Transitioning to College and Work
Part 3: Labor Market Analyses in Houston and Texas

by Brian Holzman, Ph.D., Mehreen Gul, Ph.D.,
Esmeralda Sánchez Salazar, M.S., and Camila Cigarroa Kennedy

July 2020

Appendix A. Summary Statistics by Postsecondary Degree Type

PANEL A. TEXAS

Variable	*No postsecondary degree*		*A.A. degree*		*B.A. degree or more*	
	Mean	SD	Mean	SD	Mean	SD
Age	38.40	(8.63)	38.50	(8.53)	38.60	(8.66)
Years of education	11.50	(2.42)	14.00	(0.00)	16.70	(1.27)
Years of experience	20.90	(8.84)	18.50	(8.53)	15.90	(8.61)
Hourly wage	12.90	(9.66)	16.70	(13.40)	23.50	(19.70)
Family size	3.31	(1.66)	3.00	(1.45)	2.82	(1.40)
Female (proportion)	0.45		0.51		0.48	
Non-white (proportion)	0.17		0.18		0.20	
N	45,710		6,368		17,682	

PANEL B. HOUSTON AREA

Variable	*No postsecondary degree*		*A.A. degree*		*B.A. degree or more*	
	Mean	SD	Mean	SD	Mean	SD
Age	38.30	(8.56)	38.60	(8.55)	38.50	(8.62)
Years of education	11.40	(2.52)	14.00	(0.00)	16.70	(1.28)
Years of experience	20.90	(8.74)	18.60	(8.55)	15.80	(8.64)
Hourly wage	11.60	(0.04)	18.10	(13.30)	25.20	(20.00)
Family size	3.27	(1.68)	2.98	(1.46)	2.77	(1.42)
Female (proportion)	0.44		0.45		0.47	
Non-white (proportion)	0.25		0.26		0.28	
N	9,395		1,340		4,194	

Notes: Data from CPS IPUMS, 1979-2016. All dollar values were deflated by the Consumer Price Index (CPI), 1999. CPS sampling weights were used in all calculations.

Appendix B. Summary Statistics by STEM Occupation

PANEL A. TEXAS

Variable	*STEM*		*Non-STEM*	
	Mean	SD	Mean	SD
Age	38.50	(8.56)	38.50	(8.64)
Years of education	14.20	(2.74)	12.90	(3.08)
Years of experience	18.30	(8.98)	19.50	(9.02)
Hourly wage	20.80	(15.17)	15.10	(13.80)
Family size	3.05	(1.54)	3.16	(1.60)
Female (proportion)	0.36		0.49	
Non-white (proportion)	0.21		0.17	
N	12,330		57,430	

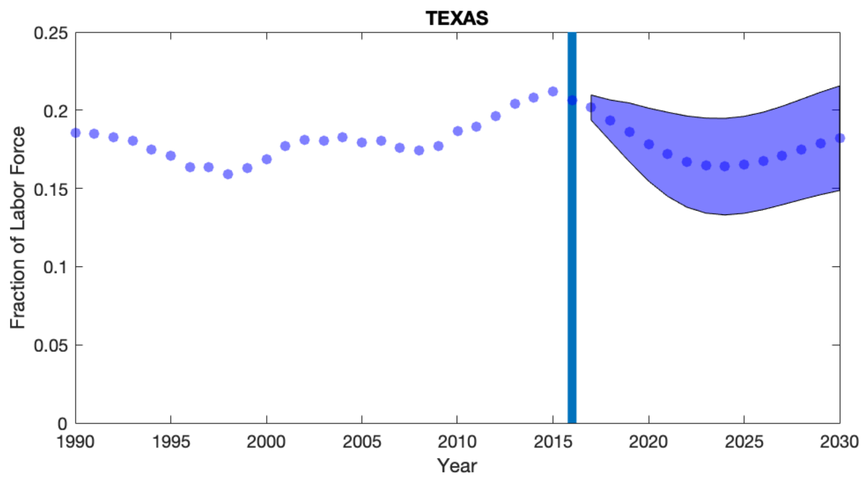
PANEL B. HOUSTON AREA

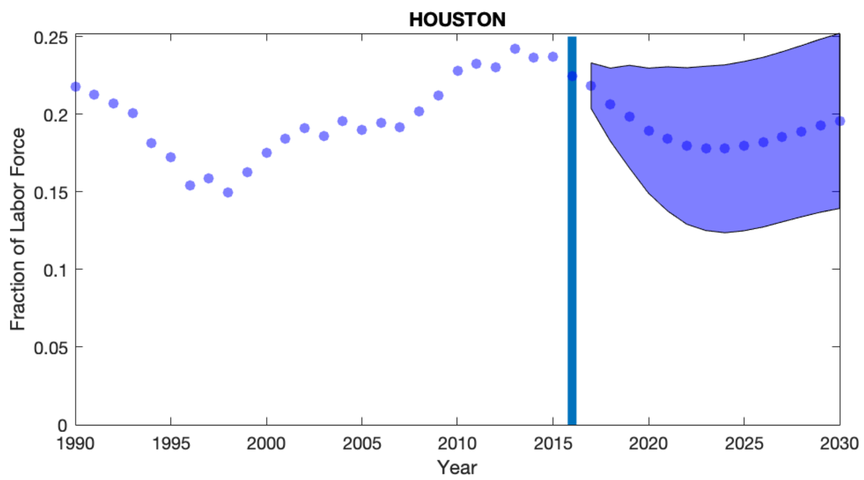
Variable	*STEM*		*Non-STEM*	
	Mean	SD	Mean	SD
Age	38.60	(8.64)	38.30	(8.60)
Years of education	14.40	(2.92)	13.00	(3.20)
Years of experience	18.20	(9.09)	19.40	(8.95)
Hourly wage	22.60	(15.20)	16.50	(15.50)
Family size	3.03	(1.60)	3.10	(1.60)
Female (proportion)	0.32		0.48	
Non-white (proportion)	0.29		0.25	
N	2,934		11,995	

Notes: Data from CPS IPUMS, 1979-2016. All dollar values were deflated by the Consumer Price Index (CPI), 1999. CPS sampling weights were used in all calculations.

Appendix C. Labor Force Composition by STEM Occupation, 1990-2030

This graph plots trends in labor force composition by STEM occupation in Texas and the Houston area between 1990 and 2016. In 2016, 20 percent of the workforce in Texas and 22 percent of the workforce in Houston was employed in a STEM occupation. Despite declines in the late 1990s and 2000s (likely due to economic recessions), the percentage of workers employed in a STEM occupation increased between 1990 and 2016. The graphs also project trends through 2030, with the shaded regions representing 95% confidence intervals. It is estimated that the percentage of workers in STEM occupations will decline in the early 2020s before increasing to 18 and 20 percent in 2030 in Texas and Houston, respectively.





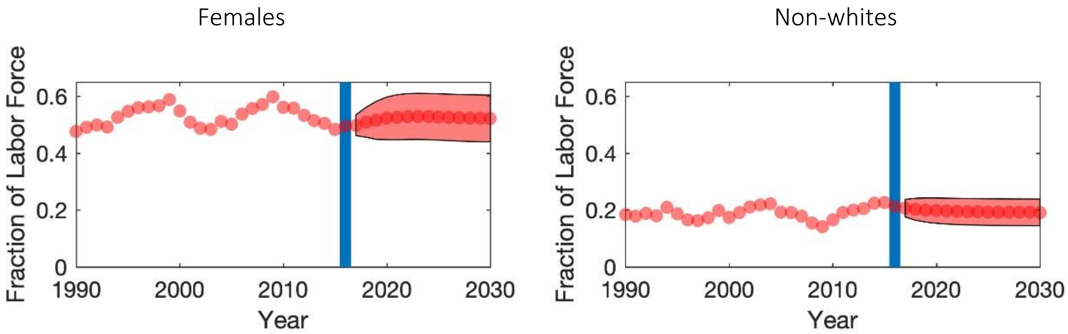
Notes: The sample was limited to individuals between 25-34 years old. All series were six-year moving averages after model estimation at yearly frequencies. The shaded regions after 2016 are 95% confidence intervals. Please see Appendix I for additional details.

Appendix D. Labor Force Composition by Postsecondary Degree Type and STEM Occupation among Females and Non-Whites, 1990-2030

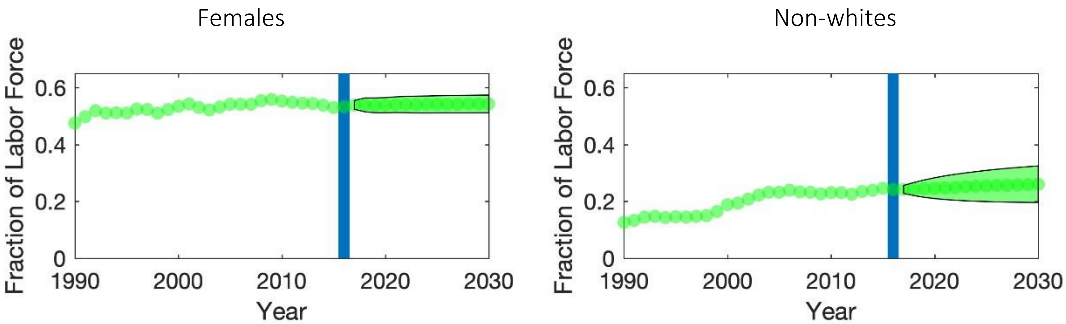
Figures D1 and D2 plot trends in labor force composition by postsecondary degree type and STEM occupation for females and non-whites in Texas and Houston. Between 1990 and 2016, female and non-white workers represented a growing share of the labor force with an associate’s degree, with a bachelor’s degree or more, and in STEM occupations. Projections through 2030 suggest that for females and non-whites, levels of educational attainment and STEM occupation participation will, for the most part, remain stable. Two exceptions are the share of non-whites with a bachelor’s degree or more and the share of non-whites working in STEM occupations at the state level, both of which may experience growth.

Figure D1. Texas

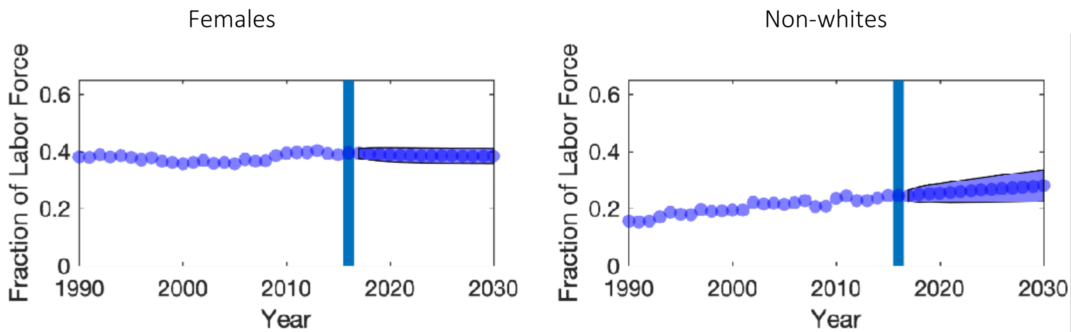
Panel A. Associate's degree



Panel B. Bachelor's degree or more



Panel C. STEM occupation

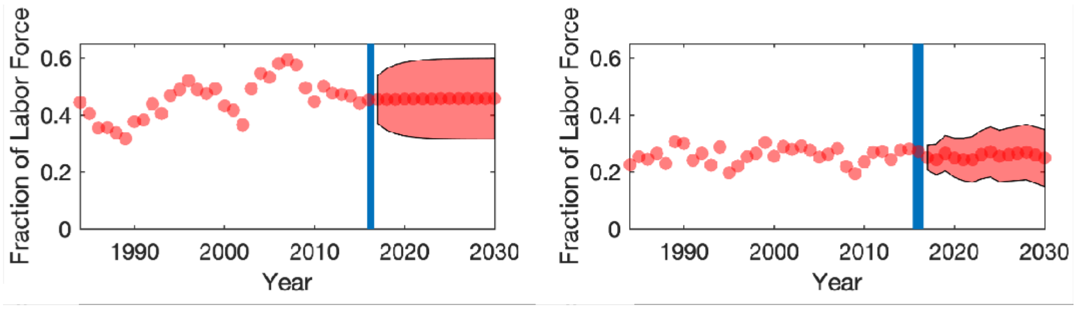


Notes: The sample was limited to individuals between 25-34 years old. All series were six-year moving averages after model estimation at yearly frequencies. The shaded regions after 2016 are 95% confidence intervals. Please see Appendix I for additional details.

Figure D2. Houston Area

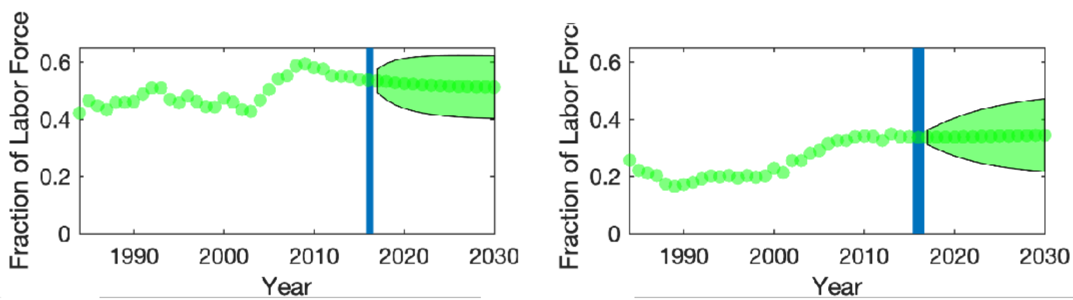
Panel A. Associate's degree

Females | Non-whites



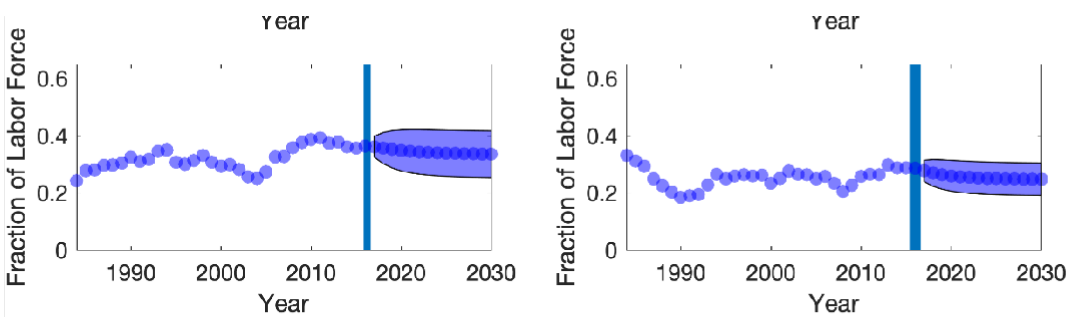
Panel B. Bachelor's degree or more

Females | Non-whites



Panel C. STEM occupation

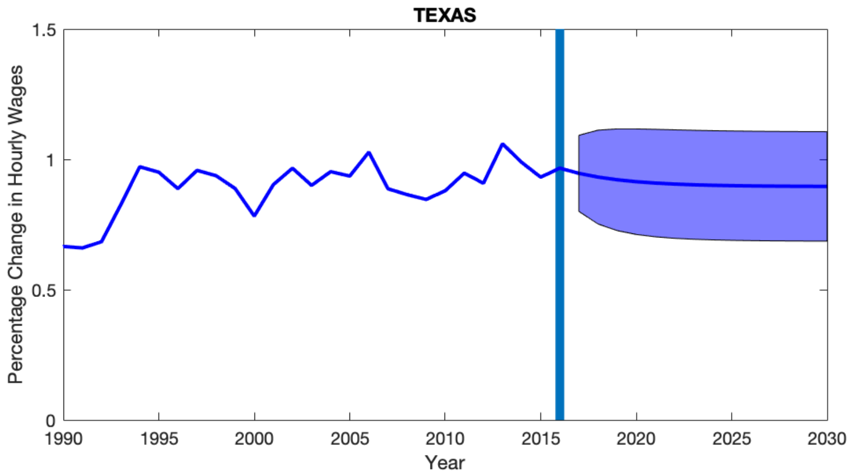
Females | Non-whites

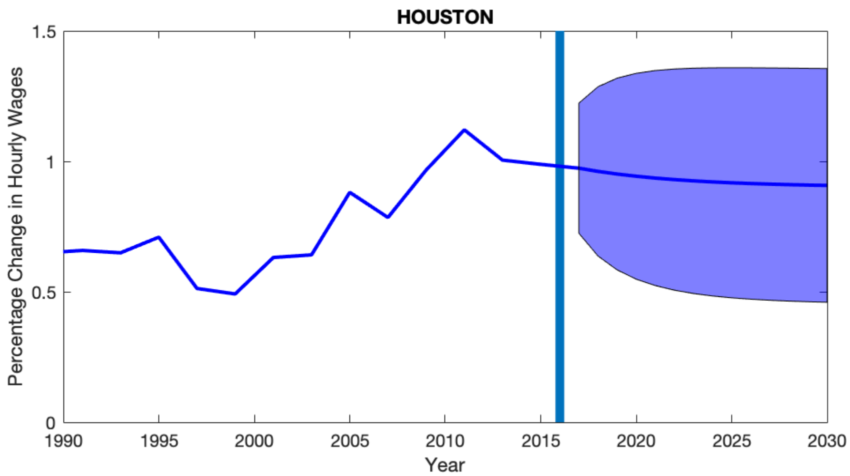


Notes: The sample was limited to individuals between 25-34 years old. All series were six-year moving averages after model estimation at yearly frequencies. The shaded regions after 2016 are 95% confidence intervals. Please see Appendix I for additional details.

Appendix E. Estimates of Wage Premiums by STEM Occupation, 1990-2030

These graphs plot trends in wage premiums for workers employed in STEM occupations relative to workers employed in non-STEM occupations. In both Texas and the Houston area, STEM workers earned higher wages than non-STEM workers in all years between 1990 and 2016. In 2016, STEM workers in Texas and Houston earned 100 percent higher wages than non-STEM workers, representing an overall increase since 1990. Projections suggest STEM workers will still earn about 100 percent more than non-STEM workers in 2030.

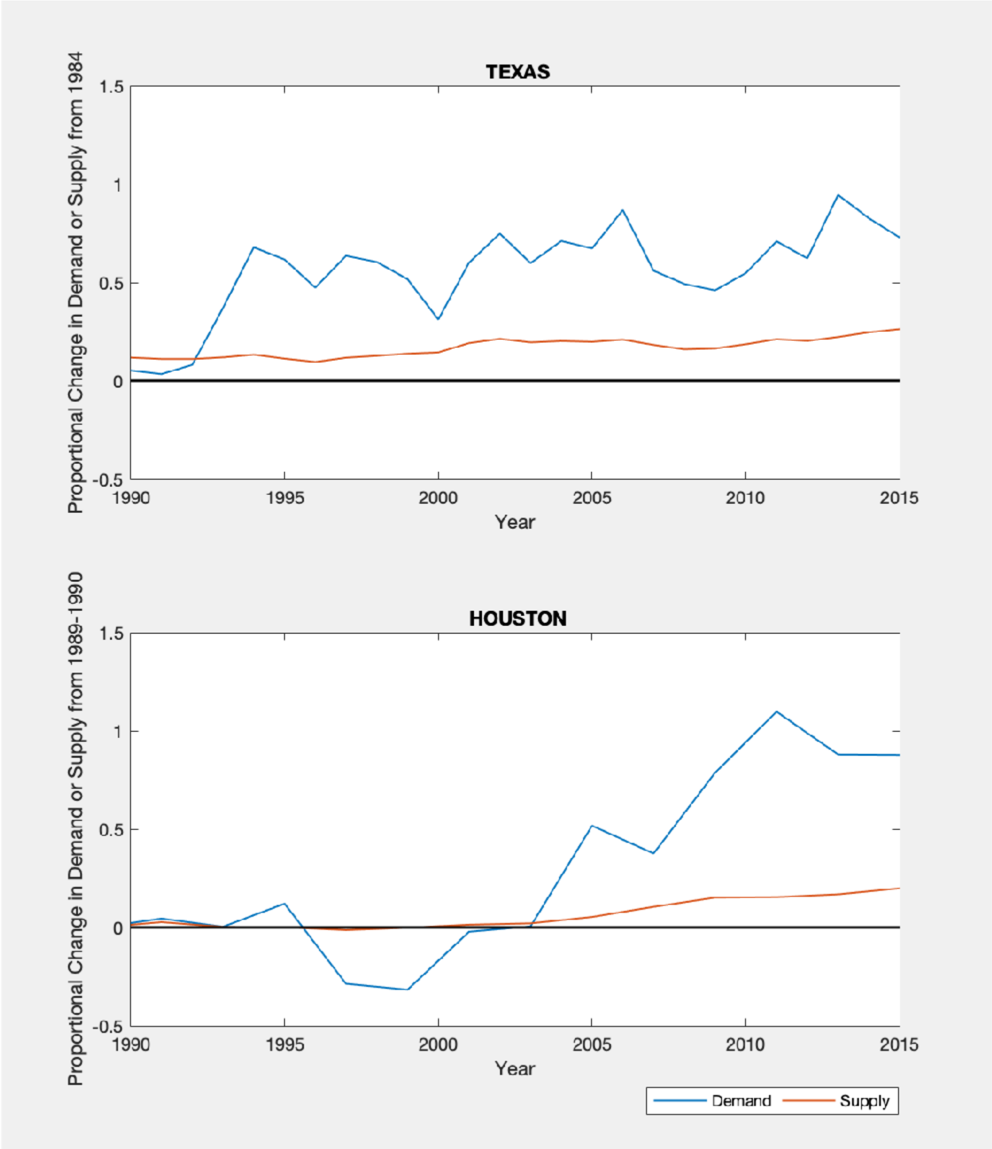




Notes: All series were six-year moving averages after model estimation at yearly frequencies. The shaded regions after 2016 are 95% confidence intervals. Please see Appendix I for additional details.

Appendix F. Change in Supply in Demand by STEM Occupation, 1990-2015

These graphs plot trends in the supply and demand for STEM workers over time. Each point on the graph represents the change relative to 1984 (Texas graphs) or 1989-1990 (Houston area graphs). At both state and regional levels, the demand for STEM workers increased faster than the supply. The figures also show that Houston, compared to Texas, had a slightly higher demand for STEM workers, but a lower supply of them in 2015.



Notes: All series were six-year moving averages after model estimation at yearly frequencies. Please see Appendix I for additional details.

Appendix G. Change in Supply and Demand for Labor and Wage Premiums by Postsecondary Degree Type and STEM Occupation, 1984-2016

This table details changes in supply and demand for workers with associate's degrees, with bachelor's degrees or more, and in STEM occupations over different time periods between 1984 and 2016. It also contains the changes in wage premiums for these groups (relative to workers with no postsecondary credential or in a non-STEM occupation), allowing one to draw a connection between changes in supply, demand, and workers' wages. In most cases, when the demand for a group of workers grew faster than the supply, the wage premium for that group increased. For example, between 2014 and 2016, the demand for workers with a bachelor's degree or more increased by 63 percent in Texas, whereas the supply for that group increased by 11 percent. This corresponded to a 35 percent increase in the wage premium. In most cases, the reverse was also true: when changes in supply outpaced changes in demand, there was downward pressure on wages and the premium decreased. This could be seen in workers in STEM occupations in 2013-2016 in Houston, where the supply increased faster than demand, leading to a slight decrease in wage premiums.

PANEL A. TEXAS

	A.A. degree			*B.A. degree or more*			*STEM occupation*		
Period	Change in supply	Change in demand	Change in wages	Change in supply	Change in demand	Change in wages	Change in supply	Change in demand	Change in wages
1984-1989	19.70	-6.50	-17.50	17.00	-17.20	-22.90	7.30	-3.50	-7.20
1990-1995	-18.80	22.20	27.30	-9.20	41.80	34.00	-0.50	42.20	28.50
1996-2001	4.20	-6.20	-6.90	1.50	7.90	4.30	9.60	11.90	1.50
2002-2007	8.10	23.10	10.00	9.60	-2.80	-8.30	-3.10	-14.90	-7.90
2008-2013	6.10	26.40	13.50	18.70	57.80	26.10	6.30	35.50	19.50
2014-2016	9.80	45.60	23.90	11.20	63.10	34.70	0.40	-3.00	-2.20

PANEL B. HOUSTON AREA

	A.A. degree			*B.A. degree or more*			*STEM occupation*		
Period	Change in Supply	Change in Demand	Change in Wages	Change in Supply	Change in Demand	Change in Wages	Change in Supply	Change in Demand	Change in Wages
1989-1994	-0.20	-1.60	-0.90	12.10	43.60	21.00	0.30	0.30	0.00
1995-2000	-20.60	-3.80	11.20	-14.20	-8.00	4.20	-0.20	-32.90	-21.80
2001-2006	4.30	24.00	13.10	-9.70	13.40	15.40	4.00	41.40	25.00
2007-2012	15.90	24.00	5.40	9.30	49.30	26.60	4.90	55.30	33.60
2013-2016	13.40	-3.60	17.00	13.30	54.20	9.10	4.50	-15.40	-1.60

Notes: The statistics reported came from the same analyses as Figure 3 in Section I and Appendix F. Additional details are available in Appendix I.

Appendix H. OLS Regression Models Predicting Log Wages (College Completers Only)

As a robustness check, a supplementary model controlling for college degree major group was estimated. The sample size reduced to 3,537 cases because the analysis focused on students who earned a postsecondary credential within six years of high school. The results showed students who majored in engineering and engineering technology, business, and health care fields earned significantly more than students who majored in general studies, the reference category. Students who majored in the natural sciences, social sciences, and humanities earned lower wages than students who majored in general studies. Like the previous models, the analysis found females earned lower wages than males and blacks earned lower wages than whites. TAKS exemption status negatively predicted wages, while earning a bachelor's degree or higher positively predicted wages.

Variable	β	S.E.	Sig.
Female	-0.12	(0.03)	***
Race/ethnicity			
(ref. = White)			
Black	-0.18	(0.05)	**
Hispanic	-0.05	(0.06)	
Asian	-0.09	(0.07)	
Economically disadvantaged	0.00	(0.04)	
11th-grade composite TAKS score	0.04	(0.03)	
Exempt from TAKS	-0.36	(0.12)	**
Course grades (in 10s)	0.05	(0.03)	
Number of college-level credits	0.01	(0.01)	
Highest degree completed			
(ref. = Certificate/diploma)			
Associate's degree	-0.06	(0.07)	
Bachelor's degree	0.28	(0.06)	***
Master's/doctorate/prof. degree	0.52	(0.10)	***
Major group			
(ref. = General studies and other)			
Computer and information sciences	0.08	(0.13)	
Engineering and engineering technology	0.33	(0.08)	***
Biological, physical, and other natural sciences	-0.29	(0.06)	***
Social sciences	-0.34	(0.05)	***
Humanities	-0.24	(0.06)	***
Health care fields	0.13	(0.06)	*
Business	0.16	(0.05)	**
Education	0.00	(0.09)	
Other applied	-0.12	(0.06)	+
Intercept	8.50	(0.28)	***

Notes: From HERC multi-year data. Sample was limited to high school seniors in fall 2006-2008 who graduated from high school the following spring, were present in the wage data seven years after high school, had non-missing data on postsecondary attainment, and had a postsecondary credential. Native American respondents were excluded due to small sample size. The model included cohort fixed-effects and standard errors were clustered at the school level.
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001 (two-tailed tests)

Appendix I. Methods for Section I

This appendix details the empirical methodology used to examine gaps between supply and demand for workers with an associate's degree[1] or a bachelor's degree or more, relative to workers with a high school diploma or less. The theoretical framework relied on previous work by Goldin and Katz (2007, 2008), who analyzed the *skills gap*, or the shortfall between supply and demand for skilled workers. This study extended their methodology to analyze a *degrees gap*, or the shortfall between supply and demand for college-educated workers. The following subsections outline the steps taken to calculate this gap.

Theoretical link between the wage premium from postsecondary attainment and the relative supply and demand for workers

First, a constant elasticity of substitution (CES) production function was set up:

$$Q_t = [\alpha_t(a_t L_{NONE,t})^\rho + \beta_t(b_t L_{AA,t})^\rho + \delta_t(c_t L_{BA,t})^\rho]^{\frac{1}{\rho}}, \tag{1}$$

where Q_t was aggregate output; $L_{NONE,t}$, $L_{AA,t}$, and $L_{BA,t}$ were three production inputs referring to the amount of employed labor with a high school diploma or less, an associate's degree, and a bachelor's degree or more at time t; α_t, β_t , and δ_t were time-varying technology parameters that added up to one and could be interpreted as the fraction of activities allocated to the multiplying factor of production; and a_t, b_t, and c_t represented labor-augmenting technological change.

Taking the ratio of first-order conditions yielded the following equations:

$$\begin{aligned} \ln\left(\frac{w_{AA,t}}{w_{NONE,t}}\right) &= \frac{1}{\sigma}\left(\sigma \ln\left(\frac{\beta_t}{\alpha_t}\right) + (\sigma - 1)\ln\left(\frac{b_t}{a_t}\right)\right) - \frac{1}{\sigma}\ln\left(\frac{L_{AAt}}{L_{NONE,t}}\right), \\ \ln\left(\frac{w_{BA,t}}{w_{NONE,t}}\right) &= \frac{1}{\sigma}\left(\sigma \ln\left(\frac{\delta_t}{\alpha_t}\right) + (\sigma - 1)\ln\left(\frac{c_t}{a_t}\right)\right) - \frac{1}{\sigma}\ln\left(\frac{L_{BA,t}}{L_{NONE,t}}\right). \end{aligned} \tag{2}$$

The first term in parentheses on the right-hand side with the coefficient $\frac{1}{\sigma}$ was the relative demand variable D_t. Relative demand increased if there was factor-augmenting technological change or an increase in the intensity of factor use. For example, in the second equation in (2), both an increase in δ_t, which represented technological change biased in favor of workers with a bachelor's degree or more, or an increase in c_t, which represented the intensity with which workers with a bachelor's degree or more were employed in the production process, increased demand for workers with a bachelor's degree or more, relative to demand for workers with a high school diploma or less. The parameter σ represented the elasticity of substitution and was equal to $\frac{1}{1-\rho}$. This elasticity could be understood as the ease with which employed labor with a high school diploma or less could be substituted for labor with some type of postsecondary degree. If $\sigma > 1$, then labor with a high school diploma or less was sufficiently substitutable

[1] It was difficult to distinguish workers with postsecondary certificates or some postsecondary education from workers with associate's degrees in the data. These workers combined to form the associate's degree category in the analyses. Please contact the authors for additional details.

for labor with a postsecondary degree. If $\sigma < 1$, then the different types of labor inputs would be treated as complements rather than substitutes.

Theoretical link between the wage premium from STEM occupations and the relative supply and demand for workers

Production technology for the analysis of relative supply and demand for workers in STEM occupations was defined similarly:

$$Q_t = [\alpha_t^{'}(a_t^{'}L_{non-STEM,t})^{\rho^{'}} + \beta_t^{'}(b_t^{'}L_{STEM,t})^{\rho^{'}}]^{\frac{1}{\rho^{'}}}. \tag{3}$$

Taking the ratio of first-order conditions yielded following equation:

$$\ln\left(\frac{w_{STEM,t}}{w_{non-STEM,t}}\right) = \frac{1}{\sigma^{'}}\left(\sigma^{'}\ln\left(\frac{\beta_t^{'}}{\alpha_t^{'}}\right) + (\sigma^{'}-1)\ln\left(\frac{b_t^{'}}{a_t^{'}}\right)\right) - \frac{1}{\sigma^{'}}\ln\left(\frac{L_{STEM,t}}{L_{non-STEM,t}}\right),$$

where $\sigma^{'} = \frac{1}{1-\rho^{'}}$.

Primary system of regression equations

The set of equations in (2) and (3) led to the following system of regression equations:

$$\ln\left(\frac{w_{AA,t}}{w_{NONE,t}}\right) = \frac{1}{\sigma}D_{1,t} - \frac{1}{\sigma}\ln\left(\frac{L_{AA,t}}{L_{HS,t}}\right) + \epsilon_t, \quad \epsilon_t \sim N(0,\sigma_\epsilon^2),$$

$$\ln\left(\frac{w_{BA,t}}{w_{NONE,t}}\right) = \frac{1}{\sigma}D_{2,t} - \frac{1}{\sigma}\ln\left(\frac{L_{BA,t}}{L_{NONE,t}}\right) + \nu_t, \quad \nu_t \sim N(0,\sigma_\nu^2),$$

$$\ln\left(\frac{w_{STEM,t}}{w_{non-STEM,t}}\right) = \frac{1}{\sigma^{'}}D_{3,t} - \frac{1}{\sigma^{'}}\ln\left(\frac{L_{STEM,t}}{L_{non-STEM,t}}\right) + \xi_t, \quad \xi_t \sim N(0,\sigma_\xi^2), \tag{4}$$

where $D_{1,t} = \sigma\ln\left(\frac{\delta_t}{\alpha_t}\right) + (\sigma-1)\ln\left(\frac{c_t}{a_t}\right)$, $D_{2,t} = \sigma\ln\left(\frac{\delta_t}{\alpha_t}\right) + (\sigma-1)\ln\left(\frac{c_t}{a_t}\right)$ and $D_{3,t} = \sigma^{'}\ln\left(\frac{\beta_t^{'}}{\alpha_t^{'}}\right) + (\sigma^{'}-1)\ln\left(\frac{b_t^{'}}{a_t^{'}}\right)$ were relative demand terms capturing the increase in demand for labor with an associate's degree relative to no postsecondary degree, demand for labor with a bachelor's degree or more relative to no postsecondary degree, and demand for labor in STEM occupations relative to non-STEM occupations.

By estimating the system of regression equations in (4), gaps in relative supply and demand for labor in terms of postsecondary education and STEM occupations could be estimated. However, before estimating this system of equations, it was necessary to estimate additional equations in which relative supply and demand for labor were entered on the right-hand side and the natural log of the wage premium, the dependent variable, was entered on the left-hand side. After constructing appropriate measures of the dependent and independent variables in (4), the system was estimated such that the coefficient on the log of relative supply was equal across the first two estimating equations in (4). The final equation in (4) was for the analysis of relative supply and demand for STEM occupations. It was estimated once measures of the wage premium from STEM occupations, the dependent variable, and the relative supply for labor in efficiency units, the independent variable, were constructed.

Estimation of the wage premium for postsecondary attainment and STEM occupations

Workers were not assigned to the three educational categories or to STEM occupations at random. For example, workers with a bachelor's degree might have had certain characteristics that predisposed them to attending a four-year college. It became necessary to account for this selection bias before calculating the wage premium for postsecondary education and STEM occupations.

This subsection outlines two Roy models: (1) wage outcomes from completing a high school diploma or less, an associate's degree, or a bachelor's degree or more and (2) wage outcomes from working in STEM and non-STEM occupations. Let

$$X = [1\ \textit{gender white exper exper}^2]$$

denote the vector of covariates that entered the outcome equations predicting logarithmic wage, and

$$Z = [1\ \textit{gender white family size}]$$

denote the vector of covariates that entered the selection equations. The variable *exper* was defined as

$$\textit{potential years of experience} = \textit{age} - 6 - \textit{years of education by the given age}.$$

The utility accruing to an individual *i* from choosing different levels of postsecondary attainment was defined as

$$U^*_{DTYPE,i} = Z_i \gamma_{DTYPE} + \varepsilon^u_{DTYPE,i},$$

while the utility accruing to an individual *i* from choosing to work in a STEM or a non-STEM occupation was defined as

$$U^*_{STEM,i} = Z_i \gamma_{STEM} + \varepsilon^u_{STEM,i}.$$

The distributional assumption for the shocks, including the taste shocks ε^u_{DTYPE} and ε^u_{STEM} are given below. Let

$$d_{DTYPE,i} = \begin{cases} 1, & \text{No Post-secondary Degree (NONE)} \;\; \text{if} \;\; -\infty < U^*_{DTYPE,i} < c_1; \\ 2, & \text{if} \;\; \text{Associate's Degree or certificate (AA)} \;\; \text{if} \;\; c_1 \le U^*_{DTYPE,i} < c_2; \\ 3, & \text{if} \;\; \text{Bachelor's Degree or more (BA)} \;\; \text{if} \;\; c_2 \le U^*_{DTYPE,i} < \infty, \end{cases} \tag{5}$$

denote three possible choices that determined the postsecondary attainment of person *i*, where c_1 was normalized to 0. The choice set in (5) demonstrated the ordered nature of the problem. Let

$$d_{STEM,i} = \begin{cases} 1, & \text{non-STEM Occupation} \;\; \text{if} \;\; U^*_{STEM,i} < 0; \\ 2, & \text{if} \;\; \text{STEM Occupation} \;\; \text{if} \;\; U^*_{STEM,i} \ge 0 \end{cases} \tag{6}$$

denote individual *i*'s choice of working in a STEM or a non-STEM occupation.

The observed wage $w_{DTYPE,i}$ was $\ln w_{DTYPE,i} = \ln w_{NONE,i}\mathbb{I}(d_{DTYPE,i}=1) + \ln w_{AA,i}\mathbb{I}(d_{DTYPE,i}=2) + \ln w_{BA,i}\mathbb{I}(d_{DTYPE,i}=3)$, while the observed wage $w_{OCC,i}$ was $\ln w_{OCC,i} = \ln w_{non-STEM,i}\mathbb{I}(d_{STEM,i}=1) + \ln w_{STEM,i}\mathbb{I}(d_{STEM,i}=2)$.

The wage equation for the analysis of the wage premium from postsecondary attainment was

$$\ln w_{DTYPE,i} = \beta_{0j} + \beta_{1j} gender_i + \beta_{2j} white_i + \beta_{3j} exper_i + \beta_{4j} exper_i^2 + \varepsilon_j,$$

while the wage equation for the analysis of the wage premium from STEM occupations was

$$\ln w_{OCC,i} = \beta'_{0j} + \beta'_{1j} gender_i + \beta'_{2j} white_i + \beta'_{3j} exper_i + \beta'_{4j} exper_i^2 + \varepsilon'_j\,. \tag{7}$$

These two outcome equations could be written more compactly as $\ln w_{DTYPE,i} = X_i\beta_j + \varepsilon_j$ and $\ln w_{OCC,i} = X_i\beta'_j + \varepsilon'_j$.

Finally, let $\varepsilon = (\varepsilon^u_{DTYPE}, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ and $\varepsilon \sim Normal(0, \Sigma)$, where $\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{u1} & \sigma_{u2} & \sigma_{u3} \\ & \sigma_1^2 & 0 & 0 \\ & & \sigma_2^2 & 0 \\ & & & \sigma_3^2 \end{bmatrix}$.

Let $\varepsilon' = (\varepsilon^u_{STEM}, \varepsilon'_1, \varepsilon'_2)$ and $\varepsilon' \sim Normal(0, \Sigma')$, where $\Sigma' = \begin{bmatrix} \sigma_{u'}^2 & \sigma_{u'1'} & \sigma_{u'2'} \\ & \sigma_{1'}^2 & 0 \\ & & \sigma_{2'}^2 \end{bmatrix}$, where σ_u^2 and $\sigma_{u'}^2$ were set equal to 1.

Estimation of the Heckman selection model

The estimation proceeded in two steps. In step one, an ordered probit model was estimated. Choice probabilities were given by:

$$Pr(d_{DTYPE,i,t} = 1 | Z_{it}) = \Phi(-\frac{Z_{it}\gamma_{DTYPE}}{\sigma_u})$$

$$Pr(d_{DTYPE,i,t} = 2 | Z_{it}) = \Phi(\frac{c_2 - Z_{it}\gamma_{DTYPE}}{\sigma_u}) - \Phi(-\frac{Z_{it}\gamma_{DTYPE}}{\sigma_u})$$

$$Pr(d_{DTYPE,i,t} = 3 | Z_{it}) = 1 - \Phi(\frac{c_2 - Z_{it}\gamma_{DTYPE}}{\sigma_u}).$$

In step two, estimates of c_2 and γ from step one were used to construct the appropriate inverse Mills ratios, which were subsequently included as regressors for consistent estimation of the wage parameters β_j in equation (7). This proceeded according to the following derivation:

$$\mathbb{E}(\ln w_{NONE,i}|d_{DTYPE,i}=1,X_i,Z_i)=X_i\beta_1+\mathbb{E}(\epsilon_1|-\infty<\frac{\varepsilon^u}{\sigma_u}\leq-\frac{Z_i\gamma_{DTYPE}}{\sigma_u},X_i,Z_i)$$

Assume: $\epsilon_j=\rho_j\epsilon_u+\nu_j$, $\mathbb{E}(\epsilon_u|\nu_j)=0$ for $j=1,2,3$. Now, $\rho_j=\frac{Cov(\varepsilon^u,\varepsilon_j)}{Var(\varepsilon_j)}=\frac{\sigma_{j,u}}{\sigma_u^2}$. Then,

$$\mathbb{E}(\ln w_{NONE,i}|d_{DTYPE,i}=1,X_i,Z_i)=X_i\beta_1+\frac{\sigma_{1u}}{\sigma_u^2}\frac{\phi(\frac{Z_i\gamma_{DTYPE}}{\sigma_u})}{\Phi(-\frac{Z_i\gamma_{DTYPE}}{\sigma_u})}. \quad (8)$$

Similarly,

$$\mathbb{E}(\ln w_{AA,i}|d_{DTYPE,i}=2,X_i,Z_i)=X_i\beta_2+\frac{\sigma_{2u}}{\sigma_u^2}\frac{\phi(\frac{c_2-Z_i\gamma_{DTYPE}}{\sigma_u})-\phi(\frac{Z_i\gamma_{DTYPE}}{\sigma_u})}{\Phi(\frac{c_2-Z_i\gamma_{DTYPE}}{\sigma_u})-\Phi(\frac{Z_i\gamma_{DTYPE}}{\sigma_u})}.$$

and

$$\mathbb{E}(\ln w_{BA,i}|d_{DTYPE,i}=3,X_i,Z_i)=X_i\beta_3-\frac{\sigma_{3u}}{\sigma_u^2}\frac{\phi(\frac{c_2-Z_i\gamma_{DTYPE}}{\sigma_u})}{1-\Phi(\frac{c_2-Z_i\gamma_{DTYPE}}{\sigma_u})}.$$

The two-step estimation of the ordered probit model first accounted for selection into different levels of postsecondary attainment, then calculated the wage premium to postsecondary attainment. These estimates were made annually for Texas and biannually for the Houston area. All estimates incorporated survey weights. A similar two-step estimation process was repeated in the analysis of STEM occupations.

Once the two-step Heckman procedure was run — annually at the state level and biannually at the metropolitan statistical area (MSA) level from 1979-2016 — the wage premium for the two postsecondary attainment categories, relative to high school or less, and for STEM occupations, relative to non-STEM occupations, was calculated as follows:

$$\mathbb{E}[\ln w_{BA,t}-\ln w_{NONE,t}|X=\overline{x}]=(\hat{\beta}_3-\hat{\beta}_1)\overline{x}$$

$$\mathbb{E}[\ln w_{AA,t}-\ln w_{NONE,t}|X=\overline{x}]=(\hat{\beta}_2-\hat{\beta}_1)\overline{x},$$

$$\mathbb{E}[\ln w_{STEM,t}-\ln w_{non-STEM,t}|X=\overline{x}]=(\hat{\beta}'_2-\hat{\beta}'_1)\overline{x}.$$

Forecasts and prediction intervals from 2017-2030 were calculated following this estimation.

Family size as an exclusion restriction in the Heckman selection model

Appendix A shows workers with bachelor's degrees or more had, on average, smaller family sizes than workers with associate's degrees, who in turn had smaller family sizes than workers with a high school diploma or less. The same relationship is observed in Appendix B between STEM and non-STEM workers. Information on family size was used an exclusion restriction in model estimation, assuming workers with lower levels of postsecondary attainment or in non-STEM occupations were more likely to come from large families. Family size was assumed to have an indirect, rather than a direct, relationship with wages. The inverse relationship between family size and educational attainment is documented in sociological literature (Blake, 1989; Steelman, Powell, Werum, & Carter, 2002), arguing that parents with more children likely devote fewer resources per child, pointing toward a quantity-quality tradeoff.

Estimation of supply and demand for postsecondary attainment and STEM occupations

Measures of relative supply by postsecondary attainment were constructed in efficiency units, a standard in the literature (e.g., Acemoglu & Autor, 2012; Autor, Katz, & Kearney, 2006; Katz & Murphy, 1992). Construction of relative supply in efficiency units essentially adjusted the total hours of work supplied by a given employee in a given year by the human capital accumulation of that person. Let

$$Y_{ij} = r_j H_i L_{ij} \;\Rightarrow\; \ln Y_{ij} = \ln r_j + \ln L_{ij} + lnH_i,$$

where Y_{ij} was the annual wage of individual i with postsecondary degree type or STEM occupational status j, r_j was the rental rate of labor to a worker with postsecondary degree type or STEM occupational status j, L_{ij} was the total hours of work individual i supplied in a year when working in category j, and H_i was the human capital of individual i. Individual i's annual wage and total hours worked were observed in the March Current Population Survey (CPS) data, but H_i was not. However, it was possible to proxy for H_i such that $H_i = \exp\{\beta_1 exper_i + \beta_2(exper_i)^2\}$, where *exper* referred to potential years of experience.

Substituting the proxy for human capital into the expression for annual income gave

$$\ln Y_{ij} = \ln r_j + \ln L_{ij} + \beta_1 exper_i + \beta_2(exper_i)^2 + u_{ij}.$$

Thus, an individual with postsecondary degree type j (none, associate's degree, or bachelor's degree or more) or occupational status j (STEM or non-STEM) supplied labor in efficiency units measured by $L_{ij}H_i$. An estimate of H_i was given by

$$\ln \hat{H}_i = \hat{\beta}_1 exper_i + \hat{\beta}_2(exper_i)^2.$$

It became possible to obtain β_1 and β_2 as sample selection-adjusted coefficients on quadratic experience using the methodology discussed earlier. Steps one and two in this context included period dummies. $L_{ij}\hat{H}_i$ was obtained by taking the exponent of both sides in the previous expression for the log of human capital, then multiplying by total hours of work supplied by individual i in category j in a given year. In this manner, $L_{i,NONE,t}\hat{H}_{i,t}$, $L_{i,AA,t}\hat{H}_{i,t}$, $L_{i,BA,t}\hat{H}_{i,t}$, $L_{i,NONSTEM,t}\hat{H}_{i,t}$, and $L_{i,STEM,t}\hat{H}_{i,t}$ were constructed for each individual at time t, given the observed postsecondary degree type and occupational status.

Following the construction of labor supply in efficiency units for each observation, data were aggregated annually for Texas and biannually for the Houston area. In order to construct relative labor supply in efficiency units, aggregates of the labor supply with associate's degrees and bachelor's degrees or more were divided by aggregates of the labor supply of the base category, high school diploma or less. The result was then logged. A similar process was used for the STEM occupation analysis. Thus,

$$\ln\left(\frac{L_{AA,t}}{L_{NONE,t}}\right) = \ln\left(\frac{\sum_{i\in\ AA,t} L_{i,AA,t}\hat{H}_{i,t}}{\sum_{i\in\ NONE,t} L_{i,NONE,t}\hat{H}_{i,t}}\right)$$

$$\ln\left(\frac{L_{BA,t}}{L_{NONE,t}}\right) = \ln\left(\frac{\sum_{i\in\ BA,t} L_{i,BA,t}\hat{H}_{i,t}}{\sum_{i\in\ NONE,t} L_{i,NONE,t}\hat{H}_{i,t}}\right)$$

$$\ln\left(\frac{L_{STEM,t}}{L_{non-STEM,t}}\right) = \ln\left(\frac{\sum_{i\in\ STEM,t} L_{i,STEM,t}\hat{H}_{i,t}}{\sum_{i\in\ non-STEM,t} L_{i,non-STEM,t}\hat{H}_{i,t}}\right) \tag{9}$$

Having constructed series of relative labor supply in efficiency units, six-year moving averages were calculated for use in graphs.

Relative demand was not observed in the CPS data. The following equations were used to estimate the demand for labor with an associate's degree, relative to high school or less, and the demand for labor with a bachelor's degree or more, relative to high school or less:

$$\hat{D}_{1,t} = \sigma \ln\left(\frac{w_{AA,t}}{w_{NONE,t}}\right) + \left(\frac{L_{AA,t}}{L_{HS,t}}\right);$$

$$\hat{D}_{2,t} = \sigma \ln\left(\frac{w_{BA,t}}{w_{NONE,t}}\right) + \left(\frac{L_{BA,t}}{L_{HS,t}}\right).$$

An estimate of the demand for labor in STEM occupations, relative to non-STEM occupations, was given by:

$$\hat{D}_{3,t} = \sigma' \ln\left(\frac{w_{STEM,t}}{w_{non-STEM,t}}\right) + \left(\frac{L_{STEM,t}}{L_{non-STEM,t}}\right).$$

These equations were a rearrangement of the equations in (4). Estimation of the wage premium series (adjusted for sample selection) and the relative labor supply series (in efficiency units) at the Texas and Houston-area levels were described earlier. The only unknowns were the elasticity of substitution parameters, σ and σ'. These parameters were calibrated to other studies like Katz and Murphy (1992) and Ciccone and Peri (2005), and were typically shown to lay between 1.5 and 3, with 1.5 as the preferred estimate of the elasticity of substitution between skilled and unskilled labor. Relative demand schedules at both geographical levels for different calibrations of σ and σ' are available upon request.

Reduced-form estimation of the wage premium

By estimating the following reduced-form model, the wage premium without adjusting for sample selection was obtained. For the analysis of postsecondary attainment, log hourly wage was regressed on dummy variables indicating whether an individual completed an associate's degree, completed a bachelor's degree or more, was female, and was non-white. For the analysis of STEM occupations, log hourly wage was regressed on dummy variables indicating whether an individual worked in a STEM occupation, was female, and was non-white.

Sample selection-unadjusted results came from the following equations:

$$\begin{aligned}\ln w_{i,t_\ell} &= \beta_0 + \beta_1 \mathbb{I}(AA)_{i,t_\ell} + \beta_2 \mathbb{I}(BA)_{i,t_\ell} \\ &+ \beta_3 exper_{i,t_\ell} - \beta_4 exper^2_{i,t_\ell} + \beta_5 gender_{i,t_\ell} + \beta_6 white_{i,t_\ell} + \epsilon, \quad \epsilon \sim N(0,\sigma^2_\epsilon),\end{aligned} \tag{10}$$

and

$$\ln w_{i,t_\ell} = \beta_0^{'} + \beta_1^{'} \mathbb{I}(STEM)_{i,t_\ell} + \beta_2^{'} exper_{i,t_\ell} - \beta_3^{'} exper^2_{i,t_\ell} + \beta_4^{'} gender_{i,t_\ell} + \beta_5^{'} white_{i,t_\ell} + \epsilon^{'}, \quad \epsilon \sim N(0,\sigma^2_{\epsilon^{'}}) \tag{11}$$

where ℓ = {1979-82, 1983-86, 1987-1990, 1991-1994, 1995-98, 1999-2002, 2003-06, 2007-10, 2011-14, 2015−16}. The equations were estimated by pooling March CPS subsamples at four-year intervals for Texas and the Houston area without accounting for selection bias. In (10), β_1 × 100% was the expected percentage hourly wage differential of workers who had an associate's degree, relative to those with no postsecondary degree, while β_2× 100% was the expected percentage hourly wage differential of workers who had a bachelor's degree or more, relative to those with no postsecondary degree. In (11), β_1 × 100% was the expected percentage hourly wage differential of STEM workers relative to non-STEM workers.

Figures I1-I3 show the sample selection-unadjusted estimates of wage premiums in Texas and the Houston area.

Figure I1: Sample selection-unadjusted estimates of the wage premium from postsecondary attainment relative to high school or less in Texas, 1979-2016

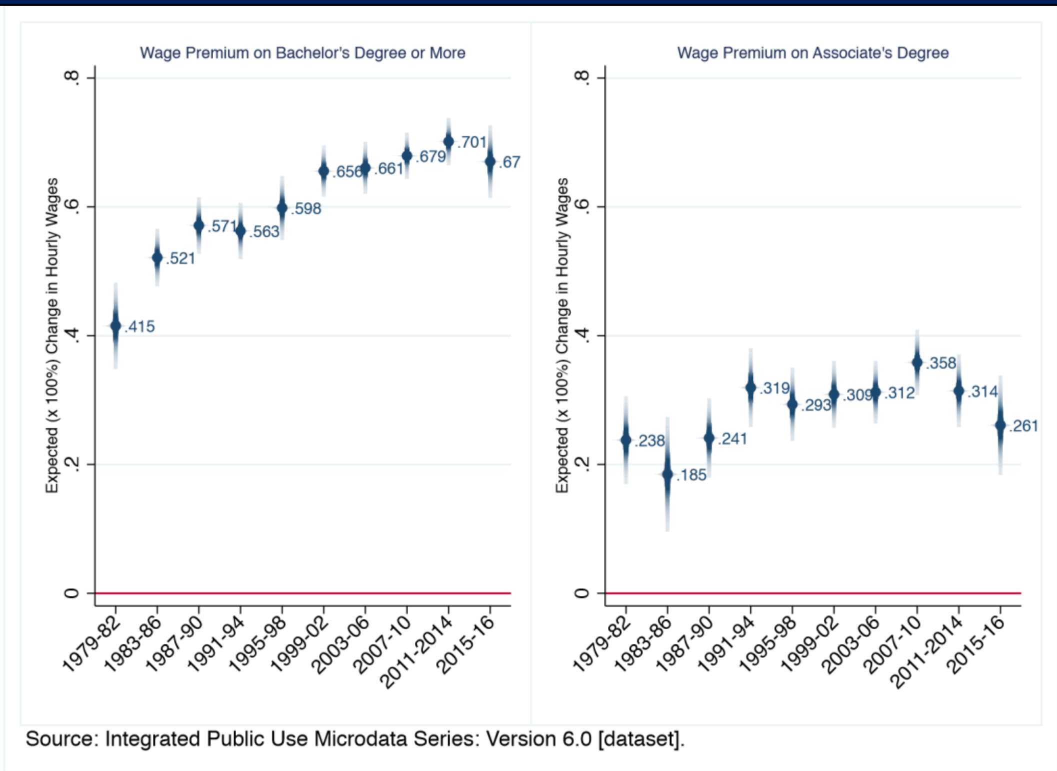
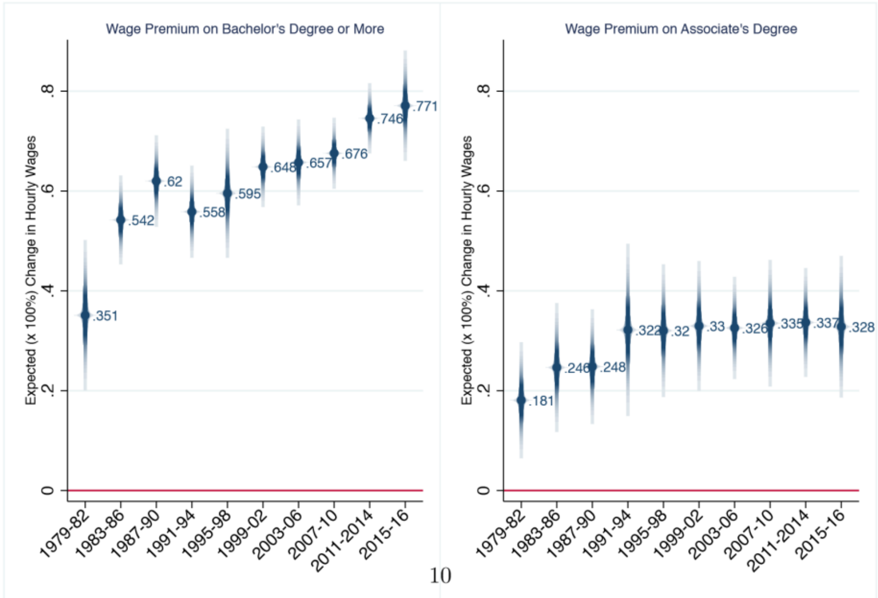
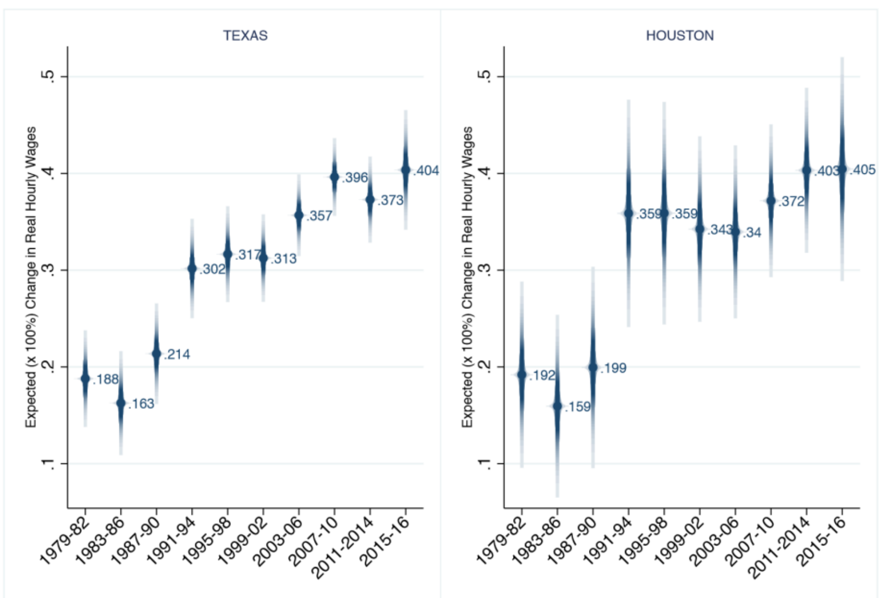


Figure I2: Sample selection-unadjusted estimates of the wage premium from postsecondary attainment relative to high school or less in the Houston area, 1979-2016



Source: Integrated Public Use Microdata Series: Version 6.0 [dataset].

Figure I3: Sample selection-unadjusted estimates of the wage premium from STEM occupations relative to non-STEM occupations in Texas and the Houston area, 1979-2016



Source: Integrated Public Use Microdata Series: Version 6.0 [dataset].

March CPS 1979-2016 data processing

March CPS data from 1979-2016 were processed following the steps taken in Autor, Katz and Kearney (2006). Top-coded wage values were multiplied by 1.5. Wage values were also deflated by the 1999 Consumer Price Index. All analyses were limited to individuals 25-55 years old who were employed, worked at least 40 weeks in the previous calendar year, and worked 35 to 45 hours per week. Individuals whose potential experience was shown to be negative or greater than 39 were considered outliers and dropped from the sample. Survey weights were incorporated in summary statistics, regression analyses, and maximum likelihood estimation.

Labor force composition by postsecondary attainment and STEM occupations, 1979-2016 and projections, 2017-2030

This subsection describes the decomposition of the labor force in Texas and the Houston area by postsecondary degree attainment and participation in STEM occupations. From 1979-2016, the data were used to estimate the percentage of workers with different levels of postsecondary attainment and the percentage of workers in STEM occupations. From 2017-2030, these estimates were forecasted, and might be subject to error. In addition to the main series, additional series examined trends by gender and race/ethnicity. All analyses were limited to workers 25-34 years old. To reduce statistical noise, a six-year moving average of the series was used in graphs. The forecasting exercise relied on the following $AR(p)$ model:

$$y_t = \rho_0 + \sum_{j=1}^{p} \rho_j y_{t-j} + \epsilon_t, \quad \forall\, t = 1984, 1985, ..., \tag{12}$$

[2] where, the stochastic process $\{y_t\}$ was defined as the six-year moving average series of the percentage of the labor force with an associate's degree, with a bachelor's degree or more, or working in a STEM occupation. The lag order, p, of the AR process for each of these series was determined by relying on standard diagnostic checks, which are available from the authors upon request.

Table I1. Lag orders of AR processes used at the Texas and Houston-area levels, 1979-2016

Series	Texas			Houston Area		
	State	Females	Non-whites	MSA	Females	Non-whites
Associate's degree	2	2	3	1	1	6
Bachelor's degree or more	3	3	3	2	2	3
STEM occupations	3	3	1	3	2	1

[2] Observations for the years 1979-1983 were dropped when calculating six-year moving averages.

Lists of high-skill and middle-skill STEM occupations based on the BLS SOC 2010 classification

The STEM occupations category relied on combining occupations that were high- or middle-skill STEM occupations based on the Bureau of Labor Statistics (BLS) Standard Occupational Classification (SOC) system, 2010.

Table I2. High-skill STEM occupations list

Code	Description
110	Computer and Information Systems Managers
300	Architectural and Engineering Managers
360	Natural Science Managers
1000	Computer Scientists and Systems Administrators
1010	Computer Programmers
1020	Software Developers, Applications
1060	Database Administrators
1100	Network and Computer Systems Administrators
1200	Actuaries
1220	Operations Research Analysts
1230	Statisticians
1240	Mathematical science occupations
1300	Architects, Except Naval
1310	Surveyors, Cartographers, and Pho
1320	Aerospace Engineers
1350	Chemical Engineers
1360	Civil Engineers
1400	Computer Hardware Engineers
1410	Electrical and Electronics Engineers
1420	Environmental Engineers
1430	Industrial Engineers, including H
1440	Marine Engineers and Naval Archit
1450	Materials Engineers
1460	Mechanical Engineers
1520	Petroleum, mining and geological
1530	Engineers, nec
1600	Agricultural and Food Scientists
1610	Biological Scientists
1640	Conservation Scientists and Fores
1650	Medical Scientists and Life Scientists
1700	Astronomers and Physicists
1710	Atmospheric and Space Scientists
1720	Chemists and Materials Scientists
1740	Environmental Scientists and Geoscientists
1760	Physical Scientists, nec
3010	Dentists
3040	Optometrists
3050	Pharmacists
3060	Physicians and Surgeons
3250	Veterinarians
3410	Health Diagnosing and Treating Practitioners, All Other
4930	Sales Engineers
4840	Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products

Table I3. Middle-skill STEM occupations list

Code	Description
205	Farmers, Ranchers, and Other Agricultural Managers
1050	Computer Support Specialists
1540	Drafters
1550	Engineering Technicians, Except Drafters
1560	Surveying and Mapping Technicians
1900	Agricultural and Food Science Technicians
1910	Biological Technicians
1920	Chemical Technicians
1930	Geological and Petroleum Technicians
1960	Life, Physical and Social Science Technicians, All Other
2900	Broadcast and Sound Engineering Technicians
3130	Registered Nurses
3150	Occupational Therapists
3160	Physical Therapists
3200	Radiation Therapists
3220	Respiratory Therapists
3260	Health Diagnosing and Treating Practitioners
3300	Clinical Laboratory Technologists
3310	Dental Hygienists
3320	Diagnostic Related Technologists
3400	Emergency Medical Technicians and
3410	Health Diagnosing and Treating Practitioners
3500	Licensed Practical and Licensed Vocational Nurses
3510	Medical Records and Health Inform
3520	Opticians, Dispensing
3530	Health Technologists and Technicians
3540	Healthcare Practitioners and Technicians
3610	Occupational Therapy Assistants
3620	Physical Therapist Assistants
3640	Dental Assistants
3650	Medical Assistants and Other Heal
4010	First-Line Supervisors of Food Preparation and Serving Workers
6005	First-Line Supervisors of Farming
6120	Forest and Conservation Workers
6355	Electricians
7000	First-Line Supervisors of Mechanics
7010	Computer, Automated Teller, and Office Machine Repairers
7020	Radio, Cellular, and Tower Equipment Installers and Repairs
7030	Avionics Technicians
7040	Electric Motor, Power Tool, and Repairers
7100	Electrical and electronics repair
7110	Electronic Equipment Installers a
7120	Electronic Home Entertainment Equipment Installers and Repairers
7125	Electronic Repairs, nec
7140	Aircraft Mechanics and Service Technicians
7200	Automotive Service Technicians an
7240	Small Engine Mechanics
7260	Vehicle and Mobile Equipment Mechanics

Table I3. Middle-skill STEM occupations list (cont.)

Code	Description
7360	Millwrights
7720	Electrical, Electronics, and Elec
7900	Computer Control Programmers and
8030	Machinists
8140	Welding, Soldering, and Brazing Workers
8250	Prepress Technicians and Workers
8630	Plant and System Operators, nec
9030	Aircraft Pilots and Flight Engineers
9410	Transportation Inspectors

Appendix J. Methods for Section II

First, Occupational Employment Statistics (OES) data from the U.S. Bureau of Labor Statistics were used. OES data listed the employment level and average annual wage for each occupation. For the analysis, the research team focused on occupations in the Houston metropolitan area in two periods, 2005-2007 and 2014-2016. Across these early and late periods, three-year averages of employment and wages for each occupation were calculated. The rate of change in employment and wages across the early and late periods was then determined, denoted as Δ_{job} and Δ_{wage}, respectively. Among all occupations, average rates of change in employment and wages were calculated, denoted as $\overline{\Delta_{job}}$ and $\overline{\Delta_{wage}}$. Each occupation was classified into four groups based on Δ_{job} and Δ_{wage} and their relationships to the average rates of change $\overline{\Delta_{job}}$ and $\overline{\Delta_{wage}}$. In the classification, occupations with job growth greater than or equal to the average rate of job growth were defined as "high supply growth rate" and occupations with job growth lower than the average rate of job growth were defined as "low supply growth rate." Occupations with wage growth greater than or equal to the average rate of wage growth were "high demand growth rate" and occupations with wage growth lower than the average rate of wage growth were "low demand growth rate." Figure J1 illustrates the classification schema. The low supply, high demand category, highlighted in red, was the focus of this analysis as it illustrated a gap between supply and demand. These occupations appeared to be increasingly needed by the Houston-area economy, and due to a shortage of individuals in these occupations, economic returns showed a high growth rate.

Figure J1. Growth of supply and demand job classification

	$\Delta_{job} < \overline{\Delta_{job}}$	$\Delta_{job} \geq \overline{\Delta_{job}}$
$\Delta_{wage} < \overline{\Delta_{wage}}$	Low supply growth rate, low demand growth rate	High supply growth rate, low demand growth rate
$\Delta_{wage} \geq \overline{\Delta_{wage}}$	Low supply growth rate, high demand growth rate	High supply growth rate, high demand growth rate

In an effort to facilitate comparisons and understand patterns, occupations were summarized using the nine-category job classification developed by the U.S. Equal Employment Opportunity Commission (EEOC):

1. Officials and managers (e.g., chief executives, sales managers)
2. Professionals (e.g., accountants, engineers)
3. Technicians (e.g., dental hygienists, pharmacy technicians)
4. Sales workers (e.g., cashiers, sales representatives)
5. Administrative support workers (e.g., legal secretaries, office clerks)
6. Craft workers (e.g., carpenters, machinists)
7. Operatives (e.g., parking lot attendants, taxi drivers)
8. Laborers and helpers (e.g., painters, plumbers)
9. Service workers (e.g., bartenders, waiters)

Using the supply and demand growth rate and EEOC classifications, the typical education required, average annual wage, and primary skills needed for each job category were described. Education and wage data came from the Texas Workforce Commission (TWC), a state agency that provides workforce development services to job seekers and employers. For each occupation, TWC listed the typical education required and average annual wage in 2015. The table for the Gulf Coast Workforce Development Area, which included Houston and its environs, was used in this analysis. The level of education required was reported in degrees, but was converted to years of education completed to simplify the analysis.[3]

To determine the primary skills needed for each occupation, an approach similar to the one advanced by economists Daron Acemolgu and David Autor (2011) was developed. First, data on abilities, work activities, and work contexts were downloaded from the Occupational Information Network (O*NET) 22.1 Database. O*NET is a database comprised of occupational definitions and information to promote greater understanding of work in the U.S. O*NET collects survey information from employees and trained job analysts about individual occupation characteristics. Employees and analysts rate each characteristic on a 1-5 Likert scale indicating how important a characteristic is to the occupation.[4,5]

Table J1. Abilities, work activities, and work contexts

Abilities: enduring attributes of the individual that influence performance
Spatial orientation
Manual dexterity
Work activities: general types of job behaviors occurring on multiple jobs
Analyzing data/information
Thinking creatively
Controlling machines and processes
Operating vehicles, mechanized devices, or equipment
Interpreting information for others
Establishing and maintaining personal relationships
Guiding, directing and motivating subordinates
Coaching/developing others
Work contexts: physical and social factors that influence the nature of work
Spend time using hands to handle, control or feel objects, tools or controls
Spend time making repetitive motions
Importance of being exact or accurate
Importance of repeating the same tasks
Structured v. unstructured work
Pace determined by speed of equipment

Source: Occupational Information Network website.

[3] Levels of education were converted to years of education in the following way: no formal educational credential (11 years); high school diploma or equivalent (12 years); postsecondary non-degree award (13 years); some college, no degree (13 years); associate's degree (14 years); bachelor's degree (16 years); master's degree (18 years); and doctoral or professional degree (22 years).
[4] Employees did not provide ratings of abilities; only analysts did. Both employees and analysts rated work activities and contexts; employee ratings had the most complete data and were used in the analysis.
[5] The characteristics analyzed had scales which indicated the importance or level required for the job. The importance and level scales were highly correlated, so in the analysis, importance scales were used.

The items listed in Table J1, which Acemolgu and Autor used to develop composite measures of occupational skills, were selected for the analyses. Using the occupation-level dataset with ratings for each of the selected abilities, work activities, and work contexts, an exploratory factor analysis revealed four separate factors.[6] Four skill indices were created by factor scoring and labeled: 1) blue-collar; 2) white-collar: routine; 3) white-collar: non-routine, analytical; and 4) white-collar: non-routine, interpersonal. These indices were similar to the skills described by Acemolgu and Autor and are defined in Table J2. The indices were not practically interpretable except higher scores meant a particular skill was more common within an occupation, while lower scores meant it was less common.

Table J2. Occupational skills definit ons

Blue-collar	• Required low levels of education • Was physically-demanding
White-collar: routine	• Required low-to-medium levels of education • Involved problem-solving and repetitive activities
White-collar: non-routine, analytical	• Required medium-to-high levels of education • Involved problem-solving and mathematical/formal reasoning
White-collar: non-routine, interpersonal	• Required medium-to-high levels of education • Involved problem-solving and in-person interactions/management

[6] Factor loadings and scores are available from the authors upon request.

Appendix K. Methods for Section III

Three restricted-use data sources were used for the analyses: HERC multi-year data, National Student Clearinghouse (NSC) data, and Texas Workforce Commission (TWC) data. Raw data files were provided by the Houston Independent School District (HISD) and used to build a dataset of high school seniors in the 2006-2007 through 2008-2009 school years. Students were tracked longitudinally, allowing measurement of labor market outcomes seven years after high school graduation.

HERC data provided measures of student characteristics, which are described later in the section. Measures of educational attainment came from the NSC, an organization that collects information on college enrollment and completion. The variables of interest — wages and unemployment insurance receipt — were made available through the TWC.

Although the initial dataset included more than 27,000 students, the analytical samples were much smaller. Table K1 summarizes sample restrictions for each analysis.

Table K1. Analytic samples for section III

Restriction	*Limited to high school seniors in fall 2006-2008*	*Limited to students who graduated high school in spring 2007-2009*[a]	*Limited to students working in Texas*[b]	*Limited to students working in Texas or who received unemployment insurance*[c]	*Limited to students with data on postsecondary attainment*	*Limited to non-Native American students*[d]
Summary statistics of wages (N = 12,434)	X		X			X
Regression models predicting wages (N = 10,996)	X	X	X		X	X
Summary statistics of unemployment insurance (N = 12,497)	X			X		X

[a] This restriction was necessary because NSC data were available for high school graduates only. While some seniors might have graduated later, this could not be observed completely: no data was available for students who completed a Certificate of High School Equivalency outside HISD.

[b] The wage data was restricted to individuals working in Texas (i.e., excluded people who were not in the labor force, unemployed, living outside the state).

[c] The sample for the unemployment insurance analysis included individuals in the wage data as well as individuals in the unemployment insurance data.

[d] There were too few Native American students to produce precise estimates for that subgroup.

A special note on the unemployment insurance analysis

Unemployment insurance receipt is not the same as unemployment. Not all unemployed persons file claims for unemployment insurance. In addition, people may still be unemployed once their unemployment benefits expire. Therefore, unemployment insurance receipt is an underestimate of unemployment overall.

Only 43 percent of the sample was working in Texas seven years after high school. The remaining 57 percent included 1) students who continued living in Texas and were either unemployed or not in the labor force (e.g., in school, stay-at-home parents) or 2) students who were living outside the state.

For the unemployment insurance analysis, summary statistics were produced. The sample size was too small to use multivariate regression (i.e., too few students received unemployment insurance).

Since one-quarter of the students in the dataset had missing information on at least one covariate, missing values were filled in using multiple imputation with chained equations. The statistics reported were averages across 10 imputed datasets, adjusting means, coefficients, and standard errors accordingly.

The key dependent variables in the analyses were early career wages and unemployment insurance receipt seven years after high school. These measures were pulled from the second fiscal quarter. To ease interpretation in the summary statistics table, second quarter wages were multiplied by four to approximate an annual wage. In regression analysis, the natural log of second quarter wages was used to normalize the wage distribution. Regression coefficients could, therefore, interpreted as percent changes in quarterly wages. The unemployment insurance variable was binary and measured whether a student received benefits.

Summary statistics of wages and unemployment insurance for the whole sample and by gender (male, female), race/ethnicity (white, black, Hispanic, Asian), and economic disadvantage (no, yes) were produced. These variables were also included in the wage regression models. The models controlled for three measures of academic performance: 11th-grade composite test scores, grades across all courses taken in the 12th grade, and the number of college-level credits earned in the 12th grade. To generate the composite test score variable, reading, mathematics, science, and social studies scores from the Texas Assessment of Knowledge and Skills (TAKS) were averaged. The measure was reported in standard deviation units. A number of students, mostly those in special education, were exempt from the TAKS, so a binary indicator that accounted for those individuals was included in the models. The course grades variable showed the average percentage grade among all courses taken in the 12th grade, while the college-level credits variable showed the number of credits earned in Advanced Placement, International Baccalaureate, and academic dual enrollment courses.[7] In terms of postsecondary attainment, a categorical variable that measured the most advanced credential earned within six years of high school was included: no postsecondary credential, certificate/diploma, associate's degree, bachelor's degree,

[7] Career & Technical Education dual enrollment courses were excluded from this measure. Each semester-long course that a student passed (grade of 69.5 and above) counted as 0.5 credits.

and master's/doctorate/professional degree. All statistical models controlled for 12th grade cohort fixed-effects (i.e., 2006-2007, 2007-2008, 2008-2009).

Finally, in a robustness check, the sample was limited to students with a postsecondary credential and controlled for their college major (see Table K2); these results are reported in Appendix H.

Table K2. College major groups
Computer and information sciences
Engineering and engineering technology
Biological and physical sciences, science technology, mathematics, and agricultural sciences
General studies and other
Social sciences
Humanities
Health care fields
Business
Education
Other applied

Note on the authors: Mehreen Gul, Ph.D. is currently vice president in the financial forecasting division at Citigroup.

About HERC. Focusing on the most pressing challenges facing the region, the Houston Education Research Consortium (HERC) is a research-practice partnership between Rice University and 11 Houston-area school districts. HERC research is developed directly alongside district leaders with findings shared with decision makers – culminating in long-term, equity-minded solutions, opportunities and growth for Houston and beyond.

RICE | HERC
Houston Education Research Consortium

Houston Education Research Consortium
a program of the Kinder Institute for Urban Research
MS-258 Rice University | Houston, Texas 77005
713-348-2532 | herc@rice.edu
kinder.rice.edu/houston-education-research-consortium