Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors

Thomas Ahn
University of Kentucky

Peter Arcidiacono
Duke University & NBER

Amy Hopson
Duke University

James Thomas
Duke University

Motivation

- Perceived shortage of STEM majors despite these fields paying well.
- *The Obama Administration stands committed to providing students at every level with the skills they need to excel in the high-paid, highly-rewarding fields of science, technology, engineering, and math (STEM).* (White House website)
- 100K in 10 program wants 100,000 more STEM teachers (primary and secondary) within the next ten years.
- The lack of STEM majors is not driven by a lack of interest:
    - 48% of those pursuing a bachelor's degree in a STEM field leave, with half going to a non-STEM field and half dropping out (NCES 2013)

Do universities want more STEM majors?

MAJOR CHOICES profiles more than 65 recent Princeton graduates who followed their intellectual passions to major in subjects they loved, subjects that in most cases bore very little obvious connection to the careers they have subsequently pursued. Its purpose is to encourage undergraduates to follow their intellectual passions and study what they love, with confidence in the fulfilling lives that lie ahead and ***the knowledge that in no way will their choice of major limit the career choices they may wish to make in the future****.*

The book focuses on many of the smaller departments in order to encourage undergraduates to be imaginative and open-minded about their choices and to take the fullest advantage of the many intellectual opportunities available to them at Princeton.

(emphasis added)

University Response

- Universities may not want to see more students in STEM fields if it comes at the expense of other departments.
- Princeton's push for enrollments in smaller departments has been somewhat successful, with the largest percentage shifts in classics, music, Slavic languages and literature, comparative literature, and religion. (Princeton website)
- But universities (or their professors....) may engage in other activities that encourage more even representation across majors:
  - differences in grading standards
  - differences in workloads

University of Kentucky

- Grading standards and workloads differ substantially across departments
- STEM instructors give lower grades than their counterparts
  - Unadjusted difference in grades is around 0.3.

  and have higher study times:
  - Students in STEM classes study an extra 45 minutes per week, off a base of 2.3 hours a week.
- These differences occur despite STEM classes drawing students with higher test scores.
- Suggests universities are actually subsidizing students to go into low paying majors...

Why are STEM classes different?

We want to separate out how much of the differences across departments are driven by:

- Student demand for courses
- Professor preferences

Why does it matter?

- Harsher grading in STEM classes affects enrollment; many come in intending to major in STEM but then attrit.
- Who switches is predictable: those with the least preparation and worst performance within a school are much more likely to switch.
  - Controlling for academic background virtually eliminates racial differences in STEM persistence (at least at Duke...).
- How courses are graded affects not only the number of STEM majors but also their composition.

Women and STEM

- But women are both less likely to begin in STEM and more likely to switch out of STEM. This is surprising because:
  - Women are just as academically prepared as men.
  - Women study significantly more than men (about 2 hours a week at both Duke and Berea)
- Higher study times by women may reflect both:
  - Lower costs of studying
  - Valuing grades more

  where the first makes STEM more attractive, the second makes STEM less attractive

Women and STEM 2

- Suppose women both have lower costs of studying and value grades more than men.
- A policy that restricts average grades to be the same across classes would result in:
    - Differences in average grades between STEM and non-STEM falling
    - STEM classes requiring even more work relative to non-STEM classes
- With average grades and workload going up in STEM, STEM becomes relatively more attractive to women.

Our approach

Using transcript and course evaluation data from the University of Kentucky

- Estimate grade production functions
- Estimate student choices of courses and effort conditional on the grade production functions
- Estimate professor preference parameters given:
    - Estimates of the grade production parameters for all courses
    - Estimates of student preferences

Data

- Administrative data from the University of Kentucky - Lexington (UK)
- Full data: Fall, 2008 - Spring, 2013
- Our Sample: Fall, 2012
  - 20,343 unique students
  - 100,811 student-course observations
- Also have data on course evaluations which we use to get average study time

Descriptive Statistics by Gender

	Men	Women
High school GPA	3.13	3.34
	(1.20)	(1.16)
ACT Score	25.2	24.4
	(4.42)	(4.18)
Fall 2012 GPA	3.02	3.24
	(0.713)	(0.665)
Fall 2012 Credits	11.7	12.0
	(4.29)	(4.22)
STEM Major	38.0%	23.8%

Note: Fall 2012 University of Kentucky undergraduate students, 9,729 men, 9,798 women. Standard deviations in parentheses.

Descriptive Statistics by Course Type

	STEM	Non-STEM
Class Size	78.1	46.3
	(101.1)	(64.0)
Average Grade	3.03	3.31
	(0.50)	(0.46)
Average Grade | Female	3.11	3.40
	(0.59)	(0.46)
Study Hours	3.61	2.70
	(1.68)	(1.12)
Percent Female	37.0%	55.9%

Note: Fall 2012 University of Kentucky courses with enrollments of 16 or more students, 379 STEM courses, 1,164 non-STEM courses. For study hours, 293 STEM courses and 793 non-STEM courses. Standard deviations in parentheses.

Regressions of Grades and Study Time on Characteristics of the Individual and/or Class

Dependent Var.	Grade	Study hours per week
STEM Class	-0.325	0.520
	(0.009)	(0.148)
Female	0.140	
	(0.008)	
Percent Female	0.395	0.547
	(0.203)	(0.191)
Average Grade		-0.635
		(0.089)
ln(Class Size)	-0.116	-0.396
	(0.004)	(0.048)
Observations	72,449	1,085

Note: Additional controls for grades regression include, minority status, freshman, STEM major, pell grant, in-state student, ACT

Grade Production

- $j = [1, \ldots, J]$ indexes courses, each course j belongs to an area of study k.
- A_{ij} is i's preparation for course j.
- Grades for i in course j, g_{ij}, are given by:

$$g_{ij} = \beta_j + \gamma_j \left(A_{ij} + \ln(s_{ij}) \right) + \eta_{ij}$$

  where s_{ij} refers to study time and η_{ij} is noise.
- β_j and γ_j are set by the professor (restricted to linear grading policies)
- s_{ij} is set by the student.

Course Utility

- Utility for choosing course j is given by:

$$U_{ij} = \phi_i \mathbb{E}\left[g_{ij}\right] - \psi_i s_{ij} + \delta_{ij}$$

- Students then solve the following maximization problem when choosing their optimal course bundle:

$$\max_{d_{i1},\ldots,d_{iJ}} \sum_{j=1}^{J} d_{ij} U_{ij} \tag{1}$$

$$\text{subject to: } \sum_{j=1}^{J} d_{ij} = n, \; d_{ij} \in \{0,1\} \forall j$$

Study Effort

The optimal study effort in course j can be found by differentiating U_{ij} with respect to s_{ij}:

$$0 = \frac{\phi_i \gamma_j}{s_{ij}} - \psi_i$$

$$s_{ij}^{\star} = \frac{\phi_i \gamma_j}{\psi_i} \qquad (2)$$

Substituting the optimal choice of study time into the utility function yields:

$$U_{ij} = \phi_i \left(\beta_j + \gamma_j \left[A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) - 1\right]\right) + \delta_{ij} \qquad (3)$$

Reduced form grade equation

Substituting the expression for optimal study time into the grade process equation yields:

$$g_{ij} = \beta_j + \gamma_j \left(A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \right) + \eta_{ij} \tag{4}$$

Professors who set relatively higher values of γ_j see more study effort because higher γ_j's induce more effort and because higher γ_j's attract students with lower study costs.

Estimation

The key equations for estimation are given by:

(i) the solution to the students maximization problem after substituting in optimal study effort,

(ii) the reduced form grade production process, and

(iii) the optimal study effort.

Parameterizations

$$
\begin{array}{rcll}
A_{ij} & = & w_i\alpha_{1k(j)} + X_i\alpha_{2k(j)} & \text{Preparation in field } k(j) \\
\delta_{ij} & = & \delta_{0j} + w_i\delta_{1k(j)} + Z_{ij}\delta_{2k(j)} + \epsilon_{ij} & \text{Preference for course } j \\
\psi_i & = & \exp\left(\psi_0 + w_i\psi_1 + X_i\psi_2\right) & \text{Study costs} \\
\phi_i & = & \phi_0 + w_i\phi_1 & \text{Preferences for grades}
\end{array}
$$

where:

- w_i is an indicator for female
- X_i includes ACT scores, other family background characteristics
- Z_{ij} includes year of the student interacted with level of the course as well as female cross female instructor
- ϵ_{ij} i.i.d. Type 1 extreme value preference shock

Grade parameters

Estimate the following by NLLS:

$$g_{ij} = \theta_{0j} + \gamma_j \left(w_i \theta_{1k(j)} + X_i \theta_{2k(j)} \right) + \eta_{ij} \tag{5}$$

where:

$$\begin{aligned} \theta_{0j} &= \beta_j + \gamma_j(\ln(\phi_0) + \ln(\gamma_j) - \psi_0) \\ \theta_{1k(j)} &= \alpha_{1k(j)} + \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 \\ \theta_{2k(j)} &= \alpha_{2k(j)} - \psi_2 \end{aligned}$$

- The variation in the data used to identify $\{\theta_{1k(j)}, \theta_{2k(j)}\}$ comes from the relationship between student characteristics and grades.
- The γ_j's are identified from how these characteristics translate into grades relative to the normalized course.

Study parameters

- We assume that the relationship between log study hours and log study effort is linear; and that log study hours is reported with measurement error ζ_{ij}:

$$
\begin{aligned}
\ln(h_{ij}) &= \mu \ln(s^*_{ij}) + \zeta_{ij} \\
&= \mu \left(\ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \right) + \zeta_{ij} \\
&= \kappa_0 + w_i\kappa_1 + X_i\kappa_2 + \ln(\gamma_j) + \zeta_{ij}
\end{aligned}
$$

where:

$$
\begin{aligned}
\kappa_0 &= \mu(\psi_0 + \ln(\phi_0)) \\
\kappa_1 &= \mu(\ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1) \\
\kappa_2 &= -\mu\psi_2
\end{aligned}
$$

Study parameters 2

- Needed to normalize one γ to one for each department in the grade estimation. We can now undo this normalization.
- Namely, we estimated $\gamma_j^N = \gamma_j / C_{k(j)}$ where $C_{k(j)}$ is a group-specific constant.
- Substituting in for $\ln(\gamma)$ with $\ln(\gamma^N) + \ln(C_{k(j)})$ yields our estimating equation:

$$\ln(h_{ij}) = \tilde{\kappa}_0 + w_i\kappa_1 + X_i\kappa_2 + \kappa_{3k(j)} + \mu \ln(\hat{\gamma}^N) + \zeta_{ij} \qquad (6)$$

  where $\kappa_{3k(j)} = \mu \ln(C_{k(j)}/C_1)$ and $\tilde{\kappa}_0 = \kappa_0 + \ln(C_1)$.
- We can then partially undo the normalization on the γ's, with $\hat{\gamma}_j^P = \hat{\gamma}_j^N \exp(\hat{\kappa}_{2k(j)}/\hat{\mu})$ now normalized with respect to one course.
- We don't actually observe linked course evaluation records but we do know year in school so our estimating equation is formed from averaging across individuals by year of school.

Utility parameters

$$\widehat{E[g_{ij}|s^*_{ij}]} = \hat{\theta}_{0j} + \hat{\gamma}^N_j \left(w_i \hat{\theta}^N_{1k(j)} + X_i \hat{\theta}^N_{2k(j)} \right)$$

$$U_{ij} = \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \widehat{E[g_{ij}|s^*_{ij}]}(\phi_0 + w_i \phi_1) + C_1 \hat{\gamma}^P_j (\phi_0 + w_i \phi_1) + \epsilon_{ij}$$

- ϕ_0 and ϕ_1 are identified from sorting into classes that reward a student's abilities more: lower ability students have a relative preference for courses with low γ's.
- With these estimates we can recover all remaining structural parameters.

Sketch of simulated maximum likelihood procedure

- Even with Type 1 extreme value errors estimation is complicated because students are choosing *bundles* of courses.
- Denote K_i as the set of courses chosen by i.
- Denote M_i as the highest payoff associated with any of the non-chosen courses:

$$M_i = \max_{j \notin K_i} \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \left(\widehat{E[g_{ij}]} - \hat{\gamma}_j \right) (\phi_0 + w_i \phi_1) + \epsilon_{ij}$$

- Suppose K_i consisted of courses $\{1, 2, 3\}$ and that the values for all the preference shocks, the ϵ_{ij}'s, were known with the exception of those for $\{1, 2, 3\}$.

Sketch of simulated maximum likelihood procedure 2

- The probability of choosing $\{1, 2, 3\}$ could then be expressed as

$$\begin{aligned} Pr(d_i = \{1,2,3\}) &= Pr(\overline{U}_{i1} > M_i, \overline{U}_{i2} > M_i, \overline{U}_{i3} > M_i) \\ &= Pr(\overline{U}_{i1} > M_i)Pr(\overline{U}_{i2} > M_i)Pr(\overline{U}_{i3} > M_i) \\ &= (1 - G(M_i - \overline{U}_{i1}))(1 - G(M_i - \overline{U}_{i2})) \\ &\quad \times(1 - G(M_i - \overline{U}_{i3})) \end{aligned}$$

where $G(\cdot)$ is the extreme value cdf and $\overline{U}_{ij}$ is the flow payoff for j net of ϵ_{ij}.

Simulated maximum likelihood procedure

- Since the ϵ_{ij}'s for the non-chosen courses are not observed, we integrate them out of the likelihood function and approximate the integral by simulating their values from the Type I extreme value distribution.
- Denoting M_{ir} as the value of M_i at the rth draw of the non-chosen ϵ_{ij}'s and R as the number of simulation draws, estimates of the reduced form payoffs come from solving:

$$\max_{\phi,\delta} \sum_i \ln \left(\left[\sum_{r=1}^{R} \prod_{j=1}^{J} \left(1 - G(M_{ir} - \overline{U}_{ij})\right)^{d_{ij}} \right] / R \right) \qquad (7)$$

Unobserved heterogeneity

- Assume individuals are one of S unobserved types.
  - In practice, two.
- Unobserved type is a component of ability and through ability affects course selection.
- Estimation of grades is easy, but the choice model is hard.
- Following Arcidiacono and Miller (2011), use the EM algorithm with a reduced-from choice model that gives estimates of:
  1. the grade parameters
  2. the conditional probabilities of being each unobserved type
- Use the conditional type probabilities as weights to recover the study and choice parameters.

Implications of Demand-side Estimates

Demand-side estimates allow us to decompose differences in course choices, grades, and study effort between males and females into parts due to:

(i) differences in preferences (both over departments and over having a female professor),

(ii) differences in value of grades (ϕ_i),

(iii) differences in study costs (ψ_{ij}),

(iv) differences in preparation ($A_{ik(j)}$),

(v) differences in grading practices ($\{\beta_j, \gamma_j\}$)

The Professor's Problem

Professor payoffs are assumed to be a function of:

(i) the total amount of learning in the course: $a(\beta, \gamma)$,

(ii) total enrollment: $b(\beta, \gamma)$, and

(iii) student study time: $c(\beta, \gamma)$.

where learning for individual i in course j is given by:

$$\begin{aligned} L_{ij}(\gamma_j) &= A_{ik(j)} + \ln(s^*_{ij}) \\ &= A_{ik(j)} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \end{aligned} \tag{8}$$

Professor Objective Function

$$
\begin{aligned}
V_j(\beta,\gamma) &= \lambda_{0j} a_j(\beta,\gamma) - \lambda_{1j} b_j(\beta,\gamma) - \lambda_{2j} c_j(\beta,\gamma) \\
&= \lambda_{0j}\left[\sum_i P_{ij}(\beta,\gamma)\left(A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)\right)\right] \\
&\quad -\lambda_{1j}\left[\sum_i P_{ij}(\beta,\gamma)\right]^2 - \lambda_{2j}\left[\sum_i P_{ij}(\beta,\gamma)\left(\frac{\phi_i\gamma_j}{\psi_i}\right)^2\right]
\end{aligned}
$$

where we normalize λ_{0j} to one.

Solution to the Professor's Problem

The choice of β_j and γ_j satisfy the two first order conditions:

$$\frac{\partial V_j}{\partial \beta_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \beta_j} - \lambda_{1j}\frac{\partial b(\beta,\gamma)}{\partial \beta_j} - \lambda_{2j}\frac{\partial c(\beta,\gamma)}{\partial \beta_j} \tag{9}$$

$$\frac{\partial V_j}{\partial \gamma_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} - \lambda_{1j}\frac{\partial b(\beta,\gamma)}{\partial \gamma_j} - \lambda_{2j}\frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \tag{10}$$

The solution to this system is then:

$$\begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \end{bmatrix} = \begin{bmatrix} \frac{\partial b(\beta,\gamma)}{\partial \beta_j} & \frac{\partial c(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial b(\beta,\gamma)}{\partial \gamma_j} & \frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial a(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix} \tag{11}$$

By substituting in the estimates of the professor's own grading practices, the estimates of the grading practices of the other professors, and the estimates of student preferences, the right hand side is known.

Counterfactuals

- How course choices, study time, and average grades would change if STEM professors had the same preferences as non-STEM professors.
- How grading practices would change in response to restriction on grading policies such as capping the fraction of high grades.

Results: Preference Parameters

Preference for:	Coeff.	Std. Error
Expected grades (ϕ)	0.4157	(0.0199)
Female $\times$ expected grades	0.0759	(0.0157)
Female $\times$ female professor	0.1455	(0.0186)
C_1 (normalizing constant)	0.9581	(0.2593)
Female preferences for Departments		
Engineering	-1.0615	(0.0737)
Econ., Fin., Acct.	-0.5091	(0.0593)
Social Sciences	-0.2862	(0.0545)
Communication	-0.1528	(0.0537)
Chemistry & Physics	-0.1482	(0.0599)
Languages	-0.1033	(0.0582)
Mathematics	-0.0072	(0.0688)
Mgmt. & Mkting.	0.1153	(0.0662)
Regional Studies	0.2216	(0.0698)
Biology	0.2546	(0.0638)
Education & Health	0.3287	(0.0581)
Psychology	0.3758	(0.0659)
English	0.4167	(0.0769)

Results: Study Effort and Returns in Median Courses

	Study Effort Coeff. ($-\psi$)	Std. Error	Department	Median γ Coeff.
Female	-0.074	-0.080	**Mathematics**	3.59
ACTR (0's)	-0.002	-0.010	**Engineering**	3.11
ACTM (0's)	-0.018	-0.011	**Biology**	2.00
HS GPA	-0.004	-0.087	English	1.80
Black	0.234	-0.172	**Chem & Phys**	1.78
Hispanic	-0.317	-0.252	Psychology	1.64
Other Min.	-0.269	-0.261	Econ., Fin., Acct.	1.49
First Gen	0.107	-0.115	Regional Studies	1.49
Unobs. Type	0.198	-0.084	Communication	1.37
	Effort elasticity		Languages	1.34
ln(γ)	0.545	(0.206)	Social Sciences	1.24
			Agriculture	0.92
			Mgmt. & Mkting.	0.80
			Education & Health	0.77

Professor Preferences

Level	Category	Disutility of: Enrollment^2 (λ_{1j}) Mean	Std. Dev.	Study Time^2 (λ_{1j}) Mean	Std. Dev.
Lower	non-STEM	0.039	0.023	0.230	0.179
Level	STEM	0.023^{*}	0.017	0.063^{*}	0.077
Upper	non-STEM	$0.048^{*,\dagger}$	0.022	$0.304^{*,\dagger}$	0.215
Level	STEM	$0.035^{*,\dagger}$	0.016	$0.136^{*,\dagger}$	0.237

Counterfactuals

Examine how the STEM gap changes given five changes to the environment:

1. No differences between men and women in their preferences for grades ($\phi_1 = 0$)
2. No differences in unobserved abilities ($\alpha_1 = 0$)
3. No differences in observed abilities ($\overline{X}_f = \overline{X}_m$)
4. No differences tastes for departments
5. No differential preference for female professors
6. ~~Equalized expected grades across courses for the average student~~

PE and GE responses to changing female tastes and abilities on share STEM

Turn off...	PE Female Increase Over Base	GE Female Increase Over Base	GE/PE	GE Male Increase Over Base
(1) grade prefs	0.94%	0.26%	27.5%	-0.43%
(2) gender ability	3.78%	1.52%	40.3%	-1.79%
(3) obs ability	0.50%	0.37%	74.4%	-0.15%
(4) tastes	1.54%	0.78%	50.2%	-0.92%
(5) female prof pref	0.40%	0.16%	39.1%	-0.22%

Baseline female: 28.40%
Baseline male: 40.02%

Supply-side Counterfactuals

Examine how changing grading practices affects the STEM gap by:

1. Setting expected grades to be the same across classes for the average student by shifting course intercepts
2. Adjust STEM professor preferences so that the averages are similar across STEM and non-STEM professors and solving for the new equilibrium grading policies

Supply-side Counterfactual Results on Share STEM

	Male Share	Female Share	Gender Ratio
Baseline	0.400	0.284	0.710
(6) Equalize exp grade for average student	0.440	0.334	0.760
(7) Change STEM prof prefs	0.235	0.126	0.536

Gender ratio=Women/Men

Conclusion

- Preliminary evidence suggests men and women respond differently to grading practices
- A cheap way of changing the STEM gap may be to change grading practices
- May be important to take into account how professors respond to university regulations on grading policies