universe of public school students in Texas from 2000 to 2015, I study high school students' response to two labor market changes: annual, industry-level wage shocks and a single-sector labor demand shock due to the mid-2000s "fracking" boom. I corroborate the existing literature that positive (negative) labor market conditions cause students to not attend (attend) college. I find that these induced high school seniors are drawn from across the achievement distribution, including students in the second quintile of standardized test score. Despite drawing from a wide distribution of high school achievement, positive shocks are on average harmful for the marginal students: students see a temporary boost in earnings which dissipate in the long-run.

Educational and earnings information comes from Texas administrative data. The data covers all students attending public school in Texas and contains detailed information on demographic characteristics and educational outcomes. I link the student data with data on Texas college attendance and unemployment insurance records, allowing me to chart the trajectory of most Texas students from their high school classrooms to their workplace. I supplement this student and worker information with national, industry-level wage data from the Quarterly Workforce Indicators (QWI) and fracking data from the U.S. Energy Information Association and the Railroad Commission of Texas.

To estimate the impact of industry-level wage shocks on college attendance and longer-run earnings, I use a shift-share estimation strategy. I construct industry-level, Bartik-style shocks, or "shifts," using year-over-year changes in non-Texas wages for 19–25 year olds for each industry. To percolate these shocks properly throughout the student population, I define "shares" for each student using an out-of-sample prediction of propensity to join the labor force in each industry based on student characteristics. This process creates an "individualized" wage shock treatment for each student, allowing for flexibility to tease out treatment effect heterogeneity across student characteristics.

I find that students are responsive to wage shocks, but there is heterogeneity in responses across the achievement spectrum. On average, a standard deviation increase in the wage shock treatment leads to a 3.2 pp decline in the probability of college matriculation. As predicted, students in the top decile of standardized test scores do not change their college attendance due to a positive shock. However, students outside the top quintile are responsive to the shocks, with the strongest responses in the middle of the test score distribution. I find that even at the lower end of the test score distribution, marginal students may have benefited from college attendance in the long-run. While exposure to positive shocks leave students better off in the short-run, their college-attending counterparts catch-up 10 years after high school graduation.

I validate these results using the mid-2000s rise of "fracking," a technological advancement in oil and gas extraction that led to sharp increase in employment in areas with shale oil and shale gas deposits. The U.S. Energy Information Association provides historical geological estimates of shale formation and the Railroad Commission of Texas holds a database of horizontal drilling permits, both of which I use to identify localities with high exposure to the fracking boom. I then compare college matriculation rates for fracking counties vs. non-fracking counties in a differences-in-differences setup. The implied elasticity of college attendance is similar to the estimates from the cross-industry analysis.

This study contributes to the literature on the impact of labor market changes and economic shocks on educational attainment. As described previously, research on mining and extraction sectors have found that an increase in labor market opportunities leads to lower educational attainment (Black et al., 2005; Morissette et al., 2015; Cascio and Narayan, 2022). In particular, Cascio and Narayan (2022) study the fracking boom and find that a 10% increase in earnings due to the fracking boom is associated with a 1.4pp decline in the high school enrollment-to-population ratio and a 0.9pp increase in male dropout. Other papers have found that increase in the minimum wage lead to reduction in college attendance (Lee, 2020; Schanzenbach et al., 2024). There is also substantial work from outside the U.S., particularly in developing country settings, which largely corroborate the negative relationship between economic shocks and educational attainment (Atkin, 2016; Shah and Steinberg, 2017, 2021).[1] My paper corroborates this literature and contributes a characterization of the types of students who are induced into and out of college due to changes in economic conditions.

A closely related literature to which I contribute is research on estimating the returns to college. Numerous studies have documented a large college wage premium, suggesting high returns to college for the average attendee (Katz and Murphy, 1992; Goldin and Katz, 2008; Acemoglu and Autor,

[1]A notable exception is Adukia et al. (2020) which finds heterogeneous effects of road construction on school attendance based on how road construction affects access to labor markets. Shah and Steinberg (2017) also find some evidence that young children may see an improvement in educational outcomes from a rainfall shock, in contrast with middle or older children.

2011; Autor, 2014). Most causal estimates of the returns to college show that marginal students have high returns to college (Card, 1995; Lemieux and Card, 2001; Bound and Turner, 2002; Stanley, 2003; Angrist and Chen, 2011; Zimmerman, 2014; Mountjoy, 2022).[2]

Several other papers bear particular mentioning. Several studies show that the fracking boom had large economic and social impacts. These papers find that by some economic measures many communities benefited from the fracking boom (Maniloff and Mastromonaco, 2017; Bartik et al., 2019). However, these benefits have come with some negative outcomes such as worse amenities (Bartik et al., 2019), worse schools (Marchand and Weber, 2020), and lower educational attainment (Zuo et al., 2019; Cascio and Narayan, 2022).[3] This paper also relates to the literature on the longer-run effects on the timing of labor market entry, which finds substantial scarring in the medium-term for workers entering the labor force during a recession (von Wachter, 2020).

2 Theoretical framework

I start with a simple theoretical model where students make decisions about whether to go to college or work after completing their last year of secondary education (graduating from high school). High school senior i in time $t = 0$ chooses two options: continuing their education through college (C) or joining the workforce (W). A student choosing to join the workforce cannot go back to college and nets utility $U_i(W)$. A student choosing to go to college nets utility $U_i(C)$ which includes both the costs of college attendance and earnings from joining the workforce with a college degree. The individual then decides to defer immediate earnings and attend college if $U_i(C) \geq U_i(W)$.

Utility is primarily determined by weighing expected earnings and the costs of college attendance. Earnings are governed simply by wages as a strictly increasing function of human capital. Through primary and secondary education, student i has human capital h_i. They may increase their human capital through college attendance, accumulating $p_i \geq 0$ additional units of human capital. For simplicity, I assume that all wage growth due to inflation or on the job human capital accumulation is governed through h_i and g_i. Thus, student i expects to earn wage $\underline{w}_i = w(h_i)$ from joining the workforce and wage $\overline{w}_i = w(h_i + g_i)$ after completing college. Since $w'(h) > 0$ is strictly in-

[2]Summary of earlier literature can be found in Oreopoulos and Petronijevic (2013).

[3]Relatedly, Schiller and Slechten (2025) find that oil and gas extraction do not directly impact ACT scores for college admissions but do find that low socioeconomic status students in school districts highly exposed to extraction do worse on ACT scores.

creasing, wages after college is weakly greater than wages after high school $(w(h_i + g_i) \geq w(h_i)\ \forall i)$. Students face two direct costs when attending college. Students pay a financial cost c_i in the form of tuition for college attendance. Students may also face a psychic cost p_i.

Student i attends college when

$$\overline{w}_i \frac{1}{1-\beta_i} - c_i - p_i \geq \underline{w}_i \frac{1}{1-\beta_i}, \tag{1}$$

where β_i is a discount factor for utility in future periods. Rearranging equation (1), we see that a student compares the college wage premium $(\overline{w}_i - \underline{w}_i)$ against the costs of college and the discount factor. College attendance probability increases with an individual's college wage premium; if a student has high human capital potential from college $(g_i \gg 0)$ they are likelier to go to college. Conversely, attendance probability decreases with college costs.

This framework highlights several factors which could contribute to over- or under-attendance of college. First, income or credit constraints may cause students to have higher discount factors. This would reduce the likelihood for attending college to earn wages and avoid incurring college costs. Second, students may misperceive several parameters in the framework. While the framework presented all objects as known to the student, only h_i and c_i are realized at the time of the decision. The remaining key parameters g_i, gains from college attendance, and p_i, the psychic costs of college, are not fully known. Students may form expectations over these parameters, with varying degrees of accuracy.

The starting wage holds a particularly important role in mediating student's decisions, leading to mismatches between students with the most to gain from colelge and those who ultimately attend. Unlike other parameters, students perceive their starting wage more directly. Risk-averse students may prefer these certain starting wages to the uncertatin college wages. Students may also perceive the starting wage level as an indicator of future wages and not account for economic conditions which may dampen or raise wages. This could lead to either overestimating or underestimating the college wage premium.

3 Data

3.1 Student data

The main observational data come from the Texas Education Research Center (ERC) database of students in the state of Texas. These data primarily contain information on students attending Texas public educational institutions, from primary school through college. These data are linked with additional data on college graduation, out-of-state college enrollment, and wages. This paper uses high school data, college data, and workforce data, with appropriate linkages between these datasets.

3.1.1 High school data

The Texas ERC data on public high school students that I use span the school years of 2003-04 to 2013-14.[4] While students in Texas attend high school for four years (grades 9 to 12), I restrict the observations to students enrolled in the last year of high school ("seniors") as of the beginning of the school year. I identify over 20 million such seniors from 2004–2014. For each senior, I obtain demographic information, attendance records, and test scores across multiple datasets.

Enrollment and attendance data The enrollment data serves as a roster of which school each senior attended as of October 31 of the school year. It also includes demographic characteristics of each student. This includes race, ethnicity, gender, socioeconomic status, Spanish proficiency, and other educational designations. Race is broken down into Asian, Black, White, and Native American.[5] Ethnicity describes whether a student is of Hispanic descent or not. Socioeconomic status is based on a student's eligibility for free- and reduced-price meals. I also observe the high school's assessment of whether a student is Spanish-speaking or not, which is important given the high proportion of Hispanic students in Texas and the potential importance of Spanish in workplace contexts. Other educational designations include indicators for students with special educational needs and participation in gifted programs.

[4]The "school year" for a student differs from the calendar year, with students starting the school year in the fall (typically August or September) and ending the school year in the spring (typically May or June). For ease of exposition, I refer to school years using the calendar year corresponding to the date of grade completion (e.g., a 9th grade student attending school from August 2008 to May 2009 is a 2009 9th grader).

[5]Race categories expand over the school years in the estimation sample. This is primarily the addition of the "two or more races" category. I exclude these students from the analysis.

The attendance data record the number of days present in school for each student. Data is available aggregated for each six-week grading period; 6 six-week grading periods make up a school year. I collect average attendance rates for all students in the school year prior to their senior year.

Testing data The testing data include individual student performance on state-administered standardized tests. For the vast majority of students, achieving a certain level of performance on high school standardized tests is a prerequisite for graduation in Texas. Therefore, most students in the sample have taken standardized tests as a high school student. Some students may have taken a particular test more than once, as the state allows students to retake exams that are required for graduation.

I obtain raw scores for each student on each standardized test and create two metrics as a measure of testing achievement. Since the state of Texas modified testing standards throughout the study period, I only use scores from the main testing instrument used to test for reading/verbal skills, which was the exit-level English Language Arts exam for the Texas Assessment of Knowledge and SKills (TAKS) and the main testing instrument mathematics/quantitative skills, which was the exit-level Mathematics exam for TAKS. Students typically took these exams in 11th grade, the penultimate school year of secondary education. I standandardize the raw scores within test administration year, which is standard in the education literature and allows for more meaningful comparison across testing cohorts. To account for students who take multiple tests, I use the latest score available prior to their enrollment in the senior year.

3.1.2 College data

The Texas ERC data on college students span the school years of 2004-05 to 2014-15.[6] For students attending public institutions of Texas, typically a two-year community college or a four-year publicly funded university, information is available on student admissions, enrollment status, and graduation. For students attending other private colleges in Texas, I only observe enrollment and graduation information.

The enrollment data contain information on student characteristics as of the census date in

[6]The school year for colleges in Texas is divided into two semesters, fall and spring, with enrollment in summer classes possible for certain institutions. Again, for ease of exposition, I refer to a school year using the calendar year corresponding to the date of the spring semester.

each semester, which typically occurs on the 12th day of class. I observe student choices for their college program, including the chosen degree for the student (e.g., associate's, bachelor's, etc.), and their college major. To complete their program and earn their degree, students are required to complete a certain number of semester credit hours ("SCH") throughout their time enrolled in college. The enrollment data reports this SCH for each student every semester. I also observe other demographic and administrative data on each student, such as their race/ethnicity, gender, tuition status, and on-campus residence status.

The graduation data reports key characteristics of every student who received a degree within a given school year. Key variables include the type of degree awarded to the student and the number of SCH hours at the time of graduation. Additional demographic variables are also available.

The admissions data include information on students who apply, are admitted, or enroll in a public, 4-year degree granting college in Texas. I observe a student's pre-college academic characteristics, such as their performance on the SAT and/or ACT—standardized tests required for admittance to most 4-year colleges in the United States—and their language fluency. The data also provide information on an applicant's family, including parental/guardian education levels and income. Finally, the data indicate an applicant's admittance decision and the type of admittance, when applicable (e.g., if the admittance was as a transfer from another college).

While this level of detailed individual information is only available for Texas college students, I supplement the ERC data with the National Student Clearinghouse ("NSC") data, which contains information on the vast majority of college students in the United States. These data allow me to observe the college choices of Texas high school graduates, in addition to limited information on their choices while in college. I obtain the institution of enrollment, length of enrollment, major, degree type, and part/full-time status from this dataset.

3.1.3 Workforce data

I use unemployment insurance wage data to document employment decisions and outcomes. The Texas Workforce Commission (TWC) collects wage reports from employers in Texas to document wages and establish unemployment insurance payouts. Data are organized by quarter and available for each job for which a person was paid wages in the state of Texas. For each job, I observe a worker's unique identification number, their county of residency, and their total wages for their job for the quarter. I also observe limited characteristics about the job and employer, including the county of the job location, the six-digit North American Industry Classification System (NAICS) code indicating the employer's industry, and the average monthly employment level for the employer. Importantly, these data do not contain the number of hours worked in a particular job.

For each worker and calendar year, I collapse the TWC wage reports to the industry level, which provides the total earnings in a calendar year by a given worker in a given year. I characterize the "industry" of a job using the job's NAICS code. Since the full, six-digit NAICS code reflects a significant amount of detail about the industry of a job, I categorize jobs using the first three-digits of the NAICS code.

3.1.4 Linking data sources

I use data across these settings—secondary school, postsecondary school, workplace—as the main observational sample. A unique identifier for each student/worker is used to link individuals observations across time and settings. Using these linkages, I construct a comprehensive database of high school seniors and their post-high school choices.

The final estimation dataset features the set of seniors in high school from the school years of 2004–2014. I restrict the sample to only those who are enrolled as seniors for the first time. For each student, I obtain their demographic characteristics, prior-year attendance rate, and verbal and quantitative test scores from the high school data. I merge each observation with college enrollment and college application information from the college data. I obtain the first year the individual is observed enrolling in college, if at all. For 2008 and onward, I am able to use the National Clearing House data to identify students who attend college outside of Texas. Finally, each observation is linked with annual wage information from the aggregated TWC data, providing a panel of earnings for each calendar year after starting high school.

Table 1 presents summary statistics of students in the estimation sample. Most students in the sample are White (42%) or Hispanic (39%); the Hispanic share is much higher than the rest of the U.S. population. Most students are not economically disadvantaged according to education records, while 34% of students are eligible for free or reduced-price school meals. 16% of the student population speak Spanish. Finally, 58% of students in the sample matriculate to college

Table 1. Summary statistics of high school student data

	Mean	SD
Race		
Indigenous	0.004	0.065
Asian	0.037	0.189
Black	0.132	0.339
Hispanic	0.392	0.488
White	0.424	0.494
Multiracial	0.010	0.098
Economic disadvantage		
Not disadvantaged	0.610	0.488
Eligible for free meals	0.272	0.445
Eligible for reduced-price meals	0.067	0.250
Other need-based	0.050	0.219
Spanish-speaking	0.164	0.370
Female	0.509	0.500
Test score		
Highest math score (standardized)	0.066	0.971
Highest reading score (standardized)	0.086	0.912
College matriculation	0.577	0.494
Observations	2, 002,566	

Notes: The table displays sample means and standard deviations of characteristics in the student sample. The sample includes all public high school students enrolled in their senior year from 2004–2014 in Texas.

immediately after their senior year.

3.2 Labor market conditions

I use data from the Quarterly Workforce Indicators (QWI) provided by United States Census Bureau from 2003-2015 to construct labor market condition indicators. The QWI contains quarterly information on earnings and employment levels for industries and across states. The QWI data are sourced from the Longitudinal Employer-Household Dynamics linked employer-employee microdata, covering over 95% of U.S. private sector jobs. Industries are broken down by the NAICS, which classifies businesses with a six-digit code based on the products or services provided by the business. I use variation in labor market conditions at the subsector-level, delineated by the "three-digit" NAICS code. Observing data at the state level is crucial, as I need to distinguish between labor market conditions measured using Texas data and conditions measured outside of Texas; I primarily use the latter to overcome the endogeneity challenge described in the subsequent section.

I make several alterations to the QWI dataset. First, data are not available for all states and all industries during the study period. To ensure that a changing sample does not impact the estimation of labor market conditions, I restrict the data to states for which I observe nearly full coverage across industries and industries for which I observe nearly full coverage across states. Second, the QWI provides data using the latest iteration of the NAICS (2022), while the Texas workforce dataset reports NAICS codes from the NAICS iteration in the year the data were recorded. While most changes during the study period occurred "within" three-digit NAICS codes, many shifts happened across three-digit codes. I remove all industries for which businesses were shifted across three-digit codes in the study period. After these modifications, the final labor market conditions data includes data from 33 states and 89 industries.

Industry-level, year-over-year wage growth for workers aged 19 to 24 serves as the primary measure of labor market conditions from my data. I first restrict the QWI data to only workers aged 19 to 24. To build the measure, I then combine aggregate quarterly earnings and employment counts in a given year. The industry-level wage for the year is simply the earnings divided by the employment counts. I calculate the year-over-year wage growth for industry k in year t as the percentage difference between the wage in year t and year $t-1$.

I assess the variation in labor market conditions from 2001 to 2019. I consider which factors contribute most to the variation in wage growth by regressing wage growth on industry fixed effects, year fixed effects, and lagged wage growth. First, I find that across this time period, industry fixed effects only account for 5.5% of the observed variation, suggesting cross-industry differences are not significant drivers of wage growth. Second, year fixed effects account for 34% of the observed variation in wage growth. This indicates that controlling for aggregate temporal shocks will be important in isolating quasi-random variation in wage growth. Finally, lagged wage growth only accounts for 1% of the observed variation in wage growth. All three factors account for 42% of the variation in wage growth.

Next, I show how predictive non-Texas labor market conditions are of Texas labor market conditions. Following the previous analysis, year fixed effects, industry fixed effects, and lagged Texas wage growth account for 49% of the variation in Texas-specific wage growth. Adding non-Texas wage growth to the model, I find that this factor accounts for more than 10 additional percent of variation in Texas wages. The coefficient on non-Texas wage growth is 0.55.

4 Estimation strategy

4.1 Shift-share IV

4.1.1 Introduction

I estimate the impact of labor market opportunities on college matriculation. Let i index students. Let W_i represent the college decision for student i and let Y_i represent the economic outcome for student i. I model college decision W_i as a function of individual characteristics $\mathbf{\Gamma}_i$, labor market conditions X_i, and a random component η_i. Thus, the outcome model is given as:

$$W_i = f(\mathbf{\Gamma}_i, X_i, \eta_i). \tag{2}$$

An important characteristic determining college decisions is the prevailing labor market condition faced by the student at the time of the decision, denoted X_i, which is higher when labor market conditions are favorable to workers (i.e., more vacancies or higher wages) and lower when conditions are less favorable. Assuming the effects of all components are additively separable, the outcome model is:

$$W_i = \theta\mathbf{\Gamma}_i + \delta X_i + \eta_i. \tag{3}$$

Although δ is the target parameter of interest—the effect of labor market conditions on college decisions—it is not straightforward to estimate. First, data on X_i is not readily available. Knowing the prevailing market conditions for student i requires information on (a) the set of jobs and careers in which the student is interested, (b) the set of firms willing to hire the student, and (c) the labor supply of other individuals at the intersection of these sets. Second, even if X_i were observable, one may be worried about X_i being endogenously determined in relation to W_i. One particular source of endogeneity is that the same economic conditions that impact labor market conditions may also impact college quality. That is, local economic conditions change both potential labor market outcomes for students and funding at local universities.

I propose using a shift-share variable, Z_i, in place of X_i to overcome this estimation challenge. I construct Z_i as a combination of a set of "relevant" industries for each individual i and changes in labor market conditions for those industries in non-Texas areas. Formally,

$$Z_i \equiv \sum_{k=1}^{K} s_{ik} g_{k,t(i)}, \tag{4}$$

where k indexes industries such that $(s_{i1}, \ldots, s_{iK})$ represent the industry weights for individual i and $(g_{1,t(i)}, \ldots, g_{K,t(i)})$ represent changes in labor market conditions for industries for year $t(i)$. If $g_{k,t(i)}$ weighted by s_{ik} is uncorrelated with η_i from equation (2), then Z_i may be considered a valid instrument for X_i, following from standard shift-share identification methods from Borusyak et al. (2023). Thus, I argue that the common "shifts" $g_{k,t(i)}$ across i are exogenous, while individualized "shares" $s_i k$ within individuals may be endogenous.

4.1.2 Labor market shifts

To ensure that Z_i meets the exclusion restriction for shift-share identification, I construct $g_{k,t(i)}$ using the QWI earnings and employment data for states in the continental United States, excluding Texas. Let $t(i)$ be the calendar year in which we observe college decisions for individual i. Then for each industry k, I estimate $\overline{w}_{k,\tilde{t(i)}}^{-Texas}$ which is the average hourly wage for industry k in calendar year $\tilde{t(i)}$ for the U.S. excluding Texas. Since data is available at the state level, I use a weighted average across states based on the number of workers in the industry for each state. Then $g_{k,t(i)}$ is defined as:

$$g_{kt(i)} = \frac{\overline{w}_{k,t(i)-1}^{-Texas} - \overline{w}_{k,t(i)-2}^{-Texas}}{\overline{w}_{k,t(i)-2}^{-Texas}}, \tag{5}$$

which is simply the year-over-year wage growth rate for the industry excluding Texas.

These shifts $g_{k,t(i)}$ are relevant to the wages earned by students in the estimation sample. For each student in their senior year, I calculate the log earnings for the primary industry they enter after high school graduation. I regress log starting earnings on the shifts, controlling for industry and year fixed effects. Table 2 presents the results of this regression. In column (1), I show that wage growth positively predicts the earnings of students immediately following high school graduation. When adjusting for the predicted earnings in the year prior to graduation, the coefficient on wage growth remains statistically significant and rises close to 1 (column (2)).

Table 2. Predictions of log post-graduation earnings for high school students

	(1)	(2)	(3)
Shifts ($g_{k,t(i)}$)	0.677	0.991	0.917
	(0.195)	(0.198)	(0.224)
Predicted log($earnings$) in $t-1$		0.601	0.736
		(0.145)	(0.199)
$g_{k,t(i)-1}$			0.063
			(0.171)
$g_{k,t(i)-2}$			0.046
			(0.139)
$g_{k,t(i)+1}$			0.645
			(0.219)
$g_{k,t(i)+2}$			0.002
			(0.244)
Observations	1,674,971	1,674,971	1,674,971

Notes: Each specification is a regression of the logarithm of post-graduation earnings on labor market "shifts." Unreported controls include industry fixed effects and year fixed effects. Standard errors reported in parentheses below coefficient estimates are clustered at the industry-year level. Industry is defined as the three-digit NAICS code of a business. The unit of observation is a student. The outcome is calculated as the natural logarithm of the earnings for an individual (a) in the year after high school graduation and (b) in their primary industry. "Shifts" are calculated at the industry-time level and correspond to the non-Texas wage growth described in Sections 3.2 and 4.1.2. "Predicted log(earnings) in $t-1$" are the predicted earnings for the student's chosen industry in the year prior to their labor market entry. Predictions are made based on the QWI data. $t(i)$ is the year of high school graduation, so each $g_{k,\cdot}$ corresponds to the shifts calculated in relation to this index year.

In column (3), I include wage growth from the two years on either of the graduation year. The coefficient on wage growth in year t remains relatively stable with the inclusion of these additional covariates, while the coefficients on wage growth in most other years remain close to 0 and are statistically insignificant. Notably, wage growth in year $t+1$ is predictive of log earnings. This is likely because earnings may rise over the course of the year, meaning some "future" growth is captured in the earnings figures. I display a binned scatter plot of column (2) in Appendix Figure A1.

Shifts are not only related to the actual wages earned by students but the post-high school choices of students. In Table A1, I show the results of regressing year-over-year employment growth for the estimation sample on shifts at the industry-year level. I show that with and without predicted earnings, the shifts are predictive of employment growth, suggesting that students are indeed selecting industries in response to these shifts.

4.1.3 Industry preference shares

The industry weights $(s_{i1},\dots,s_{iK})$ represent the propensity for individual i to work in a particular industry at baseline, after high school graduation. Lacking directly observed data on the industries preferred by or available to a particular individual, I use demographic characteristics, educational outcomes, and location to predict such likelihoods from a prior, baseline period. The prediction exercise yields $(s_{i1},\dots,s_{iK})$ such that $\sum_k^K s_{ik} = 1$ and $0 \le s_{ik} \le 1$, $\forall i$. Each s_{ik} aggregates both labor demand and labor supply factors which affect the selection of individual i into industry k. On the demand side, businesses in industry k may have a certain preference in hiring employees with particular attributes. These preferences are not only limited to instrumental factors that drive hiring decisions like student achievement, but also extend to systematic- or taste-based discrimination. On the supply side, students have preferences for working in particular industries over others. Again, these preferences may be driven by taste or by prevailing wages and amenities.

I formally estimate s_{ik} using a multinomial logit model with maximum likelihood. I begin with a model of utility for individual j choosing industry k:

$$U_{jk} = \mathbf{X}_j^T \gamma_k + \mathbf{L}_j^T \lambda_k + \epsilon_{jk}, \tag{6}$$

where $\mathbf{X}_j$ is a vector of demographic characteristics and $\mathbf{L}_j$ is a vector of location dummy variables. γ_k is a coefficient vector for demographic characteristics representing the utility contribution of selecting industry k for each demographic characteristic. λ_k is a coefficient vector representing the utility contribution of selecting industry for a particular location. Finally, ϵ_{jk} is an error term following a type 1 extreme value distribution. Letting $V_{jk} = U_{jk} - \epsilon_{jk}$, the probability that j chooses k is given by

$$P_{jk} = \frac{e^{V_{jk}}}{\sum_k e^{V_{jk}}}. \tag{7}$$

Given a sample of J students and their observed industry choices, the probability of person j their industry choice k^j under this model is $\prod_j (P_{jk})^{\mathbb{1}\{k=k^j\}}$. Assuming independence of choices across students, the log-likelihood function is:

$$\ell(\gamma_k, \lambda_k) = \sum_{j}^{J} \sum_{k} \mathbb{1}\left\{k = k^j\right\} \cdot \ln(P_{jk}). \tag{8}$$

I maximize $\ell(\cdot)$ from equation (8) using high school students' industry choices from 2002 to 2003, which is excluded from the main estimation sample. I identify all students in their last year of high school who are projected to graduate in school year t and are observed as employed at any time in year $t+1$. Each student's industry choice is defined as the industry for which they had the highest earnings. For $\mathbf{X}_j$, I include race and ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic Black, non-Hispanic White), gender, high school test scores (math and reading), an indicator for Spanish-speaking, and free and reduced-price lunch status (free, reduced-price, non-adjusted). For $\mathbf{L}_j$, I use indicators for the school's commuting zone for a student in their senior year. Schools are assigned to counties, and counties are assigned to commuting zones using definitions provided by the U.S. Department of Agriculture. Using a multinomial logistic regression with these factors, I estimate $\hat{\gamma}_k$ and $\hat{\lambda}_k$ which maximizes $\ell(\cdot)$.

I use the estimates of $\hat{\gamma}_k$ and $\hat{\lambda}_k$ from maximizing the log-likelihood function to obtain predictions for industry selection in the main estimation sample. For each student i in the estimation sample, I calculate P_{ik}, the probability of selecting industry k for individual i:

$$P_{ik} \equiv \frac{e^{\mathbf{X}_i^T \hat{\gamma}_k + \mathbf{L}_i^T \hat{\lambda}_k}}{\sum_k e^{\mathbf{X}_i^T \hat{\gamma}_k + \mathbf{L}_i^T \hat{\lambda}_k}} \tag{9}$$

where $\mathbf{X}_i$ and $\mathbf{L}_i$ is defined analogously to $\mathbf{X}_j$ and $\mathbf{L}_j$. Then industry weight s_{ik} is simply:

$$s_{ik} \equiv P_{ik} \tag{10}$$

which satisfies $\sum_k^K s_{ik} = 1$ and $0 \leq s_{ik} \leq 1, \ \forall i$.

The resulting industry weights s_{ik} represent both labor supply and labor demand forces specific for a given individual. For example, if s_{ik} is relatively high, this could be due to preference among individual i's type for jobs in industry k (e.g., relative preference for men to take construction jobs) or due to preferential selection of individual i from the industry (e.g., relative preference for bilingual speakers in service jobs). Thus, the γ_k coefficient in the utility function can best be thought of as

Table 3. 3-digit NAICS Industries with largest industry weights

Industry (k)	Average $s_{i,k}$
Food Services and Drinking Places	0.214
Administrative and Support Services	0.101
General Merchandise Retailers	0.079
Food and Beverage Stores	0.052
Clothing and Clothing Accessories Retailers	0.045
Professional, Scientific, and Technical Services	0.032
Sporting Goods, Hobby, Musical Instrument, Book Stores	0.026
Specialty Trade Contractors	0.025
Motor Vehicle and Parts Dealers	0.023
Ambulatory Health Care Services	0.022
Industry count: 89	Industry average: 0.011

Notes: The table shows the industries with the highest average industry weights/shares s_{ik} in the sample. Industries are defined at the level of the 3-digit NAICS code.

representing the degree of "match" for industry k based on demographic characteristics $\mathbf{X}_j$ and the λ_k coefficient as representing a combination of local labor supply tastes and industry availability. In most contexts, distinguishing between such forces is central or crucial to the estimation exercise. However, such a decomposition is not necessary in this context as the weights will be solely used to appropriately distribute the effect of arguably exogenous wage shifters $g_{k,t(i)}$ across individuals i.

I provide descriptive statistics of these industry shares s_{ik}. First, I show the industries with the ten largest shares across students in Table 3. The top industries are dominated by service industries, with food services and retail dominating the top industries. Over 21% of students join the food services industry, and over 10% of students join the administrative and support services industry. Close to 20% of students join some sort of retail industries, such as general merchandise retailing (e.g., department stores) or food and beverage stores (e.g., grocery stores).

Industries vary in which student characteristics explains the variation in industry weights across students. For the top 5 industries with highest average s_{ik}, I regress the industry weights on each student characteristics, and examine the share of variation each characteristic explains. Table A2 displays the R^2 for each characteristic obtained through these regressions. For these selected industries, student locality and gender are the strongest predictors of industry choice. For example, commuting zone fixed effects explain 60% of the variation in predicted exposure to industry 445 ("food and beverage stores"), while gender explains over 50% of the variation in predicted exposure to industries 722 ("food services and drinking places") and 458 ("clothing and clothing accessories retailers"). However, other student characteristics also matter for industry exposure: student race/ethnicity explains 35% of the variation in exposure to "administrative and support services." While test scores do not seem to hold a substantial amount of explanatory power for exposure to these selected industries, I note that test scores seem to matter for retail industries (455, 445, and 458). This suggests some selection over student achievement in these industries.

Finally, I observe other demographic trends in the share data. As expected, industry weights in service industries are much higher for female students than male students. I find Black students are more exposed to industries related to transportation or couriering, while Hispanic students are more likely to be in telecommunications. Spanish-speakers are more likely to be exposed to agriculture while non-Spanish-speakers are more likely to be exposed to oil and gas extraction. I also find surprising divergence across industries within a sector. For example, in the resource extraction sector, I find that the non-oil and non-gas mining industry attracts lower test score students while the petroleum product manufacturing industry attracts higher test score students.

4.1.4 Estimating equation

I estimate the impact of labor market conditions on college decisions for student i using the OLS regression

$$W_i = \theta_{d(i)} + \tau_{t(i)} + \delta Z_i + u_i \tag{11}$$

where W_i is a binary variable for attending college in the year after graduating high school and Z_i is the sum of $g_{k,t(i)}$ weighted by s_{ik}, as in equation (4). $\theta_{d(i)}$ represents fixed effects for individual i's demographic set $d(i)$, which includes all attributes in $\mathbf{X}_i$ and $\mathbf{L}_i$ used to estimate s_{ik}. This fixed effect nets out the average experienced labor market condition and college-going rate for all students in the demographic set. $\tau_{t(i)}$ represents year fixed effects, which control for temporal shocks shared across students.

Though not represented in equation (11), Z_i is implicitly an instrument for student i's experienced labor market condition X_i, which is unobserved. As such, Z_i, must meet the identifying assumptions of instrument relevance and exclusion. Since the effect of Z_i is estimated in reduced form, instrument relevance is already assumed. For exclusion, Z_i must be uncorrelated with u_i in equation (11). I argue for the exogeneity of shifters $g_{k,t(i)}$, following Borusyak et al. (2022). That is, changes in wages by industry for the U.S. excluding Texas should only affect college decisions through the industries which Texas students consider to be employment opportunities.

I identify two major general threats to the assumption that the wage growth shocks in Z_i are not correlated with u_i and outline how the controls in equation (11) account for these concerns. Consider that some industries have higher wage growth than other industries during the study period. These differential growth patterns may be correlated with student characteristics. For instance, with skill-biased technological change, certain industries may have higher wage growth and recruit more heavily students with higher school achievement. Higher school achievement is also correlated with going to college, but not necessarily through the wage shock channel. This generates a pathway in which shifters in certain industries are associated with unobserved characteristics which also affect the outcome W_i. To shut down this potential pathway, the estimation must include a control for $\bar{g}_k$, which is the average wage growth for industry k across the study period. I simply replace $g_{k,t(i)}$ with $\bar{g}_k$ in the shift-share formula to find the student sample analogue of $\bar{g}_k : \sum_k s_{ik}\bar{g}_k$. Since s_{ik} is estimated based on the characteristics given by demographic set $d(i)$, $\sum_k s_{ik}\bar{g}_k$ is simply spanned by $\theta_{d(i)}$, included in estimation equation (11).

The second way in which the wage growth shocks may be correlated with u_i is through common temporal shocks shared across industries. As discussed previously, the economic conditions which lead to worse or better local labor market conditions may also affect college funding. College funding can be a major factor that high school students consider when deciding whether to attend college. I must then account for the average wage shock in a single time period $\bar{g}_t$. Again, I simply replace this object in the shift-share formula, which yields $\sum_k s_{ik}\bar{g}_{t(i)} = \bar{g}_{t(i)} \sum_k s_{ik}$. Given that $\sum_k s_{ik} = 1$, this resolves to $\bar{g}_{t(i)}$, simply spanned by year fixed effects $\tau_{t(i)}$ included in equation (11).

I formally show evidence that equation (11) estimates the causal impact of the shift-share variable Z_i on W_i. I demonstrate that there is sufficient variation in $g_{k,t(i)}$ and balanced s_{ik} such that shocks are numerous and not entirely dependent on a particular set of industries or a particular set of years. Table **??** summarizes the relevant statistics of $g_{k,t(i)}$ and s_{ik}. I focus on the distribution of g_{kt} after it is residualized by industry and year fixed effects. The shifts appear to be well-

Table 4. Shock-level summary statistics

Distribution of $g_{k,t}$	
Mean	0
SD	.019
IQR	.0172
Largest average share $\bar{s}_{k,t}$	.0281
Number of observations	
Industries	88
Years	9
Industry-year	792
Effective N	111

$g_{k,t}$ residualized by k and t.

Notes: The table shows summary statistics of the industry-year shifts and student-industry shares at the industry-year level. The first three rows describe the distribution of shifts, $g_{k,t}$, after they are residualized by industry fixed effects and year fixed effects. The next row describes the highest observed average share $\bar{s}_{k,t}$ which is calculated as the average s_{ik} across students in a year t. Finally, the last four rows describe the number of observations at each level. Effective N is computed as the inverse HHI of $\bar{s}_{k,t}$ across kt: $1/\sum_{kt}\bar{s}^2_{k,t}$.

distributed. The standard deviation of a shift is equivalent to a 2% residualized wage growth. The IQR is slightly smaller than expected, suggesting that the the shifts are more concentrated around 0. Next, I investigate the shares s_{ik}. I compute $\bar{s}_{kt}$ which is the average share for an industry k across students in a given year t. The largest value of $\bar{s}_{kt}$ is .028, which means close to 3% of the effective "weight" of the estimation is placed in a single industry-year. To test the extent to which these large $\bar{s}_{kt}$ are skewing the estimates, I compute an "effective N" following Borusyak et al. (2022), which is calculated as the inverse Herfindahl-Hirschman Index of average shares, $1/\sum_{k,t}\bar{s}^2_{kt}$. This brings down the "effective" sample size to 111 observations, which is enough to estimate coefficients and standard errors.

Simply using OLS or standard IV regressions at the student-level does not yield appropriate standard errors for coefficients (Borusyak et al., 2022). For valid inference using shift-share variables, I convert the student-level regression from equation (2) into a regression at the "shock-level": regressions where the level of observation is the industry. First, I residualize W_i and X_i (which as described is equivalent to Z_i) on the set of controls $\theta_{d(i)}$ and $\tau_{t(i)}$, yielding residuals $W_i^{\perp}$ and $X_i^{\perp}$. Next, to convert the residualized outcome and explanatory variable at the shock-level, I take an average of each across observations, weighted by industry exposure shares:

$$\overline{C}_{kt}^{\perp} = \frac{\sum_{it} s_{ik} C_i^{\perp}}{\sum_{it} s_{ik}}, \tag{12}$$

where $C \in \{W, X\}$. Finally, I estimate the second-stage equation:

$$\overline{W}_{kt}^{\perp} = \alpha_k + \tau_t + \delta^{SSIV} \overline{X}_{kt}^{\perp} + \zeta_{kt}, \tag{13}$$

where $\overline{X}_{kt}^{\perp}$ is instrumented by g_{kt}. Borusyak et al. (2022) shows that the estimate δ^{SSIV} is equivalent to δ from directly estimating equation (2), but with appropriately adjusted standard errors.

In addition to estimating equation (11) for college matriculation, I also estimate the impact of Z_i on earnings. Positive labor market conditions may shift students into the labor force, but these decisions may be suboptimal for students. I assess the medium-run outcomes of the labor market and educational decisions by inspecting the earnings of each student, years out from high school graduation. I thus replace W_i in equation 11 with earnings for individual i in year $t(i)+v$, $Y_i^{t(i)+v}$.

5 Results

5.1 Effect on college matriculation

5.1.1 Simple estimate

Table 5 shows the results of estimating equation (11). As predicted, students are less likely to attend college when their outside option improves. Using the estimates in column (2), which includes both student characteristics controls and year fixed effects, I find that a 1 percentage point improvement in wage growth leads to a 2.4 percentage point decrease in the probability of college matriculation. Estimated in standard deviation terms, a standard deviation increase in labor market condition leads to a 3.2 percentage point decrease in the probability of enrolling in college.

To ensure there are not omitted factors or trends which explain these results, I conduct two placebo tests. First, I conduct a form of a balance test where I regress college matriculation in year t on labor market conditions in $t+1$. If the changes in $t+1$ predict decisions in t, that implies that students are not reacting to changes in their outside options, but something else which may co-vary

Table 5. Effect of labor market conditions on college matriculation

	(1)	(2)
Labor market condition (wage growth shock)	-0.738	-2.349
	(0.227)	(0.525)
Student characteristics FE		✓
Year FE		✓
Dependent variable mean	0.584	0.586
Labor market condition SD	0.0138	0.0138
Student observations	1933908	1871613
Industry-year observations	792	792
Effective N	110.6	110.6

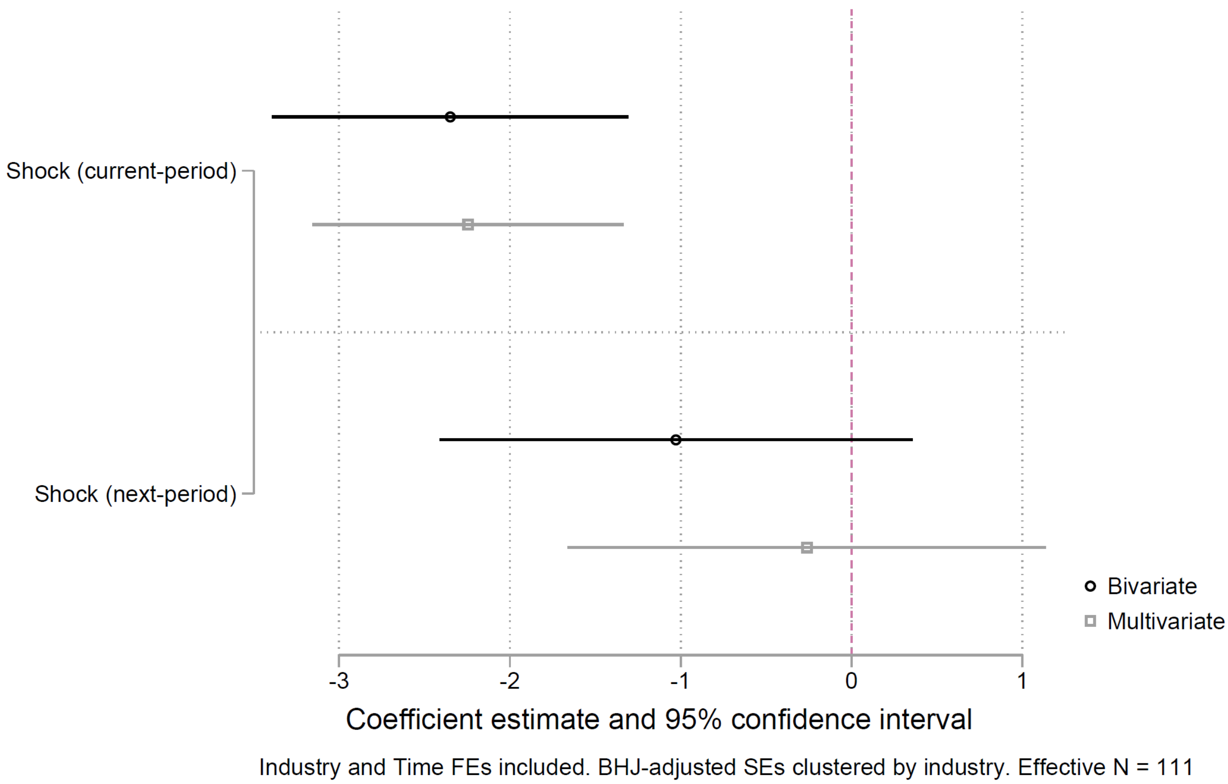
BHJ (2022)-adjusted standard errors clustered by industry in parentheses.

Notes: Each specification is a regression of college matriculation probability on the labor market conditions shift-share variable. Student characteristics FE are reported in equation (11). Standard errors reported in parentheses below coefficient estimates are (a) clustered by industry and (b) estimated at the "shock"-level following equation (13). Industry is defined as the three-digit NAICS code of a business. The unit of observation is a student. The binary outcome takes on a value of 1 if the student enrolls in college the fall immediately following high school graduation. The shift-share labor market condition variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares."

with labor market conditions. Figure 1 displays the coefficient estimates on both "current-period" labor market conditions, Z_i, and "next-period" labor market conditions, $Z_i^{t(i)+1}$. Focusing on the "bivariate" estimates, I fail to reject the null hypothesis that next-period labor market conditions are equal to zero, but the point estimate is sizable and negative. One possible explanation is that due to the construction of the shock-variable, students are effectively able to partially "observe" changes in labor market conditions which are included in $Z_i^{t(i)+1}$ by construction. To account for this possibility, I include both current-period and next-period shocks in a single regression. Figure 1 shows that while the coefficient estimate on current-period labor market conditions remain largely unchanged, the point estimate on next-period labor market conditions attenuates to zero.

I conduct an additional placebo test which randomly permutes the wage growth shifts across industries. I take each industry's time vector of shifts and permute this vector across industries. Using these new shifts, I recalculate the labor market condition shock and re-estimate equation (11). I repeat this permutation 1000 times, obtaining separate coefficient estimates each time. This exercise explores the possibility that my coefficient estimates arose from chance, potentially because it was over-reliant on wage growth shifts in a handful of particular industries. Figure A2 plots the recalculated point estimates of the effect of a standardized shock on college matriculation

Figure 1. Effect of current-period and next-period shocks on college matriculation probability

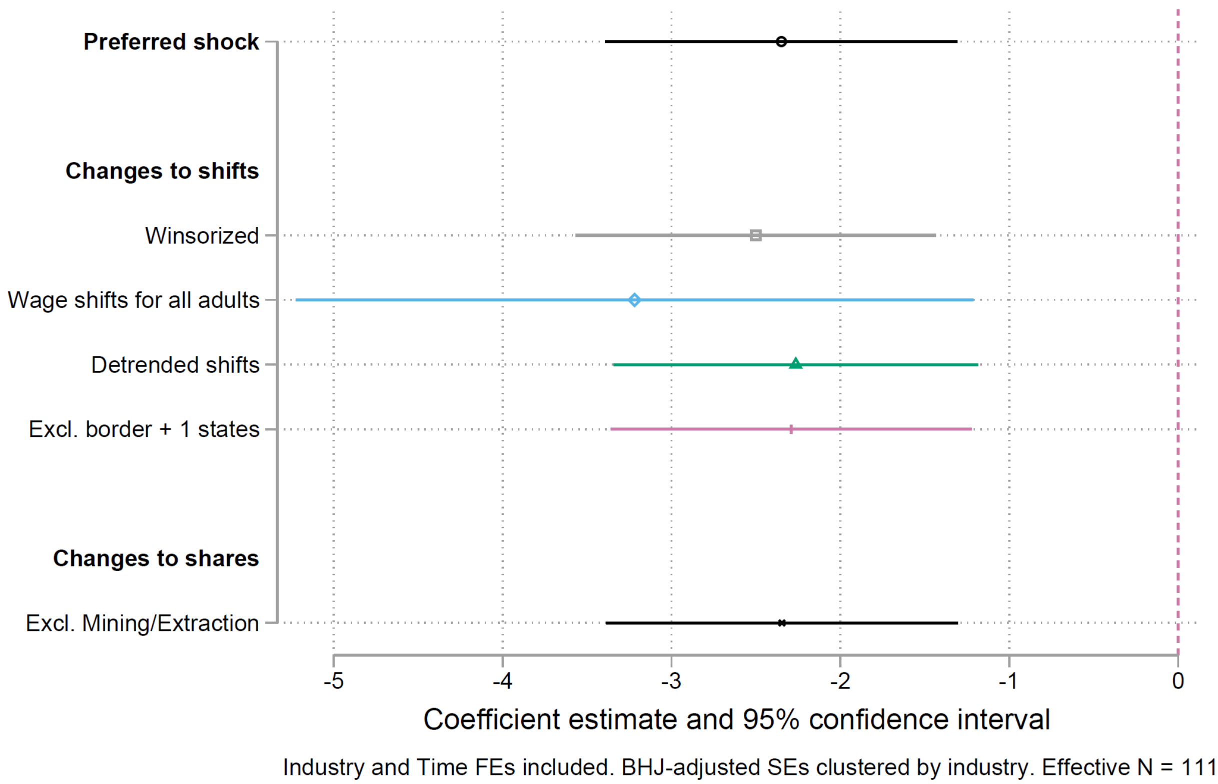


Notes: This figure represents coefficient estimates of the effect of current-period and next-period shock on college matriculation probability. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." The "bivariate" estimates regresses the outcome on each shock in separate specifications. The "multivariate" estimates regresses the outcome on both shocks in one specification.

probability. I find that only three permutations of industry shifts produce estimates as large in magnitude as the ones I obtain in my main estimation. I also observe some permutations with a very large standard deviation in the labor market condition shock. This reflects the fact that there exist some industries with relatively large swings in wage growth and some industries have relatively high exposure shares across students.

I also assess the sensitivity of my estimate to alternative ways to construct the labor market condition shocks. The coefficient estimate is relatively insensitive to both changes in the wage growth shifts and the industry propensity shares. I change shifts by reducing reliance on wage

Figure 2. Sensitivity of shock effect on college matriculation to alternatively constructed shocks



Notes: This figure represents coefficient estimates of the effect of labor market condition shocks on college matriculation probability. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained with a differently calculated shock.

growth outliers, using wage growth for all adults instead of young adults, residualizing wage growth with two-year lags of wage growth, and excluding states proximate to Texas when constructing the shifts. I change shares by excluding exposure to the mining and extraction sector, which is heavily concentrated in Texas. Figure 2 shows the coefficient estimates on labor market condition shocks adjusted in these ways. I find that point estimates are largely stable across changes. One notable exception is that the point estimate on the shock using wage shifts for all adults is higher than for the baseline shock with much larger standard errors. This may reflect the fact that wage growth for 19–24 year-old adults are measured with higher error—which would attenuate coefficient estimates—but wage growth for adults carry less relevant identifying variation.

5.1.2 Heterogeneity

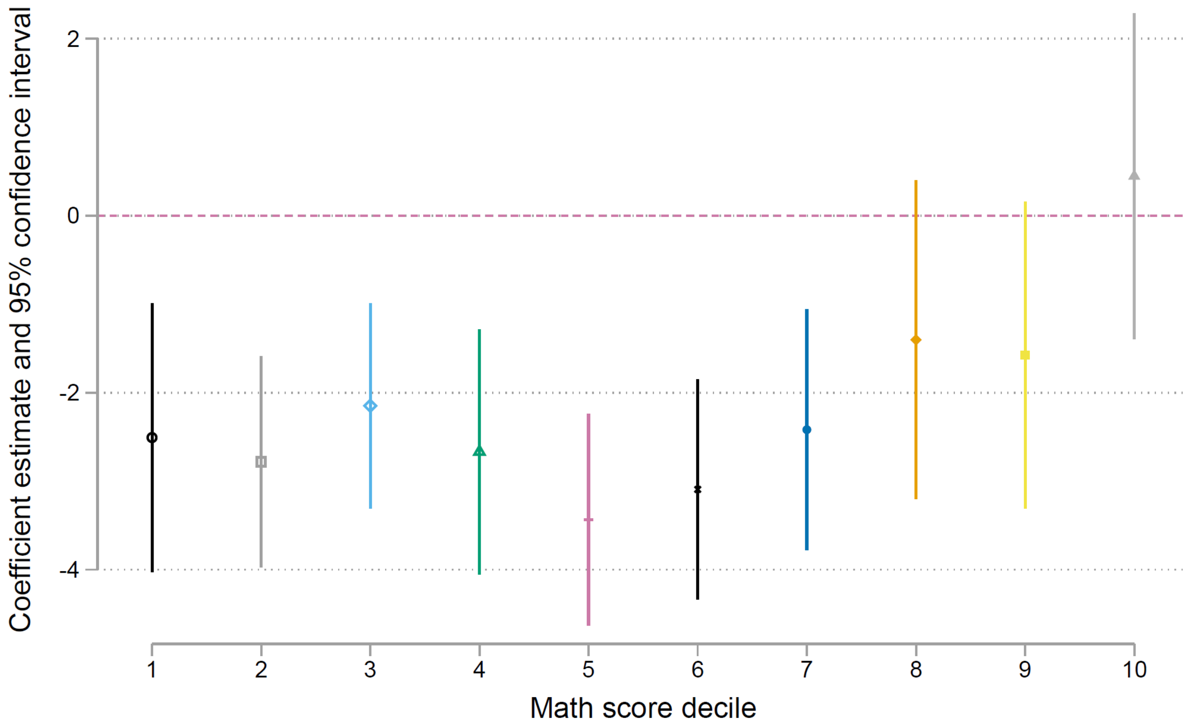
Estimating the effect of labor market condition shocks on college matriculation for the entire student sample masks potential heterogeneity in what types of students are responsive to shocks. Students with lower (higher) potential gains from college attendance may not be responsive to shocks as they are not likely (likely) to pursue higher education anyway. That is, students who can greatly improve their lifetime earnings from college attendance are not likely to be enticed by higher wages to join the labor force, while students who cannot improve their lifetime earnings from college are not likely to avoid low wages through college attendance.

Since potential gains are unobservable, I proxy for this using a student's test scores. I divide up standardized math scores by deciles and separately estimate the impact of shocks on college matriculation. One note of caution with these estimates is that since math scores are used to construct the industry exposure shares, conducting heterogeneity across this dimension purges math score variation in shares. Concerns stemming from this change are mitigated by the fact that there are two achievement proxies in constructing shares: math score and reading score. Further, Table A2 shows that test scores do not play an over-sized role in actually explaining variation across shares. I also consider heterogeneity across reading test score and predicted college matriculation. The latter variable is calculated as a prediction of college matriculation probability for each student i based on their demographic characteristics $\mathbf{X}_i$ and location $\mathbf{L}_i$. Formally, I estimate a logit regression of college matriculation on these covariates for the set-aside sample of 2002–2003 students and apply the resulting coefficient estimates to students in the estimation sample.

Figure 3 plots the coefficient estimate on labor market condition shocks separately for each decile. I find consistently sized coefficient estimates across much of the math score distribution: point estimates from deciles 1 through 7 are similar in magnitude and are significantly different from zero. While I am unable to reject the null hypothesis, estimates nearer the top of the distribution are also negative and sizable in magnitude. I find similar estimates using deciles of reading test scores (Figure A3) and deciles of predicted college matriculation (Figure 4).

These results demonstrate that changes to labor market opportunities impact a relatively broad range of students. I first find evidence that students at the very top of the achievement distribution

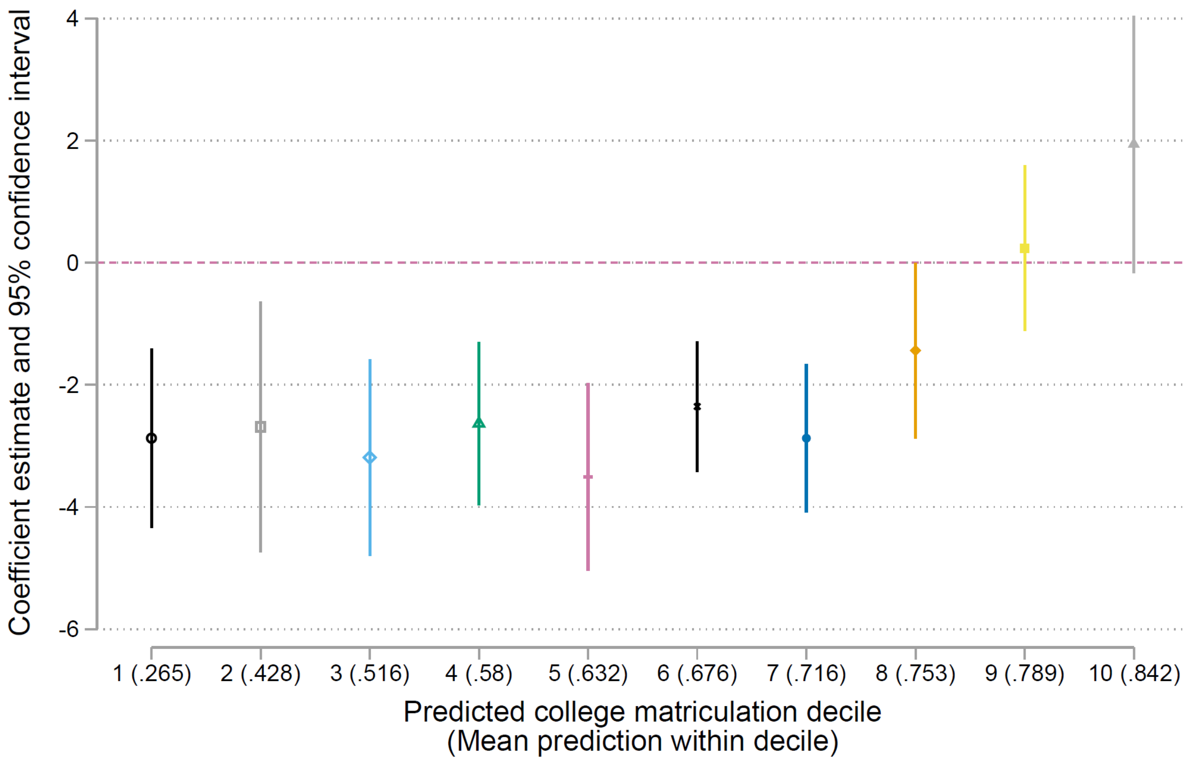
Figure 3. Effect of labor market condition shock on college matriculation by math score decile



Notes: This figure represents coefficient estimates of the effect of labor market condition shocks on college matriculation probability by math score decile. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained on a separate sample based on students' standardized math score decile.

or those very likely to go to college are unaffected or relatively unaffected by labor market shocks. These represent students who are most likely to attend college or to see earnings gains from a college degree, so their relative inelasticity is potentially unsurprising. However, I find students who are just below this threshold are responsive to labor market shocks. While potential gains are unobservable at this point, we may expect relatively high gains to college attendance for these students, which make their responsiveness more puzzling. At the lowest end of the achievement distribution, students are also responsive to labor market shocks. On one hand, this behavior is also unusual: students struggling in high school should be relatively inelastic to push or pull factors

Figure 4. Effect of labor market condition shock on college matriculation by predicted college matriculation decile



Notes: This figure represents coefficient estimates of the effect of labor market condition shocks on college matriculation probability by predicted college matriculation decile. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained on a separate sample based on students' predicted college matriculation decile. A student's predicted college matriculation is the probability of college attendance based on other students with the same demographics and in the same location.

into college. Nevertheless, insofar as policymakers seek to increase "blue collar" job opportunities, they would likely prefer those positions be given to students who may have less to gain from college attendance.

5.2 Effect on earnings

I assess whether the students who are responsive to labor market shocks are better or worse off in the medium-term. While there are non-pecuniary benefits to college attendance, I focus on the yearly earnings of students who are outside of college. If students who join the full-time labor force early would not have benefited from college, their short- to medium-term earnings would be larger than for students who decided to attend college. However, if those who opt to work would have benefited from college education, their earnings would only be better in the short-term while reverting to being worse in the medium-term.

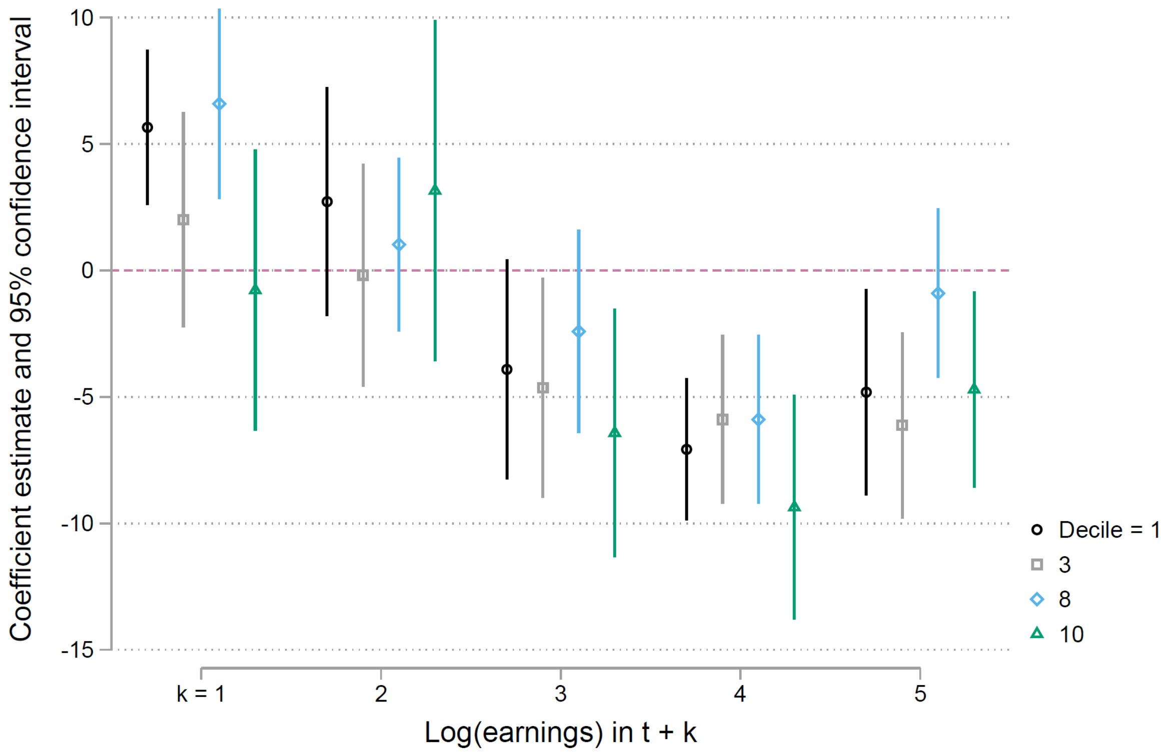
For each student, I calculate their annual earnings 1 to 5 years out from their senior year and estimate how they're affected by labor market shocks at the time of high school graduation. I use the natural logarithm of annual earnings as the main outcome measure. To account for individuals with no income, I separately obtain estimates at the extensive margin: whether an individual earns any money or not. Figure 5 plots the coefficient estimates for the impact of labor market shocks on log(earnings) 1 to 5 years out from high school separately for different deciles of high school math score. Comparing decile 1 with decile 10, earnings are relatively higher for students in decile 1 immediately after graduation, as these individuals choose to work more. However, 5 years out from graduation, the earnings gap has completely dissipated. Turning to the extensive margin, I find that contrary to the intensive margin results, individuals in decile 10 are consistently more likely to be working than those in lower deciles all years out from graduation (Figure 6).

6 Case study: The "fracking" boom

While the shift-share empirical approach has the advantage of leveraging broad labor market variation, it also holds a number of limitations. One major concern is that the wage growth shifts may not truly be conditionally exogenous. Wage growth in one industry may be correlated with wage growth in another industry. Wage changes may also be influenced by changes within Texas, which holds around 9% of the U.S. population. This leads to the possibility of an omitted variable local to Texas affecting both national wages and college-going decisions. Overall, the driving force of each of the individual industry-year wage shifts is not well-identified, which makes interpretation of the overall results challenging.

I overcome this challenge by focusing next on one specific labor market shock: the hydraulic fracturing ("fracking") boom in Texas. Due to technological advancements and fluctuations in oil and gas prices, it became feasible to use "fracking" to extract resources from previously unreachable

Figure 5. Effect of labor market condition shock on log(earnings) by math score decile

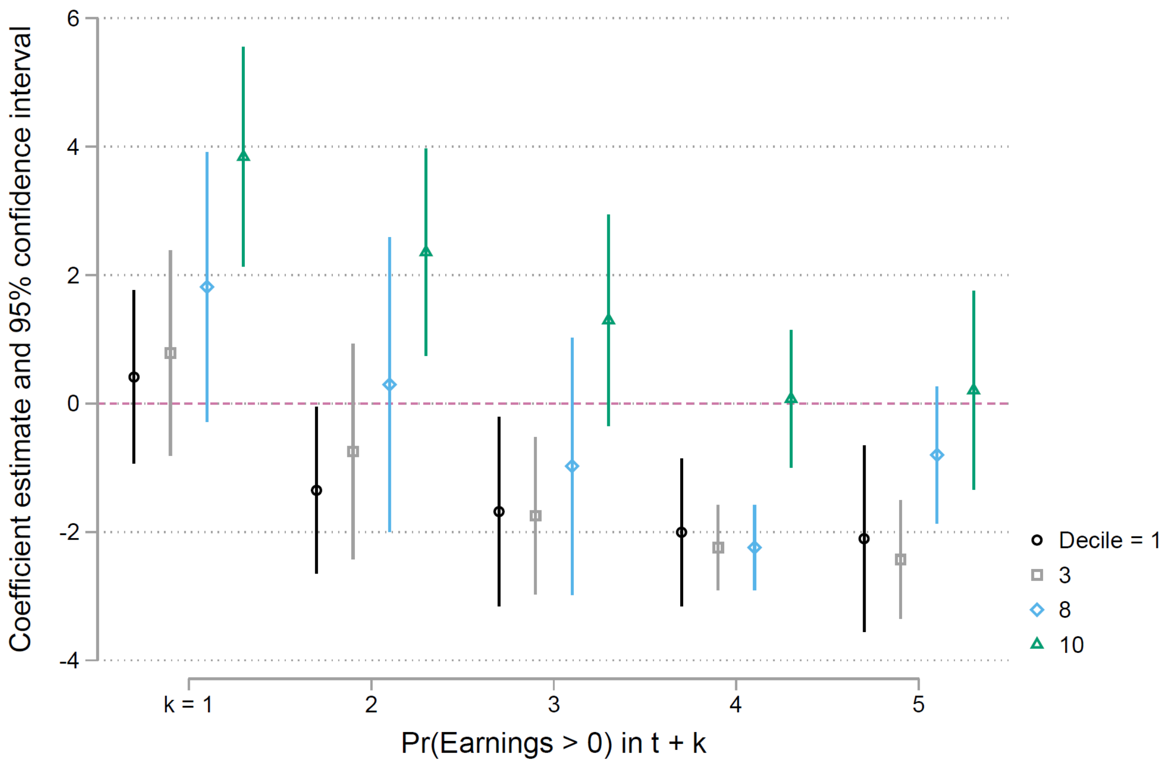


Notes: This figure presents coefficient estimates of the effect of labor market condition shocks on log(earnings) k years out from year t by predicted college matriculation decile. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is the natural logarithm of the total annual earnings for a student i in year $t(i) + k$. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained on a separate sample based on students' math score decile and k.

oil and gas deposits in shale rock. The benefit of exploring this labor market shock is due to its targeted impact. Fracking only directly affects a single sector, mining and extraction, so while it has wide-reaching spillovers to the overall economy and other sectors, its initial labor market pull is concentrated. The boom is also concentrated in particular localities where shale is previously present, providing easy comparison between "treated" and "untreated" areas. The adoption of fracking technology occurred relatively quickly, clearly delineating a "sudden" treatment time.

Comparing the fracking boom shock to the broad labor market shocks studied earlier, the effects of the fracking boom shock will be better identified at the expense of external validity. The sudden

Figure 6. Effect of labor market condition shock on employment by math score decile



Notes: This figure presents coefficient estimates of the effect of labor market condition shocks on employment k years out from year t by predicted college matriculation decile. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student student i is employed in year $t(i) + k$, measured by having positive earnings in that year. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained on a separate sample based on students' math score decile and k.

timing and spatial distribution of the fracking boom shock allows me to cleanly identify how labor market opportunities affect schooling choices and earnings outcomes. However, the specificity of the shock provides challenges in generalizing the findings to other industries or settings. For example, past research shows that men are most likely to take "fracking"-related jobs. Then the schooling responses to fracking may not be generalizable to schooling and labor force participation patterns for women.

I proceed by providing a brief overview of fracking. I then describe the additional data and empirical strategy needed to study the fracking boom. Then I present the results and reconcile them with the earlier shift-share-based analysis.

6.1 Background

Hydraulic fracturing, or "fracking," is a oil and gas extraction technique that involves injecting a high-pressure fluid mixture—primarily water, sand, and chemical additives—into rock formations to create small fractures in the rock. These fractures allow the petroleum or natural gas trapped in the formations to flow more freely and be brought to the surface. Fracking is often used in tandem with *horizontal drilling*, in which a well is drilled vertically to a certain target depth and then extended horizontally through the rock formation. This allows producers to access a much larger surface area of the rock formation for fracking than with traditional vertical wells.

Although fracking has been used in conventional vertical wells since the late 1940s, the combination of horizontal drilling and fracking only became economically viable in the late 1990s and early 2000s. Advances in drilling technology, rising natural gas prices, and successful experimentation in North Texas during the late 1990s made large-scale shale gas extraction feasible. These innovations triggered a broader shale boom in Texas and the United States in the mid-2000s, spreading the practice to other rock formations such as the Eagle Ford Shale in South Texas and the Permian Basin in West Texas.

The resulting fracking boom led to large improvements in economic outcomes, including employment. Maniloff and Mastromonaco (2017) and Bartik et al. (2019) demonstrate improvements in several economic measures across the United States in areas with more fracking. Fracking in particular sharply increased the number of jobs in a locality. This rise in labor demand pushed students to take fracking jobs (Cascio and Narayan, 2022).

6.2 Data and methodology

To assess which types of students decide not to attend college in pursuit of a fracking job, I require data on students, where the fracking boom occurred, and exactly when the boom started. For information on students, I use the ERC data described in Section 3.1. I use shale maps from the U.S. Energy Information Administration (EIA). I identify the timing of the fracking boom using horizontal drilling permit data from the Railroad Commission of Texas (RRC).

The EIA provides data on where it estimates there are shale-based oil and gas deposits across the United States. I use the EIA-provided shapefiles and Texas county boundaries to identify which counties contain any shale oil and gas deposits. Counties which contain shale oil and gas deposits are classified as "fracking counties." The EIA-based classification may not perfectly correlate with the counties which experienced booms. The EIA maps are estimates of deposits, so some "treated" counties may actually be "untreated", and vice-versa. Further, the existence of resources is necessary, but not sufficient, for oil and gas extraction. However, the EIA-based classification remains the cleanest to identify fracking counties, even if it does not align with actual drilling. Local regulatory and economic conditions that affect drilling are likely to also affect educational outcomes and decisions. Using a classification only dependent on estimated resources removes contamination from this factor. In this way, my estimate will represent an "intent-to-treat" estimate of fracking on college matriculation.

I use permitting data from the RRC to determine the timing of the fracking shock in each area. I first group fracking counties based on their shale formation: the Eagle Ford formation and the Permian Basin. Next, I determine the universe of horizontal drilling permits issued by the RRC in each county from 2000 to 2024. Then, I compare the permits issued to counties in each formation to all non-fracking counties. I fit a poisson regression by pseudo maximum likelihood

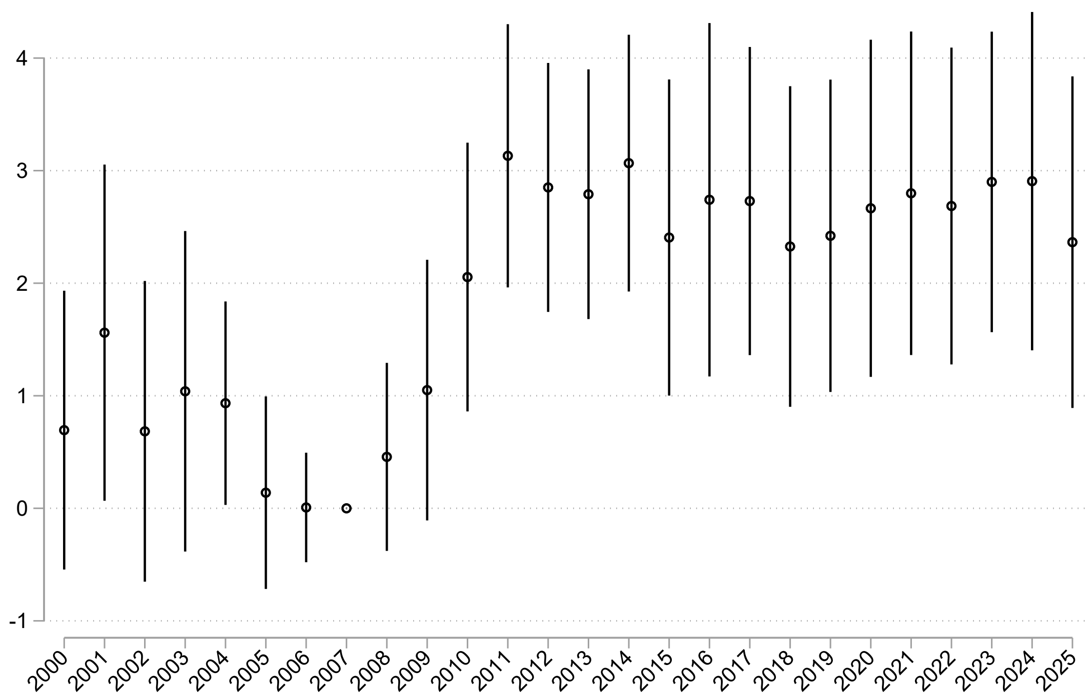
$$\mathbb{E}[N^{permits}]_{ct} = \exp\left(\alpha_c + \tau_t + \sum_{s \neq 2007} \beta^s \mathbb{1}[t = s] Fracking_c + \varepsilon_{ct}\right) \tag{14}$$

where c indexes counties and t indexes years. $\mathbb{E}[N^{permits}]_{ct}$ are the expected number of permits in county c and year t. $Fracking_c$ is an indicator of whether or not the county is a fracking county. α_c and τ_t capture common county and time trends, respectively.

I plot the results of estimating equation (14) comparing Eagle Ford counties to non-fracking counties in Figure 7. I find a gradual increase in horizontal drilling permits issued starting 2007. Relative to 2007, counties in the Eagle Ford formation saw approximately 3 additional permits per year by 2011. I find a similar jump in permitting around 2008 for Permian Basin counties.

With the affected counties and timing in hand, I estimate the impact of fracking on college matriculation through a difference-in-differences empirical strategy. I compare outcomes for students in fracking counties with students in non-fracking counties before and after fracking began in earnest. This empirical strategy relies on two core assumptions. First, I assume that fracking

Figure 7. Expected number of horizontal drilling permits in fracking counties versus non-fracking counties



Notes: This figure presents the coefficient estimates from fitting an event-study-style poisson regression of number of horizontal drilling permits on being a fracking county in a particular year. Unreported controls include county fixed effects and time fixed effects. 95% confidence intervals are obtained from standard errors clustered at the county-level. The regression is estimated via pseudo maximum likelihood. The unit of observation is a county-year. The outcome is the expected number of permits issued in a county-year. The plotted coefficients are for the interaction between year dummies and a binary indicator for fracking county status. Results for year 2007 are dropped for estimation.

and non-fracking counties would have remained the same if not for the fracking boom. Second, I assume that there are no anticipatory responses in fracking counties to the expected arrival of the boom.

I formally estimate

$$W_i = \alpha_{c(i)} + \tau_{t(i)} + college\_pred_i + \sum_{s \neq 2007} \beta^s \mathbb{1}[c(i) \in \text{Eagle Ford}]\mathbb{1}[t(i) = s] + \varepsilon_i \tag{15}$$

where $c(i)$ is the county c of student i's high school and $t(i)$ is the graduation year of student i. W_i is college matriculation, $\alpha_{c(i)}$ are county fixed effects, and $\tau_{t(i)}$ are time fixed effects. The coefficient of interest is β^s, which estimates the effect of a student being in an Eagle Ford county in year $s = t(i)$ on college matriculation. Estimates for year $s = 2007$ are dropped (set to zero) and standard errors are clustered by county. I later replace W_i with $Y_i^{t(i)+v}$ as I do in the previous analysis.

6.3 College matriculation results

Figure 8 plots the $\hat{\beta}^s$ obtained from estimating equation (15) for Eagle Ford counties. I find that Eagle Ford counties see a steady relative drop in college matriculation rates, stabilizing to a 2% decline relative to baseline from 2014 and onward. Coefficients prior to 2007 are imprecisely estimated but are stable around zero, which supports our parallel trends assumption.

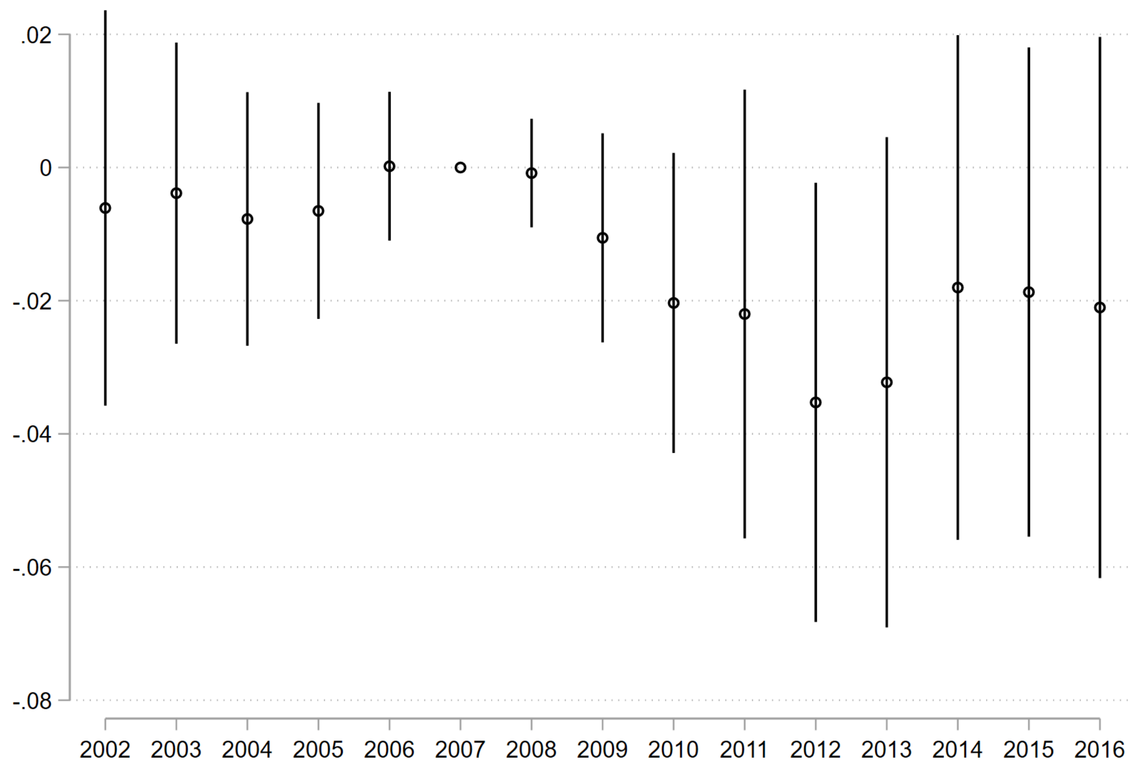
I reconcile this estimate with estimates obtained using the shift-share analysis. Earlier, I showed that a 0.01 unit increase in labor market conditions leads to a -0.023 percentage point decrease in college matriculation. Since a 0.01 unit increase in labor market condition is roughly equivalent to a 1% increase in wage growth, this implies that the fracking boom estimate is equivalent to the response one would expect from a 1% increase in wages available to high school students. I benchmark this against existing research on earnings increases due to fracking. Maniloff and Mastromonaco (2017) show that fracking "boom" counties saw 10%–29% higher increase in earnings from 2005–2011. This suggests that the response to the fracking boom is orders of magnitude lower than the shift-share estimates. One way to reconcile these results is that Maniloff and Mastromonaco (2017) find significant heterogeneity in the sectoral impact of fracking. In particular, they find that service sectors are less responsive to fracking. This may partially explain the gap in responses: the fracking boom leaves service sectors—the sectors with most exposure to high school students—relatively unchanged in wages.

6.4 Earnings results

7 Conclusion

This paper explores the impact of wage shocks and labor market conditions on college matriculation and medium-term earnings for students in Texas. I construct a shift-share variable measure of labor market condition, where the "shifts" come from non-Texas wage growth by industry and year and the "shares" come from predicted exposure of each student to every industry. This shift-

Figure 8. Event study of college matriculation probability on Eagle Ford vs. non-fracking counties across time



Notes: This figure presents the coefficient estimates from fitting an "event-study" regression of college matriculation on the interaction between Eagle Ford county (fracking) and year. Unreported controls include county fixed effects, time fixed effects, and predicted college matriculation. 95% confidence intervals are obtained from standard errors clustered at the county-level. The unit of observation is a high school senior. The outcome is whether or not the student matriculated to college in the year immediately following high school graduation. Each plotted coefficient represents the difference in college matriculation for Eagle Ford county students vs. non-fracking county students for that year, relative to the baseline year of 2007.

share measure of labor market conditions has a statistically significant, negative impact on college matriculation: students facing higher wages choose not to go to college. Students induced into the labor force cover a large portion of the achievement distribution; students from deciles 1 to 7 of high school math score are responsive to better labor market opportunities. There is little evidence to suggest that students pulled away from college are better off than their peers. While there is a temporary boost in earnings, by year 5, there are no difference in wages between those who respond to labor market conditions and those who do not.

I study these college-going decisions for a particular shock created by the fracking boom in Texas.

I compare students graduating in fracking counties vs. non-fracking counties when the fracking boom occurred. I corroborate other research which shows that college matriculation decreased in response to the fracking boom. I intend to show through additional analyses that this impact is similar to the shift-share results.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics 4B*, 1043–1171.

Acton, R. (2021). Effects of reduced community college tuition on college choices and degree completion. *Education Finance and Policy 16*(3), 388–417.

Adukia, A., S. Asher, and P. Novosad (2020). Educational investment responses to economic opportunity: Evidence from indian road construction. *American Economic Journal: Applied Economics 12*(1), 348–376.

Angrist, J. D. and S. H. Chen (2011). Schooling and the vietnam-era G.I. bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics 3*(2), 96–118.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review 106*(8), 2046–2085.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science 344*(6186), 843–851.

Bartik, A. W., J. Currie, M. Greenstone, and C. R. Knittel (2019). The local economic and welfare consequences of hydraulic fracturing. *American Economic Journal: Applied Economics 11*(4), 105–155.

Black, D. A., T. G. McKinnish, and S. G. Sanders (2005). The economic impact of the coal boom and bust. *The Economic Journal 115*(503), 449–476.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies 89*(1), 181–213.

Bound, J. and S. Turner (2002). Going to war and going to college: Did world war II and the G.I. bill increase educational attainment for returning veterans? *Journal of Labor Economics 20*(4), 784–815.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pp. 201–222. Toronto: University of Toronto Press.

Carruthers, C. K., W. F. Fox, and C. Jepsen (2023). Promise kept? free college programs and postsecondary outcomes. *American Economic Journal: Economic Policy 15*(2), 1–35. If you intended a different 2023 Carruthers–Fox–Jepsen article, adjust title/venue accordingly.

Cascio, E. U. and A. Narayan (2022). Who needs a fracking education? the educational response to low-skill-biased technological change. *ILR Review 75*(1), 56–89.

Charles, K. K., E. Hurst, and M. J. Notowidigdo (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review 108*(10), 2947–2994.

Denning, J. T. (2017). College on the cheap: Consequences of community college tuition reductions. *American Economic Journal: Economic Policy 9*(2), 155–188.

Goldin, C. and L. F. Katz (2008). *The Race between Education and Technology.* Cambridge, MA: Harvard University Press.

Goodman, J., M. Hurwitz, and J. Smith (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics 35*(3), 829–867.

Gurantz, O. (2020). Impacts of merit aid programs on postsecondary enrollment and persistence. *Education Finance and Policy 15*(3), 444–472. If you intended a different Gurantz (2020), adjust accordingly.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Lee, D. S. (2020). Minimum wages and schooling: Evidence on postsecondary enrollment. *Working Paper*. Cited in later literature as finding community-college enrollment declines after minimum wage hikes; insert full details once you select the exact version you mean.

Lemieux, T. and D. Card (2001). Education, earnings, and the "canadian g.i. bill". *Canadian Journal of Economics 34*(2), 313–344.

Maniloff, P. and R. Mastromonaco (2017). The local economic impacts of unconventional shale oil and gas development. *Resource and Energy Economics 49*, 62–85.

Marchand, J. and J. G. Weber (2020). How local economic conditions affect school finances, teacher quality, and student achievement: Evidence from the texas shale boom. *Journal of Policy Analysis and Management 39*(1), 36–63.

Morissette, R., P. C. W. Chan, and Y. Lu (2015). Wages, youth employment, and school enrollment: Recent evidence from increases in world oil prices. *Journal of Human Resources 50*(1), 222–253.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review 112*(8), 2580–2630.

National Center for Education Statistics (2025). Percentage of 18- to 24-year-olds enrolled in college, by level of institution and sex and race/ethnicity of student: 1970 through 2022. https://nces.ed.gov/programs/digest/d23/tables/dt23_302.60.asp. Accessed October 2025.

Oreopoulos, P. and U. Petronijevic (2013). Making college worth it: A review of the returns to higher education. *The Future of Children 23*(1), 41–65.

Schanzenbach, D. W., L. Bauer, S.-M. M. Stewart, et al. (2024). Raising state minimum wages, lowering community college enrollment? *NBER Working Paper* (31540). Use the finalized author list/title of the version you cite; the WP documents enrollment declines at community colleges after minimum wage increases.

Schiller, A. R. and A. Slechten (2025). Effect of natural resource extraction on school performance: Evidence from texas. *Energy Economics 144*, 107888. If you have a different title/version, update accordingly.

Shah, M. and B. Steinberg (2017). Drought of opportunities: Contemporaneous and long-run impacts of rainfall shocks on human capital. *Journal of Political Economy 125*(2), 527–561.

Shah, M. and B. Steinberg (2021). Work and schooling: An exploration of constrained optimality in developing countries. *Annual Review of Economics 13*, 535–561. If you intended a different 2021 Shah–Steinberg item, adjust accordingly.

Smith, J., J. Goodman, and M. Hurwitz (2025). (forthcoming/working title) college admissions test thresholds and degree completion. *Working Paper*. You cited "Smith, Goodman, and Hurwitz 2025"; insert the exact title/venue once available.

Stanley, M. (2003). College education and the midcentury G.I. bills. *The Quarterly Journal of Economics 118*(2), 671–708.

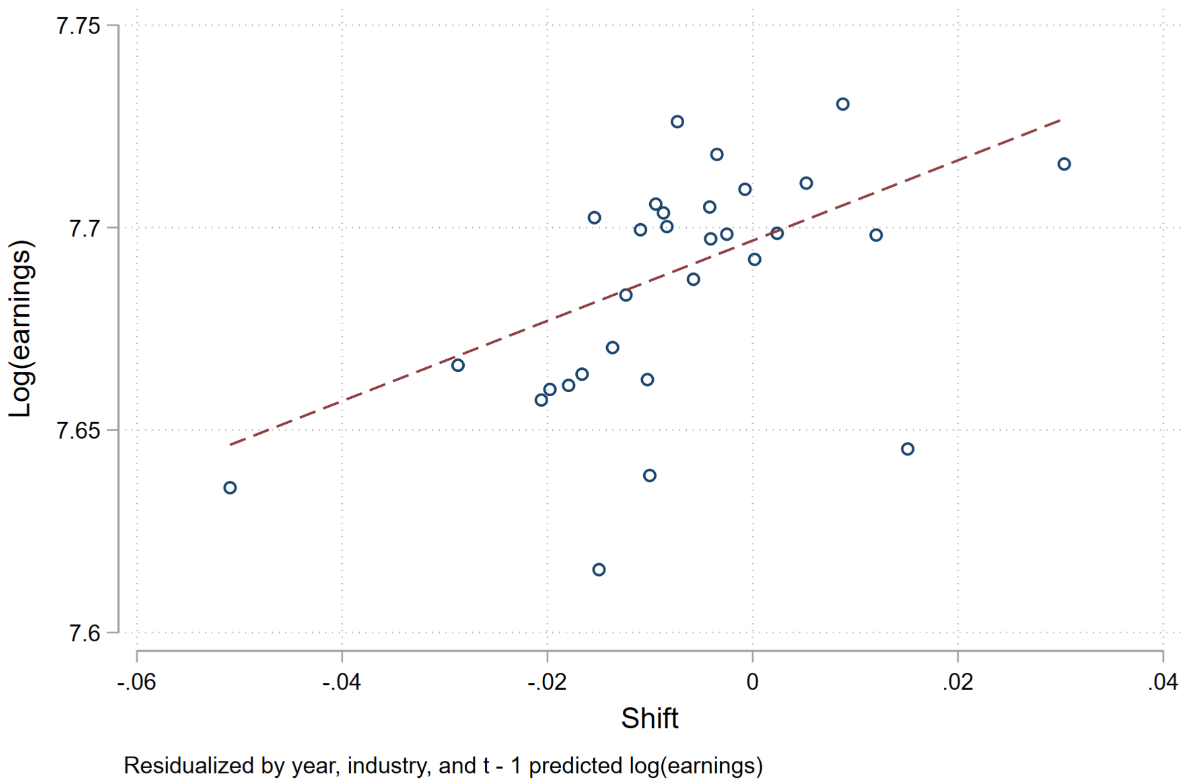
von Wachter, T. (2020). The persistent effects of initial labor market conditions for young adults and their sources. *Journal of Economic Perspectives 34*(4), 168–194.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics 32*(4), 711–754.

Zuo, N., J. Schieffer, and S. Buck (2019). The effect of the oil and gas boom on schooling decisions in the u.s. *Resource and Energy Economics 55*, 1–23.

Appendix tables and figures

Figure A1. Binscatter of log post-graduation earnings on wage growth



Notes: This figure represents a binned scatter plot of the logarithm of post-graduation earnings on labor market "shifts". The dotted red line represents a linear fit of the underlying data. The dependent and independent variables are residuzlied by year, industry, and predicted log earnings in $t-1$. Industry is defined as the three-digit NAICS code of a business. "Predicted log (earnings) in $t-1$" is the natural logarithm of predicted earnings for the industry in the year prior to high school graduation. The underlying unit of observation is a student. The outcome is calculated as the natural logarithm of the earnings for an individual (a) in the year after high school graduation and (b) in their primary industry. "Shifts" are calculated at the industry-time level and corresponds to the non-Texas wage growth described in Sections 3.2 and 4.1.2. The estimation sample includes high school seniors from 2004–2014 who report taking a job in the year following high school graduation.

Table A1. Predictions of year-over-year employment growth using labor merket shifts

	(1)	(2)
Shift $g_{k,t}$	1.083	1.136
	(0.195)	(0.206)
Predicted Log(earnings); $k, t-1$		0.093
		(0.118)
Observations	644	644

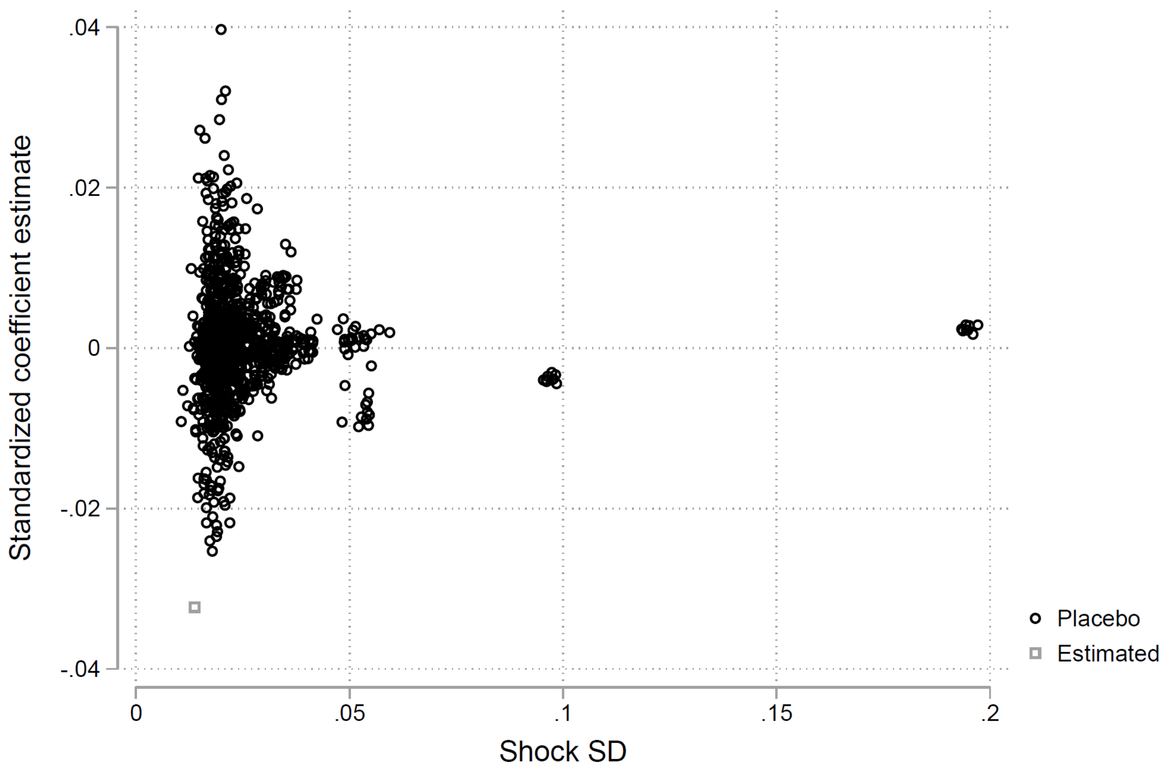
Notes: Each specification is a regression of year-over-year employment growth on labor market "shifts." Unreported controls include industry fixed effects and year fixed effects. Standard errors reported in parentheses below coefficient estimates are clustered at the industry-year level. Industry is defined as the three-digit NAICS code of a business. The unit of observation is an industry-year, and each observation is weighted by the number of employees in an industry in the prior year. The outcome is calculated as the percent change in post-high school graduation employment from year $t-1$ to year t for a given industry. "Shifts" are calculated at the industry-time level and correspond to the non-Texas wage growth described in Sections 3.2 and 4.1.2. "Predicted log(earnings) in $t-1$" are the predicted earnings for the student's chosen industry in the year prior to their labor market entry. Predictions are made based on the QWI data.

Table A2. Variation in shares explained by student characteristics for 5 most common industries

	(1) 722	(2) 561	(3) 455	(4) 445	(5) 458
Commuting zone	0.32	0.23	0.22	0.60	0.20
Gender	0.54	0.17	0.09	0.01	0.56
Race/ethnicity	0.06	0.55	0.42	0.17	0.11
Economic	0.01	0.24	0.07	0.11	0.04
Spanish-speaking	0.02	0.07	0.02	0.02	0.00
Age	0.02	0.02	0.01	0.00	0.06
Math score	0.01	0.11	0.31	0.10	0.04
Reading score	0.03	0.12	0.21	0.16	0.13
SD in $s_{i,k}$	0.05	0.04	0.02	0.01	0.02

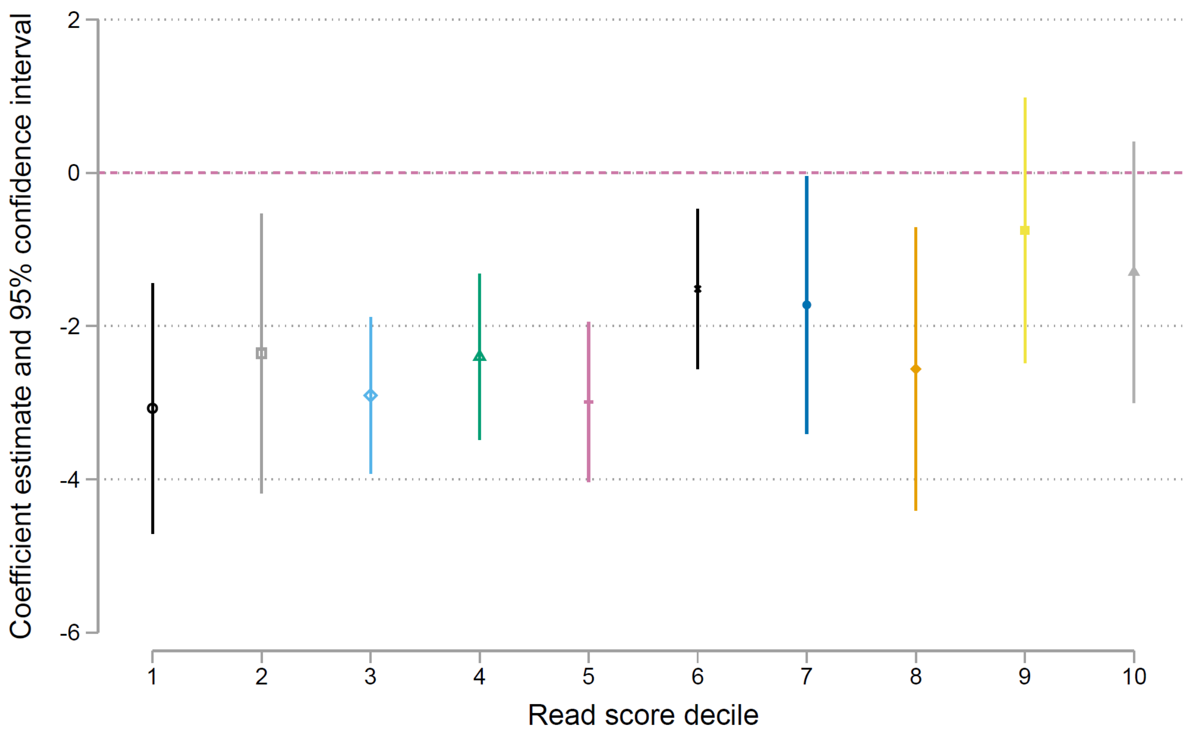
Notes: Each column represents a separate regression of shares for a given industry on fixed effects for the listed student characteristics. Each number represents the R^2 of each set of fixed effects on shares. Industries are defined at the level of the 3-digit NAICS code. Selected industries are the industries with the highest average industry weights/shares across the sample. The last row displays the standard deviation in the industry shares for each displayed industry.

Figure A2. Coefficient estimates of labor market condition shocks on college matriculation after shift permutation



Notes: Each dot is a point estimate from a separate regression of college matriculation on labor market condition shocks. Each specification includes student characteristics fixed effects and year fixed effects. The vertical axis shows the estimated point estimate, and the horizontal axis shows the standard deviation of the shock variable used. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." The shock variable is standardized. For the "placebo" coefficient estimates, each dot arises from a regression where the independent variable is replaced with labor market condition shocks constructed after shifts are randomly permuted across industry. Each permutation newly assigns an industry's time vector of shifts to another industry. The "estimated" coefficient is from estimating equation (11).

Figure A3. Effect of labor market condition shock on college matriculation by reading score decile



Notes: This figure represents coefficient estimates of the effect of labor market condition shocks on college matriculation probability by reading score decile. Each specification includes student characteristics fixed effects and year fixed effects. Reported 95% confidence intervals are derived from standard errors which are (a) clustered by industry and (b) estimated at the "shock"-level following a version of equation (13). The unit of observation is a student. The outcome is whether or not a student enrolls in college the fall immediately following high school graduation. The shock variable combines industry-year level wage growth "shifts" with student-industry level industry choice propensity "shares." Each coefficient estimate is obtained on a separate sample based on students' standardized reading score decile.