Tuition Subsidies and Overeducation

Ciprian Domnisoru
Aalto University School of Business, Helsinki GSE

Aug 17, 2021

Overeducation: policy context and motivation

- High fraction of college graduates working in jobs that typically do not require a bachelor's degree: "Overeducation"/"Education-Occupation Mismatch"/ "Underemployment" rate (NY Fed: 34%, 2019).
- Per capita undergraduate aid per year (grants, guaranteed loans) increased from 5,000$ in 1990 to about 15,000 $ in 2016 dollars. *"A student who leaves college without learning the skills required for a job in his or her field of study does not offer the same benefit to the economy—and the tax base—as a skilled graduate."* *(U.S. Senate, 2012)*
- What is the effect of tuition subsidies on the rate of overeducation *(and earnings)*?
- What explains the persistently high overeducation rate in the U.S. ?

Subsidies Welfare Persistence

Overeducation: types of indicators

- Standard Occupational Classification (SOC) level- 459 occupations.
- **Objective:** Bureau of Labor Statistics information on "typical education needed for entry," 2014-2024 Employment Matrix [43.20 %]
- **Subjective**: O*NET Education and Training Questionnaire (2003-2015): *"If someone were being hired to perform this job, indicate the level of education that would be required. (Note that this does not mean the level of education that you personally have achieved.)"*- overeducated if > 50 % of respondents in occupation indicate that a BA is not necessary (Abel et al. ,2014).[30.29 %]
- **Statistical**: overeducated if individual has 16 years of schooling or more, but average level of education in his/her occupation$+$ 1 s.d. falls below 16 years of schooling. [24.58 %]
- **Joint** : intersection of the above three indicators [23.00%]

Calculations in parantheses : US-born citizens holding a bachelor's degree, who are not enrolled in school. [2000-2014 American Community Survey, ages 24-64]

Approach and main findings

- Stylized model of sectoral choice featuring Type I EV shocks where *P(noncollege sector* | *college degree)* has a closed form solution.
- Motivating evidence from the WWII, (Vietnam) and Post 9/11 G.I.Bills
- Dynamic structural model
  1. Counterfactual tuition subsidy increases results in higher overeducation rates.
  2. Model matches the empirical overeducation rate and the declining pattern of overeducation in early career stages.
  3. The model suggests several policy actionable mechanisms for reducing overeducation.

Dynamic discrete choice models of occupational choice: Miller (1984), Keane and Wolpin (1997), Sullivan (2010), James (2011), Arcidiacono, Aucejo, Maurel and Ransom (2016), Todd and Zhang (2019). Estimation methods: Hotz and Miller (1993), Arcidiacono and Miller (2011), Arcidiacono and Miller (2019).

Dynamic structural model

Beginning at age 16, individuals choose among a set of mutually exclusive and exhaustive options, with the goal of maximizing lifetime utility:

$$\max_{c_a} E\left[\sum_{a=16}^{65} \beta^{a-16}(u[s_a, c_a] + \epsilon[c_a])\right]$$

Their decision updates the state space (s_a includes includes schooling, work experience, previous period choices), and the process is repeated at age 17 and thereafter. The choices are:

- continuing K-12 education
- enrollment in 2-year college
- enrollment in 4-year college in a STEM, Healthcare or Education major
- enrollment in 4-year college in a Liberal Arts, Social Sciences or Business major
- enrollment in 4-year college, undeclared major
- enrollment in graduate school
- employment in the college sector
- employment in the non-college sector
- home production

Variable construction .

Choice frequencies by age, men, NLSY97

Age	High school	4 yr, STEM-HE	4 yr, Non-STEM	No major	Grad sch.	2-yr coll.	College sector	Non-college sector	Home prod.
16	95.32	0	0	0	0	0.04	0.39	0.72	3.54
17	88.22	0.04	0.04	0.07	0	0.07	1.11	3.01	7.45
18	36.27	5.07	4.37	6.55	0	8.4	6.14	15.84	17.36
19	2.6	9.16	9.9	6.49	0	11.72	9.87	26.67	23.59
20	0.11	9.3	12.59	2.69	0	7.68	12.48	31.08	24.07
21	0.04	7.89	13.29	0.82	0	5.17	15.08	32.65	25.06
22	0.04	5.5	7.75	0.37	0.37	3.29	19.99	37.18	25.5
23	0	2.69	4.1	0.19	0.76	2.28	25.65	41.18	23.15
24	0	1.39	2.12	0.15	1.27	2.27	29.48	40.8	22.52
25	0.04	0.89	1.28	0.04	1.98	2.14	32.06	39.77	21.79
26	0	0.47	0.82	0.08	2	1.69	32.4	38.88	23.66
27	0	0.52	1.53	0.04	1.41	1.85	33.87	37.89	22.89
28	0	0.31	1	0.1	1.41	1.73	34.57	39.13	21.74
29	0	0.42	0.42	0	0.79	0.95	36.85	38.43	22.14
30	0	0.22	0.37	0	0.81	1.18	37.29	39.57	20.56
31	0	0.36	0.24	0	0.73	0.85	37.17	38.01	22.64

Summary statistics: NLSY97, employed men, bachelor's degree holders

	College graduates, college sector	College graduates, non-college sector (Overeducated)
Annual compensation (2019 dollars)	62,874	48,680
Cumulative GPA	3.18	3.06
AFQT percentile	73.26	67.04
Years STEM/HE ed.	1.45	1.05
Years non-STEM/HE ed.	2.00	2.22
Years no major	.28	.34
Parental income percentile	81	71
Parent attended college	70.82	66.04
College selectivity (SAT percentile)	54.85	49.31
Person-year observations	2,855	969
Age 27- %of net four-year tuition:		
Paid by family	49.5	38.6
Paid using loans	28.7	34.9
Parental household net worth	211,782	162,737

Note: Sample includes ages 24-32. Overeducation measured using the joint indicator. All differences in proportions/means are statistically significant.

Choice-specific utility functions

Choice- specific utility functions:

$$U_{ca} = \alpha_c + \gamma_c X_c + \epsilon_{ca},$$

X_c- alternative-specific vector of covariates.

- Employment choices: **expected log compensation** (net of predicted student loan repayments), experience, years in college in different majors, degrees attained, college selectivity Wage prediction.
- Education choices: **expected log tuition** Tuition prediction.
- All choices: previous period choices, unemployment rate, cumulative GPA, AFQT percentile, Black, Hispanic

ϵ_{ca}- idiosyncratic shocks, distributed Type I extreme value
α_c - choice specific constant (preference, endowment)- allowed to differ among two types in model with heterogeneity.
Variable construction.

Bellman representation

$\max_{c_a} E[\sum_{a=16}^{65} \beta^{a-16}(u[s_a, c_a] + \epsilon[c_a])]$ can be represented as:
$V_a(s_a, \epsilon(c_a)) = \max_{c_a}[u_a(s_a, c_a) + \epsilon(c_a) + \beta \int V_{a+1} f(s_{a+1} \mid s_a, c_a) dG(\epsilon(c_{a+1}))]$

Hotz and Miller (1993) representation: The conditional value function:

$$v_a(s_a, c_a) = u_a(s_a, c_a) + \beta \int V_{a+1}(s_{a+1}, \epsilon(c_{a+1})) f(s_{a+1} \mid s_a, c_a) dG(\epsilon(c_{a+1}))$$

can be replaced by:

$$v_a(s_a, c_a) = u_a(s_a, c_a) + \beta[ln[\sum_{j=1}^{J} exp(v_{a+1}(s_{a+1}, c_{a+1} = j))] + \gamma] f(s_{a+1} \mid s_a, c_a)$$

Derivation.

Model representation-CCP

Given the Type I extreme value assumption, the probability of choosing any option $\tilde{c}_a$ at age a is:

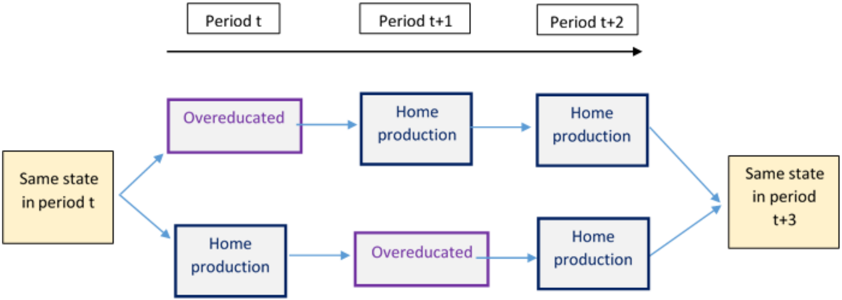
$$Pr(c_a = \tilde{c}_a \mid s_a) = \frac{exp(v_a(s_a, c_a = \tilde{c}))}{\sum_{j=1}^{J} exp(v_a(s_a, c_a = j))} = \frac{1}{\sum_{j=1}^{J} exp(v_a(s_a, c_a = j)) - exp(v_a(s_a, c_a = \tilde{c}))}.$$

The value function can be expressed as:

$$v_a(s_a, c_a = \tilde{c}) = u_a(s_a, c_a = \tilde{c}) +$$
$$\beta[ln[\sum_{j=1}^{J} exp((v_{a+1}(s_{a+1}, c_{a+1} = j) - exp(v_{a+1}(s_{a+1}, c_{a+1} = q))]f(s_{a+1} \mid s_a, c_a = \tilde{c}) +$$
$$\beta[v_{a+1}(s_{a+1}, c_{a+1} = q)]f(s_{a+1} \mid s_a, c_a = \tilde{c})$$

We can use a telescoping argument, and similarly express v_{a+1} and v_{a+2} as functions of conditional choice probabilities. However, we would still need to evaluate v_{a+3}.

Finite dependence: Arcidiacono and Miller (2011)



- In period t+3, the conditional value functions v_{a+3} and probabilities $Pr^{-1}(c_{a+3})$ will be the same, and will drop out when we take the difference in value functions at period t.

Estimation

$v_a(s_a, d_a = \tilde{c}) - v_a(s_a, d_a = H) = u_a(s_a, d_a = \tilde{c}) +$
$\beta ln[Pr^{-1}(c_{a+1} = H \mid s_{a+1})] f(s_{a+1} \mid s_a, c_a = \tilde{c}) -$
$\beta ln[Pr^{-1}(c_{a+1} = \tilde{c} \mid s_{a+1})] f(s_{a+1} \mid s_a, c_a = H) -$
$\beta[u_{a+1}(s_{a+1}, c_{a+1} = \tilde{c})] f(s_{a+1} \mid s_a, c_a = H) +$
$\beta^2 ln[Pr^{-1}(c_{a+2} = H \mid s_{a+2})] f(s_{a+2} \mid s_{a+1}, c_{a+1} = H) f(s_{a+1} \mid s_a, c_a = \tilde{c}) -$
$\beta^2 ln[Pr^{-1}(c_{a+2} = H \mid s_{a+2})] f(s_{a+2} \mid s_{a+1}, c_{a+1} = \tilde{c}) f(s_{a+1} \mid s_a, c_a = H)$

1. Estimate conditional choice probabilities using a flexible multinomial logit.
2. Calculate the scalar terms along the finite dependence path.
3. Estimate the flow utility parameters using a multinomial logit with offset terms equal to the scalars calculated in 2).
4. Heterogeneity introduced through a finite mixture model (Heckman and Singer, 1984) with two types, through an adaptation of the EM algorithm (Arcidiacono and Miller, 2011).

Identification assumptions: Assumptions imposed on the idiosyncratic shocks; Normalization of the utility of the home production option; Discount factor set to 0.9 Heterogeneity estimation.

Coefficient	K-12	2-year college	4-year, STEM/HE	4-year, non-STEM/HE	4-year, No major	Graduate school
Tuition	-	-0.161	-0.221	-0.221	-0.221	-0.144
	-	(0.393)	(0.268)	(0.268)	(0.268)	(0.795)
Cumulative GPA	3.923	0.336	0.321	0.364	-0.420	0.880
	(0.938)	(0.677)	(0.835)	(0.760)	(1.109)	(2.050)
Parent attended college	1.459	0.068	0.133	0.207	0.089	0.321
	(1.207)	(0.743)	(0.933)	(0.835)	(1.311)	(1.980)
AFQT percentile	0.089	-0.001	0.011	0.010	0.001	0.013
	(0.022)	(0.013)	(0.019)	(0.016)	(0.026)	(0.044)
Black	-0.884	-2.272	-0.033	-0.085	-0.020	-0.132
	(1.271)	(0.976)	(1.232)	(1.093)	(1.702)	(2.970)
Hispanic	-1.317	-0.025	-0.080	-0.108	0.005	-0.519
	(1.339)	(0.922)	(1.309)	(1.151)	(1.876)	(3.469)
Previous K-12	6.600	1.570	2.458	2.147	3.300	-
	(0.304)	(0.096)	(0.161)	(0.150)	(0.238)	-
Previous STEM/HE	-	1.145	4.685	3.573	4.368	0.134
	-	(0.384)	(0.198)	(0.226)	(0.430)	(0.458)
Previous Liberal	-	1.140	2.971	5.000	3.839	0.058
	-	(0.416)	(0.257)	(0.179)	(0.540)	(0.412)
Previous No major	-	0.932	4.641	5.079	5.305	-
	-	(0.931)	(0.299)	(5.079)	(5.305)	-
Previous 2-year college	-	2.326	1.017	0.936	1.758	-
	-	(0.101)	(0.291)	(0.230)	(0.572)	-
Years in 2-year college	-	-	0.164	0.304	-0.614	0.178
	-	-	(1.200)	(0.824)	(3.457)	(0.710)
Years STEM/HE	-	-0.623	-0.141	-0.456	-1.445	0.206
	-	(1.751)	(0.579)	(0.860)	(3.506)	(0.667)
Years non-STEM/HE	-	-0.837	-0.452	-0.251	-1.019	0.516
	-	(1.926)	(0.995)	(0.554)	(3.292)	(0.516)
Years no major	-	-0.379	-0.051	-0.141	-0.542	-0.178
	-	(2.972)	(1.219)	(1.230)	(2.834)	(0.473)
Unemployment rate	1.622	-0.068	-0.062	-0.061	0.044	-0.178
	(0.940)	(0.262)	(0.392)	(0.330)	(0.666)	(0.473)
Type II	-0.219	0.003	-0.014	-0.021	0.037	0.079
	(1.018)	(0.687)	(0.840)	(0.754)	(1.174)	(1.617)
Constant	-14.424	-0.273	-0.302	-0.305	-0.281	-0.669
	(0.790)	(0.401)	(0.459)	(0.424)	(0.613)	(1.044)

Coefficient	College-level skills	s.e.	Non-college level skills	s.e.
Skilled compensation	0.180	(0.370)	-	-
Unskilled compensation	-	-	0.272	(0.449)
Previous period college job	2.421	(0.067)	0.909	(0.069)
Previous period non-college job	0.930	(0.066)	1.984	(0.049)
Previous K-12	0.903	(0.098)	1.023	(0.073)
Previous STEM	1.239	(0.156)	1.178	(0.153)
Previous Liberal	1.310	(0.136)	1.107	(0.132)
Previous No major	1.093	(0.356)	1.043	(0.307)
Previous 2-year college	0.914	(0.123)	0.926	(0.103)
Experience college job	0.056	(0.173)	-0.009	(0.172)
Experience non-college job	0.090	(0.089)	0.061	(0.060)
High school degree	-2.854	(0.748)	-2.940	(0.585)
Years in 2-year college	0.203	(0.395)	0.182	(0.346)
Associate's degree	0.115	(0.114)	0.033	(0.033)
Years STEM	0.042	(0.360)	0.016	(0.391)
Years non-STEM	0.044	(0.330)	0.038	(0.339)
Years no major	0.126	(0.866)	0.113	(0.113)
Bachelor's degree	0.371	(0.370)	0.161	(0.160)
Years graduate school	0.074	(1.092)	-0.093	(1.696)
Graduate degree	-0.142	(1.757)	-0.306	(2.604)
Parent attended college	0.090	(0.485)	0.084	(0.418)
AFQT percentile	0.006	(0.009)	0.006	(0.008)
Cumulative GPA	0.248	(0.431)	0.265	(0.352)
Black	-0.052	(0.585)	-0.021	(0.492)
Hispanic	-0.102	(0.596)	-0.079	(0.503)
Unemployment rate	-0.062	(0.140)	-0.037	(0.126)
College selectivity	-0.505	(2.098)	-0.706	(1.943)
Type II	0.018	(0.430)	0.018	(0.369)
Constant	-0.410	(0.624)	-0.689	(0.921)

Model fit

Choice/ Outcome of interest	Data	Model	Static model
A. *Fraction of choices, ages 16-32*			
K-12 schooling	13.00	14.80	5.16
STEM/HE	3.39	3.65	1.66
Non-STEM/HE	4.35	5.17	2.17
No major	1.31	0.13	0.30
Jobs requiring college-level skills	22.31	21.46	11.06
Jobs not requiring college-level skills	27.76	32.42	24.93
Home production	22.95	19.04	53.82
Graduate school	0.93	1.75	0.27
Two-year college	3.94	1.56	0.62
B. *Degree attainment and overeducation*			
Bachelor's degree, age 26	23.97	19.42	11.52
Associate's degree, age 26	8.87	6.19	2.98
Overeducation rate, age 26	25.51	28.78	12.59
Overeducation rate, age 28	22.66	23.17	8.71
Overeducation rate, age 30	22.05	19.01	6.01

Notes : Frequencies are constructed by simulating the structural model 100 times for each individual.

Counterfactual analysis

- Counterfactual analysis follows Arcidiacono and Miller (*J. Econom.*, 2019) and involves re-estimating the conditional choice probabilities under the counterfactual scenario through backwards recursion, and then forward simulating starting at age 17, to obtain the counterfactual distribution of choices.
- Partial equilibrium: not modeling skill premium response to increased supply of college graduates.

Counterfactual analysis .

Counterfactual analysis-Men

		Decrease in tuition		
	Baseline	-25%	-50%	-95%
A.Overeducation rate				
Age 26	28.8	29.6	30.7	34.2
Age 28	23.2	24.1	25.1	29.2
Age 30	19.0	19.5	20.6	24.5
B.Characteristics of college graduates				
Parent attended college	0.74	0.72	0.70	0.59
% Black	0.09	0.10	0.11	0.15
% Hispanic	0.08	0.09	0.10	0.13
AFQT score percentile	69	68	66	57
HS GPA	3.23	3.20	3.17	2.97
College selectivity (SAT entering class)	53	53	52	48

Notes: Frequencies are constructed by simulating the structural model 100 times for each individual included in the estimating sample.

NPV analysis, tuition subsidy scenarios

		Decrease in net tuition:		
	Baseline	25%	50%	75%
Fraction holding BA, age 28	19.24	22.62	27.97	38.17
Overeducation rate	23.51	24.11	25.06	26.72
Earnings-age 28 (matched)	64,016	63,027	62,144	61,484
Earnings-age 28 (overeducated)	42,525	42,087	41,657	41,362
Earnings-age 28 (non-graduates)	32,142	32,226	32,337	32,533
Average subsidy (four years) (all college-goers)	17,008	18,707	20,891	24,031
Tuition subsidy change NPV	-	979	5,281	-3,504

Note: Averages shown for 100 simulations of the counterfactual scenario.

Conclusions

- G.I. Bill subsidies have fueled the overeducation rate in the U.S.
- Increases in tuition subsidies are likely to raise the overeducation rate. Moderate increases in subsidies may still result in positive NPV, given the high college premium.
- Overeducation is an explanatory channel for previous results that indicated low lifetime earnings gains for students induced to attend college through increases in subsidies (e.g. Keane and Wolpin, 1997; Arcidiacono, 2005)
- Addressing the efficiency of tuition subsidies after enrollment, regardless of initial ability/endowments:
  - (early) choice of college major. Providing better academic counseling and access to professional networks for first-generation college goers- in particular in regards to college major choice, college selectivity. Example: Expanding College Opportunities project (Hoxby and Turner, 2015).
  - reducing match costs for first generation college graduates. Unemployment benefits for recent graduates who lack work histories.
  - credit constraints for graduate school attendance (Fu et al., 2019)
  - initial overeducated employment shows persistence- the "overeducation trap". Countercycical policies and the starting jobs of college graduates (Rothstein, 2019).