Online Appendix A: Additional Figures and Tables

Natural Resource Booms, Human Capital, and Earnings: Evidence from Linked Education and Employment Records

Alina Kovalenko

Table A.1: Summary Statistics

Variable	Levels, 1995			Changes, 1995-2001		
	Below median	Above median	p-value	Below median	Above median	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Educational outcomes						
High school graduation rate	0.759	0.744	0.23	0.023	0.041	0.09
Absence rate	0.054	0.056	0.34	-0.003	-0.004	0.43
Grade repetition rate	0.190	0.203	0.29	-0.017	-0.034	0.04
Enrollment in public 4-year college	0.224	0.201	0.19	-0.013	0.009	0.02
Enrollment in community college	0.387	0.371	0.49	0.057	0.050	0.60
Graduation from public 4-year college	0.147	0.136	0.23	-0.004	0.007	0.16
Graduation from community college	0.025	0.021	0.28	0.001	0.001	0.92
Panel B. Labor market outcomes						
Employed, age 14-18	0.792	0.775	0.44	-0.058	-0.037	0.12
Quarterly earnings (excl. 0s), age 14-18	1,504.562	1,547.294	0.21	-141.173	-180.662	0.21
Quarterly earnings (incl. 0s), age 14-18	480.232	480.828	0.98	-100.938	-111.198	0.66
Quarterly earnings (excl. 0s), age 24-25	8,101.262	8,174.381	0.71	51.684	-5.853	0.72
Quarterly earnings (incl. 0s), age 24-25	5,418.106	5,689.343	0.12	4.608	-46.448	0.74
Panel C. Student demographics						
Male	0.491	0.487	0.40	-0.000	0.004	0.56
White	0.586	0.545	0.47	-0.050	-0.030	0.02
Black	0.080	0.076	0.87	0.003	-0.002	0.17
Hispanic	0.325	0.371	0.49	0.047	0.029	0.04
Gifted	0.123	0.107	0.10	-0.002	0.008	0.24
Special education	0.063	0.065	0.84	-0.015	-0.012	0.69
Economically disadvantaged	0.365	0.396	0.38	0.054	0.035	0.11
Panel D. Commuting zone characteristics						
Population density	68.449	46.177	0.39	8.438	4.875	0.44
Share hispanic	0.285	0.323	0.53	0.037	0.029	0.03
Share black	0.070	0.067	0.84	0.000	0.003	0.25
Share male	0.495	0.497	0.62	0.004	0.007	0.25
Unemployment rate	6.275	6.795	0.67	-1.237	-1.454	0.76
Median household income	26,603.072	25,684.573	0.46	5,494.597	5,436.667	0.89
Number of commuting zones	31	31	–	31	31	–

Notes: The table reports means for student and commuting zone characteristics in 1995 and their changes from 1995 to 2001. The sample includes 9th grade cohorts enrolled in public schools in the state of Texas. Columns (1)-(2) and (4)-(5) split the sample into commuting zones with below and above median oil and gas reserves. Columns (3) and (6) report the p-values associated with the null hypothesis of equivalent means across the groups. The data are drawn from the Texas Education Research Center and the U.S. Census Bureau.

Table A.2: The Effect of the Fracking Boom on Student Employment at Age 14-18, by Industry

	Full sample	Men	Women	Quartile of grade 6 test score distribution			
				Q1 (Bottom)	Q2	Q3	Q4 (Top)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Retail trade employment							
Fully treated	2.756	2.758	2.733	2.881	2.888	2.779	2.059
	(1.232)	(1.276)	(1.191)	(1.030)	(1.146)	(1.362)	(1.494)
Partially treated	0.929	0.738	1.107	1.022	0.801	1.217	0.600
	(0.734)	(0.700)	(0.778)	(0.751)	(0.670)	(0.836)	(0.714)
Baseline mean	32.1	32.4	31.9	28.9	33.6	34.5	31.5
Panel B. Accommodation and food services employment							
Fully treated	2.346	2.426	2.263	2.784	2.477	1.678	1.749
	(1.104)	(1.176)	(1.035)	(0.973)	(0.978)	(1.262)	(1.203)
Partially treated	0.704	0.796	0.615	0.363	0.945	0.546	0.925
	(0.667)	(0.652)	(0.701)	(0.813)	(0.599)	(0.721)	(0.646)
Baseline mean	37.4	35.7	39.1	38.7	39.9	38.1	33.2
Panel C. Oil and gas employment							
Fully treated	0.081	0.118	0.030	0.105	0.073	0.112	-0.068
	(0.070)	(0.140)	(0.019)	(0.124)	(0.060)	(0.057)	(0.106)
Partially treated	0.270	0.506	0.020	0.171	0.499	0.236	0.135
	(0.073)	(0.151)	(0.029)	(0.204)	(0.108)	(0.077)	(0.088)
Baseline mean	0.37	0.59	0.16	0.46	0.40	0.32	0.32
Observations	5,357,850	2,664,414	2,693,436	1,339,456	1,339,464	1,339,461	1,339,469

Notes: This table reports difference-in-differences estimates of the effect of the fracking boom on high school students employment at age 14-18. The unit of analysis is at the student-cohort-commuting zone level. In columns (4)-(7), the ability quartiles are assigned based on 6th grade test scores on the state standardized exam. “Partially treated" and “Fully treated" rows report coefficients on the interaction terms between predicted shale reserves and an indicator variable for entering high school in 2001-2004, and an indicator variable for entering high school in 2005 or later, respectively. Commuting zone fixed effects, cohort fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Coefficients are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.3: Sensitivity of Main Results to Alternative Specifications

	Absence	Repetition	HS grad.	Employed	Log earnings	Community college		Public university	
						Enrol.	Grad.	Enrol.	Grad.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Baseline model									
Fully treated	0.215	1.041	-1.102	3.243	3.706	-0.138	-0.099	-0.044	-0.363
	(0.089)	(0.441)	(0.394)	(1.660)	(1.553)	(0.482)	(0.052)	(0.279)	(0.147)
Partially treated	0.038	0.413	-0.465	1.169	0.435	-0.351	-0.109	0.080	-0.132
	(0.060)	(0.309)	(0.457)	(1.066)	(0.898)	(0.360)	(0.049)	(0.262)	(0.176)
Observations	5,357,850	5,357,850	5,357,850	5,357,850	3,482,389	4,724,806	4,724,806	4,724,806	3,501,707
Panel B. Dropping controls									
Fully treated	0.258	1.242	-1.219	3.902	4.879	0.408	-0.015	0.461	-0.135
	(0.100)	(0.537)	(0.477)	(1.877)	(1.892)	(0.564)	(0.045)	(0.341)	(0.169)
Partially treated	0.089	0.391	-0.612	2.044	1.311	-0.128	-0.067	0.084	-0.210
	(0.064)	(0.233)	(0.438)	(1.351)	(1.093)	(0.354)	(0.036)	(0.226)	(0.170)
Observations	5,357,850	5,357,850	5,357,850	5,357,850	3,482,389	4,724,806	4,724,806	4,724,806	3,501,707
Panel C. Areas with non-zero reserves per capita									
Fully treated	0.161	0.689	-0.678	3.329	3.927	0.540	0.044	0.090	-0.457
	(0.083)	(0.412)	(0.317)	(1.915)	(1.820)	(0.662)	(0.056)	(0.296)	(0.176)
Partially treated	-0.002	0.450	-0.487	1.146	0.374	-0.001	-0.037	-0.025	-0.097
	(0.054)	(0.364)	(0.498)	(1.164)	(0.912)	(0.351)	(0.037)	(0.267)	(0.167)
Observations	3,043,665	3,043,665	3,043,665	3,043,665	2,070,347	2,682,700	2,682,700	2,682,700	1,986,446
Panel D. Differential timing by shale play									
Fully treated	0.205	0.952	-0.916	3.463	3.981	0.077	-0.093	0.055	-0.294
	(0.084)	(0.389)	(0.318)	(1.718)	(1.665)	(0.492)	(0.055)	(0.228)	(0.140)
Partially treated	0.011	0.495	-0.517	1.571	1.199	-0.217	-0.070	-0.119	-0.094
	(0.046)	(0.301)	(0.403)	(1.115)	(1.134)	(0.302)	(0.034)	(0.300)	(0.165)
Observations	5,357,850	5,357,850	5,357,850	5,357,850	3,482,389	4,724,806	4,724,806	4,724,806	3,501,707

Notes: This table checks the sensitivity of the main difference-in-differences estimates to alternative specifications. The unit of analysis is at the student-cohort-commuting zone level. Panel A reports estimates from the main specification. Panel B reports estimates without student and commuting zone covariates. Panel C focuses on a subsample of commuting zones with non-zero shale reserves. Panel D explores alternative assumptions about the timing of the boom following Bartik et al (2019). "Partially treated" and "Fully treated" rows report coefficients on the interaction terms between predicted shale reserves and an indicator variable for entering high school in 2001-2004, and an indicator variable for entering high school in 2005 or later, respectively. Commuting zone fixed effects, cohort fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Coefficients are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.4: Sensitivity of Main Results to Commuting Zone Assignment

	Absence	Repetition	HS grad.	Employed	Log earnings	Community college		Public university	
						Enrol.	Grad.	Enrol.	Grad.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Baseline model									
Reserves × Post	0.215	1.041	-1.102	3.243	3.706	-0.138	-0.099	-0.044	-0.363
	(0.089)	(0.441)	(0.394)	(1.660)	(1.553)	(0.482)	(0.052)	(0.279)	(0.147)
Observations	5,357,850	5,357,850	5,357,850	5,357,850	3,482,389	4,724,806	4,724,806	4,724,806	3,501,707
Panel B. Subsample: Same commuting zone since grade 1									
Reserves × Post	0.091	0.475	-0.690	2.393	3.825	0.394	0.032	-0.391	-0.368
	(0.072)	(0.355)	(0.369)	(0.967)	(1.222)	(0.374)	(0.078)	(0.287)	(0.185)
Observations	2,328,273	2,328,273	2,328,273	2,328,273	1,467,708	2,132,227	2,132,227	2,132,227	1,402,950

Notes: This table checks the sensitivity of the main difference-in-differences estimates to assignment of commuting zone to students. The unit of analysis is at the student-cohort-commuting zone level. Panel A reports estimates from the baseline model. Panel B reports estimates from a model which includes high school students from cohorts 2001-2015 who were in the same commuting zone in grade 1 and grade 9. Commuting zone fixed effects and year fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Coefficients estimates are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.5: The Effect of the Fracking Boom on Returns to a High School Degree, Age 18-22

	Full Sample	Men	Women
	(1)	(2)	(3)
Reserves×Post	-0.078	0.391	-0.372
	(0.941)	(0.901)	(0.721)
Observations	1,237	1,237	1,237

Notes: This table reports difference-in-differences estimates of the effect of the fracking boom on the expected earnings gap/premium between high school graduates and dropouts. The dependent variable is the difference in log quarterly earnings of 18-22 y.o. individuals with and without a high school degree. The unit of analysis is at the commuting zone-year level. Commuting zone fixed effects and year fixed effects are included in all specifications. Coefficients estimates are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.6: The Effect of the Fracking Boom on School Resources

	Log Revenue per student	Log Expenditure per student	Teachers		
			% Advanced degrees	% Experience < 5 years	Log Earnings
	(1)	(2)	(3)	(4)	(5)
Reserves×Post	0.027	0.032	-0.006	0.006	-0.004
	(0.012)	(0.012)	(0.004)	(0.004)	(0.004)
Baseline mean	9.69	9.74	0.27	0.30	10.77
Observations	1,426	1,426	6,955,926	6,955,926	6,955,926

Notes: This table reports difference-in-differences estimates of the effect of the fracking on school resources. Columns (1) and (2) use school district financial data and columns (3)-(5) use individual-level data on teacher characteristics from the TEA for 1995-2018. Commuting zone fixed effects and cohort fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.7: The Effect of the Fracking Boom on Educational and Labor Market Outcomes, by Shale Play

	Absence	Repetition	HS grad.	Employed	Log earnings	Community college		Public university	
						Enrol.	Grad.	Enrol.	Grad.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Oil shales									
Fully treated	0.235	1.192	-1.302	3.429	3.920	-0.145	-0.094	-0.051	-0.287
	(0.101)	(0.517)	(0.492)	(1.968)	(1.852)	(0.584)	(0.056)	(0.331)	(0.163)
Partially treated	0.044	0.581	-0.749	1.221	0.581	-0.527	-0.107	0.066	-0.111
	(0.064)	(0.356)	(0.562)	(1.244)	(1.079)	(0.429)	(0.053)	(0.273)	(0.190)
Observations	3,441,271	3,441,271	3,441,271	3,441,271	2,189,546	3,035,482	3,035,482	3,035,482	2,251,352
Panel B. Gas shales									
Fully treated	0.050	0.208	1.994	2.708	2.713	4.086	-0.241	-0.615	-0.139
	(0.238)	(1.126)	(1.583)	(2.481)	(3.244)	(2.501)	(0.343)	(1.530)	(0.781)
Partially treated	0.119	0.835	2.488	3.148	0.690	2.496	-0.150	2.070	1.499
	(0.211)	(1.098)	(1.855)	(1.928)	(2.808)	(0.821)	(0.337)	(1.142)	(0.863)
Observations	4,230,764	4,230,764	4,230,764	4,230,764	2,704,885	3,731,430	3,731,430	3,731,430	2,765,616

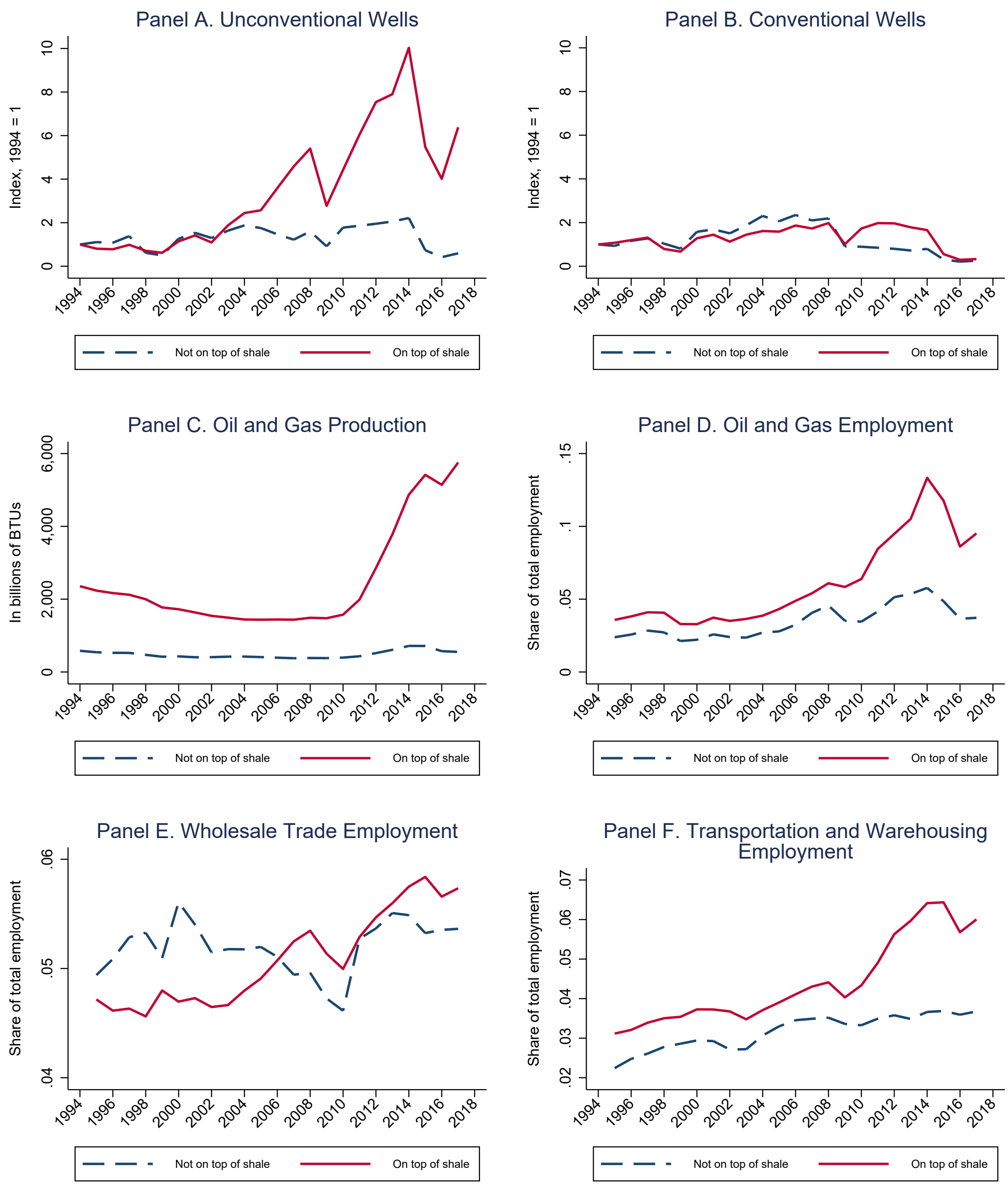
Notes: This table reports the estimated effects of the fracking boom by the type of shale play. The unit of analysis is at the student-cohort-commuting zone level. Panel A reports estimates for oil shale plays (Eagle Ford and Permian) and Panel B reports estimates for gas shale plays (Barnett and Haynesville). “Partially treated" and “Fully treated" rows report coefficients on the interaction terms between predicted shale reserves and an indicator variable for entering high school in 2001-2004, and an indicator variable for entering high school in 2005 or later, respectively. Commuting zone fixed effects, cohort fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Coefficient estimates are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Table A.8: The Effect of the Fracking Boom on Education and Labor Market Outcomes, by Race/Ethnicity

	Absence	Repetition	HS grad.	Employed	Log earnings	Community college		Public university	
						Enrol.	Grad.	Enrol.	Grad.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. White									
Fully treated	0.160	0.977	-0.731	3.095	4.245	0.269	-0.179	0.150	-0.295
	(0.077)	(0.315)	(0.279)	(1.695)	(1.835)	(0.924)	(0.071)	(0.340)	(0.198)
Observations	2,129,514	2,129,514	2,129,514	2,129,514	1,525,034	1,939,683	1,939,683	1,939,683	1,544,090
Panel B. Hispanic									
Fully treated	0.216	0.913	-1.182	2.966	2.947	-0.035	-0.067	-0.021	-0.225
	(0.100)	(0.492)	(0.482)	(1.552)	(1.444)	(0.455)	(0.074)	(0.349)	(0.149)
Observations	2,317,062	2,317,062	2,317,062	2,317,062	1,380,677	1,986,172	1,986,172	1,986,172	1,376,974
Panel C. Black									
Fully treated	0.350	1.084	-0.839	4.885	3.139	0.756	0.035	0.411	-0.607
	(0.199)	(0.805)	(0.969)	(3.763)	(2.460)	(1.386)	(0.135)	(0.584)	(0.294)
Observations	689,758	689,758	689,758	689,758	463,104	613,293	613,293	613,293	461,572
Panel D. Asian									
Fully treated	0.151	0.595	-0.680	4.540	13.670	0.962	0.529	3.017	3.682
	(0.227)	(1.551)	(1.198)	(4.176)	(7.151)	(1.912)	(0.517)	(1.577)	(2.447)
Observations	160,576	160,576	160,576	160,576	74,805	138,928	138,928	138,928	98,960

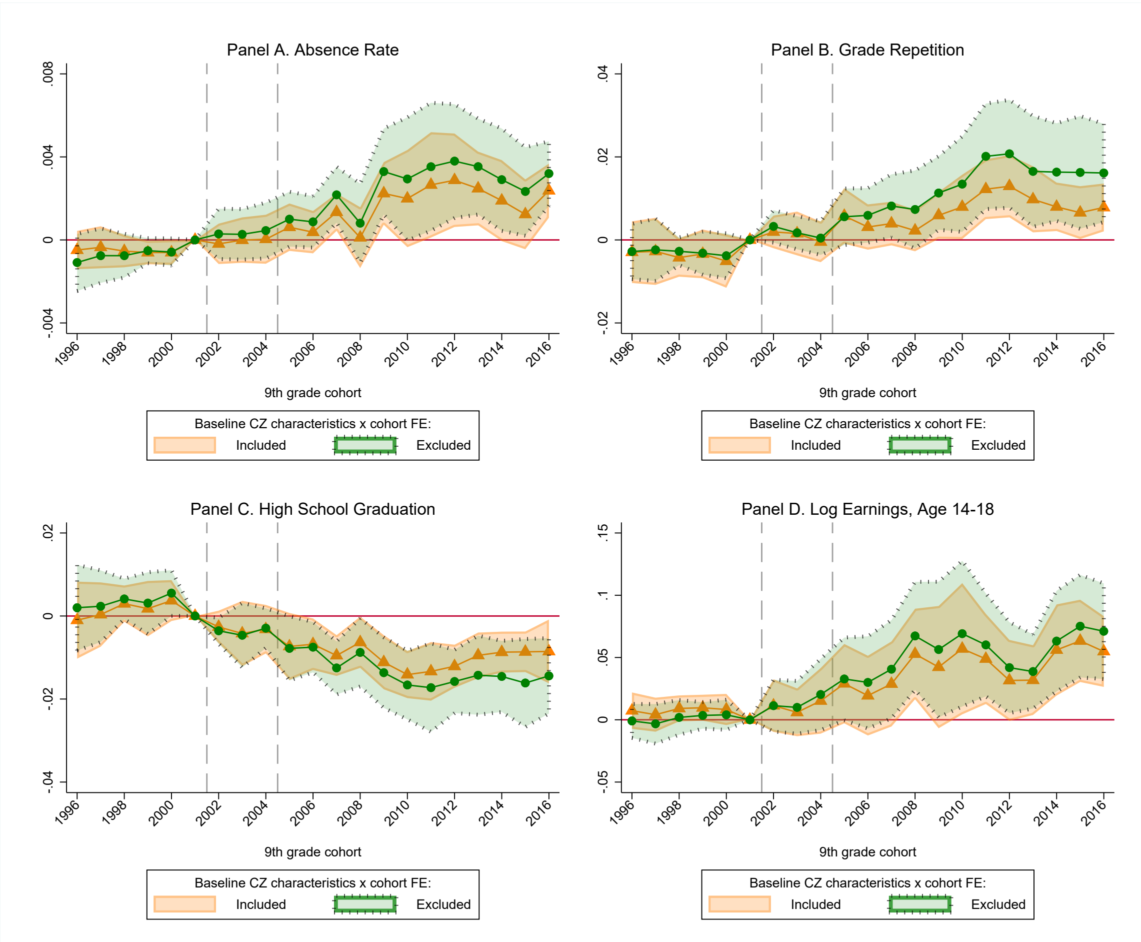
Notes: This table reports the effects of the fracking boom by race/ethnicity. The unit of analysis is at the student-cohort-commuting zone level. The reported estimates are for "fully treated" cohorts who entered high school in 2005 or later. Commuting zone fixed effects, cohort fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. Coefficient estimates are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

Figure A.1: Oil and Gas Drilling, Production and Employment Share in Texas



Notes: This figure displays the average number of new unconventional wells drilled in Texas in panel A, new conventional wells drilled in Texas in panel B, oil and gas production in Texas in panel C, share of employment in oil and gas industry (NAICS 211) in Texas in panel D, share of employment in wholesale trade (NAICS 42) in Texas in panel E, and share of employment in transportation and warehousing (NAICS 48-49) in Texas in panel F. These statistics are presented separately for commuting zones that lie on top of shale formations and those that do not. The data are from Enverus, Texas Railroad Commission, and the Quarterly Workforce Indicators.

Figure A.2: The Effect of the Fracking Boom on Educational Achievement and Earnings, Age 14-18

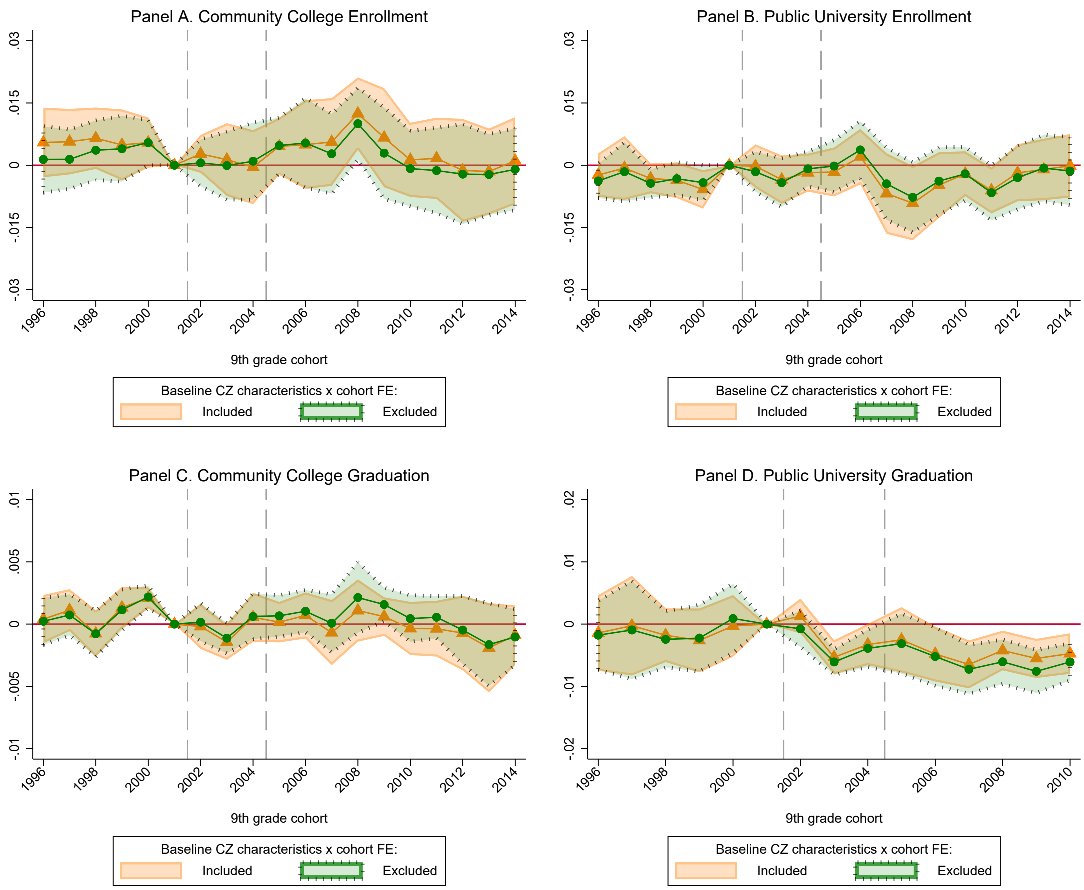


Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1) both with (shown in orange) and without (in green) 1995 commuting zone characteristics interacted with cohort fixed effects. The dependent variables are the absence rate, grade repetition rate, high school graduation, and log quarterly earnings. Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, and cohort fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.3: The Effect of the Fracking Boom on Employment Outcomes at Age 14-18 and Earnings at Age 24-25

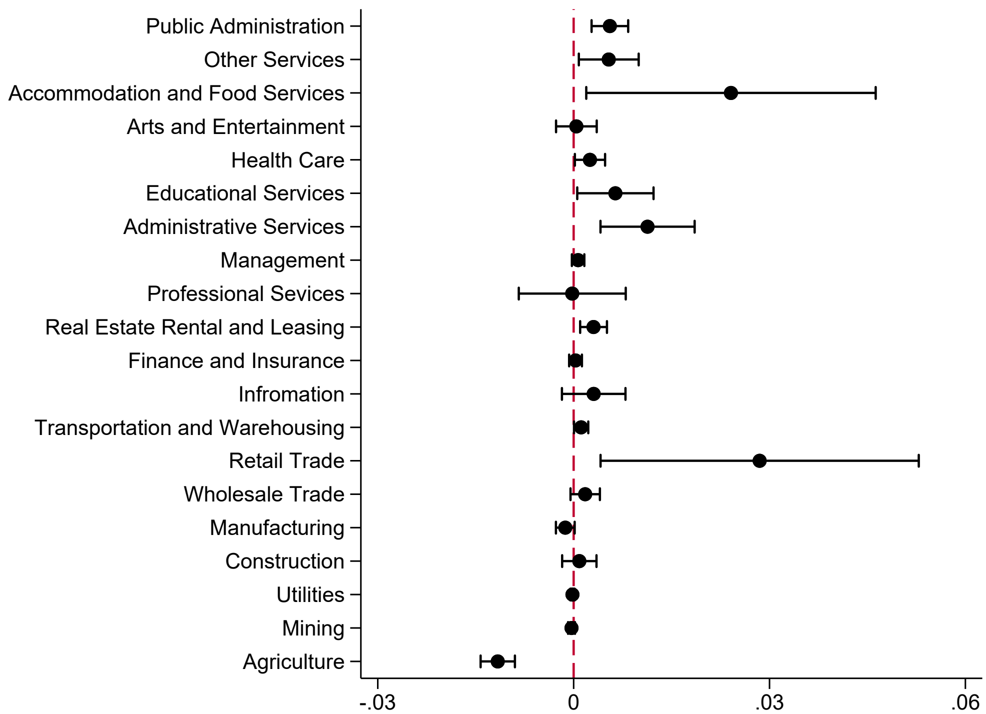
Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1) both with (shown in orange) and without (in green) 1995 commuting zone characteristics interacted with cohort fixed effects. The dependent variables are any employment, employment in the accommodation and food services industry, employment in oil and gas industry, and log quarterly earnings at age 24-25. Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, and commuting zone fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.4: The Effect of the Fracking Boom on College Outcomes



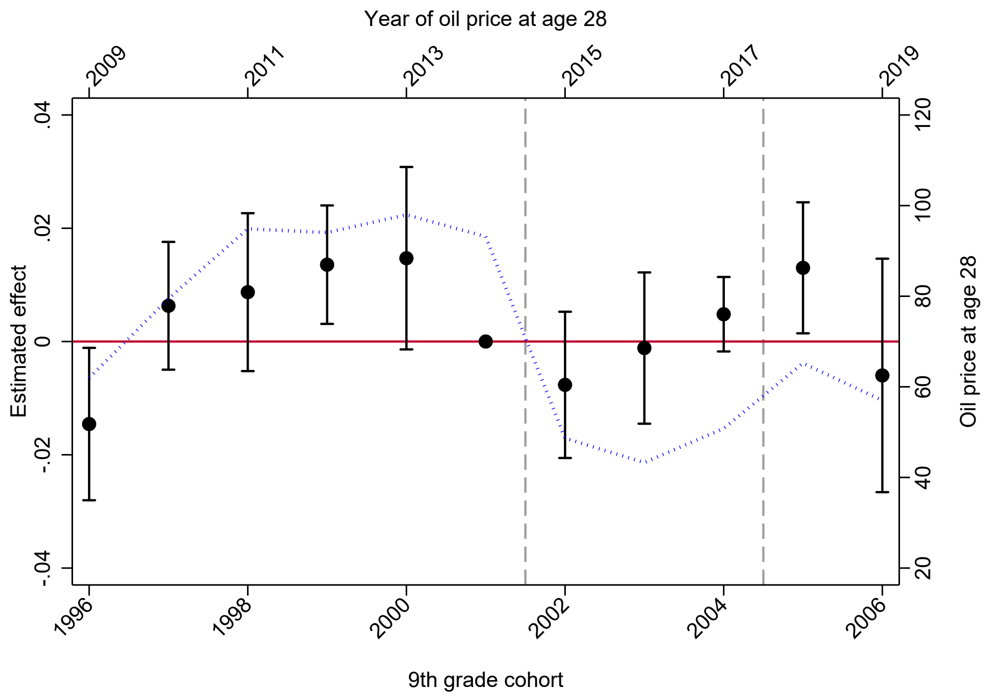
Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1) both with (shown in orange) and without (in green) 1995 commuting zone characteristics interacted with cohort fixed effects. The dependent variables are community college enrollment, public university enrollment, community college graduation, and public university graduation. Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, and commuting zone fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.5: The Effect of the Fracking Boom on Student Employment, by Industry



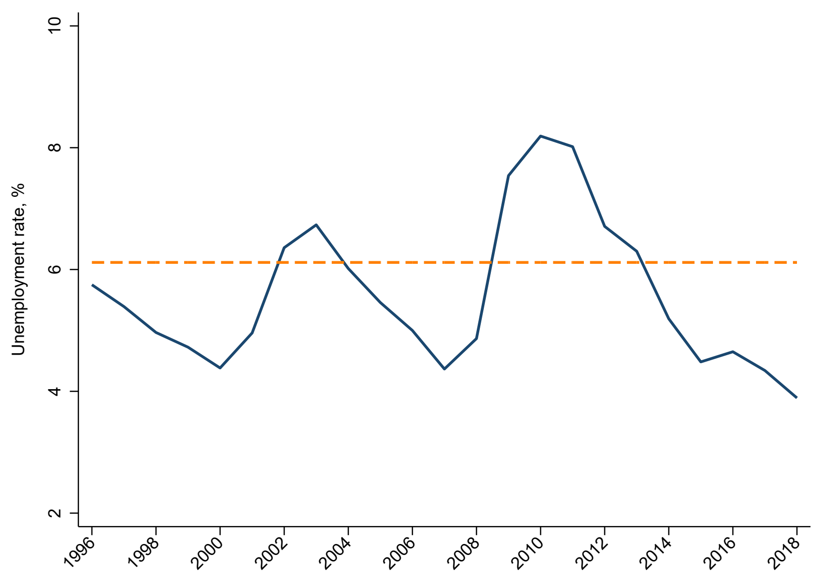
Notes: This figure reports difference-in-differences estimates of the effect of the fracking boom on the probability of employment of high school students by industry. Student demographic controls, commuting zone fixed effects, year fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.6: The Effect of the Fracking Boom on Quarterly Earnings at Age 28



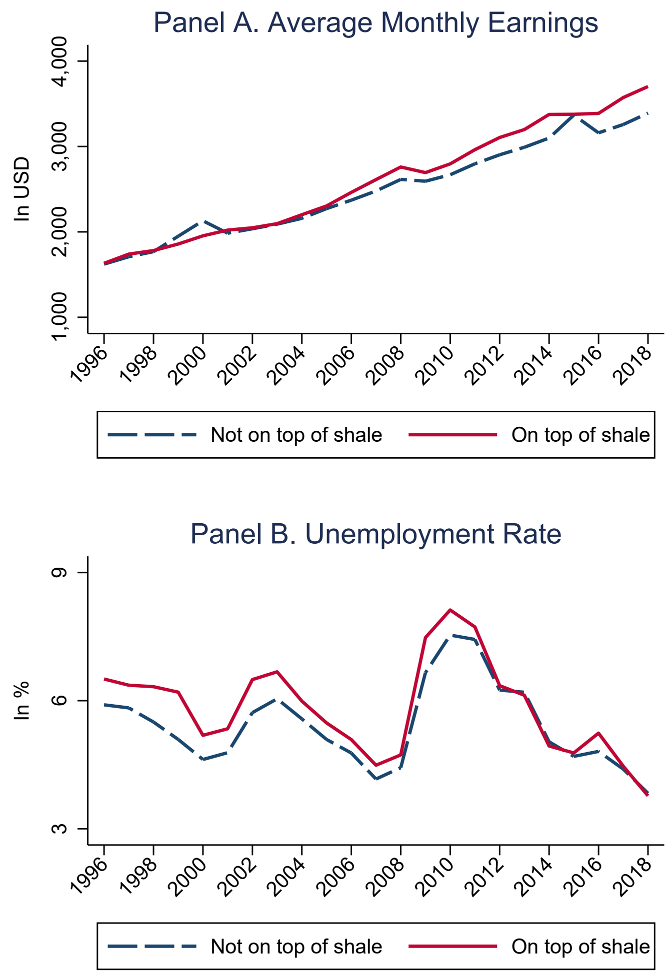
Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1). The dependent variable is the natural logarithm of quarterly earnings at age 28. Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, commuting zone fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects. The sample in columns (2)-(8) includes individuals with non-missing earnings in at least half of the quarters at age 28. The oil price is measured by the West Texas Intermediate (WTI) crude oil price and is plotted at age 28 for each cohort. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency, Texas Workforce Commission, and U.S. Energy Information Administration.

Figure A.7: Annual Unemployment Rate in Texas



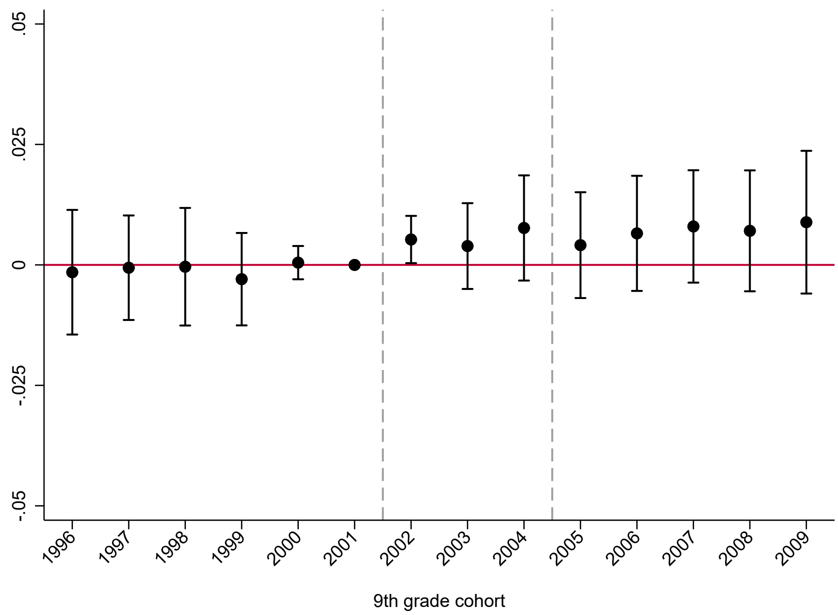
Notes: This figure plots the annual unemployment rate in Texas. The dotted line represents an average unemployment rate in Texas from 1976 to 2007. The data are from the Bureau of Labor Statistics.

Figure A.8: Average Monthly Earnings and Annual Unemployment Rate in Texas



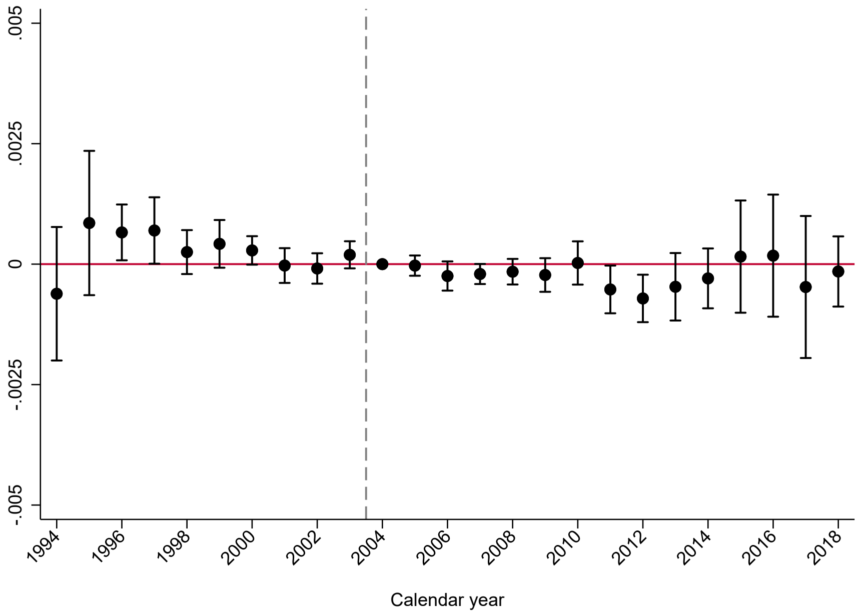
Notes: This figure plots the average monthly earnings and annual unemployment rate in Texas. These statistics are presented separately for commuting zones that lie on top of shale formations and those that do not. The data are from the Bureau of Labor Statistics and Quarterly Workforce Indicators.

Figure A.9: The Effect of the Fracking Boom on the Probability of Observing Students in College or Earnings Records Through Age 25



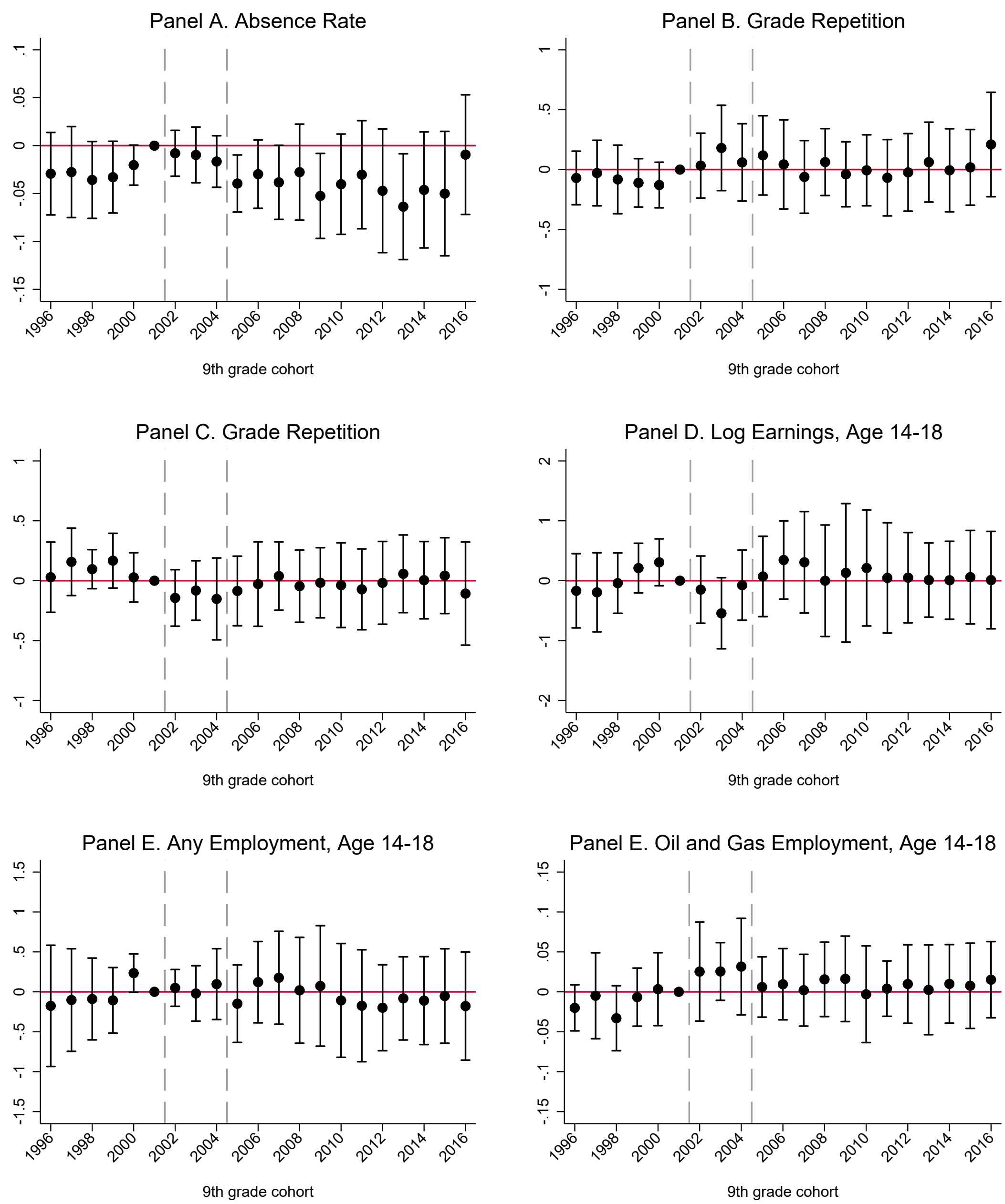
Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1). The dependent variable is the indicator equal to one if a student is observed in college or earnings records by age 25. Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, commuting zone fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.10: The Effect of the Fracking Boom on Out-of-state Migration Rate in TX



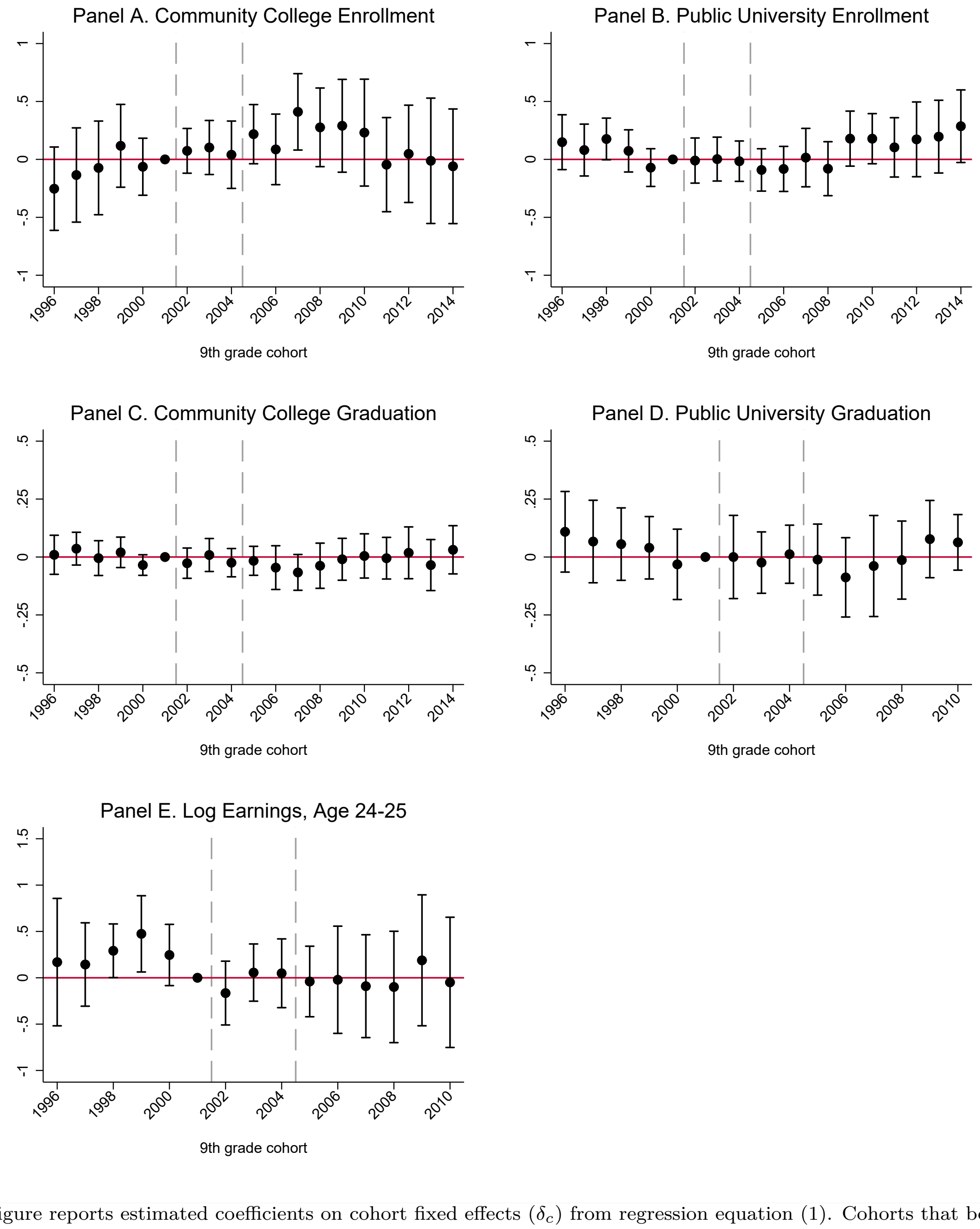
Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1). The dependent variable is out-migration rate, calculated by dividing the number of out-of-state migrants by the total population in 1995. Year 2004, the last year before the beginning of the boom, is the omitted category. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the IRS SOI Tax Stats.

Figure A.11: The Effect of the Fracking Boom on High School and Labor Market Outcomes of Untreated Cohorts, Age 14-18



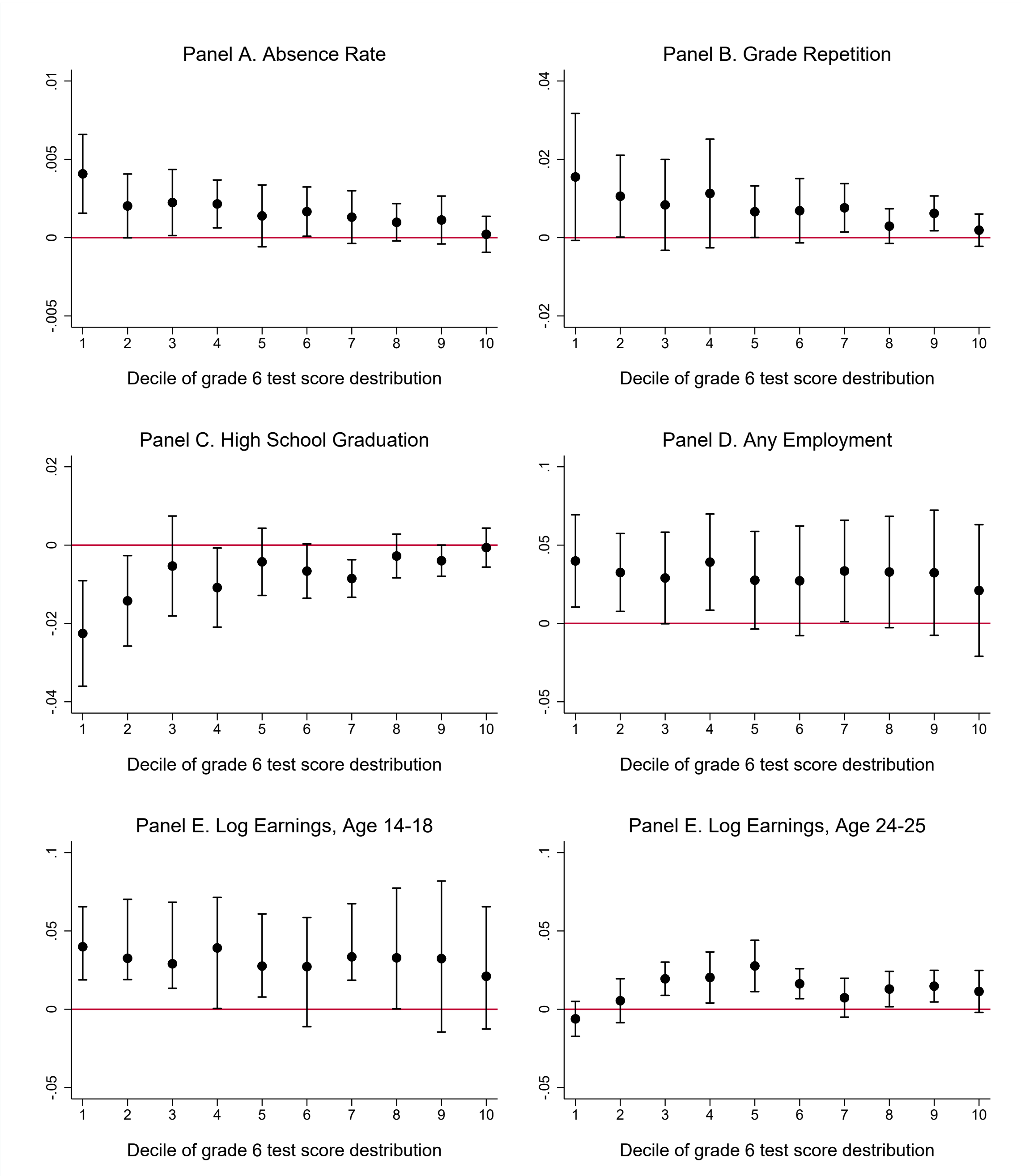
Notes: This figure reports estimated coefficients on cohort fixed effects (δ_c) from regression equation (1). Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, commuting zone fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.12: The Effect of the Fracking Boom on College Outcomes and Long-Term Earnings of Untreated Cohorts



Notes: This figure reports estimated coefficients on cohort fixed effects (δ_c) from regression equation (1). Cohorts that begin grade 9 in 2001 are the omitted category. Cohorts of students that begin high school in 2005 or later are considered fully treated, while cohorts that begin high school before 2001 are considered untreated. The region between two dashed vertical lines represents cohorts that are partially treated. The regression also includes individual-level demographic controls, cohort fixed effects, commuting zone fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Figure A.13: The Effect of the Fracking Boom on Educational and Labor Market Outcomes, by Decile of Grade 6 Test Score Distribution



Notes: This figure reports estimated coefficients on interactions between year indicators and predicted shale oil and gas reserves per capita (β_k) from regression equation (1) for each decile. Students are split into deciles based on their grade on a state standardized test for math and English. 95% confidence intervals for standard errors clustered at commuting zone level are displayed around each point estimate. The data are from the Texas Education Agency and the Texas Workforce Commission, provided by the Texas Education Research Center.

Online Appendix B: Literature Comparison

Natural Resource Booms, Human Capital, and Earnings: Evidence from Linked Education and Employment Records

Alina Kovalenko

1 Overview

This Appendix provides more detailed comparisons between the methodologies and results in my work and several other important papers from the literature.

1.1 Marchand and Weber (2020)

One of the papers closest to mine in terms of question, methodology, and data is Marchand and Weber (2020). In this section, I closely examine the similarities and differences between our findings. Our results overlap substantially, suggesting that they are driven by similar underlying mechanisms. While they do differ somewhat for a few outcomes, I show that these differences are mostly driven by differences in data definitions, time horizon, methodology, and sample construction. First, I include a summary of the differences in main findings between our papers below in Table B.1. After that, I discuss the differences in our approaches in more detail for many of the outcomes for which our papers overlap.

Table B.1: Comparison of Results to Marchand and Weber (2020)

Outcome	Results		Potential source of differences
	This paper	Marchand and Weber (2020)	
Attendance rate	Negative effect	Negative effect	Similar results despite differences in the underlying data
HS completion	Negative effect	No effect	Different variable definitions
Teachers with < 5 years of experience	No effect	Positive effect	Differences in timeframe, regions analyzed, and methodology
Teachers with advanced degree	No effect	No effect	The results are similar
Teacher earnings	No effect	No effect	Similar results despite differences in the underlying data
Student composition	No effect	Negative effect on economically disadvantaged students	The results are similar but effects in this study are less precisely estimated due to the differences in the underlying sample
Local finances	Positive effect	Positive effect	The results are similar

Most of the differences between our papers are due to four factors: differences in education and labor market data, differences in the areas being considered, differences in econometric methodology, and differences in the timing of the boom. I describe the main differences below.

- Labor market outcomes: Due to the nature of the data available to them, Marchand and Weber (2020) cannot track students into the labor market, and as a result only look at aggregate labor market outcomes. Thus while we find similar qualitative effects on these outcomes, my estimates are likely to more accurately reflect the short- and long-run effects for students because I can track them explicitly into the labor market rather than relying on population averages.

- Educational achievement: We find similar effects on attendance, though we use slightly different measures: Marchand and Weber (2020) rely on aggregate measures of daily attendance from grade 1 through 12, whereas I focus on daily attendance of high school students, i.e. grades 9 through 12. Similarly, Marchand and Weber (2020) look at aggregate test passing rates which include students in grades 3 through 12. I am directly able to observe student test scores in a given grade, which allows me to better distinguish between ability and attendance. In addition, I look at high school grade retention, as well as college enrollment and graduation, separately by institution type. Marchand and Weber (2020) cannot track students into college but they instead analyze the percentage of high school students taking college entrance exams.

- High school completion: We define high school completion differently. Publicly available TEA rates are calculated using a cohort-based approach that in general will overestimate graduation rates because it only includes students observed at the end of the sample period in the denominator; the TEA itself reports that one advantage of their measure is that "districts have more time to encourage dropouts to return to school before being held accountable."[1] In contrast, I measure graduation rates using a simpler approach that tracks whether a student had obtained a high school degree in Texas within four years of starting high school.[2] While my measure has the shortcoming of not being able to track outcomes for students who leave Texas[3], I believe it more accurately captures students who do not in

[1] "Secondary School Completion and Dropouts in Texas Public Schools 1998-99", Texas Education Agency, January 2001. Link: https://tea.texas.gov/sites/default/files/DropComp_1998-99.pdf

[2] This approach is consistent with other papers that relied on the same data; for example, see Denning (2017) and Ballis and Heath (2021).

[3] I provide evidence in online Appendix A that selective migration does not appear to be a major concern in my setting.

fact obtain their diploma, even if they do not end up being classified by the TEA as official dropouts at the time. Another important factor is the difference in timing. High school graduation rates had started to fall during the beginning of the boom in the early 2000s. This period represents the pre-treatment period for Marchand and Weber (2020), whereas I am able to track students all the way back to 1996. To the extent that the fracking boom started to affect educational outcomes for the students I consider to be partially treated, I believe my pre-treatment group does a better job of reflecting educational outcomes prior to the boom.

- Teacher outcomes: Our findings differ somewhat when comparing teacher experience; Marchand and Weber (2020) find a positive and statistically significant increase in the share of teachers with fewer than five years of experience, wheras I estimate a positive but statistically insignificant effect. These differences are driven by a combination of data, timing, and methodological differences. First, Marchand and Weber (2020) consider only oil areas, whereas I look at both oil and gas regions. Finally, they use a measure of fracking exposure based on shale depth, whereas I measure it based on the volume of reserves. For comparison, I can adjust my methodology to more closely match theirs by restricting my sample to 2001-2014 and including only oil shales. When I use this approach, my point estimate for the increase in the share of inexperienced teachers is little changed, but becomes statistically significant at the 10% level. While these effects are still modestly smaller than the ones documented in Marchand and Weber (2020), they suggest that both of our approaches generate qualitatively similar results when the methodologies are brought closer together.

- Gas vs. oil shales: Marchand and Weber (2020) find that the labor market effects of fracking were concentrated in oil-rich regions. Thus to provide a closer comparison, I estimate my main specification separately using two subsamples of fracking areas: one including only the oil plays (Permian and Eagle Ford shales), and one including only the gas plays (Barnett and Haynesville shales). The results of this exercise are shown in Table A.7 in online Appendix A. In Panel A, which shows the effects for oil shale plays, I obtain estimated effects that show the same sign as my main results, but with slightly larger magnitudes for most outcomes. However, in the gas plays, shown in Panel B, I show that there were no statistically significant effects on educational or labor market outcomes. This is consistent with the findings of Marchand and Weber (2020), who find that oil plays also experienced larger booms in local labor markets. The fact that estimated education outcomes were also largest in these areas provides further support for the idea that changes in labor market conditions were the primary channel through which student outcomes were affected.

1.2 Emery et al. (2012)

Emery et al. (2012) show that students in Canada exposed to the fracking boom experienced reductions in college graduation rates. However, they also find that these effects were not permanent, and that the gap between treated and untreated cohorts closed as individuals exposed to fracking around the time of high school returned to college later in life. Here I analyze the degree to which this effect might be happening in my setting. While I can track individuals across time, I am limited by the timing of the boom in my ability to track educational outcomes of treated individuals in their late 30s. There are several cohorts I can track through age 28, however, and based on what I can see, the reduction in educational attainment for individuals exposed to the fracking boom seems to be quite persistent.

In Table B.2 below, I compare the effects of fracking on contemporaneous and later-life educational outcomes for the subset of students who I can track up to age 28. This sample includes three "partially treated" cohorts and two "fully treated" cohorts. I use my baseline difference-in-difference model to analyze five outcomes: high school graduation, community college enrollment and graduation, and public college enrollment and graduation. In panel A, I use the same horizon for calculating graduation and enrollment rates as in my main analysis.[4] The estimated effects of fracking on short-run educational outcomes are overall very similar to my main results (shown in Table 2 of the main paper) that include all students.

In Panel B, I present results that look at the same outcome variables, but instead extend the horizon for evaluating each outcome through age 28. For example, if a student first enters high school at age 14 but doesn't graduate until age 21, they would be counted as a non-graduate in Panel A but a graduate in Panel B. If students were returning to school later in life, we would expect the causal effects of fracking to be much smaller in the long run, as fracking would simply be delaying, rather than reducing, educational attainment. Instead, I find that the results in this panel are very similar in magnitude to Panel A, which suggests that the gap in educational outcomes remains relatively stable through age 28. These results are limited by the fact that I cannot track student outcomes as far as in Emery et al. (2012), so I cannot definitively rule out that some of these students may ultimately return to seek more education in the future. Nonetheless,

[4] High school graduation status is measured four years after starting high school. Enrollment in either a community college or public university is measured within two years of expected high school graduation. Community college graduation is measured within four years of expected high school graduation. Public university graduation is measured within six years of expected high school graduation.

Table B.2: The Effect of the Fracking Boom on Educational Outcomes in the Short and Long Run

	High school	Community college		Public university	
	Graduation	Enrollment	Graduation	Enrollment	Graduation
	(1)	(2)	(3)	(4)	(5)
Panel A. Short-run educational outcomes					
Fully treated	-0.907	-0.001	-0.038	0.241	-0.252
	(0.536)	(0.529)	(0.107)	(0.228)	(0.146)
Partially treated	-0.476	-0.338	-0.111	0.053	-0.138
	(0.466)	(0.375)	(0.049)	(0.260)	(0.176)
Panel B. Educational outcomes by age 28					
Fully treated	-0.800	0.168	-0.161	-0.365	-0.265
	(0.465)	(0.622)	(0.298)	(0.194)	(0.162)
Partially treated	-0.555	-0.230	-0.095	-0.412	-0.055
	(0.433)	(0.473)	(0.276)	(0.145)	(0.180)
Observations	2,450,439	2,450,439	2,450,439	2,450,439	2,450,439

Notes: This table reports difference-in-differences estimates of the effect of the fracking on educational outcomes in the short (Panel A) and long run (Panel B). The unit of analysis is at the student-cohort-commuting zone level. “Partially treated" and “Fully treated" rows report the coefficients on the interactions between predicted shale reserves and an indicator variable for entering high school between 2001 and 2004 and an indicator variable for entering high school in 2005 or later, respectively. Commuting zone fixed effects, year fixed effects, and 1995 commuting zone characteristics interacted with cohort fixed effects are included in all specifications. The sample includes cohorts that start high school in 1996-2006. Coefficients are multiplied by 100 for readability. Standard errors, shown in parentheses, are clustered at commuting zone level.

the data suggest that the reductions in educational attainment caused by the fracking boom in Texas were quite persistent.

1.3 Lee (2015)

- Data sources: Lee (2015) uses county-by-year observations of income and employment, while my data follow individuals. Tracking students through school and into the workforce is important in my setting because my paper analyzes how the fracking boom affects the tradeoffs between education and employment for individual students; even if the fracking boom raises aggregate wages in a county, this by itself does not guarantee that these benefits will accrue to students on the margin of dropping out of school. On the other hand, my data do not allow me to speak directly to aggregate labor market conditions, because only

Texas students are included in my sample. Thus to the extent that the total county-level income and employment effects of Lee (2015) are coming from workers who finished school before my data start or moved to work there from out of state, my results will not capture them.

- Estimation methodology: My approach identifies variation in the intensity of the fracking boom through pre-existing variation in the geological features of shales. I use this geological approach to measure exposure to the boom rather than other activity measures such as drilled wells because, as Lee (2015) points out, these activity measures are likely to be endogenously related to other factors that can affect education and labor market outcomes. This model treats areas with greater shale reserves as having greater exposure to the fracking boom. In contrast, Lee (2015) focuses primarily on actual drilling activity as their measure of fracking. Lee (2015) also reports similar results using the percentage of a county covering a shale as an instrument for drilling activity, which is conceptually closer to my approach. Because I analyze individuals, my specification can also control for a much richer set of characteristics that could affect education and labor market outcomes.

- Industries: My paper focuses primarily on labor market outcomes for high school students, very few of whom would be qualified (or legally allowed, given age constraints on many oil and gas jobs) to work directly in energy extraction. As shown in Table 1 of the main paper, I find virtually no effect on oil and gas employment for these cohorts, and instead estimate large effects on employment in industries such as retail or food services. In contrast to my work, Lee (2015) focuses on estimating "employment multipliers" which ask how the direct effects of increased drilling on oil and gas employment map to broader employment effects in the rest of the economy. Also, Lee (2015) points out that the employment effects of legacy wells may be different than new wells. The long-run effects may be muted if they do not lead to the drilling of new wells, which can potentially explain why my long-run effects are smaller than short-run effects of the boom.

- Timing of the analysis: Our papers rely on different time periods in our analysis. Lee (2015) includes data from 2009-2014, whereas I look at a longer time period from 1996-2018. This is useful in my setting because it allows me to analyze how the short- and long-run effects differ for cohorts whose later-life earnings correspond to periods of lower oil prices and drilling activity.

1.4 Cai et al. (2019)

- Data sources: Cai et al. (2019) use pooled cross-sectional observations from the American Community Survey (ACS). This is a richer level of detail than the data used in Lee (2015) and allows them to compare labor market outcomes for different demographic groups. Cai et al. (2019) find that the fracking boom led to improvements in labor markets that are generally consistent with my findings. However, because they cannot follow individuals, their results cannot link these labor market outcomes to individual education decisions, which is a crucial feature of my analysis.

- Estimation methodology: Cai et al. (2019) use several different estimation approaches. They initially consider an OLS specification in which fracking intensity is measured directly by oil and gas employment shares in a given location. However, citing endogeneity concerns with this approach, most of their results focus on a shift-share IV which interacts the oil and gas employment share in each Texas county in the year 2000 (which predates the boom) with the total oil and gas employment in the U.S. (excluding Texas) over time as their instrument.

- Industries: Cai et al. (2019) analyze both the direct effects of the fracking boom on employment and earnings in the oil and gas sector, as well as the indirect effects on other industries. They find that greater exposure to the fracking boom led to statistically and economically significant increases in employment and earnings for workers outside the oil and gas sector, which is consistent with my findings. They also find that these improvements were broad-based across different demographic groups. While they analyze outcomes across a much broader range of demographic groups, their findings are consistent with my results in Table 1 of the main paper and Table A.8 of online Appendix A that do not show significant differences in education or employment across black/white or male/female students. While my results are largely driven by gains in food and retail services, however, they find negative (though generally not statistically significant) effects for retail. One possible explanation for this discrepancy is that Cai et al. (2019) focus explicitly on workers 18 and older, while I look primarily at students aged 14-18. It is likely that teenage employees filled the job openings created when adults previously employed in retail moved into other jobs created by the boom, especially since Cai et al. (2019) show that many of these other jobs experienced higher wages.

1.5 Methodologies of Bartik et al. (2019) and Feyrer et al. (2017)

The identification strategy of Bartik et al. (2019) is the closest paper to my own in terms of research design. Both of our papers use an identification strategy that is based on geological variation and time variation in the start of the fracking boom. While my data provide information regarding the total level of reserves in shales, however, I cannot directly observe the distribution of reserves within them, and as a result I assign treatment based on the size of each commuting zone that lies on top of a shale. Bartik et al. (2019) use a proprietary dataset that tracks variation of resources *within* shales, which allows them to track metrics like shale thickness, depth, and thermal maturity that can more precisely determine which areas have the greatest potential benefits from drilling. Overall, their methodology can be considered a similar (but more accurate) measure than the one I use. If reserves were uniformly distributed across each shale, both of our measures should give the same results, but to the extent that they are more concentrated in some areas, the Bartik et al. (2019) approach will better reflect the true potential benefits of fracking.

In my robustness analysis, I perform two additional exercises incorporating elements of the Bartik et al. (2019) methodology.[5] First, instead of assigning a common boom start year for all areas in Texas, I use Bartik et al. (2019) variation in the start of the fracking boom by shale play. Second, I restrict my analysis to areas that only lie on top of shale. I find results that are similar in magnitude and significance to my main estimates.

Feyrer et al. (2017) also shares many similarities with my paper, but they differ by using an IV approach in which geological features serve as an instrument for the value of new oil and gas production. They predict production by estimating it as a function of county fixed effects and shale play by year fixed effects. Although I don't use production directly, in Figure 2 of my main paper I show that my measure of reserves is closely related to the number of new wells, suggesting that our results are driven by similar mechanisms.

1.6 Black et al. (2005) and Cascio and Narayan (2020)

The two papers that provide the most direct comparison to my work regarding graduation rates are Black et al. (2005), who look at the coal boom, and Cascio and Narayan (2020), who look at the fracking boom. One difficulty here is that we all focus on different outcomes: Black et al. (2005) use grade 9-12 enrollment counts divided by age 15-19 population, while Cascio and

[5] I would ideally have liked to directly test their approach in my paper, but unfortunately the data are prohibitively expensive.

Narayan (2020) use grade 11-12 enrollment counts divided by age 17-18 population as well as high school dropout rates constructed using self-reported information from the ACS. To compare the magnitude of my results to dropout effects in Cascio and Narayan (2020), I first redefine my outcome to measure dropout rates instead of graduation and restrict my sample to cover only males from 2000-2015 cohorts in order to better match the specification in Cascio and Narayan (2020). I then estimate a regression model in which I use cohorts that start high school in 2000 and 2001 as a control group, and calculate separate treatment effects for cohorts that start high school in 2002-06 and 2007-2015. I find that male students who were exposed to fracking in 2007-15 had 1.5 p.p. higher dropout rates. Evaluated at the baseline average dropout rate of 25.5%, this translates into a 5.9% increase in dropout rates. I also estimate that male students exposed to the boom had a 4% increase in earnings. A back-of-the-envelope calculation thus suggests that a 10% increase in earnings would be associated with a 14.8% increase in dropout rates, which is larger than the 8.5% increase in high school dropout rate resulting from a 10% increase in earnings that Cascio and Narayan (2020) calculates.[6] This difference is likely to be driven by the fact that estimates in Cascio and Narayan (2020) include 14 states that experienced fracking, rather than just Texas, which experienced a very strong boom.

Separately, Black et al. (2005) and Cascio and Narayan (2020) each estimate the effect on high school enrollment of a 10% increase in earnings: Black et al. (2005) find that this increase leads to a decline in high school enrollment rates of 5-7%, while Cascio and Narayan (2020) find a more modest decrease of 1.9%. Because I do not directly analyze enrollment as an outcome, in order to facilitate comparison to these papers, I assume that fracking causes the same percentage change in high school enrollment as it does in high school graduation, and scale my estimates accordingly. As shown in Tables 1 and 2 of the main paper, I estimate that exposure to an average level of reserves leads to an increase in male earnings of 4.0% and a decrease in graduation rates of 1.92% (the latter calculated by comparing the coefficient estimate of -1.38 p.p. to the mean graduation rate of 72%). From these numbers, a back-of-the-envelope calculation suggests that a 10% increase in earnings leads to a 4.8% decline in graduation rates. Thus my results are larger than past studies analyzing the fracking boom, and similar to the low end of the range of estimates calculated from the coal boom.

[6] This is calculated by taking the 2SLS coefficient corresponding to the effect of a 10% increase in earnings on dropout rates in Table 4 in Cascio and Narayan (2020), which is 0.89, and dividing by the mean dropout rate of 10.45.

References

Ballis, Briana and Katelyn Heath, “The long-run impacts of special education,” *American Economic Journal: Economic Policy*, 2021, *13* (4), 72–111.

Bartik, Alexander W., Janet Currie, Michael Greenstone, and Christopher R. Knittel, “The Local Economic and Welfare Consequences of Hydraulic Fracturing,” *American Economic Journal: Applied Economics*, 2019, *11* (4), 105–55.

Black, Dan A., Terra G. McKinnish, and Seth G. Sanders, “Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust,” *ILR Review*, 2005, *59* (1), 3–16.

Cai, Zhengyu, Karen Maguire, and John V Winters, “Who Benefits from Local Oil and Gas Employment? Labor Market Composition in the Oil and Gas Industry in Texas and the Rest of the United States,” *Energy Economics*, 2019, *84*, 104515.

Cascio, Elizabeth U. and Ayushi Narayan, “Who Needs a Fracking Education? The Educational Response to Low-Skill-Biased Technological Change,” *ILR Review*, 2020, *forthcoming.*

Denning, Jeffrey T, “College on the cheap: Consequences of community college tuition reductions,” *American Economic Journal: Economic Policy*, 2017, *9* (2), 155–88.

Emery, JC Herbert, Ana Ferrer, and David Green, “Long-Term Consequences of Natural Resource Booms for Human Capital Accumulation,” *ILR Review*, 2012, *65* (3), 708–734.

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote, “Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution,” *American Economic Review*, 2017, *107* (4), 1313–34.

Lee, Jim, “The Regional Economic Impact of Oil and Gas Extraction in Texas,” *Energy Policy*, 2015, *87*, 60–71.

Marchand, Joseph and Jeremy G. Weber, “How Local Economic Conditions Affect School Finances, Teacher Quality, and Student Achievement: Evidence from the Texas Shale Boom,” *Journal of Policy Analysis and Management*, 2020, *39* (1), 36–63.