Essays on the economics of higher education

by

James Thomas

Department of Economics
Duke University

Date: ______________________

Approved:

____
Peter Arcidiacono, Supervisor

____
Robert Garlick

____
V. Joseph Hotz

____
Hugh Macartney

____
Arnaud Maurel

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016

ABSTRACT

Essays on the economics of higher education

by

James Thomas

Department of Economics
Duke University

Date: ____________________
Approved:

____
Peter Arcidiacono, Supervisor

____
Robert Garlick

____
V. Joseph Hotz

____
Hugh Macartney

____
Arnaud Maurel

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2016

Abstract

At least since the seminal works of Jacob Mincer, labor economists have sought to understand how students make higher education investment decisions. Mincer's original work seeks to understand how students decide how much education to accrue; subsequent work by various authors seeks to understand how students choose where to attend college, what field to major in, and whether to drop out of college.

Broadly speaking, this rich sub-field of literature contributes to society in two ways: First, it provides a better understanding of important social behaviors. Second, it helps policymakers anticipate the responses of students when evaluating various policy reforms.

While research on the higher education investment decisions of students has had an enormous impact on our understanding of society and has shaped countless education policies, students are only one interested party in the higher education landscape. In the jargon of economists, students represent only the 'demand side' of higher education—customers who are choosing options from a set of available alternatives. Opposite students are instructors and administrators who represent the 'supply side' of higher education—those who decide which options are available to students.

For similar reasons, it is also important to understand how individuals on the

supply side of education make decisions: First, this provides a deeper understanding of the behaviors of important social institutions. Second, it helps policymakers anticipate the responses of instructors and administrators when evaluating various reforms. However, while there is substantial literature understanding decisions made on the demand side of education, there is far less attention paid to decisions on the supply side of education.

This dissertation uses empirical evidence to better understand how instructors and administrators make decisions and the implications of these decisions for students.

In the first chapter, I use data from Duke University and a Bayesian model of correlated learning to measure the signal quality of grades across academic fields. The correlated feature of the model allows grades in one academic field to signal ability in all other fields allowing me to measure both 'own category' signal quality and 'spillover' signal quality. Estimates reveal a clear division between information rich Science, Engineering, and Economics grades and less informative Humanities and Social Science grades. In many specifications, information spillovers are so powerful that precise Science, Engineering, and Economics grades are more informative about Humanities and Social Science abilities than Humanities and Social Science grades. This suggests students who take engineering courses during their Freshman year make more informed specialization decisions later in college.

In the second chapter, I use data from the University of Central Arkansas to understand how universities decide which courses to offer and how much to spend on instructors for these courses. Course offerings and instructor characteristics directly affect the courses students choose and the value they receive from these choices. This chapter reveals the university preferences over these student outcomes which best explain observed course offerings and instructors. This allows me to assess whether university incentives are aligned with students, to determine what alternative university choices would be preferred by students, and to illustrate how a revenue neutral tax/subsidy policy can induce a university to make these student-best decisions.

In the third chapter, co-authored with Thomas Ahn, Peter Arcidiacono, and Amy Hopson, we use data from the University of Kentucky to understand how instructors choose grading policies. In this chapter, we estimate an equilibrium model in which instructors choose grading policies and students choose courses and study effort given grading policies. In this model, instructors set both a grading intercept and a return on ability and effort. This builds a rich link between the grading policy decisions of instructors and the course choices of students. We use estimates of this model to infer what preference parameters best explain why instructors chose estimated grading policies. To illustrate the importance of these supply side decisions, we show changing grading policies can substantially reduce the gender gap in STEM enrollment.

Contents

List of Tables

Acknowledgements

I am deeply indepted to my advisor Peter Arcidiacono and committee members Robert Garlick, V. Joseph Hotz, Hugh Macartney, and Arnaud Maurel without whom this work would not be possible. I would also like to thank Thomas Ahn and Amy Hopson who are collaborating authors on the third chapter. Finally, I would like to thank Ken Spenner for access to the Campus Life and Learning survey data and members of the Arkansas Research Center for access to the Arkansas Higher Education administrative data.

1

The signal quality of grades across academic fields

1.1 Introduction

Students arrive on university campuses with limited information about their academic abilities making it difficult to navigate academic life. Previous literature identifies earned grades as an important signal of academic abilities to students yet there is little evidence on which grades reveal abilities most efficiently.[1] Grading methods and standards vary widely across academic fields (Johnson (2003)) which suggests the information quality of grades may vary across fields as well.

It is important to know which fields offer the most informative grades because missing information about academic abilities can be quite burdensome. Missing information about academic abilities may lead to delayed graduation, delayed spe-

[1] Arcidiacono (2004), Stinebrickner and Stinebrickner (2012, 2014b,a) , Zafar (2011). See Altonji et al. (2012) for a review. An exception is Stinebrickner and Stinebrickner (2014b) which includes reduced form evidence suggesting Science grades are more informative about unobserved abilities than non-Science grades.

cialization, or specialization mismatch—whereby chosen field differs from the field that would be chosen if full information were available.[2] This is especially important for under-privileged students. These students are generally less informed when they arrive on campus (Hoxby and Avery (2013)) and are more likely to face financial constraints which make delaying graduation infeasible (Manski (1992)).

This paper compares the information quality of grades across academic fields. I find evidence of sizable differences: Natural Science, Engineering, and Economics courses are significantly more informative than Humanities and Social Science courses. In a preferred specification, one Engineering grade has the same average information content as three Humanities grades.[3]

There are a number of ways universities can improve information diffusion: First, curricula can be modified to reduce the number of noisy courses students are obligated to take. At the university I examine, students must take 13-15 Humanities or Social Science courses but only 6-8 Natural Science, Engineering or Economics courses to satisfy graduation requirements.[4] My results suggest these requirements

[2] See Bound et al. (2012) on delayed graduation. See Bordon and Fu (2015) and Malamud (2010, 2011) on the relative importance of delayed specialization and specialization mismatch. See Allen and Van der Velden (2001) and Nordin et al. (2010) on the effects of specialization mismatch in the labor market.

[3] In a preferred specification, one Engineering grade reduces uncertainty in beliefs about Humanities ability by 24%, Social Science ability by 20%, Natural Science ability by 36%, Engineering ability by 37%, and Economics ability by 33%. Three Humanities grades reduce uncertainty in Humanities ability by 42%, Social Science ability by 31%, Natural Science ability by 22%, Engineering ability by 28%, and Economics ability by 25%. This corresponds to a 30% average reduction from the Engineering grade and a 29% average reduction from the three Humanities grades.

[4] The university examined is Duke University. Requirements are: Arts, Literatures, and Performance (2 courses); Civilizations (2 courses); Natural Sciences (2 courses); Quantitative Studies (2 courses); Social Sciences (2 courses, includes Economics); Cross-Cultural Inquiry (2 courses); Ethical Inquiry (2 courses); Science, Technology and Society (2 courses); Foreign Language (3 courses); Writing (3 courses).

oppose information diffusion. Second, grading policies may be regulated to improve the information quality of grades. In certain circumstances, instructors have incentives to inflate grades.[5] Because grades are typically capped at A or A+, grade inflation becomes grade compression which limits the information quality of grades (Mansfield (2001)). My results corroborate this—fields with higher average grades provide less informative grade signals. Policies which encourage instructors to assign more disperse grades will improve the information quality of grades.

I measure differences in the information quality of grades across academic fields using a Bayesian model of correlated learning. The correlated learning feature allows signals from each academic field to have 'own field' signal quality and 'spillover' signal quality to all other fields. Allowing for information spillovers across fields turns out to be quite important. In some specifications, these spillovers are so strong that high quality signals from one field reveal more about academic abilities in another field than signals from that field. Incorporating these spillovers provides a complete measure of signal quality across academic fields.

The model also allows for unobserved differences in the initial confidence and initial certainty of students. Because these prior beliefs shape the formation of future beliefs, this feature allows the ability revelation process to vary by unobserved characteristics. This also turns out to be quite important: unobserved characteristics generate substantial differences in initial confidence and certainty; additionally, students who were initially less confident and less certain experience much weaker

[5] Achen and Courant (2009) argue grades may be inflated to increase enrollments, reduce complaints about low marks, or as a result of personal relationships with students. Johnson (2003) argues grades may be inflated to improve student evaluations of courses. Chan et al. (2007) and Wikström and Wikström (2005) argue grade inflation may arise from competition between schools.

information spillovers across fields.

I estimate my model using course level transcript data from Duke University and an adaptation of the Expectation-Maximization algorithm. The data are well suited for two important reasons: First, course level data allows me to clearly match grade signals to academic fields. This is necessary for comparing signal quality across academic fields. Second, transcript data allows me to see the same student earning multiple grades in several academic fields. This variation is important for identifying information spillovers across fields. In theory, all parameters of my model could be estimated by maximum likelihood; however, the large number of confounding unobserved belief variables would make this computationally infeasible. The Expectation-Maximization algorithm is an effective method for reducing the computational burden of maximizing a likelihood with a large number of unobserved variables.

I begin by illustrating my main finding with a simpler case in which all students are assumed to have the same initial beliefs about their unobserved abilities. The results portray a stark difference between information rich Science, Engineering, and Economics (henceforth, SEE) grades and less precise Humanities and Social Science (henceforth, HuSS) grades. In a preferred specification, one Engineering grade provides a 30% average reduction in belief uncertainty across fields while one Humanities grade provides only a 13% average reduction. These results also demonstrate the importance of information spillovers across academic fields. In many cases, spillovers are so powerful that precise SEE grades are more informative about HuSS abilities than HuSS grades.

I later consider a more general case which allows for unobserved heterogeneity in initial beliefs about abilities. In this scenario, unobserved characteristics make students differ in both their initial confidence and certainty. While most students experience information patterns similar to those described above, a small minority of students process grade information very differently. These students still find SEE courses to be more informative on average but they experience more modest information spillovers across fields. These students also have higher uncertainty in their initial expectations. This implies less certain students are missing a lot of field specific information while more informed students are only missing information which is general to all academic fields.

Why are SEE grades more informative than HuSS grades? Two theories seem most plausible: First, the difference in signal quality may arise from differing degrees of grade compression. In theory, Duke instructors may assign 11 possible grades to students; however, in practice lower grades are rarely assigned.[6] Compression is especially strong for HuSS grades where 94.2% of grades are B- or above and 35.3% of grades are A (the maximum). Comparatively, 82.1% of SEE grades are B- or above and 29.7% of SEE grades are A. By reducing the number of grades which are assigned in practice and compressing more students into the highest category HuSS instructors are reducing the information quality of their grades. Second, it may be more difficult to assign precise grades in courses where assessment is more subjective. SEE courses typically assess student performance using multiple-choice or short-answer format questions with well-defined correct answers which limits the idiosyncrasy of grading. Conversely, HuSS courses typically assess student performance with papers or essay

[6] Possible grades and corresponding grade point values are A (4.0), A- (3.7), B+ (3.3), B (3.0), B- (2.7), C+ (2.3), C (2.0), C- (1.7), D+ (1.3), D (1.0), F (0.0).

questions where the idiosyncratic preferences of the grader play a larger role. Future research may examine the relative importance of these and other mechanisms in explaining why SEE grades are more informative than HuSS grades.

This paper relates to a long and rich literature on human capital and revelation of missing information.[7] Seminal works by Jovanovic (1979) and Miller (1984) consider the search patterns of workers if occupation match quality is only revealed from occupation specific experience. More recently, Arcidiacono (2004), Arcidiacono et al. (2014), Stinebrickner and Stinebrickner (2012, 2014b,a), and Zafar (2011) consider the role of missing information about academic abilities in the decisions of college students. Arcidiacono (2004) and Arcidiacono et al. (2014) use data from the National Longitudinal Study to estimate structural models of major choice and dropout which include ability revelation through grades. The authors find missing information about academic abilities plays an important role in decisions to drop out or change majors; however, a comparison of information quality across academic fields is not the focus of these papers.

Stinebrickner and Stinebrickner (2012, 2014a), and Zafar (2011) use data on elicited beliefs to analyze how students form beliefs and how information revealed by grades affects major choice and dropout decisions. These papers provide an important bridge between theoretical models of belief formation and actual data on beliefs in the context of higher education; however, they do not analyze how

[7] Notable contributions not mentioned in the body include but are not limited to Shaw (1987), McCall (1990), Farber and Gibbons (1996), Jovanovic and Nyarko (1997), Neal (1999), Altonji and Pierret (2001), Gibbons et al. (2005), Altonji (2005), Pavan (2011), Antonovics and Golan (2012), and Papageorgiou (2014). There also exists a rich literature examining the role of missing information in consumer decisions (Ackerberg (2003); Crawford and Shum (2005); Cai et al. (2009) and health behaviors (Downs et al. (2009); Dupas (2011, 2014); Jalan and Somanathan (2008)).

courses chosen affect the speed of ability revelation. Stinebrickner and Stinebrickner (2014b) briefly analyses the information quality of grades across academic fields and finds evidence that Science grades are more informative than non-Science grades. My paper builds upon this analysis by estimating a full Bayesian learning model which provides complete measures of information quality across academic fields. Overall this literature acknowledges the importance of missing information to college students and identifies grades as an important signal of this missing information; however, there have been few attempts to measure differences in the speed of ability revelation across academic field. This paper contributes by providing such an analysis.

The remainder of the paper is organized as follows: Section 1.2 introduces the Duke University transcript data and discusses why these data are ideal for my analysis. Section 1.3 presents descriptive statistics illustrating heterogeneity in signal quality across fields and information spillovers between fields. Section 1.4 presents the correlated learning model which I use to develop measures of signal quality across academic fields. Section 1.5 describes the adaptation of the EM algorithm I use to estimate the correlated learning model. Section 1.6 reports estimation results. Section 2.7 concludes.

1.2 Duke University Transcript Data

For my empirical analysis, I employ administrative data from Duke University which include full academic transcripts and information used to make admissions decisions. These data are available for students who participated in the Campus Life and Learning (CLL) Survey which followed sub-samples of the 2001 and 2002 entering undergraduate cohorts at Duke University.

These data are well suited for my analysis for two important reasons: First, full transcript data allows me to assign grade signals to academic fields. This is very important for comparing the signal quality of grades across academic fields. Previous literature often uses aggregate signals such as semester average GPA.[8] Without additional data these aggregate signals cannot be assigned to academic fields making it difficult to evaluate heterogeneity in signal quality. Second, full transcript data allows me to observe the same student taking many classes in multiple fields. Repeated observations of the same student in the same field are important for separating unobserved student abilities from grading noise; outcomes for the same student across multiple fields are necessary for measuring information spillovers across academic fields.

The CLL Survey originally contacted 1,536 students in the 2001 and 2002 entering undergraduate cohorts at Duke University. Of these students, 1,132 ever gave consent to have their confidential records used for research purposes. After removing five students who were missing education measures for both parents I arrive at a sample of 1,127 students. Arcidiacono et al. (2012) show non-respondents have lower SAT scores, have better educated parents, are more likely to be from private schools, and have slightly lower grade point averages than respondents. However, they note the differences are quite small and thus conclude non-response bias is minimal. To account for this non-response bias (and for the stratified sampling of the CLL) I use survey weights based on race and cohort to improve the representativeness of the sample.

After removing independent study courses and courses not taken for a grade, I

[8] Arcidiacono (2004), Arcidiacono et al. (2014), Zafar (2011).

observe 37,255 course observations. Courses are labeled as either Humanities, Social Science (excluding Economics), Natural Science, Engineering, and Economics using Duke's internal definition for these fields. The 37,255 course observations include 12,076 from Humanities, 10,314 from Social Science, 10,605 from Natural Science, 1,612 from Engineering, and 2,648 from Economics.

For the empirical analysis, I construct a measure of course difficulty level based on course number. I label all courses with a course number less than 100 as beginner and all courses with a course number equal to or greater than 100 as advanced. With this definition, 78.3% of classes taken during the freshman academic year are beginner, 42.8% of classes taken during the sophomore academic year are beginner, 18.9% of classes taken during the junior academic year are beginner, and 12.6% of the classes taken during the senior academic year are beginner.

1.3 Descriptive Statistics

A number of factors—both observed and unobserved—-affect which courses students choose. This non-random selection makes it challenging to estimate the information quality of grades across academic fields. In Section 1.6 I report estimates of a structural model of ability revelation which controls for sorting on observed and unobserved variables. In this section, I present simple descriptive evidence on the information quality of grades across academic fields and course levels. These estimates are confounded by non-random sorting but the patterns are qualitatively similar to those described in Section 1.6: first, there is substantial heterogeneity in grade distributions across academic fields and course levels; second, Science, Engineering, and Economics grades identify high performing students more consistently than Humanities and Social Science grades; third, high marks in one field are correlated with high marks in other fields. These findings suggest there are important differences in the informativeness of grades across academic fields and demonstrate the need to allow for information spillovers across academic fields.

Panel A of Table 1.1 reports estimates of the mean and variance of assigned grades in beginner and advanced courses for each academic field. On average, grades are lower and more dispersed in Science, Engineering, and Economics (henceforth, SEE) and higher and more homogenous in Humanities and Social Science (henceforth, HuSS). These descriptive statistics likely overstate expected performance and understate variation in performance as a result of sorting on field specific abilities. They may also mischaracterize the difference in expected performance between SEE and HuSS if sorting on expected grades is stronger in some fields relative to others.

We can make some inferences about sorting behavior by comparing grade distributions in advanced courses to grade distributions in beginner courses. Students who select into advanced courses in a specific field have probably learned through past course work that they have strengths in this field—as such, students enrolling in advanced courses are likely to be more selected than students enrolling in beginner courses. As expected, we see grade distributions are generally higher and more homogenous in advanced courses relative to beginner courses. These distributional differences between advanced and beginner courses are generally larger in SEE courses and are especially large in Economics where average grades increase by .38 grade points and variance decreases by .28 squared grade points. This provides suggestive evidence that expected performance affects how students choose courses and justifies the need to control for selection on observed and unobserved variables.

Table 1.1: Grade Means, Variances, and Correlations

Panel A: Means and Variances of Grades					
Beginner	Humanities	Social Science	Science	Engineering	Economics
Means	3.488	3.330	3.061	3.241	3.029
Variances	0.3815	0.3883	0.6977	0.6525	0.6872
Course Observations	5648	2244	4774	642	996
Advanced					
Means	3.5706	3.4879	3.2632	3.4318	3.4129
Variances	0.2800	0.3258	0.6029	0.4540	0.4074
Course Observations	6483	8131	5864	980	1670

Panel B: Variance Decomposition (% of field variance)					
Beginner	Humanities	Social Science	Science	Engineering	Economics
Across Students	45.1%	61.0%	58.1%	68.9%	79.1%
Within Student	54.9%	39.0%	41.9%	31.1%	20.9%
Advanced					
Across Students	43.2%	42.0%	50.1%	41.7%	45.4%
Within Student	56.8%	58.0%	49.9%	58.3%	54.6%

Panel C: Within-student across-field correlations					
	Humanities	Social Science	Science	Engineering	Economics
Humanities	1.0000	0.6428	0.6198	0.4930	0.4697
		1087	*1131*	*226*	*642*
Social Science		1.0000	0.5671	0.3473	0.5233
			1086	*186*	*607*
Science			1.0000	0.7373	0.5711
				226	*641*
Engineering				1.0000	0.5291
					130
Economics					1.0000

Variation in assigned grades exists for two reasons: first, there are differences in the true performance of different students in the same course; second, there is noise in the evaluation process which implies the mapping from true performance to assigned grades is not one-to-one. To assess the relative importance of these two mechanisms in each academic field Panel B of Table 1.1 additively decomposes field and level specific variances into within-student and across-students components.[9] Within student variance occurs when one student earns different grades within the same field and this reflects noise in the evaluation process. Across student variance occurs when two students have different average grades in the same field and this reflects differences in true parameters across students.

For beginner Engineering and Economics courses, a large share of variance comes from differences across individuals (68.9% and 79.1% respectively) compared to a smaller share for beginner Humanities courses (45.1%). This suggests that in beginner courses there are important differences in the informativeness of grades across academic fields. While the distribution of beginner Humanities grades is tighter, the same student earns grades spanning a large section of the distribution in different beginner Humanities courses. Conversely, beginner Engineering and Economics grades are more dispersed but the same student usually earns similar grades in different

[9] Suppose Y_{it} is an outcome that is observed T_i times for each individual i. The variance of Y_{it} can be decomposed as follows:

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(Y_{it}-\bar{Y}\right)^2}{\mathbb{N}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(Y_{it}-\bar{Y}_i\right)^2}{\mathbb{N}} + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\bar{Y}_i-\bar{Y}\right)^2}{\mathbb{N}}$$

where $\bar{Y}_i$ is individual i's average outcome, $\bar{Y}$ is the population average, and $\mathbb{N}$ is the total number of observations. The left hand side represents overall variance in Y_{it}. The first term on the right hand side represents within-individual variance. The second term on the right hand side represents across-individual variance.

beginner Engineering and Economics courses. Beginner Engineering and Economics grades thus appear to reveal the relative performance of students more clearly than beginner Humanities grades.

In advanced courses, a larger share of variance arises from idiosyncratic noise in the evaluation process. The difference is especially stark in Engineering and Economics where 58.3% and 54.6% of variation in advanced grades arises from differences within students compared with 31.1% and 20.9% in beginner grades. This change appears to mostly result from selection into advanced courses. Across student variance drops dramatically between beginner and advanced courses: from .45 to .19 squared grade points in Engineering and from .54 to .18 squared grade points in Economics. Within student variance due to noise in the evaluation process also increases slightly between beginner and advanced courses—from .20 to .26 squared grade points in Engineering and from .14 to .22 squared grade points in Economics—but the shift is small when compared to the change in across student variance. This suggests the tighter grade distributions in advanced courses arise because the population is more selected and not because grades are less discriminating.

The results in Panel B of Table 1.1 illustrate the capacity of grades to signal own-field academic abilities but they do not capture 'spillover' signals to other fields. To illustrate the importance of information spillovers across fields Panel C of Table 1.1 reports correlations between an individual's average grade for each pair of academic fields. Intuitively, this measures the degree to which high marks in one field predict high marks in another. The results show high correlations in performance across all academic fields. Even fields which are generally considered to have little in common—such as Humanities and Natural Science—offer higher grades to the same students.

This suggests the same skills and behaviors lead to strong performance in all fields. As such, new information about performance in one field should also inform students about performance in all other fields. These information spillovers across fields must be measured carefully to obtain a complete measure of signal quality.

1.4 Model of Ability Revelation

In this section, I present a model of the ability revelation process. I model how students use grade signals from each academic field to update their beliefs about their unobserved abilities in all academic fields allowing for both 'own field' ability revelation and 'spillover' ability revelation to other fields. I also model how students select courses given observed characteristics and unobserved beliefs about abilities. This provides the framework for adjusting for selection in the empirical analysis.

I begin by introducing the primitives of my model; next, I describe how grades depend on observed and unobserved characteristics; following this I show how grade residuals are used to update beliefs about abilities; finally, I model how these beliefs and other observed characteristics affect course choices.

1.4.1 Primitives

Index students by $i = 1, 2, \ldots, N$, academic semesters by $t = 1, 2, \ldots, T$, and courses within semester t by $c = 1, 2, \ldots, C_{it}$. All courses are categorized into exactly one academic field $k \in \{1, \ldots, K\}$ and difficulty level $l \in \{1, \ldots, L\}$.[10] In my most general empirical application, I allow important student parameters to vary across

[10] In the empirical analysis, academic fields are humanities, social sciences, natural sciences, engineering, and economics and difficulty levels are beginner and advanced.

student 'types' $\tau = 1, \ldots, \mathscr{T}$ which are known by students but unobserved by the econometrician (Heckman and Singer (1984)).

1.4.2 Grade Production

Earned grades play a crucial role in this model as they are the sole mechanism through which students learn about their unobserved abilities. I model the production of this important variable as:

$$g_{itklc} = X_{itkl}\theta_k + \alpha_{ik} + \eta_{itklc} \tag{1.1}$$

where X_{itkl} are observed covariates, α_{ik} is unobserved ability in academic field k, and η_{itklc} is idiosyncratic grading noise with distribution $\eta_{itklc} \overset{iid}{\sim} N\left(0,\, \sigma^2_{kl}\right)$. The heteroskedasticity of grading noise by academic field and course level is one mechanism by which information quality is allowed to vary across academic fields.

Individuals draw field specific abilities as a vector from a type specific multivariate normal distribution $\alpha_i \sim N\left(\gamma_\tau,\, \Delta_\tau\right)$. When students enroll as Freshmen, they observe the distribution parameters γ_τ and Δ_τ but do not observe their realization α_i. As such, the distribution parameters γ_τ and Δ_τ characterize the distribution of initial prior beliefs about abilities. This is a second important mechanisms by which information quality is allowed to vary across academic fields.

In a setting without information spillovers, the values of $\Delta\left(k,k\right)$ and σ^2_{kl} can be used to construct signal-to-noise ratios which measure the own field signal quality of grades. The signal-to-noise ratio for a field k level l grade signal is given by:

$$SNR_{kl} = \frac{\Delta\left(k,k\right)}{\Delta\left(k,k\right) + \sigma^2_{kl}} \tag{1.2}$$

. If $\Delta\left(k,k\right)$ is large relative to σ^2_{kl} then there substantial uncertainty in initial beliefs but very little noise in grade signals. In this scenario, SNR_{kl} approaches one implying grade signals are very informative. Conversely, if $\Delta(k,k)$ is small relative to σ_{kl}^2 then there is very little uncertainty and only noisy signals available. In this scenario, SNR_{kl} approaches zero implying grade signals have little value.

Section 1.3 suggests there are substantial information spillovers across fields which make own field signal-to-noise ratios an incomplete measure of information content. To allow for these spillovers, the variance-covariance matrix Δ_τ is permitted to be non-diagonal. Off-diagonal elements of Δ_τ measure how students initially believe components of α_i are related. Large values for $\Delta(k,k')$ imply grade signals from field k are highly informative about abilities in field k'(and vice versa). This is a third important mechanism by which information quality is allowed to vary across academic fields.

The linear structure additively decomposes earned grades into an observed (both by the student and the econometrician) component $X_{itkl}\theta_k$, an initially unobserved (both to the student and the econometrician) component α_{ik} which is revealed over time, and an idiosyncratic noise term η_{itklc}. The interpretation of α_{ik} changes dramatically when different variables are included in X_{itkl}: a rich specification assumes students are very knowledgeable about the grade production process while a parsimonious specification assumes they know very little. In the empirical analysis, I consider alternative choices for X_{itkl} which range from very parsimonious to very rich. This allows me to compare the signal quality of grades across academic fields under varying assumptions about student information sets.

1.4.3 *Ability Revelation*

While the realization of α_i is unknown, students can use earned grades (and their knowledge of the grade production process) to refine their beliefs about this variable. The central focus of this paper is to compare the effectiveness of grades from different academic fields in improving the precision of these beliefs.

When a student initially enrolls, she knows $\alpha_i \sim N(\gamma_\tau, \Delta_\tau)$ but does not observe her specific realization α_i. When a student earns a grade, she extracts the grade residual:

$$\begin{aligned} z_{itklc} &= g_{itklc} - X_{itkl}\theta_k \\ &= \alpha_{ik} + \eta_{itklc} \end{aligned} \tag{1.3}$$

From the student's perspective, $z_{itklc} \sim N(\alpha_{ik}, \sigma_{kl}^2)$ implying z_{itckl} can be used as a noisy signal of ability α_{ik}.

Define the vector of average signals received prior to semester t as

$$\bar{z}_{it}(k) = \begin{cases} \frac{1}{n_{itk}} \sum_{t'=1}^{t-1} \sum_{c=1}^{C_{it'}} z_{it'klc} & n_{itk} > 0 \\ 0 & n_{itk} = 0 \end{cases} \tag{1.4}$$

where n_{itk} is the number of field k classes taken prior to semester t. Students combine this vector of average signals with initial prior beliefs to form posterior beliefs. The distribution of posterior beliefs is given by $b_{it\tau} \sim N(\mu_{it\tau}, \delta_{it\tau})$ (DeGroot (1970)) where:[11]

$$\mu_{it\tau} = \gamma_\tau + \left(\Delta_\tau^{-1} + \sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)^{-1}\left(\sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)(\bar{z}_{it} - \gamma_\tau) \tag{1.5}$$

$$\delta_{it\tau} = \left(\Delta_\tau^{-1} + \sum_{l=1}^{L} D_{il}\Phi_l^{-1}\right)^{-1} \tag{1.6}$$

The mean belief vector $\mu_{it\tau}$ represents the students' 'best guesses' about the value of their abilities at the beginning of period t. These best guesses are a weighted average of initial prior beliefs γ_τ and new information from grade signals $\bar{z}_{it} - \gamma_\tau$ where the weights depend on the number and quality of signals received and the uncertainty in initial prior beliefs.

The diagonal elements of the belief covariance matrix $\delta_{it\tau}$ measure the uncertainty in students' best guesses about their abilities.[12] As such, these diagonal elements provide a complete and intuitive measure of how informed students are. Large values indicate substantial uncertainty about future performance while small values suggest students know their abilities well.

These formulas demonstrate the important role of initial belief parameters γ_τ and

[11] Φ_l is a $K \times K$ matrix defined as,

$$\Phi_l\left(k,\, k'\right) = \begin{cases} \sigma_{kl}^2 & k = k' \\ 0 & k \neq k' \end{cases}$$

and D_{itl} is a $K \times K$ matrix defined as,

$$D_{itl}\left(k,\, k'\right) = \begin{cases} n_{itkl} & k = k' \\ 0 & k \neq k' \end{cases}$$

where n_{itkl} is the number of field k level l courses taken by individual i prior to semester t.

[12] The off diagonal elements of the matrix $\delta_{it\tau}$ measure the relatedness of uncertainty across academic fields.

Δ_τ in determining current beliefs. Students with different initial beliefs process grade signals very differently possibly resulting in differences in the relative information quality of grades across academic fields. To allow for heterogeneity in initial prior beliefs, I allow the distribution parameters γ_τ and Δ_τ to vary by unobserved type. This allows unobserved characteristics to influence students' initial confidence and certainty in their academic abilities.

1.4.4 Course Choices

As I detail in Section 1.5, if the distribution parameters γ_τ and Δ_τ vary by unobserved student type the estimation procedure requires a model for course choices. The goal of this model is not to provide choice parameters with economic meaning but rather to approximate the choice process in a way that fits observed choices well. For this reason, I employ a flexible quasi-structural form following Bernal and Keane (2010).

I assume utility from a field k level l course flexibly depends on expected grade $\tilde{g}_{it\tau kl}$, the number of beginner and advanced courses completed in this field n_{itk1} and n_{itk2}, a dummy variable for whether the student initially intended to major in field k I_{ik}, and intercepts which vary by whether the student is an upperclassmen or underclassmen.[13]

[13] Expected grades given current information sets consisting of observed covariates X_{itkl}, parameter values θ_k, and belief variables $\mu_{it\tau}$ and $\delta_{it\tau}$ are given by:

$$\tilde{g}_{itkl} = \mathbb{E}\left[g_{itckl} \,|\, X_{itk}, \theta_k, \mu_{it\tau}, \delta_{it\tau}\right] = X_{itkl}\theta_k + \mu_{it\tau}$$

Intended major data are from the CLL pre-college questionnaire. Respondents are given free space to write their intended major. Free responses are manually coded as Humanities, Social Science, Natural Science, Engineering, Economics, Undecided, and Missing. For respondents listing expected majors in two separate fields the first field is used. Coded responses are 56 Humanities, 183 Social Science, 267 Natural Science, 176 Engineering, 81 Economics, 315 undecided, and 54 missing. See Arcidiacono et al. (2012) for a further discussion of elicited expected major data in the CLL.

Utility from a field k level l course is given by:

$$\begin{aligned} U_{it\tau kl} &= \omega_{1kl}\tilde{g}_{it\tau kl} + \sum_{l'=1}^{L} \omega_{2kl}^{l'} n_{itkl'} + \omega_{3klp(t)} I_{ik} + \omega_{4klp(t)} + \epsilon_{itkl} \\ &= u_{it\tau kl} + \epsilon_{itkl} \end{aligned} \quad (1.7)$$

This model includes several important features: First, the term $\omega_{1kl}\tilde{g}_{it\tau kl}$ lets the importance of expected grades vary by field and level to allow students to place more value on higher marks in courses where they have more future value. Second, the term $\sum_{l'=1}^{L} \omega_{2kl}^{l'} n_{itkl'}$ lets the importance of past coursework in own-field beginner and advanced courses vary by field and level to flexibly account for pre-requisites and allow for path dependence in course choices. Third, the term $\omega_{3klp(t)} I_{ik}$ may increase the desirability of courses in the intended field of specialization and allows the effect to vary by field, difficulty level, and whether the student is an upperclassmen. Finally, the term $\omega_{4klp(t)}$ allows the general attractiveness of courses in each field-level pair to vary based on whether the student is an upperclassmen.

1.5 Expectation-Maximization Estimation Procedure

This section describes the adaptation of the Expectation-Maximization (EM) algorithm I use to estimate my ability revelation model. The EM Algorithm is an iterative approach to solve for the parameter values which maximize a likelihood function Dempster et al. (1977). It is useful for cases in which unobserved variables

Upperclassmen/underclassmen status is indicated by $p(t)$. Terms prior to the Fall semester of Junior year are assigned to underclassmen while terms after and including the Fall semester of Junior year are labeled upperclassmen.

make it computationally difficult to numerically maximize the likelihood function all at once. This adaptation of the EM algorithm is similar to the one used in James (2012).

I begin this section by deriving the full likelihood function which includes both the likelihood of observing course choices and the likelihood of observing grades given these choices. Next I provide a general overview of the EM algorithm. After this, I describe how my adaptation of the EM algorithm solves for the parameter values which maximize this likelihood function in the simplified case in which unobserved ability distribution parameters γ and Δ are the same for all students. I later relax this assumption and describe the algorithm in a general case where γ_τ and Δ_τ vary by unobserved student type τ.

1.5.1 Likelihood Function

Each individual's likelihood contribution depends on observed course choices and observed grades in chosen courses. To build the full likelihood function, I begin by presenting grade likelihood contributions conditional on ability α_i and choice likelihood contributions conditional on unobserved student type τ. I derive the full likelihood function by combining these conditional grade and choice contributions and integrating over unobserved ability and student type.

Grade Likelihood Conditional on Ability

The likelihood of observing grade g_{itklc} given ability α_i and parameter values θ_k and σ_{kl} is:

$$l\left(g_{itklc} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) = \frac{1}{\sqrt{2\pi\sigma_{kl}^2}} \exp\left(-\frac{\left(g_{itklc} - X_{itkl}\theta_k - \alpha_{ik}\right)^2}{2\sigma_{kl}^2}\right) \tag{1.8}$$

where the structure arises from the assumption that grading noise is normally distributed.

Because transitory grading noise η_{itklc} is independent across courses, the likelihood of observing the grade vector $\mathbf{g_i}$ conditional on ability α_i and parameter values θ_k and σ_{kl} is given by:

$$\mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) = \prod_{t=1}^{T}\prod_{c=1}^{C_{it}}\prod_{k=1}^{K}\prod_{l=1}^{L} l\left(g_{itklc} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right)^{\left(d_{itc}=\{k,l\}\right)} \tag{1.9}$$

This represents individual i's grade likelihood contribution conditional on her ability α_i.

Choice Likelihood Conditional on Student Type

To develop choice likelihood contributions conditional on unobserved type I begin by assuming preference shocks ϵ_{itklc} are iid type 1 extreme value. The probability that individual i chooses a field k level l course in semester t given student type, grade parameter values γ_τ, θ_k, σ_{kl}, Δ_τ, and choice parameter values ω is then given by:

$$\Pr\left(d_{itc} = \{k,l\} \mid \tau, \gamma_\tau, \theta_k, \sigma_{kl}, \Delta_\tau, \omega\right) = \frac{\exp\left(u_{itkl}\right)}{\sum_{k'=1}^{K}\sum_{l'=1}^{L}\exp\left(u_{itk'l'}\right)} \tag{1.10}$$

where u_{itkl} is the deterministic component of choice utility presented in subsection 1.4.4.

After conditioning on unobserved student type, past grades, and observed characteristics, choices are independent of α_i. This is because course utility depends on expected grades $\tilde{g}_{it\tau kl}$ at the beginning of period t. These expectations are formed using current beliefs about α_i rather than true α_i. Following formula 1.5, beliefs about

α_i are a function of type specific initial prior beliefs, model parameters, and grade residuals. As such, unobserved student type, past grades, observed characteristics, and model parameters contain all information necessary to capture how beliefs about unobserved α_i influence course choices (James (2012)).

Because preference shocks are independent across courses, the likelihood of observing the choice vector $\mathbf{d_i}$ conditional on type, grade parameter values γ_τ, θ_k, σ_{kl}, Δ_τ, and choice parameter values ω is given by:

$$\mathscr{L}_i^c\left(\mathbf{d_i}\,|\,\tau,\Theta\right) = \prod_{t=1}^{T}\prod_{c=1}^{C_{it}}\prod_{k=1}^{K}\prod_{l=1}^{L}\Pr\left(d_{itc} = \{k,l\}\ |\ \tau,\Theta,\omega\right)^{\left(d_{itc}=\{k,l\}\right)} \tag{1.11}$$

where Θ contains grading parameters γ_τ, θ_k, σ_{kl}, Δ_τ, and ω.

Full Likelihood

At this point, I have constructed grade likelihood contributions conditional on unobserved ability α_i and choice likelihood contributions conditional on unobserved student type. To form individual likelihood contributions I combine grade likelihood contributions conditional on ability and choice likelihood contributions conditional on type and integrate over unobserved type and unobserved abilities conditional on type. Because choice likelihood contributions do not depend on unobserved abilities the integration over unobserved abilities conditional on type only pertains to the grade component of the likelihood. This yields the following individual likelihood contributions:

$$\mathscr{L}_i = \sum_{\tau=1}^{\mathscr{T}}\pi_\tau\mathscr{L}_i^c\left(\mathbf{d_i}\,|\,\tau,\Theta\right)\left[\int_{\alpha_i}\mathscr{L}_i^g\left(\mathbf{g_i}\,|\,\alpha_i,\theta_k,\sigma_{kl}\right)f\left(\alpha_i\,|\,\gamma_\tau,\Delta_\tau\right)d\alpha_i\right] \tag{1.12}$$

where π_τ is the unconditional probability of being type τ and $f(\cdot)$ is the K-dimensional normal PDF.

The full log-likelihood is then given by:

$$\ln \mathscr{L} = \sum_{i=1}^{N} \ln \mathscr{L}_i \tag{1.13}$$

1.5.2 EM Algorithm: Overview

The K-dimensional integration of α_i for every student i as each possible unobserved type τ makes it computationally burdensome to calculate the log-likelihood function value at specific parameter values. This means numerical maximization techniques which rely on repeated evaluations of the maximand will be unfeasibly slow. The EM algorithm is a useful method for maximizing likelihood functions such as this one which contain a large number of unobserved variables. Arcidiacono and Jones (2003) show the algorithm can be adapted to restore additive separability within the likelihood function making it possible to solve for maximizing parameter values sequentially rather than jointly. This greatly reduces computational burden at the cost of modest efficiency losses.

In general, the EM algorithm involves iteratively repeating two steps: In the E step, observed data and guesses for parameter values are used to estimate probability distributions for each individual's unobserved variables. Intuitively, these probability distributions are estimated by evaluating how likely the individual's observed data are to have occurred for all possible values of the unobserved variables. These probability distributions are then used to construct an expected likelihood function given these distributions. In the M step, this expected likelihood function is maximized with respect to all parameter values. The E step is then repeated using these updated parameter estimates and the algorithm iterates until sequential parameter estimates become arbitrarily close.

1.5.3 EM Algorithm: Homogenous Distributions of Unobserved Ability

I now describe my adaptation of the EM algorithm in the special case where all students draw unobserved abilities as $\alpha_i \sim N(\gamma,\, \Delta)$. This is equivalent to assuming all students enter college with the same beliefs about the unobserved component of their academic abilities.[14]. In Section 1.5.4, I describe the algorithm in a general case where γ_τ and Δ_τ vary by unobserved student type τ.

Likelihood Function: Homogenous Distributions of Unobserved Ability

Without unobserved variation in initial beliefs, the full log likelihood is given by:

$$\ln \mathscr{L} = \sum_{i=1}^{N} \ln \mathscr{L}_i^c \left(\mathbf{d_i} \,|\, \Theta\right) + \sum_{i=1}^{N} \ln \left[\int_{\alpha_i} \mathscr{L}_i^g \left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma, \Delta\right) d\alpha_i \right] \tag{1.14}$$

While it is less efficient, the additive separability allows the grade parameters γ, θ_k, σ_{kl}, and Δ to be consistently estimated by maximizing the grade component of the likelihood only. This obviates the need to model course choices and reduces the computational burden of the estimation routine. However, maximizing the grade component of the likelihood still requires K-dimensional integration of α_i for ev-

[14] The plausibility of this assumption depends on the richness of student information sets at the time of enrollment relative to the richness of controls included in X_{itkl}. If entering students can predict their academic performance more accurately than the econometrician's prediction $X_{itkl}\hat{\theta}_k$ then students initially possess private information about α_i which violates this assumption of identical initial beliefs. As such, this assumption is likely to be violated when X_{itkl} includes few controls but is more plausible when X_{itkl} includes rich controls.

ery student which is computationally challenging. I use an adaptation of the EM algorithm to avoid this computational burden.

EM Algorithm: Homogenous Distributions of Unobserved Ability

In this case, the unobserved variable confounding estimation is α_i. As such, the E step involves estimating the distribution of α_i given observed grades and parameter estimates from the previous iteration. From a Bayesian perspective, the most informed distribution of α_i can be derived by updating iteration m estimates of the initial prior distribution $\alpha_i \sim N(\gamma^m, \Delta^m)$ with grade signals from every grade earned by student i. This results in iteration m terminal posterior distributions $\alpha_i \sim N(\xi_i^m, \Upsilon_i^m)$. The exact formulas for ξ_i^m and Υ_i^m are given in Appendix A. These distributions represent student beliefs at the time of graduation and thus represent the most informed distributions achievable with grade data.

Next, I use these iteration m individual specific distributions to form the following expected log-likelihood function conditional on individual specific distribution parameters ξ_i^m and Υ_i^m:

$$Q_{hom}\left(\Theta_g \,|\, \Theta_g^m\right) = \sum_{i=1}^{N} \int_{\alpha_i} \ln \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \xi_i^m, \Upsilon_i^m\right) d\alpha_i \qquad (1.15)$$

The M step maximizes this expected log-likelihood over the set of grade parameters $\Theta_g = \{\gamma, \Delta, \theta_k, \sigma_{kl}\}$ taking iteration m estimates as given. The algorithm is then repeated using the solution to this m^{th} iteration maximization as iteration $m+1$ parameter estimates. Final estimates are obtained when the algorithm converges such that sequential iteration estimates are arbitrarily close. In this special case of homogenous ability distributions, all components of the solution to the maximization problem have simple closed form expressions which are presented in Appendix A.

1.5.4 EM Algorithm: Heterogeneous Distributions of Unobserved Ability

I now describe my adaptation of the EM algorithm in the general case in which a student's unobserved type τ determines the distribution her unobserved abilities are drawn from. I allow type specific distributions to have both type specific means γ_τ and type specific covariance matrices Δ_τ. This allows for unobserved differences in initial confidence and in initial certainty about beliefs. A student's type is observed by the student but is not observed by the econometrician.

In this general setting, I solve for the parameters which maximize the full log-likelihood using a nested EM approach (James (2012)). The 'outer' EM component deals with unobserved student types and allows the full likelihood to be additively separated into choice and grade contributions. The 'inner' EM component addresses the K-dimensional integration over unobserved abilities. The inner EM is analogous to the EM algorithm used in the simplified case of homogenous distributions of unobserved ability described in Section 1.5.3.

The E step of the outer EM algorithm involves estimating the probability that each student is each possible type conditional on observed data and guesses for parameter values. Intuitively, these type probabilities are derived by comparing the likelihood of observing individual i's actual grades and choices if she were type τ to the likelihood of observing her grades and choices if she were other types. Denote the m iteration estimates of these probabilities by $q_{i\tau}^m$. The exact formula for estimating $q_{i\tau}^m$ is given in Appendix A.

These conditional type probabilities are then used to form the following expected log-likelihood function conditional on type probabilities $q_{i\tau}^m$:

$$\tilde{Q}\left(\Theta \,|\, \Theta^m\right) = \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \left\{ \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \Theta\right) \left[\int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau^m, \Delta_\tau^m\right) d\alpha_i \right] \right\} \tag{1.16}$$

Notice this function can be additively separated into:

$$\begin{aligned} \tilde{Q}\left(\Theta \,|\, \Theta^m\right) &= \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \Theta\right) \\ &\quad + \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \int_{\alpha_i} \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \gamma_\tau^m, \Delta_\tau^m\right) d\alpha_i \end{aligned} \tag{1.17}$$

The M step of the outer EM algorithm involves maximizing $\tilde{Q}\left(\Theta \,|\, \Theta^m\right)$ over parameters $\Theta = \{\gamma_\tau, \Delta_\tau, \theta_k, \sigma_{kl}, \omega\}$ taking $q_{i\tau}^m$ as given. While the log operator can be used to greatly simplify the choice contribution to the expected likelihood, the K-dimensional integration over unobserved abilities still makes it challenging to numerically maximize $\tilde{Q}\left(\Theta \,|\, \Theta^m\right)$. Including the inner EM component addresses this issue.

To include the inner EM component, the E step is expanded to include estimates of type specific probability distributions for each individual's unobserved ability vector α_i conditional on all earned grades. Similar to the case of homogenous distributions of unobserved ability, the distributions are estimated by updating type specific prior distributions $\alpha_i \sim N\left(\gamma_\tau, \Delta_\tau\right)$ with signals from all grades earned by individual i to form individual-type specific posterior distributions $N\left(\xi_{i\tau}^m, \Upsilon_{i\tau}^m\right)$. The exact formulas for $\xi_{i\tau}^m$ and $\Upsilon_{i\tau}^m$ are given in Appendix A. These estimated type specific distributions for α_i are then combined with conditional type probabilities $q_{i\tau}^m$ to form the following expected log-likelihood function conditional on estimates of ability distributions and type probabilities:

$$\begin{aligned} Q_{het}\left(\Theta \,|\, \Theta^m\right) &= \sum_{i=1}^{N}\sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \ln \mathscr{L}_i^c\left(\mathbf{d_i} \,|\, \tau, \Theta\right) \\ &\quad + \sum_{i=1}^{N}\sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \int_{\alpha_i} \ln \mathscr{L}_i^g\left(\mathbf{g_i} \,|\, \alpha_i, \theta_k, \sigma_{kl}\right) f\left(\alpha_i \,|\, \xi_{i\tau}^m, \Upsilon_{i\tau}^m\right) d\alpha_i \quad (1.18) \end{aligned}$$

Once again, the M step entails maximizing $Q_{het}\left(\Theta \,|\, \Theta^m\right)$ over parameters $\Theta = \{\gamma_\tau, \Delta_\tau, \theta_k, \sigma_{kl}, \omega\}$ taking $q_{i\tau}^m$, $\xi_{i\tau}^m$, and $\Upsilon_{i\tau}^m$ as given. As before, the algorithm is repeated using the solution to $\text{argmax}_\Theta Q_{het}\left(\Theta \,|\, \Theta^m\right)$ as iteration $m+1$ parameter estimates. Final estimates are obtained when the algorithm converges such that $\Theta^\star = \text{argmax}_\Theta Q_{het}\left(\Theta \,|\, \Theta^\star\right)$. The advantage of this approach is many components of the solution to $\text{argmax}_\Theta Q_{het}\left(\Theta \,|\, \Theta^m\right)$ have simple closed form expressions. Whereas the full likelihood is nearly impossible to numerically maximize, solutions to $\text{argmax}_\Theta Q_{het}\left(\Theta \,|\, \Theta^m\right)$ can be computed in a matter of minutes. A full characterization of the solution to $\text{argmax}_\Theta Q_{het}\left(\Theta \,|\, \Theta^m\right)$ is given in Appendix A.

1.6 Results

In this section, I use estimates of my correlated learning model to compare expected grades and signal quality across academic fields. I begin by presenting results which assume there is no unobserved heterogeneity in the distribution parameters γ and Δ. These are important parameters for analyzing signal quality because they characterize initial prior beliefs about the unobserved component of grade production. I later relax this assumption by allowing for type specific unobserved heterogeneity in γ and Δ with two unobserved types.

As I discussed in Section 1.4, information parameters are sensitive to which variables are included as observed covariates in the grade production function. This is because the model assumes students arrive at Duke knowing how these covariates affect grade production. Students use this information to form initial beliefs about their performance and also to extract grade residuals used to update these beliefs. Intuitively, a richer set of observed covariates imply students have less unobserved ability to reveal. To address sensitivity to this empirical choice, I report four specifications: specification 1 only controls for whether the course is beginner or advanced; specification 2 also includes a quadratic for number of courses taken in this field previously and dummy variables for contemporaneous course load; specification 3 adds race and gender dummy variables; and specification 4 adds math and verbal SAT scores and five measures of the student's application quality.[15] The parsimonious specification 1 assumes students know little about the effects of covariates on grade production and can only inform their beliefs with grade signals—this likely understates the richness of student information sets. Conversely, the rich specification 4 assumes students understand the role of covariates quite well. A careful analysis by Arcidiacono et al. (2011) suggests this specification overstates the richness of student information sets. Results which allow for unobserved heterogeneity in γ and Δ also address sensitivity to this choice by including unobserved characteristics in student information sets.

[15] Admissions officers at Duke grade applications along five dimensions: academic achievement, high school curriculum difficulty, essay quality, personal qualities, and letters of recommendation. Grading scale is from 1 to 5; average grade is 3.88, standard deviation is .81.

1.6.1 Homogenous Ability Distributions

Expected Grades by Academic Field and Course Level

Table 1.2 reports expected grades across academic fields and course levels conditional on observed covariates for four possible specifications.[16] These results are computed using estimates of my correlated learning model which controls for selection on observed covariates and unobserved beliefs about abilities; as such, they are relevant to the entire student population (conditional on observed covariates) rather than a selected population. The results imply a student with specified characteristics expects to earn substantially higher grades in Humanities and Social Science (HuSS) courses relative to Science, Engineering, and Economics (SEE) courses. In beginner courses, the difference between harshest grading Economics and most lenient Humanities ranges from .45 - .53 grade points.[17] In advanced courses, the difference between harshest grading Science and most lenient Humanities ranges from .29 - .40 grade points.

The results for specification 1 in Table 1.2 can be directly compared to the descriptive results in panel A of Table 1.1 to assess the importance of sorting on information revealed through grades. For beginner courses, expected grades are similar with the exception of Engineering where controlling for sorting on learned ability decreases expected grades by .1 grade points; for advanced courses, controlling for sorting on learned ability decreases expected grades by .09 grade points for both Engineering and Science courses. This indicates students who have performed well in the past are

[16] When covariates are included student is white, male, taking four classes, has no prior experience for beginner courses and five courses of own category experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures.

[17] For reference, this is greater than the difference between a B+ and an A- (.4 grade points).

Table 1.2: Expected Grades by Academic Field and Course Level

	(1)		(2)		(3)		(4)	
	Beg	Adv	Beg	Adv	Beg	Adv	Beg	Adv
Humanities	3.507	3.567	3.474	3.534	3.469	3.529	3.480	3.538
	0.014	*0.011*	*0.015*	*0.013*	*0.024*	*0.022*	*0.023*	*0.022*
Soc Science	3.369	3.527	3.340	3.512	3.347	3.517	3.358	3.522
	0.016	*0.010*	*0.017*	*0.011*	*0.025*	*0.021*	*0.025*	*0.021*
Science	3.095	3.169	3.059	3.105	3.169	3.214	3.176	3.211
	0.021	*0.020*	*0.022*	*0.022*	*0.038*	*0.037*	*0.037*	*0.038*
Engineering	3.140	3.342	3.165	3.320	3.215	3.377	3.256	3.417
	0.036	*0.031*	*0.037*	*0.034*	*0.052*	*0.045*	*0.061*	*0.057*
Economics	3.025	3.373	2.973	3.338	3.060	3.423	3.057	3.422
	0.029	*0.022*	*0.031*	*0.032*	*0.040*	*0.043*	*0.042*	*0.052*
Course Level	Yes		Yes		Yes		Yes	
Experience	No		Yes		Yes		Yes	
Course Load	No		Yes		Yes		Yes	
Gender	No		No		Yes		Yes	
Race	No		No		Yes		Yes	
SAT Scores	No		No		No		Yes	
App quality	No		No		No		Yes	
Course Obs	37255		37255		37255		37255	
Student Obs	1127		1127		1127		1127	

more likely to take beginning or advanced Engineering courses and advanced Science courses. This is consistent with existing literature which finds early academic performance determines whether students persist in STEM fields (Arcidiacono (2004); Arcidiacono et al. (2012); Stinebrickner and Stinebrickner (2014b)).

Decomposition of Variance in Grades

Table 1.3 reports variance in grades across academic fields and course levels and decomposes these variances into variance from observed covariates, variance from unobserved ability, and variance from idiosyncratic grading noise to assess signal quality across academic fields and levels.[18] These results are produced using estimates of the richest specification of my correlated learning model which includes controls for experience, course load, gender, race, SAT scores, and application quality measures. Once again, the correlated learning model controls for non-random course selection which implies the results represent variance in the entire student population rather than variance in a selected population.

The results show there is substantially higher variance in SEE grades relative to HuSS grades for both beginner and advanced courses. For beginner courses, the most variable field is Science which has an interquartile range of 1.09 grade points; the least variable field is Social Science which has an interquartile range of .80 grade points. For advanced courses, the most variable field is Science which has an interquartile range of 1.05 grade points; the least variable field is Humanities which has an interquartile range of .72 grade points. This is similar to the pattern in descriptive variances presented in Table 1.1 Panel A. This demonstrates high variance in SEE grades and low variance in HuSS grades is not a result of selection on observed covariates or unobserved beliefs about abilities.

The decomposition analysis shows idiosyncratic grading noise contributes the

[18] To facilitate a clear comparison across fields and levels, I exclude variance in grades which arises because students have different experience or course loads. Let $\tilde{X}_{itkl}$ represent observed covariates other than experience and course load and let $X_{itkl}\backslash\tilde{X}_{itkl}$ represent these experience and course load covariates. If $X_{itkl}\backslash\tilde{X}_{itkl}$ are fixed, variance in g_{itklc} can be additively decomposed as:

$$\text{Var}\left(g_{itklc}\right) = \text{Var}\left(\tilde{X}_{itkl}\tilde{\theta}_k\right) + \Delta\left(k,k\right) + \sigma_{kl}^2$$

where $\tilde{\theta}_k$ are the coefficients corresponding to $\tilde{X}_{itkl}$. The first term represents variance due to differences in observed covariates, the second term represents variance due to differences in unobserved abilities, and the third term represents variance due to idiosyncratic grading noise.

Table 1.3: Variance Decomposition (Specification 4)

Beginner	Hum	Soc Sci	Sci	Eng	Econ
Total Var	0.3562	0.3476	0.6490	0.5985	0.5652
Std. Error	*0.0144*	*0.0134*	*0.0161*	*0.0403*	*0.0293*
Share of variance due to					
Obs Covariates	7.5%	7.6%	10.9%	16.1%	9.5%
Unobs Ability	18.0%	20.9%	29.1%	30.8%	23.8%
Grading Noise	74.6%	71.5%	60.0%	53.1%	66.7%
Advanced	Hum	Soc Sci	Sci	Eng	Econ
Total Var	0.2842	0.3122	0.6101	0.5850	0.4799
Std. Error	*0.0124*	*0.0144*	*0.0189*	*0.0437*	*0.0272*
Share of variance due to					
Obs Covariates	9.3%	8.4%	11.5%	16.4%	11.2%
Unobs Ability	22.5%	23.3%	31.0%	31.5%	28.0%
Grading Noise	68.2%	68.3%	57.5%	52.0%	60.8%

majority of variance (53.1% - 74.6%), unobserved ability contributes the second most variance (18% - 31.5%), and observed covariates contribute the least variance (7.5% - 16.4%). This has several important implications: First, it shows that even with a rich set of controlling covariates there is substantial information to be revealed by grades. The interquartile ranges for field specific unobserved abilities are .34 grade points for Humanities, .36 grade points for Social Science, .59 grade points for Science, .58 grade points for Engineering, and .49 grade points for Economics. This implies that an incoming student with a very rich information set still believes that with 50% probability her expected grades in Science courses will be outside a range of .59 grade points. This is substantial uncertainty which makes it difficult to make informed academic decisions about which major to choose and which courses to enroll in. This highlights the importance of grades as a mechanism for revealing this missing information. Second, it shows that even a rich set of covariates do a poor job of capturing persistent heterogeneity across students. As such, research which aims to assess the effects of academic choices on later outcomes should be careful to control for confounding selection on unobserved factors.

The decomposition analysis also shows HuSS grades are noisier than SEE grades. In beginner courses, 71.5% - 74.6% of variation in HuSS grades is due to grading noise versus 53.1% - 66.7% for SEE grades; in advanced courses, 68% of variation in HuSS grades is due to grading noise versus 52% - 60.8% for SEE grades. As in Section 1.3, this shows that while HuSS grades have a tight distribution, the same student earns HuSS grades which span the full distribution. Conversely, SEE grades have a more disperse distribution but the same student typically earns similar grades in different SEE courses. This suggests SEE grades reveal missing information more efficiently than HuSS grades.

Signal Quality by Academic Field

The variance decomposition in Table 1.3 shows HuSS grades are generally noisier than SEE grades but does not fully measure signal quality across academic fields because it ignores information spillovers to other fields. If unobserved abilities are correlated, grade signals in one field provide information about unobserved abilities in all fields; this spillover information must be included to fully measure information quality.

To achieve this, Table 1.4 reports initial uncertainty across fields and the reduction in uncertainty which occurs after earning one grade from each field. Initial uncertainty is measured by the diagonal elements of Δ which represent each individual's variance in initial beliefs about α_i. Reductions in uncertainty are measured by comparing variances in initial beliefs to variances in beliefs after one field k grade signal is received. These variances in revised beliefs are given by the diagonal elements of $\delta(e_k)$ where:[19]

$$\delta(e_k) = \left(\Delta^{-1} + \text{diag}(e_k)\,\Phi_1^{-1}\right)^{-1} \tag{1.19}$$

Notice variance in revised beliefs depend on signal parameters Δ and σ_{kl} but do not depend on grades earned or individual characteristics. As such, the reduction from $\text{diag}(\Delta)$ to $\text{diag}(\delta(e_k))$ provides an intuitive measure of field k information quality which is relevant for the entire population of students.

The variance decomposition in Table 1.3 shows there is substantial information left to be revealed by grades—results in Table 1.4 suggest SEE grades reveal this missing information more efficiently than HuSS grades. Engineering consistently offers the largest average reduction in uncertainty while Humanities offers the smallest average reduction. Moreover, Engineering signals are so informative that they almost always reduce uncertainty in beliefs about abilities in other categories more than grades from these categories. For example, in specification 4, one engineering grade reduces uncertainty in beliefs about humanities ability by 24% while one humanities grade only reduces uncertainty in beliefs about humanities ability by 19%;

[19] Φ_1 is a $K \times K$ diagonal matrix containing noise variances in beginner classes:

$$\Phi_1(k, k') = \begin{cases} \sigma_{k1}^2 & k = k' \\ 0 & k \neq k' \end{cases}$$

$\text{diag}(e_k)$ is a sparse $K \times K$ matrix defined by:

$$\text{diag}(e_k)(k, k') = \begin{cases} 1 & k = k' \\ 0 & k \neq k' \end{cases}$$

Table 1.4: Learning through Grade Signals

Panel A: No Covariates		Reduction (%) after one beginner grade (by field)				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.095	26	22	25	35*	21
Social Science	0.102	20	29	24	30	26
Science	0.277	16	17	41	44	30
Engineering	0.281	19	18	39	47	29
Economics	0.200	17	22	37	41*	33
Average		20	22	33	39	28
Panel B: Exp and Load		Reduction (%) after one beginner grade (by field)				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.095	26	24	26	37*	23
Social Science	0.106	21	30	23	31	27
Science	0.282	16	17	42	45	31
Engineering	0.283	21	20	40	47	31
Economics	0.204	18	23	38	42*	34
Average		20	23	34	40	29
Panel C: + Demographics		Reduction (%) after one beginner grade (by field)				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.085	24	22	23	32*	20
Social Science	0.093	19	27	20	26	24
Science	0.235	15	15	38	41	28
Engineering	0.229	18	17	36	42	28
Economics	0.168	16	21	34	38*	31
Average		18	20	30	36	26
Panel D: + Scores		Reduction (%) after one beginner grade (by field)				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.063	19	17	17	24	15
Social Science	0.072	14	23	14	20	20
Science	0.187	10	10	33	36	23
Engineering	0.180	13	12	32	37	23
Economics	0.135	11	17	28	33	26
Average		13	16	25	30	21

one engineering grade reduces uncertainty in beliefs about social science ability by 20% compared to 23%; for science the reduction is 35% relative to 33%; and for economics the reduction is 32% relative to 27%.[20] This paradoxical result shows the importance of allowing for spillovers across fields—these spillovers are so powerful that students can learn more about their unobserved abilities in one field by taking a course in another field with more precise grade signals.

1.6.2 Heterogeneous Ability Distributions

In this subsection, I present results which allow for type specific unobserved heterogeneity in the distribution parameters γ_τ and Δ_τ. These distribution parameters characterize initial prior beliefs about the unobserved component of grade production. In a Bayesian setting, initial prior beliefs inform all subsequent beliefs making the information revelation process very sensitive to the distribution parameters γ_τ and Δ_τ. For example, larger values in γ_τ imply students initially expect to perform better than their observed characteristics suggest—grades which do not meet these high expectations result in downward revisions in beliefs. Additionally, smaller diagonal values in Δ_τ imply students have more certain initial beliefs and will not update their beliefs as much when they receive grades. With two unobserved types I find one type has much lower initial expectations and higher uncertainty in these expectations. These students also experience fewer information spillovers across academic fields than their classmates with higher and more certain expectations.

[20] In Table 1.4, (*) indicates cases in which non-field signals are statistically more informative than own field signals at 5% significance. For example: in Panel A, the 35% reduction in humanities uncertainty after receiving one engineering grade is statistically greater than the 26% reduction from one humanities grade.

Expected Grades and Type Characteristics

Panel A of Table 1.5 reports expected grades across academic field and levels conditional on observed covariates for each unobserved type[21]. Panel B of Table 1.5 reports type frequencies, average course choices by type, and average baseline characteristics by type. For brevity, I only report results for specification 1 and specification 4.[22] As in Table 1.1 Panel A and Table 1.2, students of both unobserved types expect to earn significantly lower grades in SEE courses relative to HuSS courses. This shows the differences in grading leniency between SEE and HuSS courses is not driven by sorting on initial beliefs about unobserved abilities.

The results also show 10-13% of students—labeled as Type 1—arrive at Duke with substantially lower expected grades in all five academic fields. The smallest difference is in Social Science courses where Type 2 students expect to perform .50 - .60 grade points better than Type 1 students. The largest difference is in Engineering courses where Type 2 students expect to perform .65 - .87 grade points better than Type 1 students. Type 1 students also have larger differences between expected grades in SEE courses and expected grades in HuSS courses. In beginner courses, the difference between harshest grading Economics and most lenient Humanities is .53 - .63 grade points for Type 1 students relative to .40 - .48 grade points for Type 2 students. In advanced courses, the difference between harshest grading Science and most lenient Humanities is .47 - .52 grade points for Type 1 students relative to .31 - .39 grade points for Type 2 students. This shows Type 2 students initially expect

[21] When covariates are included student is white, male, taking four classes, has no prior experience for beginner courses and five courses of own field experience for advanced courses (0 and 5 are the respective modes), and has average SAT scores and application quality measures.

[22] Results for specifications 2 and 3 are available upon request.

Table 1.5: Expected Grades and Type Characteristics

	Specification 1				Specification 4			
	Type 1		Type 2		Type 1		Type 2	
	Beg	Adv	Beg	Adv	Beg	Adv	Beg	Adv
Panel A: Expected Grades								
Humanities	2.977	3.036	3.589	3.649	2.997	3.059	3.517	3.579
	0.071	*0.071*	*0.026*	*0.021*	*0.070*	*0.072*	*0.027*	*0.022*
Soc Science	2.850	3.007	3.448	3.606	2.881	3.051	3.383	3.553
	0.104	*0.103*	*0.023*	*0.016*	*0.083*	*0.083*	*0.026*	*0.022*
Science	2.450	2.520	3.194	3.264	2.551	2.587	3.234	3.270
	0.123	*0.122*	*0.040*	*0.037*	*0.109*	*0.113*	*0.038*	*0.040*
Engineering	2.389	2.588	3.256	3.454	2.641	2.807	3.290	3.456
	0.152	*0.157*	*0.050*	*0.045*	*0.136*	*0.134*	*0.062*	*0.050*
Economics	2.447	2.791	3.114	3.458	2.364	2.729	3.118	3.482
	0.164	*0.162*	*0.037*	*0.031*	*0.125*	*0.131*	*0.046*	*0.043*
Panel B: Type Characteristics								
% of Students	13.1%		86.9%		10.4%		89.6%	
Coursework								
Humanities	11.37		10.31		11.66		10.31	
Soc Science	11.32		8.73		10.45		8.91	
Science	7.73		9.52		8.20		9.41	
Engineering	1.07		1.66		1.38		1.61	
Economics	2.43		2.36		2.34		2.37	
Characteristics								
HS: Religious	14.6%		10.7%		12.6%		11.1%	
HS: Private	16.6%		21.0%		17.4%		20.8%	
Parent: College	21.7%		15.7%		21.2%		16.0%	
Parent: Some grad	35.9%		33.7%		35.4%		33.8%	
Parent: Prof.	26.2%		42.7%		31.3%		41.6%	
AP scores	3.19		4.73		3.77		4.62	
Course Obs	37255				37255			
Student Obs	1127				1127			

to possess both an absolute advantage in all fields and a relative advantage in SEE courses when compared to Type 1 students.

To analyze the relationship between initial expectations and course choices Panel B of Table 1.5 reports the average number of courses taken in each field for each unobserved type. Results show Type 2 students take an average of 1.8 more Science courses and .6 more Engineering courses than Type 1 students. This suggests higher expected grades and an expected relative advantage in SEE courses is associated with taking more Science and Engineering courses. This is consistent with existing literature which concludes expected grades largely determine whether a student pursues Science or Engineering (Arcidiacono (2004); Arcidiacono et al. (2012); Stinebrickner and Stinebrickner (2014b)).

To understand what baseline covariates are associated with differences in initial expectations, Panel B of Table 1.5 also reports average high school type, education of the most educated parent, and number of Advanced Placement (AP) scores reported for each unobserved type.[23] Results show Type 2 students are more likely to have attended private high schools (instead of public or religious high schools), are more likely to have at least one parent with a Professional degree (instead of a Bachelor's degree or some graduate coursework), and report more AP scores on average than Type 1 students. This suggests higher socioeconomic status and more exposure to college level coursework are associated with higher initial expectations and an expected relative advantage in SEE courses.

[23] AP courses are nationally standardized college level courses taken in high school. After taking an AP course, students have the option to take a standardized test of the material. Scores on these tests may be used to bolster college applications and earn college credit for the coursework. These are typically the most advanced courses taken by high school students and they are more widely available at high schools with many high achieving students.

Signal Quality by Academic Field

To understand how these differences in initial expectations relate to ability revelation Table 1.6 conducts the ability revelation analysis of Table 1.4 separately for each unobserved type. This analysis reports the initial uncertainty in beliefs about unobserved abilities as well as the reductions in uncertainty following each possible grade signal. Once again, I focus on specification 1 and specification 4.[24]

The initial uncertainty results suggest low expectation type 1 students have substantially higher uncertainty in their initial expectations. In the richest specification, interquartile ranges for beliefs range from .44 to .71 grade points for type 1 students and from .23 to .50 grade points for type 2 students. One reason why low expectation students might also have high uncertainty is noise in the admissions process. Intuitively, lower ability students will only be admitted if they produce unusually strong applications which is more likely to occur if there is more uncertainty about their abilities. Conversely, higher ability students will only be rejected if they produce unusually weak applications which is more likely to occur if there is more uncertainty about their abilities. Appendix B formalizes this intuition with a simple model.

Alternatively, a negative relationship between expectations and uncertainty may exist because certain pre-college experiences both increase expectations and reduce uncertainty. Panel B of Table 1.5 shows high expectation and low uncertainty Type 2 students have higher socio-economic status and more exposure to college coursework during high school. This suggests having more educated parents and taking more AP courses may increase expected performance and reduce uncertainty about

[24] Results for specifications 2 and 3 are available upon request

Table 1.6: Learning through Grade Signals

Panel A1: Spec 1 Type 1		Reduction (%) after one beginner grade (by field)				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.1419	35	11	2	2	2
Social Science	0.1592	10	39	10	6	14
Science	0.2564	2	9	39	37	10
Engineering	0.2516	1	5	33	45	6
Economics	0.1652	2	19	13	9	30
Average		10	17	19	20	12
Panel A2: Spec 1 Type 2		**Reduction (%) after one beginner grade (by field)**				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.0408	13	11	26*	30*	22*
Social Science	0.0429	11	14	19*	22*	21*
Science	0.2075	10	8	35	35	25
Engineering	0.1921	11	9	32	38	24
Economics	0.1546	10	11	31	33	28
Average		11	11	29	32	24
Panel B1: Spec 4 Type 1		**Reduction (%) after one beginner grade (by field)**				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.1206	31	10	0	6	2
Social Science	0.1733	7	41	3	14	3
Science	0.2039	0	4	34	39	4
Engineering	0.2797	4	12	29	47	3
Economics	0.1060	3	5	6	7	22
Average		9	14	14	23	7
Panel B2: Spec 4 Type 2		**Reduction (%) after one beginner grade (by field)**				
	Initial	Human	Social Sci	Science	Engineer	Econ
Humanities	0.0290	10	9	15*	20*	11
Social Science	0.0328	8	12	10	14	11
Science	0.1357	6	4	26	28	15
Engineering	0.1349	7	5	25	30	15
Economics	0.0800	6	7	22*	26*	18
Average		7	7	20	24	14

performance.

The reduction percentage results suggest there is substantial heterogeneity in how students process information. Low expectation type 1 students experience modest information spillovers between Humanities and Social Science, Social Science and Economics, and Natural Science and Engineering while high expectation type 2 students experience substantial spillovers between all fields. For type 1 students, modest spillovers imply own field signals are always a much more efficient way to reveal abilities: In the richest specification a Humanities signal decreases Humanities uncertainty by 31% versus 0% - 10% from other signals; Social Science grades yield an own field reduction of 41% compared to 3% - 14%; Natural Science yields a 34% reduction relative to 0% - 39%; Engineering yields a 47% reduction versus 3% - 29%; and Economics yields a 22% reduction compared to 3% - 7%. Conversely, for type 2 students, spillovers are so strong that precise SEE signals always reveal more about HuSS abilities than HuSS signals. In the richest specification an Engineering grade reduces Humanities uncertainty by 20% and Social Science uncertainty by 14% whereas Humanities and Social Science grades reduce these uncertainties by 10% and 12% respectively.

This finding paints an interesting picture of how students at Duke process grade information. Most Duke students expect to perform well and are fairly confident in these expectations. These students are mostly uncertain about factors which are general to the college experience such as their ability to adjust to new instruction formats or their ability to develop effective study habits at Duke. These students are more likely to have attended a private high school, more likely to have highly educated parents, and have completed more advanced placement courses prior to arriving at Duke. These pre-college experiences may have revealed their aptitudes for different academic fields leaving only uncertainty about general aspects of the transition to college. For these students, precise grade signals from SEE classes reveal this general information most efficiently.

There is also a small minority of students who expect to perform much worse but are uncertain about these expectations. For these students, a large share of the uncertainty comes from factors which are specific to academic fields. These students must take courses in each academic field to reveal this field specific information.

1.7 Conclusion

This paper compares the signal quality of grades across academic fields by estimating a correlated learning model using transcript data from Duke University. The model allows for both 'own field' signal quality and 'spillover' signal quality and controls for selection on unobserved beliefs about abilities. Estimates from specifications with no unobserved heterogeneity in distribution parameters suggest grades in Science, Engineering, and Economics (SEE) courses are substantially more informative than grades in Humanities and Social Science (HuSS) courses. In some of these specifications, cross-field spillovers are so powerful that grades in Engineering courses reveal abilities in all other fields more efficiently than grades in these fields.

Estimates from specifications which allow for unobserved heterogeneity in distribution parameters tell a more nuanced story: While most students experience the strong information spillovers described previously, a small minority of students experience only modest spillovers. These students have low initial expectations and high uncertainty in these initial expectations. SEE courses still offer more overall information to these students but the difference comes from more efficient within-field signaling rather than spillovers across fields. This suggests less informed students are missing a lot of information which is field specific while more informed students are only missing information which is general across all academic fields.

These findings have several implications for university policies: First, they suggest graduation requirements should be adjusted to improve information diffusion. Currently, all Duke students must take 13-15 HuSS courses but only 6-8 SEE courses to satisfy graduation requirements.[25] A multitude of factors influence how degree requirements are chosen; however, it is worth noting that the current degree requirement structure at Duke University opposes efficient information diffusion. Second, the results suggest policies to curb grade inflation may also improve the information quality of grades. Grade inflation compresses the grade distribution making it difficult for instructors to assign grades which separate students. In my analysis, less informative Humanities and Social Science fields also have higher average grades suggesting grade inflation may be reducing the information quality of grades in these fields.

One concern with my empirical analysis is it reports results which are only relevant to Duke University. This is only a concern if universities are heterogeneous. Moreover, if universities are heterogeneous estimates which represent a large set of schools are not relevant to any specific school. Most policy decisions regarding curricula are made at the university level (or below). As such, the empirical results

[25] Requirements are: Arts, Literatures, and Performance (2 courses); Civilizations (2 courses); Natural Sciences (2 courses); Quantitative Studies (2 courses); Social Sciences (2 courses, includes Economics); Cross-Cultural Inquiry (2 courses); Ethical Inquiry (2 courses); Science, Technology and Society (2 courses); Foreign Language (3 courses); Writing (3 courses).

which should inform these policy decisions should come from analyzing the population of students at one university. This paper provides a theoretical framework and methodology for comparing the signal quality of grades across academic fields at a university and presents results of this analysis for one specific university. Policy decisions at other institutions should be informed by analyses which use data from these institutions.

While it is beyond the scope of this paper, the results also have important implications for employer-employee matching. Just as students use grade signals to inform specialization decisions, firms use grade signals to inform hiring decisions. If firms are risk averse (as in Greenwald et al. (1990)), they will favor students with a large number of SEE grades to those with mostly HuSS grades. This mechanism may partially explain the labor market premium for SEE majors documented in Altonji et al. (2012). Further research may investigate how firms value expected ability by field and uncertainty in expectations by field.

2

What do classroom spending decisions reveal about university preferences?

2.1 Introduction

Every semester, a university decides which courses to offer and how much to spend on instructors for these courses. These classroom spending decisions have important implications for students as they may affect the courses students choose, the welfare they receive from these choices, and other potential outcomes. This paper asks: Can researchers reveal a university's preferences for these student outcomes by observing their classroom spending decisions?

Preferences are fundamental to all economic analyses; however, very little evidence exists on the preferences of universities.[1] This is surprising because universities are very important social institutions. There is abundant evidence that

[1] Two notable exceptions are Bhattacharya et al. (2014) and Turner (2014).

post-secondary outcomes have lasting effects on students in the labor market.[2] Furthermore, there is growing evidence that institutional choices have important effects on post-secondary outcomes.[3] Largely due to the value of undergraduate education in the labor market, large sums are spent on undergraduate education every year. In 2011, spending on post-secondary education comprised 2.7% of the United States gross domestic product (OECD (2014)). A better understanding of university preferences could be used to devise policies which lead universities to make decisions which benefit students and save money for taxpayers, families, and donors.

In this paper, I develop several tools for inferring university preferences from classroom spending decisions. I begin by developing a theoretical framework for analyzing classroom spending decisions. The framework casts these spending decisions as a sequential game between universities and students. In the first stage, universities observe constraints and the composition of the student body and decide which courses to offer and how much to spend on instructors for offered courses. In the second stage, students observe course offerings and spending on instruction and choose courses to maximize their utility. In a classical rational choice framework, the utility students achieve from these choices can be directly interpreted as a measure of welfare.[4] This framework thus provides a link between classroom spending decisions and outcomes which may enter university objectives such as student course choices

[2] See Altonji et al. (2012) and Oreopoulos and Petronijevic (2013) for reviews.

[3] Stinebrickner and Stinebrickner (2014b) argue that grading policies affect specialization decisions of students. Figlio et al. (2015) and Bettinger and Long (2010) provide mixed results about the effects of instructor characteristics on specialization decisions of students.

[4] There are many alternative settings where choice value does not fully capture welfare. For example, if students are myopic or uninformed choice value is an incomplete measure of welfare. Future research may adapt my theoretical framework to incorporate alternative measures of student welfare or other student outcomes.

and welfare.

I use this theoretical framework to develop three methods for inferring university preferences from observed classroom spending decisions. First, I propose a method for statistically testing whether classroom spending decisions maximize student welfare. To develop this test, I derive the tangency conditions which characterize the spending decisions of a university which maximizes total student welfare giving equal weight to all students. I then show estimates of a course choice model and observed data on spending on instruction can be used to statistically test whether these tangency conditions hold for an observed university. If they do not hold, I reject the hypothesis that the observed university is maximizing total student welfare.

Next, I develop two methods for estimating preference parameters of a more general structure for university objectives. The structure allows the university to value total student welfare and the type of courses students choose. University preferences for the type of courses students choose may arise because the university it trying to internalize social externalities, because certain courses increase alumni donations, because university administrators have personal preferences for certain fields, or for many other possible reasons. In all cases, these institutional preferences result in classroom spending decisions which are not aligned with student preferences.[5]

The first estimation method relates to a university's intensive margin decision of how much to spend on instructors for offered courses. First, I derive the tangency conditions which define how much this university would spend on instructors for offered courses. I then propose a variance minimization routine which solves for

[5] In Appendix G, I extend the structure to include welfare weights which allow the university to favor some students more than others. Whether institutional preferences increase or decrease total student welfare is beyond the scope of this paper.

the parameter values which come closest to satisfying these tangency conditions. These parameter estimates thus represent the values which best explain how much a university is observed spending on instructors for different courses. Intuitively, this method measures positive (negative) institutional preferences for course types by identifying courses where the marginal value of spending on instruction is low (high) from the perspective of students. This indicates the university is over (under) investing in instruction in these courses.

The second estimation approach focuses on the university's extensive margin decision of which courses to offer. I propose a maximum likelihood estimator which solves for the parameter values which best explain why observed course offerings were preferred to all other feasible course offerings. Intuitively, this method measures positive (negative) institutional preferences for course types by identifying course types which are offered more (less) than student preferences and market costs would suggest. The two alternative methods employ different empirical variation and have complementary strengths and weaknesses providing researchers with multiple tools for analysis.

I apply my inference methods using administrative data from the University of Central Arkansas. University of Central Arkansas (UCA) is a large public university in central Arkansas whose primary focus is teaching.[6] UCA's teaching focus makes analyzing the preferences underlying its classroom spending decisions especially in-

[6] UCA's teaching focus is apparent in their vision statement:

> The University of Central Arkansas aspires to be a premier learner-focused public university, a nationally recognized leader for its continuous record of excellence in undergraduate and graduate education, scholarly and creative endeavors, and engagement with local, national, and global communities.

teresting. The administrative data include information on all offered courses and information on the instructors teaching these courses between 1993 and 2013. Furthermore, the data include demographic information and full academic records for all students enrolled between 2004 and 2013. Importantly, the data include instructor salaries and fraction of salaries paid for teaching. This allows me to connect costs of instruction to choices and outcomes of students—a crucial link for inferring university preferences from observed classroom spending decisions.

The first stage of my empirical analysis is to estimate a multinomial choice model of students choosing courses. These estimates measure student preferences for course characteristics and estimate how much the desirability of a course increases when it is taught by a higher salaried instructor. To avoid issues of unobserved choice set heterogeneity, my analysis focuses on choices of introductory courses.[7] Estimates show introductory humanities courses are most popular with first year students while introductory business courses are most popular with sophomores, juniors and seniors. The estimates also show that students with higher ACT scores are relatively more attracted to introductory STEM courses. This corroborates existing literature which finds that initial preparation is an important determinant of whether a student pursues a STEM education (Arcidiacono (2004); Stinebrickner and Stinebrickner (2014b)). Finally, the estimates show that higher salaried instructors generally increase an introductory course's desirability but only to a small degree. This finding has important implications for universities: it implies that the vast amounts of resources spent hiring higher salaried instructors has relatively small effects on student

[7] Many advanced courses have prerequisite restrictions implying they are only in the choice sets of students who have satisfied these pre-requisites.

course choices and student welfare.

The second stage of my analysis is to estimate university preference parameters taking student parameters as given. Using my maximum likelihood estimation method, I find UCA has institutional preferences for decreasing enrollment in introductory business courses and increasing enrollments in introductory humanities and STEM courses.[8] This suggests UCA over invests in introductory STEM and humanities courses and under invests in introductory business courses relative to a university that is purely maximizing student welfare.

To place these estimates in context and to examine university behaviors under alternative constraints, I also develop a Marginal Improvement Algorithm (MIA) for simulating classroom spending decisions under alternative preferences and constraints (Chade and Smith (2006)). A university's choice set is typically so large that it is intractable to solve its true maximization problem. The MIA reduces dimensionality by breaking the full maximization problem into a series of smaller maximization problems where the objective of each problem is to maximize marginal improvements to the university's payoff.

I use the MIA to examine two counter-factual simulations: First, I solve for counter-factual course offering costs which lead UCA to offer courses which maximize student welfare. The simulation shows decreasing the minimum cost of offering business courses and increasing the minimum cost of offering other introductory courses prices out UCA's preferences for course types and induces UCA to offer courses which are in line with student preferences. These counterfactual costs which

[8] I do not use the variance minimization method in my empirical application due to a lack of identifying variation in the data.

maximize student welfare can be achieved with a revenue neutral tax and subsidy policy.

Second, I simulate course offerings and excess spending decisions which produce welfare efficiently in the absence of contractual constraints. This simulation shows UCA could achieve the same student welfare at 38.5% of original costs in the absence of contractual constraints. These savings are achieved by vastly reducing spending on instruction and by changing the course composition to include more introductory business and social sciences courses and fewer introductory STEM and humanities courses. While these scenarios may be undesirable for other reasons, it is useful to see how a revenue neutral policy could be used to benefit students and it is striking to see that students could receive the same benefit at drastically lower costs with changes in instructors and course composition.

Two papers which use observed behaviors to make inferences about university preferences are Bhattacharya et al. (2014) and Turner (2014). Bhattacharya et al. (2014) examines the admissions decisions of a selective British university and finds the university has lower admission thresholds for female and private school applicants. This suggests the university is interested in increasing the number of female and private school students in attendance. Turner (2014) examines the financial aid decisions of US colleges and finds schools are willing to pay an additional $284 to have less privileged students attend their institution. This suggests the university is interested in increasing the number of less privileged students matriculating at their institution. My analysis uses completely different observed behaviors to infer preferences and thus complements these other works nicely.

The remainder of this paper proceeds as follows: Section 2.2 presents the theoretical framework for analyzing classroom spending decisions, Section 2.3 describes the Arkansas Department of Higher Education administrative data used in my empirical application, Section 2.4 presents the methods for inferring university preferences from classroom spending decisions, Section 2.5 describes my empirical analysis and presents estimates of student and university parameters, Section 2.6 describes the Marginal Improvement Algorithm and reports classroom spending decisions under counter-factual preferences and constraints, and Section 2.7 concludes.

2.2 Theoretical Framework for Inferring University preferences

In this section, I present a theoretical framework for analyzing how universities make classroom spending decisions and the implications of these decisions for students. The framework describes the setting as a sequential game between a university and students. In the first stage, the university chooses which courses to offer and how much to spend on instruction in these courses to maximize their expected payoff subject to budget and contract constraints. In the second stage, students observe the university's decisions, choose one course to maximize their utility, and derive welfare from this choice. This provides a direct link between classroom spending decisions and student course choices and welfare. In Section 2.4, I show how this link can be used to reveal university preferences over these student outcomes using data on observed classroom spending decisions.

2.2.1 Primitives

Index students by $i = 1, \dots, N$ and potential courses by $j = 1, \dots, J$. When specified, academic semesters are indexed by $t = 1, \dots T$. However, most of my analysis considers a static setting of one academic semester; therefore, semester subscripts are generally suppressed.

Suppose a university has an endowment E to spend on undergraduate instructors. The university chooses spending on instruction c_j for every potential course $j \in J$ subject to the budget constraint $\sum_{j=1}^{J} c_j \leqslant E$. If spending exceeds a course specific minimum cost m_j, an instructor is hired and the course is offered; otherwise, the course is not offered. Let $\mathbf{d} \subset J$ denote the set of offered courses. Formally:

$$j \in \mathbf{d} \; iff \; c_j \geqslant m_j \tag{2.1}$$

Define excess spending e_j as spending on instruction which exceeds minimum costs

$$e_j = \begin{cases} c_j - m_j & c_j > m_j \\ 0 & c_j \leqslant m_j \end{cases} \tag{2.2}$$

and let $\mathbf{e} = \begin{bmatrix} e_1 & \cdots & e_J \end{bmatrix}'$ represent the full vector of excess spending decisions.

Spending in excess of fixed cost may change unobserved instructor quality I_j following:

$$I_j = \begin{cases} \phi_j \left(e_j \right) & e_j > 0 \\ 0 & e_j = 0 \end{cases} \tag{2.3}$$

where the quality of a baseline instructor in course j is normalized to zero. Excess spending may increase instructor quality either because these funds are used to hire a more talented instructor or because increases in compensation motivate the same instructor to perform better. I assume the production function $\phi_j(\cdot)$ is differentiable but allow it to vary across courses.

Universities often negotiate long term contracts with instructors which limit the institution's capacity to make short run classroom spending decisions. To correctly reveal university preferences from classroom spending decisions in one semester, it is crucial to identify which decisions were made to maximize an objective and which were made to satisfy preexisting contracts. To incorporate these short term contracts, let $K \subset J$ denote the subset of potential courses which must be offered by contract. If $j \in K$, then j must be included in every feasible offering vector $\mathbf{d}$ and e_j is set by contract.

2.2.2 Student Utility

Suppose student utility from enrolling in course j depends on student characteristics X_i, course characteristics Z_j (possibly including expected class size), and unobserved instructor quality I_j following a general additively separable structure:

$$U_{ij} = u_{ij}\left(Z_j\left(\mathbf{e}, \mathbf{d}\right),\, I_j,\, X_i\right) + \epsilon_{ij} \tag{2.4}$$

where ϵ_{ij} is assumed to follow a Generalized Extreme Value (GEV) distribution (McFadden (1978)) and the deterministic utility function $u_{ij}\left(\cdot\right)$ is differentiable in I_j and is allowed to vary across individuals and courses.

If student utility depends on expected class size, course offering and excess spending decisions affect the utility of each course indirectly through their effects on expected class sizes.[9] To emphasize the importance of these general equilibrium effects, the dependence of Z_j on decision vectors $\mathbf{e}$ and $\mathbf{d}$ is made explicit.

[9] Using expected class size is equivalent to assuming each student's idiosyncratic preferences are private information. Bayer and Timmins (2007) use an alternative justification for integrating out idiosyncratic prefences which is to assume there is a continuum of individuals with different unobserved preferences for each vector of observed characteristics.

In this framework, student choice value is given by:

$$V_i(\mathbf{e},\mathbf{d}) = \max_{j\in\mathbf{d}} \{u_{ij}(Z_j(\mathbf{e},\mathbf{d}), \phi_j(e_j), X_i) + \epsilon_{ij}\} \tag{2.5}$$

For clarity, I consider a simplified setting where all offered courses are included in every student's choice set. This framework can easily be extended to accommodate observed choice set heterogeneity arising from prerequisite restrictions or other mechanisms. For simplicity, I also use a standard multinomial choice framework in which students choose exactly one course.

2.2.3 Timing

University and student decisions proceed as follows:

1. The university observes all parameters, minimum costs m_j for every potential course $j \in J$, observed student characteristics for every enrolled student i, and observed course characteristics for every potential course $j \in J$.

2. The university observes the set of contracted courses K and predetermined excess spending levels in these courses.

3. The university makes a two-tiered decision:

   (a) The university decides which courses to offer by choosing the offering vector $\mathbf{d}$ where $K \subset \mathbf{d}$

   (b) The university chooses excess spending e_j for non-contract offered courses $j \in \mathbf{d}\backslash K$

4. Students observe $\mathbf{d}$, $\mathbf{e}$, observed student characteristics for every enrolled student i, observed course characteristics for every offered course, and their own idiosyncratic preferences for offered courses, and choose one offered course to maximize their utility.[10]

2.2.4 University's Problem

Denote the university's expected payoff from choosing offering vector $\mathbf{d}$ and excess spending vector $\mathbf{e}$ as $\mathbb{E}[\Pi \,|\, \mathbf{e}, \mathbf{d}]$.[11] The university's problem is to choose a spending vector $\mathbf{c} = \begin{bmatrix} c_1 & \cdots & c_J \end{bmatrix}'$ to maximize the expected value of this payoff subject to a budget constraint and contractual constraints. Formally:

$$\mathbf{c}^{\star} = \text{argmax}_{\mathbf{c}} \{\mathbb{E}[\Pi \,|\, \mathbf{e}, \mathbf{d}]\} \quad \text{s.t.} \sum_{j=1}^{J} c_j \leqslant E, \tag{2.6}$$

At a solution to the university's problem, all non-contracted courses where spending exceeds fixed cost must satisfy the following tangency conditions:

$$\frac{d\mathbb{E}[\Pi \,|\, \mathbf{e}, \mathbf{d}]}{de_j} = \frac{d\mathbb{E}[\Pi \,|\, \mathbf{e}, \mathbf{d}]}{de_{j'}} \tag{2.7}$$

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j \text{ and } c_{j'} > m_{j'}$$

Intuitively, the tangency conditions described in Equation 2.7 impose that the marginal payoff of an additional dollar of spending must be equal across all non-contract

[10] If class size does not affect utility, students do not need to observe characteristics of other students.

[11] Expectations are taken over idiosyncratic shocks to student course preferences which are not observed by the university.

courses where spending exceeds the minimum cost. Importantly, the tangency conditions do not apply to courses where the university only pays the minimum cost at a solution or courses where excess spending is determined by contract.

2.3 Data for Inferring University Preferences

I will apply my methods for inferring university preferences using administrative data from University of Central Arkansas (UCA). University of Central Arkansas is a large public teaching focused university located in central Arkansas. The administrative data include information on all offered courses and information on the instructors teaching these courses for all courses offered between 1993 and 2013. Importantly, these course and instructor data include instructor salaries and fraction of salaries paid for teaching.[12] Furthermore, the data include demographic information and full academic records for all students who were enrolled between 2004 and 2013. These detailed course and student data allow me to connect costs of instruction to choices and outcomes of students—a crucial link for inferring university preferences from observed classroom spending decisions.

Generally speaking, the inference methods presented in this article involve revealing student preferences from student course choices given offered courses and revealing university preferences from classroom spending decisions given anticipated responses of students. To reveal student preferences, data must include student

[12] The observed measure of fraction of salary paid for teaching is based on the number of credit hours an instructor teaches relative to UCA's definition of a full time instructor. At a university where faculty have multiple duties this definition may not accurately reflect fraction of salary paid for instruction. At UCA, 96% of all student course observations are in courses taught by instructors who earn at least 95% of their salary from instruction. This implies almost all of UCA's instruction is provided by full time instructors meaning this observed measure of fraction of salary paid for instruction provides little variation.

characteristics, characteristics of offered courses, and student choices given available alternatives. Importantly, characteristics of offered courses must include excess spending on instruction e_j to directly link the university's excess spending decisions to student choices. To reveal university preferences given student preferences, data must include information on preexisting contracts which constrain short run decisions, minimum costs of offering courses m_j, and characteristics of all potential courses in J.

Using the administrative data, I directly observe student characteristics, characteristics of offered courses other than excess spending on instruction, and student choices given available alternatives.[13] Important components of the theoretical framework which are not directly observed and must be estimated from data are excess spending on instruction e_j, minimum costs of offering courses m_j, and characteristics of potential courses which are not offered.

With limited assumptions, total spending on instruction c_j can be constructed for every offered course using observed data on instructor salaries and fraction of salaries paid for teaching.[14] To estimate minimum costs m_j using c_j, I assume minimum costs are the same for all introductory courses in the same academic field. I then estimate these minimums as the fifth percentile of the distribution of c_j for

[13] Observed student characteristics used in my analysis are race, cohort, and ACT test scores. Observed course characteristics used in my analysis (other than excess spending on instruction) are academic field and difficulty level. Possible academic fields are: STEM, humanities and arts, social science, occupation, and business. Possible difficulty levels are: introductory and advanced. Only introductory courses are included in my empirical analysis.

[14] In many cases, instructors will be paid one salary to teach multiple courses. In these cases, I allocate an instructor's teaching salary to specific courses based on the number of credit hours each course is worth.

introductory courses in a specific field.[15] Excess spending on instruction e_j is then defined using Equation (2.2).

To construct characteristics of potential introductory courses which are not offered I assume introductory courses are fully defined by academic fields and that infinitely many potential introductory courses exist in every field. This implies non-offered introductory STEM courses have the same minimum cost and intrinsic popularity as offered introductory STEM courses and that the university can always offer additional introductory STEM courses. This provides a computationally simple method for constructing characteristics of non-offered courses; however, this approach has important limitations: First, it ignores unobserved heterogeneity in the desirability and cost of courses. One may expect non-offered courses to be unobservably worse—from the perspective of the university—than observationally equivalent observed courses. Second, the assumption that infinitely many potential introductory courses exist in every field is clearly unrealistic. At a certain point, recruiting costs and facility constraints will limit the number of introductory courses a university can offer in any given field.[16]

2.4 Methods for Inferring University Preferences

In this section, I present various methods for making empirical inferences about university preferences. First, I propose a method for statistically testing whether

[15] I use the fifth percentile instead of the sample minimums to reduce sensitivity to outliers with unrealistically low values for c_j.

[16] An alternative approach for constructing the characteristics of non-offered courses uses panel data to define the set of potential courses J as all courses which were ever offered in any semester present in the panel. While it would be computationally challenging, this approach could theoretically estimate unobserved course attributes for every course in J.

Table 2.1: Descriptive Staistics

	Fall, 2007	Spring, 2008	Fall, 2008	Spring, 2009
Introductory course offerings				
STEM	279	242	249	226
Social Sciences	272	245	257	258
Humanities	433	407	418	368
Occupational	125	114	107	99
Business	65	57	61	58
Median spending per introductory course				
STEM	$8,410	$9,098	$8,632	$8,597
Social Sciences	$6,868	$6,354	$6,516	$6,292
Humanities	$6,547	$5,996	$5,802	$5,718
Occupational	$7,266	$6,824	$6,324	$5,531
Business	$9,480	$7,642	$7,391	$7,279
Share of total spending on intro courses by field				
STEM	26.3%	27.9%	27.6%	28.5%
Social Sciences	21.7%	21.3%	22.2%	24.5%
Humanities	34.1%	33.8%	34.5%	31.6%
Occupational	10.6%	11.3%	9.1%	8.4%
Business	7.3%	5.7%	6.6%	7.0%
Share of total intro student-course observations by field				
STEM	22.2%	20.9%	21.8%	21.7%
Social Sciences	26.8%	27.0%	27.5%	28.2%
Humanities	31.6%	32.1%	32.2%	30.7%
Occupational	11.3%	11.3%	9.4%	9.6%
Business	8.1%	8.6%	9.2%	9.7%
Total Cost	$9,784,463	$8,369,421	$7,930,221	$7,109,921
Total Courses	1174	1065	1092	1009

classroom spending decisions maximize student welfare. To develop this test, I derive the tangency conditions which characterize the excess spending decisions of a university which maximizes total student welfare giving equal weight to all students. I then show estimates of a course choice model and observed data on spending on instruction can be used to statistically test whether these tangency conditions hold for an observed university. If they do not hold, I reject the hypothesis that the observed university is maximizing total student welfare.

Next, I develop two methods for estimating preference parameters of a more general structure for university objectives. The structure allows the university to value total student welfare and the type of courses students choose. The first estimation method relates to a university's intensive margin decision of how much to spend on instructors for offered courses. First, I derive the tangency conditions which define how much this university would spend on instructors for offered courses. I then propose a variance minimization routine which solves for the parameter values which come closest to satisfying these tangency conditions. These parameter estimates thus represent the values which best explain observed excess spending decisions. Intuitively, this method measures positive (negative) institutional preferences for course types by identifying courses where the marginal value of spending on instruction is low (high) from the perspective of students. This indicates the university is over (under) investing in instruction in these courses.

The second estimation approach focuses on the university's extensive margin decision of which courses to offer. I propose a maximum likelihood estimator which solves for the parameter values which best explain why observed course offerings were preferred to all other feasible course offerings. Intuitively, this method measures positive (negative) institutional preferences for course types by identifying course types which are offered more (less) than student preferences and market costs would suggest. The two alternative methods employ different empirical variation and have complementary strengths and weaknesses providing researchers with multiple tools for analysis.

In Section 2.4, I perform an empirical application of the maximum likelihood estimation method described in Subsection 2.4.3 using administrative data from the University of Central Arkansas (UCA). Methods described in Subsections 2.4.1 and 2.4.2 were found to be inappropriate at UCA due to a lack of identifying variation; however, these methods may be applicable in alternative settings.

2.4.1 Are excess spending decisions consistent with utilitarian student welfare maximization?

In this subsection, I examine the special case of a university whose goal is to maximize student welfare giving equal weight to all students. I refer to this baseline school as a utilitarian student welfare maximizing (U-SWM) university. I derive the tangency conditions described in Equation (2.7) for this university and show how observed data on spending on instruction and estimates of a student course choice model can be used to statistically test whether observed spending decisions satisfy the tangency conditions of this U-SWM university. This is equivalent to testing whether the incentives of an observed university are aligned with its students. I begin by describing the statistical test for the baseline setting in which students only value fixed course characteristics and instructor quality. Following this, I describe the statistical test in the general equilibrium setting in which students also value class size. Finally, I show how panel data can be used to reduce reliance on functional form assumptions about student utility. I conclude with a short discussion of these methods.

Statistical test of U-SWM without class size effects

A U-SWM university's payoff from decision vectors **d** and **e** is the sum of choice values over all students. The U-SWM problem is then given by:

$$\mathbf{c}^{\star} \quad = \quad \text{argmax}_{\mathbf{c}} \left\{ \sum_{i=1}^{N} \mathbb{E}\left[V_i\left(\mathbf{e}, \mathbf{d}\right)\right] \right\} \quad \text{s.t.} \sum_{j=1}^{J} c_j \leqslant E \tag{2.8}$$

Because student preference shocks are assumed to follow a GEV distribution, a convenient property can be used to simplify the tangency conditions for a U-SWM university. For any θ_j which affects deterministic utility in course j , $\frac{d\mathbb{E}[V_i(\mathbf{e},\mathbf{d})]}{d\theta_j} = \left(\frac{du_{ij}}{d\theta_j}\right) P_{ij}\left(\mathbf{e}, \mathbf{d}\right)$ where $P_{ij}\left(\mathbf{e}, \mathbf{d}\right)$ is the probability individual i chooses course j given excess spending vector **e** and offering vector **d**.[17] This property implies the U-SWM version of the general tangency conditions given in Equation (2.7) is given by :

$$\sum_{i=1}^{N} \left(\frac{\partial u_{ij}}{\partial I_j}\right) \left(\frac{\partial \phi_j}{\partial e_j}\right) P_{ij}\left(\mathbf{e}, \mathbf{d}\right) = \sum_{i=1}^{N} \left(\frac{\partial u_{ij'}}{\partial I_{j'}}\right) \left(\frac{\partial \phi_{j'}}{\partial e_{j'}}\right) P_{ij'}\left(\mathbf{e}, \mathbf{d}\right) \tag{2.9}$$

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j \text{ and } c_{j'} > m_{j'}$$

As before, note the tangency conditions only relate the marginal values of spending across non-contract courses where spending exceeds minimum costs.

[17] $P_{ij}\left(\mathbf{e}, \mathbf{d}\right) = 0$ if course j is not offered.

With functional form assumptions on the structure of $u_{ij}(\cdot)$ and $\phi_j(\cdot)$ and distribution of ϵ_{ij}, researchers can estimate the parameters of a multinomial course choice model in which students choose one course to maximize utility defined in Equation (2.4). The parameters of this choice model can then be used to construct estimates of the tangency condition components $\left(\frac{\partial u_{ij}}{\partial I_j}\right)$, $\left(\frac{\partial \phi_j}{\partial e_j}\right)$, and $P_{ij}(\mathbf{e},\mathbf{d})$ for all students i and offered courses $j \in \mathbf{d}$

These estimates can be used to form test statistics which are the empirical analogs of the tangency conditions:

$$\hat{t}_{jj'} = \left\{\sum_{i=1}^{N}\left(\frac{\partial \hat{u}_{ij}}{\partial I_j}\right)\left(\frac{\partial \hat{\phi}_j}{\partial e_j}\right)\hat{P}_{ij}(\mathbf{e},\mathbf{d})\right\} - \left\{\sum_{i=1}^{N}\left(\frac{\partial \hat{u}_{ij'}}{\partial I_{j'}}\right)\left(\frac{\partial \hat{\phi}_{j'}}{\partial e_{j'}}\right)\hat{P}_{ij'}(\mathbf{e},\mathbf{d})\right\} \quad (2.10)$$

For the observed spending vector to be consistent with the goal of maximizing student welfare, $\hat{t}_{jj'}$ must be statistically indistinguishable from zero for every course pair $j, j' \in \mathbf{d}\backslash K$ for which $c_j > m_j$, and $c_{j'} > m_{j'}$. If $\hat{t}_{jj'}$ is statistically positive (negative), it implies the welfare return on an additional dollar of spending is significantly higher (lower) in course j relative to course j'. This would be inconsistent with the goal of maximizing student welfare because welfare could be increased by marginally increasing (reducing) spending in course j and reducing (increasing) spending in course j'. Formally, the testing procedure is as follows:

$$\begin{aligned} H_0 &: \text{Excess spending decisions maximize student welfare} \\ H_a &: \text{Excess spending decisions do not maximize student welfare} \end{aligned}$$

Testing procedure:

1. Identify the set of courses $\tilde{J} = \{j \in \mathbf{d}\backslash K \mid c_j > m_j\}$.

2. Use a bootstrap algorithm to estimate the distribution of the $\frac{\tilde{J}\left[\tilde{J}-1\right]}{2}$ dimensional random vector $\hat{\mathbf{t}} = \left[\hat{t}_{12} \quad \ldots \quad \hat{t}_{\tilde{J}-1\tilde{J}}\right]$.[18]

3. Test the joint hypothesis: $H_0 : \hat{t}_{jj'} = 0$ for all pairs of offered courses $j, j' \in \tilde{J}$.

To implement the first step of this procedure, researchers can use observed data on c_j and estimates of m_j obtained as described in Section 2.3. To reduce sensitivity to estimation error in $\hat{m}_j$, researchers may use a stricter set: $\tilde{J}_\delta = \{j \in \mathbf{d} \backslash K \mid c_j > m_j + \delta\}$where $\delta > 0$. Choosing a large δ guarantees that spending exceeds minimum costs implying the tangency conditions must bind. However, as δ increases, the set of courses shrinks which reduces power to reject the null hypothesis.

Statistical test of U-SWM with class size effects

When class size affects choice utility, excess spending in course j has direct effects on choice utility for course j but also has indirect effects on choice utility for all courses through changes in class sizes. These general equilibrium effects make simplifying the general tangency conditions in Equation 2.7 somewhat more difficult. A general version of the GEV property used previously is helpful: For any θ affecting deterministic utility in any course, $\frac{d\mathbb{E}[V_i(\mathbf{e},\mathbf{d})]}{d\theta} = \sum_{j \in \mathbf{d}} \left(\frac{du_{ij}}{d\theta}\right) P_{ij}\left(\mathbf{e},\mathbf{d}\right)$. This yields the following general equilibrium U-SWM tangency conditions:

$$\sum_{i=1}^{N} \sum_{k \in J_i(\mathbf{d})} \left(\frac{du_{ik}}{de_j}\right) P_{ik}\left(\mathbf{e},\mathbf{d}\right) = \sum_{i=1}^{N} \sum_{k \in J_i(\mathbf{d})} \left(\frac{du_{ik}}{de_{j'}}\right) P_{ik}\left(\mathbf{e},\mathbf{d}\right) \tag{2.11}$$

[18] In theory, one could derive the true asymptotic distribution of $\hat{\mathbf{t}}$ since this random vector is a function of a maximum likelihood estimator which is asymptotically multivariate normal. However, the function is extremely complicated even for very simple utility structures which makes this derivation impractical.

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j \text{ and } c_{j'} > m_{j'}$$

where

$$\frac{du_{ik}}{de_j} = \begin{cases} \frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j} + \frac{\partial u_{ij}}{\partial \tilde{n}_j}\frac{d\tilde{n}_j}{de_j} & k = j \\ \frac{\partial u_{ik}}{\partial \tilde{n}_k}\frac{d\tilde{n}_k}{de_j} & k \neq j \end{cases} \tag{2.12}$$

Estimates of the parameters of a general equilibrium course choice model can be used to estimate the tangency condition components $P_{ik}(\mathbf{e}, \mathbf{d})$, $\frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j}$, and $\frac{\partial u_{ik}}{\partial \tilde{n}_k}$; however, they cannot be used to directly construct the effects of spending on class sizes given by $\frac{d\tilde{n}_k(\mathbf{e},\mathbf{d})}{de_j}$.[19] These are complicated effects because they depend on the effects of spending on course utility and these effects depend on the effects of spending on class sizes. In Appendix C, I show how this recursive relationship can be unraveled to yield a closed form expression when ϵ_{ij} follows a type 1 extreme value distribution.

As before, these estimates can be used to construct test statistics which are the empirical analogs of the general equilibrium tangency conditions:

$$\hat{t}_{jj'} = \sum_{i=1}^{N}\sum_{k\in\mathbf{d}}\left(\widehat{\frac{du_{ik}}{de_j}}\right)\hat{P}_{ik}(\mathbf{e}, \mathbf{d}) - \sum_{i=1}^{N}\left(\sum_{k\in\mathbf{d}}\left(\widehat{\frac{du_{ik}}{de_{j'}}}\right)\hat{P}_{ik}(\mathbf{e}, \mathbf{d})\right) \tag{2.13}$$

These test statistics can then be used to test whether observed spending is consistent with the goal of utilitarian student welfare maximization following the same procedure described in Subsection 2.4.1.

[19] In this general equilibrium setting, class sizes are correlated with unobserved course attributes by construction. To estimate parameters of such a general equilibrium choice model, researchers may adapt the iterative instrumental variables approach described in Bayer and Timmins (2007).

Statistical test of U-SWM with panel data

One concern with the baseline test described in Subsection 2.4.1 is the results may be sensitive to the functional form of student utility. In this subsection, I develop a complementary panel data test which is more robust to functional form assumptions. The idea behind this test is that the researcher can solve for the return on spending parameters which exactly satisfy the U-SWM first order conditions in a given semester. If the university is U-SWM in every semester then these implied returns should be statistically similar for the same course in different semesters. In this subsection only, academic semesters are indexed by $t = 1, \ldots, T$.

Suppose student utility falls into a general class of models in which utility from enrolling in course j in semester t depends on student characteristics X_{it}, course characteristics Z_{jt}, and excess spending on instruction e_{jt} in the following additively separable manner:

$$U_{ijt} = \theta_j \ln e_{jt} + \psi_{ijt} \left(Z_{jt},\, X_{it} \right) + \epsilon_{ijt} \tag{2.14}$$

where ϵ_{ijt} is assumed to follow a Generalized Extreme Value distribution (McFadden (1978)) and the deterministic utility function $\psi_{ijt}(\cdot)$ is allowed to vary across individuals, courses, and semesters. While this class is generally quite flexible it places some restrictions on how spending on instruction affects utility. Most notably, although returns on spending are allowed to vary across courses they are not allowed to vary across individuals. Additionally, this structure imposes that concavity is generated by the natural logarithm function.

For this class of utility structures, the tangency conditions describing a U-SWM university's solution are:

$$\frac{\sum_{i=1}^{N_t} \theta_j P_{ijt}\left(\mathbf{e_t}, \mathbf{d_t}\right)}{e_{jt}} = \frac{\sum_{i=1}^{N_t} \theta_{j'} P_{ij't}\left(\mathbf{e_t}, \mathbf{d_t}\right)}{e_{j't}} \tag{2.15}$$

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j \text{ and } c_{j'} > m_{j'}$$

Rearranging yields:

$$\frac{\theta_j}{\theta_{j'}} = \left(\frac{e_{jt}}{e_{j't}}\right)\left(\frac{\sum_{i=1}^{N_t} P_{ij't}\left(\mathbf{e_t}, \mathbf{d_t}\right)}{\sum_{i=1}^{N_t} P_{ijt}\left(\mathbf{e_t}, \mathbf{d_t}\right)}\right) \tag{2.16}$$

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j \text{ and } c_{j'} > m_{j'}$$

Notice the left hand side of this expression is invariant across semesters. This implies that if the U-SWM first order conditions are satisfied in both semesters t and t' then the following conditions must hold:

$$\left(\frac{e_{jt}}{e_{j't}}\right)\left(\frac{\tilde{n}_{j't}\left(\mathbf{e_t}, \mathbf{d_t}\right)}{\tilde{n}_{jt}\left(\mathbf{e_t}, \mathbf{d_t}\right)}\right) = \left(\frac{e_{jt'}}{e_{j't'}}\right)\left(\frac{\tilde{n}_{j't'}\left(\mathbf{e_{t'}}, \mathbf{d_{t'}}\right)}{\tilde{n}_{jt'}\left(\mathbf{e_{t'}}, \mathbf{d_{t'}}\right)}\right) \tag{2.17}$$

$$\forall j, j' \in \mathbf{d} \backslash K \text{ s.t. } \min\left\{c_{jt}, c_{jt'}\right\} > m_j \text{ and } \min\left\{c_{j't}, c_{j't'}\right\} > m_{j'}$$

where $\tilde{n}_{jt}\left(\mathbf{e_t}, \mathbf{d_t}\right) = \sum_{i=1}^{N} P_{ijt}\left(\mathbf{e_t}, \mathbf{d_t}\right)$ is expected enrollment in course j in semester t given decision vectors $\mathbf{e_t}$ and $\mathbf{d_t}$. Intuitively, this panel condition states that a U-SWM university responds to changes in relative intrinsic popularity by spending more in courses which are becoming more popular.

The benefit of the panel condition given in Equation (2.17) is it depends on excess spending e_{jt} and equilibrium choice probabilities $P_{ijt}\left(\mathbf{e_t}, \mathbf{d_t}\right)$ but does not depend on other parameters. Because no specific parameters are required to construct (2.17), researchers may conduct this test using flexible reduced form utility structures which are robust to functional form assumptions.[20]

As before, estimates of equilibrium choice probabilities and observed data can be used to construct the empirical analogs of (2.17):

$$\hat{t}_{jj',tt'} = \left(\frac{e_{jt}}{e_{j't}}\right)\left(\frac{\hat{\tilde{n}}_{j't}\left(\mathbf{e_t},\mathbf{d_t}\right)}{\hat{\tilde{n}}_{jt}\left(\mathbf{e_t},\mathbf{d_t}\right)}\right) - \left(\frac{e_{jt'}}{e_{j't'}}\right)\left(\frac{\hat{\tilde{n}}_{j't'}\left(\mathbf{e_{t'}},\mathbf{d_{t'}}\right)}{\hat{\tilde{n}}_{jt'}\left(\mathbf{e_{t'}},\mathbf{d_{t'}}\right)}\right) \tag{2.18}$$

These test statistics can then be used to test whether observed spending is consistent with the goal of utilitarian student welfare maximization following the same procedure described in Subsection 2.4.1.

Discussion of tangency condition inference methods

The tangency condition inference methods have several strengths: First, they provide a clear statistical test of whether observed behavior is consistent with utilitarian student welfare maximization. The methods test a specific structure of university objectives—rather than imposing a structure and estimating parameters assuming that structure is true—and they handle sampling error in estimates of student choice parameters appropriately. Furthermore, the trio of a baseline test, general equilibrium test, and panel data test offers researchers several inference tools which apply to a variety of settings and can be used to assess the robustness of results.

While these inference methods are desirable for their clarity and rigor, the tradeoff is they only offer narrow inferences about university preferences. Specifically, the tests can only reject or fail to reject that observed spending is consistent with

[20] Estimates of $P_{ijt}\left(\mathbf{e_t},\mathbf{d_t}\right)$ must approximate how the university believes individual i will choose courses. As such, researchers should avoid using student data which is not observed by the university or utility structures which are impractical.

utilitarian student welfare maximization. If the null hypothesis of U-SWM is rejected, these methods do not offer a preferable alternative. In subsequent sections, I introduce and discuss complementary inference methods for estimating university preference parameters of a more general objective structure.

The other limitation of these methods is that statistical power depends on the size of $\tilde{J}_\delta$. $\tilde{J}_\delta$ contains non-contract courses where spending on instruction exceeds minimum costs. In many settings—including UCA—compensation for instructors hired on short term contracts almost always represents the minimum cost of instruction. This results in a $\tilde{J}_\delta$ which is too small for meaningful analysis.

2.4.2 What university preferences best explain observed spending decisions

The preceding subsection demonstrates how estimates of student choice parameters and minimum course costs can be used to test whether observed spending decisions are consistent with student welfare maximization. While these tests are a useful place to start, they can only reject or fail to reject that spending maximizes student welfare. If welfare maximization is rejected, it would be useful to understand the alternative motives which are driving classroom spending decisions.

In this subsection, I present methods for estimating preference parameters of a more general structure for university objectives. The structure allows the university to value total student welfare and the type of courses students choose. First, I derive the tangency conditions which define how much this university would spend on instructors for offered courses. I then propose a variance minimization routine which solves for the parameter values which come closest to satisfying these tangency conditions. These parameter estimates thus represent the values which best explain observed excess spending decisions. Intuitively, this method measures positive (negative) institutional preferences for course types by identifying courses where the marginal value of spending on instruction is low (high) from the perspective of students. This indicates the university is over (under) investing in instruction in these courses.

With this more general objective structure, university expected payoffs are an additively separable function of total student welfare and institutional preferences for the type of courses students choose.[21] Formally

$$\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right] = \sum_{i=1}^{N} \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] + \sum_{j=1}^{J} \gamma_j \tilde{n}_j \left(\mathbf{e}, \mathbf{d}\right) \tag{2.19}$$

where $\sum_{i=1}^{N} \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right]$ represents expected total student welfare given university choices for $\mathbf{d}$ and $\mathbf{e}$, $\tilde{n}_j \left(\mathbf{e}, \mathbf{d}\right)$ represents the expected number of students choosing course j given university choices for $\mathbf{d}$ and $\mathbf{e}$, and γ_j represent institutional preferences for students choosing j. University preferences γ_j may exist because the university it trying to internalize social externalities, because certain courses increase alumni donations, because university administrators have personal preferences for certain fields, or for many other possible reasons. In all cases, these institutional preferences result in classroom spending decisions which are not aligned with student preferences. The objective of this method—and my paper more generally—is to estimate γ_j to reveal university preferences which drive classroom spending decisions and thus have consequences for student course choices and welfare.

[21] In Appendix G, I extend the structure to include welfare weights which allow the university to favor some students more than others.

The university's problem is then given by:

$$\mathbf{c}^{\star} \quad = \quad \text{argmax}_{\mathbf{c}} \left\{ \sum_{i=1}^{N} \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] + \sum_{j=1}^{J} \gamma_j \tilde{n}_j\left(\mathbf{e}, \mathbf{d}\right) \right\} \quad \text{s.t.} \sum_{j=1}^{J} c_j \leqslant E \qquad (2.20)$$

With this structure, the university's tangency conditions are given by:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} \quad = \quad \frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_{j'}} \qquad (2.21)$$

$$\forall j, j' \in \mathbf{d} \backslash K \, \text{s.t.} \, c_j > m_j \, \text{and} \, c_{j'} > m_{j'}$$

In the baseline setting where students do not value class sizes, $\frac{d\mathbb{E}[\Pi \,|\, \mathbf{e},\mathbf{d}]}{de_j}$ is given by:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} = \sum_{i=1}^{N} \left(\frac{\partial u_{ij}}{\partial I_j}\right) \left(\frac{\partial \phi_j}{\partial e_j}\right) P_{ij}\left(\mathbf{e}, \mathbf{d}\right) + \sum_{k=1}^{J} \gamma_k \left(\frac{\partial \tilde{n}_k}{\partial e_j}\right) \qquad (2.22)$$

With functional form assumptions on the structure of $u_{ij}(\cdot)$ and $\phi_j(\cdot)$ and distribution of ϵ_{ij}, researchers can use estimates of a multinomial course choice model to construct $\left(\frac{\partial u_{ij}}{\partial I_j}\right)$, $\left(\frac{\partial \phi_j}{\partial e_j}\right)$, $P_{ij}(\mathbf{e}, \mathbf{d})$, and $\left(\frac{\partial \tilde{n}_k}{\partial e_j}\right)$ for all students i and offered courses j s.t. $d_j = 1$.

In a general equilibrium setting where class size affects choice utility, $\frac{d\mathbb{E}[\Pi \,|\, \mathbf{e},\mathbf{d}]}{de_j}$ is given by:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} = \left\{ \sum_{i=1}^{N} \sum_{k \in \mathbf{d}} \left(\frac{du_{ik}}{de_j}\right) P_{ik} \right\} + \sum_{k=1}^{J} \gamma_k \left(\frac{d\tilde{n}_k}{de_j}\right) \qquad (2.23)$$

With functional form assumptions on the structure of $u_{ij}(\cdot)$ and $\phi_j(\cdot)$ and distribution of ϵ_{ij}, researchers can construct $\frac{du_{ik}}{de_j}$ and $\frac{d\tilde{n}_k}{de_j}$ with estimates of a general equilibrium sorting model.[22]

To identify preference parameters which best explain observed spending, I propose solving for values of γ_j which come closest to satisfying these tangency conditions at observed spending levels. For a single academic semester, the university's excess spending tangency conditions state that the marginal returns $\frac{d\mathbb{E}[\Pi \,|\, \mathbf{e},\mathbf{d}]}{de_j}$ must be equal for all courses where spending exceeds fixed costs. To solve for the parameter values which come closest to satisfying this condition at observed spending levels, I propose solving

$$\hat{\gamma} = \text{argmin}_{\gamma} \left\{ \text{Var}_{j \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j} \left(\frac{d\mathbb{E}\left[\Pi \,|\, \gamma, \tilde{\mathbf{e}}, \tilde{\mathbf{d}}\right]}{de_j} \right) \right\} \tag{2.24}$$

where $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{d}}$ represent observed excess spending. If the tangency conditions are satisfied for all pairs of courses $j, j' \in \mathbf{d} \backslash K$ for which $c_j > m_j$, and $c_{j'} > m_{j'}$then this objective variance is exactly zero. As such, the parameter values γ which minimize this variance represent university preferences which best explain observed excess spending decisions.

As in Subsection 2.4.1, this inference method requires identifying the set of courses $\tilde{J} = \{j \in \mathbf{d} \backslash K \,|\, c_j > m_j\}$. As before, researchers may use a stricter set:

$$\tilde{J}_{\delta} = \{j \in \mathbf{d} \backslash K \,|\, c_j > m_j + \delta\} \tag{2.25}$$

where $\delta > 0$ to reduce sensitivity to error in estimates of m_j. Choosing a large δ guarantees that spending exceeds minimum costs implying the tangency conditions

[22] In this general equilibrium setting, class sizes are correlated with unobserved course attributes by construction. To estimate parameters of such a general equilibrium choice model, researchers may adapt the iterative instrumental variables approach described in Bayer and Timmins (2007).

must bind. However, as δ increases, the set of courses shrinks which reduces precision in estimates of γ.

Identification and estimation

For identification and estimation, it is necessary to place some restrictions on γ_j. In my empirical application, I restrict γ_j to be equivalent within academic fields.[23] In Appendix D, I show this restriction means γ_j are over-identified as long as the number of course types is less than the total number of courses $j \in \mathbf{d} \backslash K$ such that $c_j > m_j$. Intuitively, identification of γ_j comes from differences in the marginal effects of excess spending across academic fields. For instance, if student parameter estimates imply increasing spending in STEM courses has larger effects on student welfare than increasing spending in humanities courses this implies the university is over-spending in humanities courses and under-spending in STEM courses. This reveals an institutional preference for drawing students out of STEM and into humanities courses.

The fact that γ_j are over-identified provides a useful over-identification test of functional form assumptions. If all functional form assumptions are correct, the objective variance:

$$\mathrm{Var}_{j \in \mathbf{d} \backslash K \text{ s.t. } c_j > m_j} \left(\frac{d\mathbb{E}\left[\Pi \mid \gamma, \tilde{\mathbf{e}}, \tilde{\mathbf{d}}\right]}{de_j} \right) \tag{2.26}$$

should be exactly zero at true parameter values γ_j. If the objective variance is statistically positive at estimates of γ_j, we can jointly reject the structures of student utility and university objectives. As with any over-identification test, failure to reject

[23] Academic fields are STEM, humanities and arts, business, social science, and occupational.

the null hypothesis does not validate the model structure.

Discussion of marginal return variance minimizing inference methods

The strength of the variance minimizing methods described in this subsection is that they provide estimates of preference parameters for an objective structure which is more general than utilitarian student welfare maximization. Specifically, they quantify institutional preferences over the type of courses students choose. These institutional preferences measure the extent to which student and university preferences are not aligned.

The main limitation of these methods is that the set $\tilde{J}_\delta$ may be too small to conduct inference with sufficient power. $\tilde{J}_\delta$ contains non-contract courses where spending on instruction exceeds minimum costs. In many settings—including UCA—compensation for instructors hired on short term contracts almost always represents the minimum cost of instruction. This results in a $\tilde{J}_\delta$ which is too small for meaningful analysis.

2.4.3 What university preferences best explain observed course offerings?

In the preceding subsection, I presented an algorithm for estimating university preference parameters by solving for parameter values which best explain observed excess spending decisions in non-contract courses. One shortcoming of this method is it relies on variation from non-contract courses with positive excess spending. This will be a small set of courses if salaries for non-contract instructors are at the bottom of the pay distribution as is often the case.

In this subsection, I propose complementary methods which estimate university preference parameters which best explain observed course offering and excess spending decisions. Intuitively, this method solves for the parameter values which maximize the likelihood that observed course offerings are preferred to all feasible alternatives.

To illustrate these methods more concretely, note that it is often possible to combine the university's conditions given by Equation (2.21) with the budget constraint to solve for the optimal excess spending vector for each offering vector.[24] Denote these optimal excess spending vectors as: $\mathbf{e}(\mathbf{d})^{\star}$. The university's problem can then be restated to focus on extensive margin decisions:

$$\mathbf{d}^{\star} = \text{argmax}_{\mathbf{d}} \left\{ \mathbb{E}\left[\Pi \,|\, \mathbf{e}(\mathbf{d})^{\star}, \mathbf{d}; \gamma\right]\right\} \text{ s.t. } \sum_{j=1}^{J} d_j m_j \leqslant E \tag{2.27}$$

If the objective structure is correct, the observed offering vector $\hat{\mathbf{d}}$ is the optimal vector $\mathbf{d}^{\star}$ and the observed excess spending vector $\hat{\mathbf{e}}$ is the optimal spending vector given the optimal offering vector $\mathbf{e}(\mathbf{d}^{\star})^{\star}$ and the true parameter values γ.

This implies that the observed decision vectors $\hat{\mathbf{d}}$ and $\hat{\mathbf{e}}$ must provide the highest university payoff at the true parameter values γ if the objective structure is correct. Formally,

$$\mathbb{E}\left[\Pi \,|\, \hat{\mathbf{e}}, \hat{\mathbf{d}}; \gamma\right] \geqslant \mathbb{E}\left[\Pi \,|\, \mathbf{e}(\mathbf{d})^{\star}, \mathbf{d}; \gamma\right] \tag{2.28}$$

$$\forall \mathbf{d} \text{ s.t. } \sum_{j=1}^{J} d_j f_j \leqslant E \text{ and } K \subset \mathbf{d}$$

at the true parameter values γ.

[24] See Appendix E for an illustration.

This suggests one method for estimating γ: Choose the values of γ which maximize the number of times the observed decision vector yields a higher payoff than an alternative feasible vector. Formally,

$$\hat{\gamma} = \operatorname{argmax}_{\gamma} \left\{ \sum_{\mathbf{d} \in \mathbf{D}(E)} \mathbf{1} \left\{ \mathbb{E} \left[\Pi \,|\, \hat{\mathbf{e}}, \hat{\mathbf{d}}; \gamma \right] > \mathbb{E} \left[\Pi \,|\, \mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}; \gamma \right] \right\} \right\} \tag{2.29}$$

where $\mathbf{D}(E) = \left\{ \mathbf{d} \,|\, \sum_{j=1}^{J} d_j f_j \leqslant E \text{ and } K \subset \mathbf{d} \right\}$ is the set of feasible offering vectors given endowment E. Importantly, every feasible set of offered courses must contain the courses K which are set by preexisting contracts.

It is theoretically possible to point identify γ when the set $\mathbf{D}(E)$ generates a connected space of welfare and class size outcomes in a neighborhood around the outcomes generated by the observed choices $\hat{\mathbf{d}}$ and $\hat{\mathbf{e}}$. In this case, there is at most one point in the parameter space at which observed choices $\hat{\mathbf{d}}$ and $\hat{\mathbf{e}}$ yield a higher payoff than all feasible alternatives in $\mathbf{D}(E)$. This is clearly illogical because it requires the set $\mathbf{D}(E)$ to be infinitely large; however, this illustrates how increasing the number of feasible alternatives in $\mathbf{D}(E)$ shrinks the parameter subspace in which observed choices $\hat{\mathbf{d}}$ and $\hat{\mathbf{e}}$ yield a higher payoff than all feasible alternatives.

To provide a unique solution with a finite number of alternatives in $\mathbf{D}(E)$, I propose treating expected payoffs as quantities which are measured with error and solving the stochastic analog of (2.29). Suppose true expected payoffs are given by:

$$\mathbb{E} \left[\Pi \,|\, \mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}; \gamma \right] = \hat{\Pi} \left(\mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}, \gamma \right) + \zeta_{\mathbf{d}} \tag{2.30}$$

where $\hat{\Pi} \left(\mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}, \gamma \right)$ is observed up to values for γ and $\zeta_{\mathbf{d}}$ is an unobserved error term. In practice, welfare and class size outcomes will be estimated using estimates of a student choice model. As such, a theoretical justification for including $\zeta_{\mathbf{d}}$ is the presence of standard sampling error in estimates of these outcomes. The stochastic analog of (2.29) is then given by:

$$\hat{\gamma} = \operatorname{argmax}_{\gamma} \left\{ \Pr \left(\hat{\Pi} \left(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \gamma \right) + \zeta_{\mathbf{d}} \geqslant \hat{\Pi} \left(\mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}, \gamma \right) + \zeta_{\mathbf{d}} \, \forall \mathbf{d} \in \mathbf{D} \left(E \right) \right) \right\} \tag{2.31}$$

Because the term in braces is a probability, (2.31) represents a a maximum likelihood estimator for γ. Intuitively, the estimator solves for the parameter values which maximize the likelihood that observed classroom spending decisions are preferred to all feasible alternatives.

Implementing (2.31) requires an assumption about the distribution of the error term $\zeta_{\mathbf{d}}$. In my empirical application, I assume $\zeta_{\mathbf{d}}$ follows a type 1 extreme value distribution. This yields the following likelihood function:

$$\mathscr{L} \left(\hat{\mathbf{d}}; \gamma \right) = \frac{\exp \left(\hat{\Pi} \left(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \gamma \right) \right)}{\sum_{\mathbf{d} \in \mathbf{D}(E)} \exp \left(\hat{\Pi} \left(\mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}, \gamma \right) \right)} \tag{2.32}$$

If researchers have access to panel data, the likelihood function given by (2.32) can be adapted to include multiple semesters of classroom spending decisions. Let $\hat{\mathbf{d}}_t$ and $\hat{\mathbf{e}}_t$ represent observed decisions in period t and let $\mathbf{D}_t \left(E_t \right)$ represent the set of feasible offerings in semester t. With T semesters of observed classroom spending decisions, the likelihood function is given by:

$$\mathscr{L} \left(\hat{\mathbf{d}}_t; \gamma \right) = \prod_{t=1}^{T} \left[\frac{\exp \left(\hat{\Pi} \left(\hat{\mathbf{e}}_t, \hat{\mathbf{d}}_t, \gamma \right) \right)}{\sum_{\mathbf{d} \in \mathbf{D}_t(E_t)} \exp \left(\hat{\Pi} \left(\mathbf{e} \left(\mathbf{d} \right)^{\star}, \mathbf{d}, \gamma \right) \right)} \right] \tag{2.33}$$

With panel data, this estimator converges to the true parameter values as T increases and as the size of $\mathbf{D}_t(E_t)$ increases. When T increases, the estimator approaches true values because there are a larger number of choices which must be explained by parameter values. This is akin to increasing the number of individuals in a standard discrete choice framework. When the size of $\mathbf{D}_t(E_t)$ increases, the estimator approaches true values because each semester's decision is more difficult to explain. For the theoretical estimator in (2.29), increasing the number of feasible alternatives shrinks the parameter subspace in which observed choices $\hat{\mathbf{d}}$ and $\hat{\mathbf{e}}$ yield a higher payoff than all feasible alternatives. For the stochastic maximum likelihood estimator in (2.32), increasing the number of feasible alternatives reduces reliance on specific realizations of the error terms for generating a unique solution.

Identification and estimation

As in Subsection 2.4.2, it is necessary to place some restrictions on γ_j. In my empirical application, I restrict γ_j to be equivalent within academic fields.[25] Intuitively, identification of γ_j comes from the share of courses offered by the observed university relative to the welfare maximizing composition of courses. For instance, if student parameter estimates imply students would prefer a university which offered more STEM and fewer humanities courses this implies the observed university is offering too many humanities courses and too few STEM courses. This reveals an institutional preference for drawing students out of STEM and into humanities courses.

One challenge with implementing the estimation algorithm described in this subsection is that the set of feasible vectors $\mathbf{D}(E)$ may be so large that summing over

[25] Academic fields are STEM, humanities and arts, business, social science, and occupational.

all $\mathbf{d} \in \mathbf{D}(E)$ is impractical. To address this, researchers can use a subset of $\mathbf{D}(E)$ rather than the full set. In this case, the estimator chooses parameter values which make the observed offering vector yield a higher payoff than a subset of alternative feasible vectors. This provides a computationally feasible algorithm at the cost of modest efficiency losses.

As before, this method also comes with a useful specification test. If all functional form assumptions are correct, the observed offering vector $\hat{\mathbf{d}}$ should yield a higher payoff than all alternative offering vectors at true student and university parameter values. If there are a significant number of alternatives which are preferred to $\hat{\mathbf{d}}$ at parameter estimates we can jointly reject the university objective structure and other functional form assumptions. Once again, failure to reject the null hypothesis does not validate the model structure.

Discussion of best offering vector methods

The strength of the variance minimizing methods described in this subsection is that they provide estimates of preference parameters for an objective structure which is more general than utilitarian student welfare maximization. Specifically, they quantify institutional preferences over the type of courses students choose. These institutional preferences measure the extent to which student and university preferences are not aligned.

The main disadvantage of these methods is that they depend on the characteristics of non-offered courses. As discussed in Section 2.3, researchers rarely observe non-offered courses; this implies that strong assumptions about non-offered courses are required to implement this method. Researchers can somewhat address this issue by using a subset of $\mathbf{D}(E)$ which only contains offering vectors which are deemed reasonable. Furthermore, these methods are very sensitive to estimates of minimum costs m_j because these estimates determine which offering vectors are feasible and how much residual money is left for excess spending. Any error in estimates of m_j may lead to spurious conclusions about university preference parameters.

2.5 Empirical Application

In this section, I describe estimation details and present estimates of student and university preference parameters. Student preference parameters are estimated using a nested logit course choice model where nests are defined by academic fields. Estimates show introductory humanities courses are most popular with first year students while introductory business courses are most popular with sophomores, juniors and seniors. The estimates also show students with higher ACT scores are relatively more attracted to introductory STEM courses. Finally, the estimates suggest higher salaried instructors generally increase an introductory course's desirability but only to a small degree.

University parameters are estimated using the maximum likelihood estimator presented in Subsection 2.4.3 using student parameter estimates as inputs.[26] Estimates suggest UCA has institutional preferences for decreasing enrollment in introductory business courses and increasing enrollments in introductory humanities and STEM courses. This suggests UCA over invests in introductory STEM and humanities courses and under invests in introductory business courses relative to a university

[26] I do not use the variance minimization method in my empirical application due to a lack of identifying variation in the data.

that is purely maximizing student welfare.

2.5.1 Estimation Details

In the sequential game between a university and students described in Section 2.2, university decisions about which courses to offer and how much to spend on instructors hinge on how the university expects students to respond to these decisions. As such, it is crucial to obtain credible estimates of how student course choices depend on the set of offered courses and spending on instructors for these courses.

To obtain these estimates, I use a multinomial nested logit model of student course choices where nests are defined by academic fields.[27] The nesting structure relaxes the independence of irrelevant alternatives assumption by allowing for correlation in unobserved preferences for courses of the same field. To avoid issues of unobserved heterogeneity in choice sets, I focus on introductory course choices only and assume that all introductory courses are in the choice sets of all enrolled students.[28] As such, the estimation method should be viewed as a conditional nested logit in which students choose which introductory courses to take conditional on already choosing to take some introductory course.

In this analysis, I consider the baseline setting in which class sizes do not affect the desirability of a course. Future research may employ a general equilibrium framework in which class size affects course desirability.

[27] See McFadden (1978). Academic fields are: STEM, social science, humanities and arts, occupational, and business.

[28] Many advanced courses have prerequisite restrictions implying they are only in the choice sets of students who have satisfied these pre-requisites.

Student utility, choice probabilities, and expected welfare

I assume the deterministic utility of introductory course j for student i depends on observed student characteristics X_i and excess spending on instruction in course j e_j as:

$$u_{ij}\left(e_j\right) = X_i\beta_{f(j)} + \theta_{f(j)}\ln\left(e_j + 1\right) \tag{2.34}$$

where $f\left(j\right)$ indicates the academic field of introductory course j. The logarithmic structure is included to make the marginal utility of excess spending diminish as spending increases.[29] X_i includes gender and ACT scores to capture heterogeneous preferences for academic fields by gender and initial preparation. Furthermore, X_i includes cohort dummy variables to allow for changes in relative preferences for introductory courses of different fields over the course of college.

I assume stochastic utility is given by deterministic utility with an additively separable error:

$$U_{ij}\left(e_j\right) = u_{ij} + \epsilon_{ij} \tag{2.35}$$

where ϵ_{ij} follows a nested logit structure in which nests are defined by academic fields.

With this structure, the probability student i chooses introductory course j conditional on choosing an introductory course is given by:

$$P_{ij}\left(\mathbf{e},\mathbf{d}\right) = \frac{\exp\left(\frac{u_{ij}}{\rho}\right)\left[\sum_{j'\in f(j)}\exp\left(\frac{u_{ij'}}{\rho}\right)\right]^{\rho-1}}{\sum_{f=1}^{F}\left[\sum_{j'\in f}\exp\left(\frac{u_{ij'}}{\rho}\right)\right]^{\rho}} \tag{2.36}$$

[29] I add 1 to excess spending to make the marginal utility of excess spending finite over the entire support of excess spending: $e_j \in [0,\infty)$.

where ρ is the nesting parameter.[30] Furthermore, with this structure, student i's expected welfare from her choice of which introductory course to take is given by:

$$\mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] = \ln\left\{\sum_{f=1}^{F}\left[\sum_{j' \in f}\exp\left(\frac{u_{ij'}}{\rho}\right)\right]^{\rho}\right\} + \gamma \tag{2.37}$$

where γ is the Euler-Mascheroni constant. Importantly, choice probabilities and expected welfare both depend on the university's choice of which courses to offer $\mathbf{d}$ and how much to spend in excess of minimum costs to increase instructor quality $\mathbf{e}$.

Maximum Likelihood Estimation

Let C_{it} represent the number of introductory courses taken by individual i in semester t and index these courses by c. Let y_{itcj} indicate whether individual i chooses introductory course j for choice c in academic semester t. The conditional likelihood that student i chooses her observed introductory course for choice c in academic semester t is given by:

$$\mathscr{L}_{itc} = \prod_{j \in \mathbf{d_t}} P_{ij}\left(\mathbf{e_t}, \mathbf{d_t}\right)^{y_{itcj}} \tag{2.38}$$

where $\mathbf{d}_t$ and $\mathbf{e}_t$ represent course offerings and excess spending in semester t respectively. Taking products over courses within semesters, students, and semesters, the conditional likelihood of observing the observed introductory course choices is given

[30] $\rho \in (0, 1]$. $1 - \rho$ can be viewed as an indication of the correlation in unobserved preferences within the same academic field (McFadden, 1978).

by:[31]

$$\mathscr{L}\left(\mathbf{y};\beta,\theta,\rho\right)=\prod_{t=1}^{T}\prod_{i=1}^{N_t}\prod_{c=1}^{C_{it}}\prod_{j\in\mathbf{d_t}}P_{ij}\left(\mathbf{e_t},\mathbf{d_t};\beta,\theta,\rho\right)^{y_{itcj}} \tag{2.39}$$

The log conditional likelihood is then given by:

$$\ln\mathscr{L}\left(\mathbf{y};\beta,\theta,\rho\right)=\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{c=1}^{C_{it}}\sum_{j\in\mathbf{d}_t}y_{itcj}\ln P_{ij}\left(\mathbf{e_t},\mathbf{d_t};\beta,\theta,\rho\right) \tag{2.40}$$

Estimates of student utility parameters β, θ, and ρ are obtained by numerically solving for the parameter values which maximize this log conditional likelihood.

2.5.2 Estimates of student utility parameters

Table 2.2 compiles estimates of the multinomial logit choice model. The estimates imply a first year male student with average ACT scores is most attracted to introductory humanities and arts courses followed by STEM, occupational, social science and business. First year female students with average scores are also most attracted to introductory humanities and arts courses followed by social science, occupational, STEM, and business. While introductory business courses are unpopular with freshmen, they are quite popular with more advanced students. With one exception, male and female sophomores, juniors, and seniors with average ACT scores favor introductory business courses to all other courses. Comparatively, while introductory humanities and arts courses are popular with Freshmen, they are rarely taken by more advanced students—sophomores, juniors and seniors with average ACT scores are least interested in taking introductory humanities and arts courses.

[31] This framework approximates choices by students within academic semesters as C_{it} independent choices. An alternative framework models students as choosing the best bundle of C_{it} courses from all feasible bundles. I abstract from this complication to focus on university preferences.

Table 2.2: Nested Logit Coefficient Estimates

	STEM	Social Science	Humanities	Occupational	Business
Intercept	0.2754***	0.098***	0.4986***	0.2188***	*omitted*
	0.0090	*0.0091*	*0.0096*	*0.0143*	
Female	0.514***	0.8093***	0.456***	0.6694***	*omitted*
	0.0112	*0.0115*	*0.0128*	*0.0187*	
ACT Z-score	0.1518***	0.0005	-0.0049	-0.0677***	*omitted*
	0.0062	*0.0057*	*0.0094*	*0.0146*	
Missing ACT	-0.222***	-0.055***	-0.1696***	0.0017	*omitted*
	0.0200	*0.0181*	*0.0187*	*0.0253*	
Sophomore	-1.7746***	-1.6281***	-2.3829***	-1.9234***	*omitted*
	0.0191	*0.0173*	*0.0176*	*0.0274*	
Junior	-2.2214***	-2.2204***	-3.2888***	-2.003***	*omitted*
	0.0251	*0.0233*	*0.0257*	*0.0317*	
Senior	-0.933***	-1.1566***	-2.3669***	-0.7185***	*omitted*
	0.0321	*0.0313*	*0.0358*	*0.0397*	
ln(spending)	0.0218***	0.052***	0.0432***	0.0067***	-0.0021
	0.0010	*0.0011*	*0.0012*	*0.0018*	*0.0019*

The estimates also imply students with higher ACT scores are relatively more likely to enroll in introductory STEM courses and slightly less likely to enroll in introductory occupational courses. For example, while a first year male student with average ACT scores prefers taking introductory humanities and arts courses, a first year male student whose ACT scores are 1.5 standard deviations above the mean is approximately indifferent between introductory STEM and humanities courses. The finding that students with higher ACT scores are relatively more likely to enroll in introductory STEM courses is consistent with existing literature which shows initial preparation is an important determinant of whether a student pursues a STEM education (Arcidiacono (2004); Stinebrickner and Stinebrickner (2014b)).

The return on excess spending results show that excess spending on instruction has a positive and significant effect on course desirability for all fields except business. However, the magnitudes of these estimates suggest the effects are quite small relative to non-spending preferences for fields. Social science courses appear most sensitive to excess spending but even these coefficients are small. The distribution of $\ln(e_j + 1)$ has mean 7.71 and standard deviation 2.10. This implies first year male students with average ACT scores are approximately indifferent between an introductory social science course with $\ln(e_j + 1)$ that is 3.67 standard deviations above zero and an introductory humanities course with zero excess spending. Put simply, even in the field most sensitive to excess spending it takes large increases in spending to overcome latent differences in preferences.

This finding has important implications for universities. It implies that the vast amounts of resources spent hiring instructors who cost more than minimally qualified teachers has relatively small effects on student course choices and student welfare. To see this more concretely, consider the comparison of three simple hypothetical universities given in Table 2.3. The baseline university offers one course from each field and hires minimally qualified instructors to teach these courses. This costs $14,420 and yields 61,829 units of welfare. Now suppose the university has surplus funds and wishes to either increase student welfare or increase enrollment in social science courses. In the "More Courses" alternative, the university spends its additional funds hiring another minimally qualified instructor to teach one additional social science course. Under this alternative, total cost is $17,124, social science enrollment is 28.3% of total enrollment, and student welfare is 66,282 units. In the "More Spending" alternative, the university spends the same amount of funds hiring a more qualified instructor to teach its one social science course. Under this alternative, total cost is still $17,124, but social science enrollment is only 24.2% of the total, and student welfare is only 64,509. This implies that even for social science—where excess spending has the largest effects on utility—it is more efficient to increase student welfare or change student course choices by offering additional courses rather than hiring more qualified instructors.

Table 2.3: Adding Courses vs Spending on Instruction

	Baseline	One more course	More spending
Introductory course offerings			
STEM	1	1	1
Social Science	1	2	1
Hum and Arts	1	1	1
Occupational	1	1	1
Business	1	1	1
Spending per course			
STEM	$2,819	$2,819	$2,819
Social Science	$2,704	$2,704	$5,408
Hum and Arts	$2,976	$2,976	$2,976
Occupational	$2,663	$2,663	$2,663
Business	$3,258	$3,258	$3,258
Total Cost	$14,420	$17,124	$17,124
Share of total student-course observations by field			
STEM	16.7%	14.4%	15.3%
Social Science	17.6%	28.3%	24.2%
Humanities	15.4%	13.2%	14.0%
Occupational	18.6%	16.0%	17.0%
Business	31.8%	28.1%	29.5%
Student Welfare	61,829	66,282	64,509

This result provides an interesting complement to existing literature which examines the effects of instructor qualifications on student learning. Figlio et al. (2015) use data from Northwestern University and find students learn relatively more from non-tenure track instructors—who generally have lower salaries—than tenure track instructors. Additionally, Bettinger and Long (2010) use data from public four year colleges in Ohio and find non-tenure track instructors make students more likely to take subsequent courses in a field. While my result only applies to introductory courses at University of Central Arkansas, it does corroborate existing literature which suggests the vast amounts of resources universities spend hiring more expensive instructors have small (or possibly negative) effects on the academic experiences of students.

2.5.3 Estimates of university parameters

Table 2.4 compiles estimates of university preference parameters obtained using the maximum likelihood estimation procedure discussed in Subsection 2.4.3.[32] Intuitively, this estimation algorithm solves for parameter values which best explain why observed introductory course offerings were preferred to alternative feasible offering vectors. To be feasible, an alternative offering vector must contain introductory courses which must be offered to honor preexisting contracts and must satisfy the budget constraint.[33]

[32] These are preliminary estimates which assume $\mathbf{e}(\mathbf{d})^{\star} = \mathbf{0}$ for all $\mathbf{d}$. Results which use the true $\mathbf{e}(\mathbf{d})^{\star}$ are in progress; however, as discussed in Subsection 2.5.2, estimates of student parameters imply these optimal excess spending levels will always be close to zero. As such, these preliminary estimates should be similar to the final estimates.

[33] As discussed in Subsection 2.4.3, the set of alternative feasible offering vectors is typically unfeasibly large making it necessary to draw a sample of alternative feasible offering vectors. I sample 1000 feasible offering vectors for each academic semester.

Recall from Section 2.4 that university payoffs are modeled as:

$$\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right] = \sum_{i=1}^{N} \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] + \sum_{j=1}^{J} \gamma_{f(j)} \tilde{n}_j \left(\mathbf{e}, \mathbf{d}\right) \tag{2.41}$$

Table 2.4 contains estimates of $\gamma_{f(j)}$. These parameters represent institutional preferences for the type of courses students choose. If $\gamma_f = 0$ for all f then the university's objective is to maximize total student welfare giving equal weight to all students. Non-zero estimates for γ_f reveal relative institutional preferences for the type of courses students choose. These institutional preferences lead to classroom spending decisions which are not in line with student preferences.

Table 2.4: University Preferences for the Type of Courses Students Choose

	Preferences for enrollments
STEM	0.778***
	0.038
Social Science	0.651***
	0.031
Humanities and Arts	1.229***
	0.058
Occupational	0.472***
	0.029
Business	*omitted*
Share of alternatives preferred to chosen option	
Fall 2007	0.5%
Spring 2008	0.1%
Fall 2008	0.0%
Spring 2009	0.0%
Alternatives per semester	1000
Number of semesters	4

The results in table 2.4 imply University of Central Arkansas (UCA) values enrollment in humanities and arts courses more than all other fields. After humanities and arts, UCA values STEM enrollment, social science enrollment, occupational enrollment and business enrollment. This implies UCA over-invests in humanities and arts courses and under-invests in business courses relative to a university whose objective is to maximize student welfare.

To directly see UCA's relative preference for humanities and arts enrollment, note that descriptive statistics in Table 2.1 show humanities courses comprise 31.6% - 34.5% of introductory course spending but only 30.7% - 32.2% of introductory course enrollment. This outsize investment reflects UCA's desire to increase enrollment in introductory humanities courses. Comparatively, business courses comprise only 5.7% - 7.3% of introductory course spending but make up 8.1% - 9.7% of introductory course enrollment. This under investment is consistent with an objective to draw students away from introductory business courses and into other fields.

Table 2.4 also includes measures assessing goodness of fit for the model of classroom spending decisions. As discussed in Subsection 2.4.3, model fit can be evaluated by calculating the fraction of alternative feasible offering vectors which are preferred to the chosen vector in the estimated model. If a substantial fraction of alternative feasible offering vectors yield larger university payoffs than the chosen vector then the model is explaining university decisions poorly. The results of this analysis show that at most 0.5% of alternative feasible offerings are preferred to the chosen vector in any given semester. In two of the four semesters, all alternative feasible offerings yield lower university payoffs than the chosen vector. This indicates that the fitted model is explaining university choices remarkably well.

2.6 Counter-factual Simulations

In Section 2.5, I presented estimates of university preference parameters which show University of Central Arkansas (UCA) has a relative preference for increasing enrollments in introductory humanities courses and decreasing enrollments in introductory business courses. To place estimates of university parameter values in context and to examine university behaviors under alternative constraints, I develop a Marginal Improvement Algorithm (MIA) for simulating classroom spending decisions under alternative preferences and constraints (Chade and Smith (2006)). A university's choice set is typically so large that it is intractable to solve its true maximization problem. The MIA reduces dimensionality by breaking the full maximization problem into a series of smaller maximization problems where the preference of each problem is to maximize marginal improvements to the university's payoff.

I use the MIA to examine two counterfactual scenarios: First, I solve for counterfactual minimum costs m_j which induce UCA to offer courses which maximize student welfare. The simulation suggests that a revenue neutral tax and subsidy policy which increases the cost of offering introductory humanities, STEM, social science, and occupational courses and decreases the cost of offering business courses leads UCA to offer courses which maximize student welfare. Second, I simulate course offerings and excess spending decisions which produce welfare efficiently in the absence of contractual constraints. This simulation shows UCA could achieve the same student welfare at 38.5% of original costs in the absence of contractual constraints.

2.6.1 Marginal Improvement Algorithm

This subsection presents a Marginal Improvement Algorithm (Chade and Smith (2006)) for approximating the course offerings and excess spending decisions of a university with a known objective function. Broadly speaking, the algorithm iteratively adds single courses which best complement previously selected courses until the budget constraint is satisfied or no marginal improving courses exist.

To illustrate these methods more concretely, note that it is often possible to combine tangency conditions given by Equation (2.7) with the budget constraint to solve for the optimal excess spending vector for each offering vector.[34] Denote these optimal excess spending vectors as: $\mathbf{e}(\mathbf{d})^{\star}$. The university's problem can then be restated to focus on extensive margin decisions:

$$\mathbf{d}^{\star} = \operatorname{argmax}_{\mathbf{d}} \left\{ \mathbb{E}\left[\Pi \,|\, \mathbf{e}(\mathbf{d})^{\star}, \mathbf{d}\right] \right\} \;\; \forall \mathbf{d} \,\text{s.t.} \sum_{j=1}^{J} d_j f_j \leqslant E \tag{2.42}$$

Because the offering vector $\mathbf{d}$ is discrete, Lagrange methods cannot be used to characterize properties of the extensive margin solution. Furthermore, because the number of feasible $\mathbf{d}$ is typically very large, directly solving the problem is impractical.[35] To solve for $\mathbf{d}^{\star}$, the Marginal Improvement Algorithm starts by selecting the single course offering which delivers the greatest expected payoff to the university.[36]

[34] See Appendix E for an illustration.

[35] In my empirical application, a conservative lower bound for the number of feasible choices for $\mathbf{d}$ is 2.95×10^{13}.

[36] I describe the algorithm for a setting in which the university has no contractual constraints. To incorporate contractual constraints, the algorithm should start with the set of courses which must be offered by contract. A variation of the algorithm selects courses which yield the greatest marginal improvement per minimum cost. This variation outperforms the standard version in settings where variation in minimum costs is large relative to marginal utilities of excess spending.

Denote this course by $j_1^\star$. Finding $j_1^\star$ requires computing the university's payoff for every potential course. Following this, the algorithm selects the best course to offer alongside $j_1^\star$. This entails calculating the university's payoff for every offering vector which includes $j_1^\star$ and one other potential course. The algorithm continues adding marginally improving course until marginal effects turn negative or until the constraint $\sum_{j=1}^{J} d_j m_j \leqslant E$ binds. Technical details on this algorithm are provided in Appendix F.

2.6.2 Counter-factual costs which yield welfare maximizing course offerings

Table 2.5 reports market minimum costs and counterfactual minimum costs m_j which induce UCA to offer courses which maximize student welfare. The counterfactual setting increases minimum costs of offering introductory STEM, humanities, social science, and occupational courses and decreases the minimum costs of offering introductory business courses. These counterfactual costs can be achieved with a tax and subsidy policy that is approximately revenue neutral.

Column 1 reports optimal course offerings for a welfare maximizing university with a 1.9 million dollar endowment and no contractual constraints facing market minimum costs estimated from data. The results suggest introductory business and occupational courses yield the best value to students while introductory STEM and humanities courses are less desirable. Column 2 reports offerings for an unconstrained university with the same endowment and minimum costs but with estimated institutional preferences reported in Table 2.4. These results show UCA's preference for students choosing humanities and STEM courses relative to occupational and business courses lead to many more humanities and STEM course offerings and fewer

Table 2.5: Counterfactual costs which yield Welfare Maximizing Course Offerings

	Welfare Max at Market Costs	Estimated Obj at market costs	Estimated Obj at counter costs
	(1)	(2)	(3)
		Minimum costs	
	Market costs	Market costs	Counter costs
STEM	\$2,819.0	\$2,819.0	\$4,081.6
Social Science	\$2,704.0	\$2,704.0	\$3,563.9
Humanities	\$2,976.0	\$2,976.0	\$5,560.0
Occupational	\$2,663.0	\$2,663.0	\$3,033.3
Business	\$3,258.0	\$3,258.0	\$2,571.4
		Simulated Optimal Course Offerings	
STEM	34	160	34
Social Science	86	78	86
Humanities	21	423	21
Occupational	171	10	171
Business	338	0	338
Total Cost	\$1,947,463	\$1,947,430	\$1,949,873

occupational and business courses.

Column 3 shows the welfare maximizing offerings presented in column 1 will be chosen by a university with estimated institutional preferences under counter-factual minimum costs. The counter-factual costs represent a 87% increase in the minimum cost of introductory humanities courses, a 45% increase for STEM courses, a 32% increase for social science courses, a 14% increase in occupational courses, and a 21% decrease in the minimum cost of introductory business courses.

These counterfactual costs could be achieved with a fixed tax on instruction spending in introductory humanities, STEM, social science, and occupational courses and a fixed subsidy for spending on introductory business courses. The cost of such a policy would be:

$$\tau = \sum_{j \in \mathbf{d}} \left(m_j - m'_j \right) \tag{2.43}$$

where m_j represents market minimum costs and m'_j represents counterfactual minimum costs. In my application, $\tau = \$2,410$ or 0.12% of the university's endowment making it effectively revenue neutral.

This demonstrates that an approximately revenue neutral tax and subsidy policy which modifies the relative minimum costs of offering different courses can induce UCA to offer courses which maximize student welfare. While such a policy may be impractical or undesirable for other reasons it is interesting to see what counterfactual minimum costs would price out institutional preferences for enrollments in different fields.

2.6.3 Welfare maximizing classroom spending decisions at University of Central Arkansas

Table 2.6 compares UCA's observed course offerings in the Fall semester of 2007 to the course offerings which produce welfare efficiently in a scenario with no contractual constraints. The differences are quite striking. In the counterfactual scenario, UCA achieves the same student welfare at 38.5% of original costs. These savings are primarily due to reductions in spending on instruction. As discussed in Section 2.5, the student utility parameter estimates suggest spending on instruction has relatively small effects on course desirability. Because students place little value on spending on instruction, a counterfactual university which produces welfare efficiently spends 54.5% - 66.5% less on instruction in median courses by field. This saves vast sums of money and only modestly decreases student welfare.

Table 2.6: Counterfactual costs which yield Welfare Maximizing Course Offerings

	Observed	Welfare Maximizing
Introductory course offerings		
STEM	279	38
Social Sciences	272	518
Humanities	433	54
Occupational	125	110
Business	65	521
Median spending per course		
STEM	$8,410	$2,819
Social Sciences	$6,868	$2,704
Humanities	$6,547	$2,976
Occupational	$7,266	$2,663
Business	$9,480	$3,258
Share of students choosing each field		
STEM	22.2%	3.0%
Social Sciences	26.8%	38.7%
Humanities	31.6%	4.5%
Occupational	11.3%	7.9%
Business	8.1%	45.9%
Relative Cost	1	0.385
Welfare	221955	224251
Relative Welfare	1.000	1.010
Total Courses	1174	1241

The results also show that UCA could alter its introductory course composition to better serve students. Observed course offerings at UCA include 433 introductory humanities courses, 279 introductory STEM courses, 272 introductory social science courses, 125 introductory occupational courses, and only 65 introductory business courses. Comparatively, the counterfactual welfare maximizing university offers 521 introductory business courses, 518 introductory social science courses, 110 introductory occupational courses, and only 54 introductory humanities courses and 38 introductory STEM courses.

Importantly, although the counter-factual scenario is very different from observed course offerings the total number of courses offered is similar in both scenarios. This suggests the welfare maximizing offerings would not require large changes in facilities which would introduce costs not included in my analysis. While such a vastly different university may be undesirable for other reasons, it is striking to see that students could obtain the same welfare with drastically lower costs and interesting to note what alternative classroom spending decisions achieve these savings.

2.7 Conclusion

Despite the importance of universities in our society, very little is known about how universities make decisions. A better understanding of how universities make decisions could lead to policies which benefit students and reduce financial burdens on taxpayers, families, and donors.

In this paper, I develop tools for revealing university preferences from decisions of which courses to offer and how much to spend on instructors for these courses. The methods include a statistical test of whether classroom spending decisions maximize student welfare and two methods for estimating a university's relative preferences for student welfare and the type of courses students choose. I apply these methods to administrative data from University of Central Arkansas (UCA) and find UCA has institutional preferences for decreasing enrollment in introductory business courses and increasing enrollments in introductory humanities and STEM courses.

In addition to discussing methods for inferring the preferences of an observed university, I also present a method for simulating the classroom spending decisions of a university with alternative preferences or facing counter-factual constraints. I use this method to run two simulations: First, I show that a revenue neutral tax and subsidy policy which reduces the cost of offering introductory business courses and increases the cost of offering other introductory courses can induce UCA to offer courses which maximize student welfare. Second, I show that UCA could achieve the same student welfare at 38.5% of original costs in the absence of contractual constraints. The savings primarily result from hiring less expensive instructors but are also generated by offering more introductory business courses and fewer introductory STEM and humanities courses. While these scenarios may be undesirable for other reasons, it is useful to see how a revenue neutral policy could be used to benefit students and it is striking to see that students could receive the same benefit with drastically lower costs with changes in instructors and course composition.

Future work may build upon my framework and methods in many possible ways. First, empirical applications which allow class size to affect utility and estimate non-utilitarian weights on student welfare should be conducted. I describe methods for these extensions but do not include these features in my empirical application.

Second, my framework and methods could be extended to include a larger set of university decisions. For clarity and tractability, I focus on revealing university preferences from observed classroom spending decisions for non-contract introductory courses. Data on prerequisite restrictions could be used to extend my methods to include non-contract advanced courses. Moreover, a dynamic university model could potentially model decisions of which instructors to hire on long term contracts. Such an analysis may explain why UCA and other schools spend vast amounts of resources hiring more expensive instructors despite evidence that these sums have small effects on student choices and welfare.

Third, my model for student course choices could be expanded to create a richer link between classroom spending decisions and student outcomes. My model for student course choices is static and reduced form. While this is a tractable way to relate classroom spending decisions to student outcomes it has many limitations. My specification assumes students derive utility from spending on instruction. In fact, students derive utility from instructor characteristics which are correlated with spending on instruction. A more structural student choice model which captures the relationship between instructor characteristics and utility could be used to extend the university's problem to a three tier decision where the university chooses which courses to offer, how much to spend on instruction, and what instructor characteristics to rent with these funds. This more structural model would yield a deeper understanding of university decisions and the implications for students.

Furthermore, a richer dynamic model of student course choices could provide a more complete portrait of the relationship between university decisions and student outcomes. For example, a dynamic model which includes major choices, dropout decisions, or labor market outcomes could be used to reveal university preferences for these outcomes. A deeper understanding of institutional preferences could be used to determine whether institutional preferences are socially beneficial phenomena which internalize externalities or whether they are socially detrimental factors which hurt the very students universities exist to serve.

3

Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors

Additional authors for this chapter: Thomas Ahn, Peter Arcidiacono, and Amy Hopson

3.1 Introduction

Even after accounting for selection, substantial earnings differences exist across majors. Majors in engineering and the sciences, as well as economics and business, pay substantially more than other fields.[1] Further, earnings disparities across majors have increased substantially over time (Altonji et al. (2014) and Gemici and Wiswall (2014)). Despite their value in the marketplace, STEM (Science, Technology, Engineering, and Mathematics) fields are perceived to be undersubscribed.

But many more students enroll in college expecting to major in a STEM field

[1] See Altonji et al. (2012) and Altonji et al. (2015) for reviews.

than actually finish in a STEM field (Arcidiacono (2004), Arcidiacono et al. (2015), Stinebrickner and Stinebrickner (2014b)). This is not just due to students dropping out: many students switch from STEM to non-STEM fields, particularly in comparison to those who switch from non-STEM to STEM fields. Further, it is predictable who will switch. Those who have relatively weak academic preparation (e.g. SAT scores or HS grades) are much more likely to leave STEM fields. While relatively high levels of academic preparation are associated with persisting in STEM majors, there is little evidence that high levels of academic preparation are more rewarded in the labor market for STEM majors than for non-STEM majors. Women are also more likely to switch: Arcidiacono et al. (2012) show with data from Duke University, that differences in academic preparation can account for the large differences in switching behavior across races but is unable to explain the substantial gender gap.

A potential channel for influencing the number and composition of STEM majors are grading policies. Should grading policies prove to be an important predictor of major choice, they may serve as a relatively cheap way of increasing STEM majors. While other means such as increasing pre-college academic preparation or the share of underrepresented groups in STEM fields may also be effective, these methods are also very costly with the benefits coming much later. Shifting the way teachers teach the sciences and introducing more laboratory-based curricula are both expensive. Altering training, hiring, and promotion in academia, government agencies, and firms is also costly, as are adjusting long-standing cultural attitudes in the home, school, and workplace.

There is evidence that grades affect sorting into majors. The same majors that pay well also give (on average) significantly lower grades (Sabot and Wakeman-Linn

(1991), Johnson (2003)) and are associated with more study time (Brint et al. (2012), Stinebrickner and Stinebrickner (2014b)). Lower grades and higher study times deter enrollment. Sabot and Wakeman-Linn (1991) show that the absolute level of grades was a far more important indicator of taking further courses in the subject than their ranking within the class. Butcher et al. (2014) showed that Wellesley's policy of capping the fraction of A's given resulted in shifts towards science classes and science majors. There is also evidence that students enter unaware of the extent of cross-department differences in grading standards. Stinebrickner and Stinebrickner (2014b) show that the over-optimism regarding performance at Berea College is primarily driven by students over-predicting their performance in the sciences. As students take more classes, students generally revise their expected performance in the sciences downward. This holds true even for students who persist in the sciences who ought to have received relatively positive grade realizations.

With students responding to grading practices through their choice of courses, departments may set their grading policies in order to deter or encourage enrollment. Those with low enrollments may find it difficult to increase or maintain their faculty size. Hence, incentives exist to raise grades in order to encourage enrollment in these departments. On the other hand, departments that are flush with students may have incentives to lower grades to keep their enrollments to a more manageable size. Within any given department, individual professors may also seek to influence enrollment up or down for his or her class to minimize teaching effort or maximize student learning.[2]

[2] This issue is becoming even more salient as more universities move toward a fiscal model where departmental budgets are more directly determined by enrollment size or credits generated.

Differences in grading policies may have differing effects for males and females. In principle female students should be particularly interested in STEM fields. Women report studying substantially more than men (Stinebrickner and Stinebrickner (2014b), Arcidiacono et al. (2012)), and they should be undeterred by the higher study requirements of these classes. Yet, females are substantially less likely to graduate with a STEM major than males.[3] Ideas for why this might happen have been numerous, including role model effects (Rask and Bailey (2002), Hoffmann and Oreopoulos (2009), Carrell et al. (2010)) and future labor market considerations (Gemici and Wiswall (2014), Bronson (2014)) among many others. In addition to these channels, women may study more in part because they value the benefits of studying–higher grades–more than their male counterparts (Rask and Bailey (2002), Rask and Tiefenthaler (2008)). Good grades may yield direct psychic benefit, or they may impact time to graduation or ability to qualify for grants and scholarships. For example, if female students are more risk averse or pessimistic about attrition probability compared to their male counterparts, grades may hold more value. Again, the advantage to focusing on grading policies is that it may be relatively cheap to do so compared to alternative programs.

We propose to estimate an equilibrium model of student course enrollment and effort decisions as well as professor decisions regarding grading standards. How professors set grades affects enrollment and how much students study, though differentially for men and women. The professor objective function includes enrollments, so part of how professors set grades is determined by course demand. With the esti-

[3] The gender gap is not uniform across STEM fields. Indeed, in some STEM fields, such as biosciences, women receiving BA's actually outnumber men.

mates of the equilibrium model, we will be able to evaluate how differences in grading practices across fields affect, partly as a result of demand, the share of courses taken in different fields. Further, we can see whether cross-departmental differences in professor preferences over enrollment either exacerbate or mitigate the differences in grading across fields.

3.2 Data

Estimating such a model requires rich data on student course taking, study hours, and grades. We use a detailed student enrollment data set from the University of Kentucky (UK). UK, the state's flagship public post-secondary institution, has a current undergraduate enrollment of approximately 21,000. The school was ranked 119 out of approximately 200 'National Universities' by U.S. News & World Report (U.S. News & World Report 2013). This places UK in the middle of the distribution of large post-secondary institutions, and the student body serves as a good cross-section of college students nationwide.

The data set contains student demographic and enrollment information, spanning Fall 2008 to Spring 2013. Each semester, the entire student body's course selections and grades are recorded by the Registrar's Office. Enrollment data can be linked across semesters to provide a complete panel data of every student's academic trajectories across the ten semesters. This yields approximately 1.4 million student/class observations. This data set is particularly valuable because every student outcome in every class is captured, allowing us to estimate a rich model of student and professor interactions. Furthermore, we can analyze course selection and performance in the context of a 'class bundle.' For this study, we focus on student enrollment observations from one semester, Fall 2012.

In addition, we have access to class evaluation surveys completed by students at the end of the semester. We note that coverage is not complete, as some departments chose not to make evaluation data available. Data from classes with a small number of student respondents are deleted, to prevent possible identification. For our Fall 2012 sample, we are able to link 1,086 classes to the enrollment data, which represents a 76 percent successful match rate. We use classes with at least a 70 percent response rate. The survey asks 20 questions on the value of the course and instructor to the student on a five-point Likert scale. Each student reveals what year of school he or she is in, how many hours per week spent studying for this course, expected final grade, and whether the course was a major requirement. Evaluation data cannot be linked to individual students; we use class average data.

Restricting the sample to Fall 2012 yields 89,582 student/class observations. There are 19,527 unique undergraduates, implying that on average, each student enrolls in (but not necessarily completes) four to five courses.[4] Table 3.1 provides demographic summary statistics, separated by gender. Overall, women and men look similar when entering college. Women have slightly higher high school grades and slightly lower standardized ACT scores.[5] Women also have higher grades while in college. Sharp differences show up in major selection. While women comprise a slight majority at UK overall, the ratio between men and women in STEM majors is approximately 1.6.

[4] We also observe withdrawal data.

[5] SAT scores are converted to equivalent ACT scores, and the math and verbal sections are averaged.

Table 3.2 summarizes class-level characteristics separated by STEM-status of the course. STEM classes are substantially larger and give significantly lower grades compared to non-STEM courses. As implied by Table 3.1, female students are the minority in STEM classes. This is despite the fact that they perform better, on average, than their male counterparts in these courses. On average, each STEM course requires one more hour of study time per week (or 30 percent more time/effort) than a non-STEM course.

Table 3.3 presents simple OLS results showing the relationship between individual and class characteristics with grades and study hours after controlling for a large number of academic background measures.[6] The grades regression sample is at the student/class level, and the study hours per week regression sample is at the class level. The first column gives the results for grades. The patterns are consistent with those in Table 3.2, STEM classes give lower grades and females have higher grades. Classes that have a higher fraction of female students also give higher grades. This is consistent with there not being a grade curve that is common across STEM or non-STEM departments else the higher grades females receive would translate into lower grades for everyone else. Class size has a negative effect on grades. The coefficient on class size confounds two effects that work in opposite directions. On the one hand, students prefer higher grades so higher enrollments should be associated with higher grades. On the other, courses that have high demand for reasons besides grades may have lower grades since these courses do not need to have high grades to attract students.

The second column on Table 3.3 shows regressions of study hours on the average

[6] We restrict our sample to standard classes with at least 16 students.

Table 3.1: 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.

characteristics of the class. STEM classes are associated with an extra half hour of study, slightly less than what is seen in the descriptive statistics. This suggests that STEM classes are attracting students who are willing to study more. Classes that have more women also study more, consistent with the previous literature. But perhaps the most interesting coefficient is that on average grades. Courses that give higher grades have less study time, suggesting grades should be interpreted as relative, not absolute, measures of accomplishment, as well as suggesting grade inflation may have negative consequences for learning.

3.3 Model

Individual i chooses n courses from the set $[1, \ldots, J]$. Let $d_{ij} = 1$ if j is one of the n courses chosen by student i and zero otherwise. The payoff associated with a bundle of courses is given by the sum of the payoffs for each of the individual courses where the payoffs do not depend on the other courses in the bundle. We specify the payoff

Table 3.2: 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.

Table 3.3: 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 score, HS gpa, percent minority, percent freshman.
Additional controls in study hours regression include percent freshmen, percent STEM major, percent pell grant, percent in-state, average ACT score, average HS gpa, percent minority.

for a particular course j as depending on student i's preference for the course, δ_{ij}, the amount of study effort the individual chooses to exert in the course, s_{ij}, and the expected grade conditional on study effort, $\mathbb{E}\left[g_{ij} \mid s_{ij}\right]$:

$$U_{ij} = \phi_i \mathbb{E}\left[g_{ij} \mid s_{ij}\right] - \psi_i s_{ij} + \delta_{ij} \tag{3.1}$$

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

$$\max_{d_{i1},\dots,d_{iJ}} \quad \sum_{j=1}^{J} d_{ij} U_{ij} \tag{3.2}$$

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

The grade student i receives in course j, g_{ij}, depends on the academic preparation of student i for course j, A_{ij}, the amount of study effort put forth by the student in the course, s_{ij}, the grading policies of the professor, and a shock that is unknown to the individual at the time of course enrollment, η_{ij}. We specify the grading process as:

$$g_{ij} = \beta_j + \gamma_j \left(A_{ij} + \ln(s_{ij})\right) + \eta_{ij} \tag{3.3}$$

Grading policies by the professors are then choices over an intercept, β_j, and a return to academic preparation and effort, γ_j.[7] Gains from study effort enters in as a log to capture the diminishing returns to studying. Along with the linear study effort cost defined in the utility function, this ensures an interior solution for the optimal amount of study time.

[7] For example, if there is a university-wide (or department-level) mandated/recommended grade distribution, we will be able to capture such a policy, as β_j and γ_j will have lower variance.

Students are assumed to know the professors' grading policies.[8] Substituting in for expected grades in (3.1) yields:

$$U_{ij} = \phi_i \left(\beta_j + \gamma_j \left[A_{ij} + \ln(s_{ij})\right]\right) - \psi_i s_{ij} + \delta_{ij} \tag{3.4}$$

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

$$\begin{aligned} 0 &= \frac{\phi_i \gamma_j}{s_{ij}} - \psi_i \\ s^{\star}_{ij} &= \frac{\phi_i \gamma_j}{\psi_i} \end{aligned} \tag{3.5}$$

Substituting the optimal choice of study time into (3.4) yields:

$$U_{ij} = \phi_i \left(\beta_j + \gamma_j \left[A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) - 1\right]\right) + \delta_{ij} \tag{3.6}$$

Those who have lower study costs, low ψ_i, and higher levels of academic preparation, high A_{ij}, find courses with higher γ_j's relatively more attractive all else equal. Those who place a relatively high weight on expected grades, high ϕ_i, study more conditional on choosing the same course, but are more attracted to courses with higher grade intercepts, high β_j.

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

$$g_{ij} = \beta_j + \gamma_j \left(A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)\right) + \eta_{ij} \tag{3.7}$$

[8] Students have a number of formal and informal resources to learn about grading policies. Informally, they may rely on friends who have previously taken the course and other information social networks. Professors may send out preemptive signals by posting syllabi online. More formally, course evaluations, which also reveal the (anonymous) responders' own expected final course grades, are on-line and publicly available. In addition, several websites curate online "reviews" of professors and courses.

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.

The key equations for estimation are then given by:

(i) the solution to the students maximization problem where (3.6) is substituted into (3.2),

(ii) the grade production process given in (3.7), and

(iii) the optimal study effort given in (3.5).

The next section describes the parameterizations used to estimate the model as well as the assumptions necessary to overcome the fact that our measures of study effort from the course evaluations are not linked to the individual's characteristics.

3.4 Estimation

3.4.1 Parameterizations

To estimate the model, we need to place some structure on course preferences, δ_{ij}, the value of grades, ψ_i, and the cost of effort, ϕ_i. Further, we must relate academic preparation, A_{ij}, to what we see in the data. Denote $w_i = 1$ if individual i is female and zero otherwise. Denote X_i as a row vector of explanatory variables such as ACT scores, high school grades, race, etc.[9] Denote Z_i a a row vector of explanatory variables that affect preferences for particular departments or levels of courses within departments. Hence Z_i includes gender as well as year in school,

[9] The majority of students at the University of Kentucky submit ACT scores in their college applications.

allowing women to have a preference for classes in particular departments and the attraction of upper-division versus lower-division classes to vary by department and year in school. Preference shocks for courses are represented by ϵ_{ij}. Finally, we partition courses into K departments, $K < J$, where $k(j)$ gives the department for the jth course. We then parameterize the model as follows:

$$A_{ij} = w_i\alpha_{1k(j)} + X_i\alpha_{2k(j)} \tag{3.8}$$

$$\delta_{ij} = \delta_{0j} + w_i\delta_{1k(j)} + Z_i\delta_{2k(j)} + \epsilon_{ij} \tag{3.9}$$

$$\psi_i = \exp\left(\psi_0 + w_i\psi_1 + X_i\psi_2\right) \tag{3.10}$$

$$\phi_i = \phi_0 + w_i\phi_1 \tag{3.11}$$

There is no intercept in A_{ij} as it can not be identified separately from the β_j's. Note that the same variables enter into academic preparation, preferences, and effort costs, only with different coefficients. Preferences for courses allow for both course fixed effects as well as students with particular characteristics preferring courses in particular departments, $\delta_{1k(j)}$. Note also that the effort costs are exponential in the explanatory variables. This ensures that effort costs are positive. Finally, preferences for grades are only allowed to vary by gender. In principle, we could allow them to vary with X_i as well, but this would substantially complicate the model.

3.4.2 Estimation without Unobserved Heterogeneity

Grade parameters

Substituting the parameterizations for academic preparation, A_i, the value of grades, ϕ_i, and study costs, ψ_i, into (3.7) yields the following reduced form grade equation:

$$g_{ij} = \theta_{0j} + \gamma_j\left(w_i\theta_{1k(j)} + X_i\theta_{2k(j)}\right) + \eta_{ij} \tag{3.12}$$

where:

$$\theta_{0j} = \beta_j + \gamma_j(\ln(\phi_0) + \ln(\gamma_j) - \psi_0) \quad (3.13)$$

$$\theta_{1k(j)} = \alpha_{1k(j)} + \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 \quad (3.14)$$

$$\theta_{2k(j)} = \alpha_{2k(j)} - \psi_2 \quad (3.15)$$

We estimate the reduced form parameters $\{\theta_{0j}, \theta_1, \theta_2\}$ as well as the structural slopes, the γ_j's, using nonlinear least squares. One γ_j must be normalized to separately identify the remaining parameters so we set one $\gamma_j = 1$ for each department.[10] The variation in the data used to identify $\{\theta_1, \theta_2\}$ comes from the relationship between student characteristics and grades. The variation in the data used to identify the γ_j's is how these characteristics translate into grades relative to the normalized course.

Study parameters

We next turn to recovering some of the study effort parameters as well as undoing the normalization made on the γ's. To do so, we use (3.5). The issue with using (3.5) is that we do not directly observe study effort. However, the course evaluation data give reported study hours for each individual in the classroom. This information cannot be linked to the individual data on grades, academic preparation, and course choices. But the evaluation data does provide information about the year in school of the evaluator (e.g., freshman, sophomore, junior, or senior).

To link study hours to study effort, we assume that effort translates into hours linearly at rate μ but is reported with multiplicative measurement error ζ_{ij}:

[10] The study effort analysis allows us to recover the normalizations for all the departments but one, as we will show in section 3.4.2.

$$h_{ij} = \mu_1 (s^*_{ij})^{\mu_2} \exp(\zeta_{ij}) \tag{3.16}$$

Taking logs and substituting in for s^*_{ij} yields:

$$\begin{aligned} \ln(h_{ij}) &= \mu_2 \left(\ln(\mu_1) + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)\right) + \zeta_{ij} && (3.17) \\ &= \kappa_0 + w_i \kappa_1 - X_i \kappa_2 + \ln(\gamma_j) + \zeta_{ij} && (3.18) \end{aligned}$$

where:

$$\begin{aligned} \kappa_0 &= \mu_2(\ln(\mu) - \psi_0 + \ln(\phi_0)) && (3.19) \\ \kappa_1 &= \mu_2(\ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1) && (3.20) \\ \kappa_2 &= \mu_2 \psi_2 && (3.21) \end{aligned}$$

The coefficient on $\ln(\gamma_j)$ then gives the curvature of the relationship between study effort and hours.

Recall that we had to normalize one γ_j for every department in the grade equation. Denote the normalized values of γ as γ^* which relates to the unnormalized returns through $\gamma^*_j = \gamma_j / C_{k(j)}$.[11] Given our estimate $\hat{\gamma}^*$ from the grade equation and rearranging terms we have:

$$\ln(h_{ij}) = \kappa_0 + w_i \kappa_1 - X_i \kappa_2 + \kappa_{3k(j)} + \mu_2 \ln(\hat{\gamma}^*) + \zeta_{ij} \tag{3.22}$$

where $\kappa_{3k(j)} = \mu_2 \ln(C_{k(j)})$.

Since we can only link characteristics of the students to the evaluation data by year in school, the observations we use in estimating the study parameters are at

[11] Note that one γ_j is still normalized to one, just not one γ_j in each department.

the class-year level. Let l_i indicate the year in school of student i. Our estimating equation for students of level l is then:

$$\frac{\sum_i (l_i = l) d_{ij} \ln(h_{ij})}{\sum_i (l_i = l) d_{ij}} = \kappa_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{3k(j)} + \mu_2 \ln(\hat{\gamma}^*) + \zeta_{jl} \tag{3.23}$$

where w_{jl} and X_{jl} are the averages of these characteristics for those of year level l enrolled in course j.

Our estimate of the unnormalized returns to study and ability and then be recovered using $\hat{\gamma}_j = \hat{\gamma}^*_j \exp(\hat{\kappa}_{2k(j)}/\hat{\mu}_2)$. The department-specific weights on each of the observed characteristics (with the exception of gender) can now be recovered given $\hat{\gamma}$ and the estimates of the study costs, $\hat{\psi}_2 = \hat{\kappa}_2/\hat{\mu}_2$. Namely, $\hat{\alpha}_{2k(j)} = \hat{\psi}_2 + \hat{\theta}_{2k(j)} \exp(-\hat{\kappa}_{2k(j)}/\hat{\mu}_2)$.

Disentangling female preparation in each of the departments, $\hat{\alpha}_{1j}$, and female study costs, $\hat{\psi}_1$, first requires estimates of female preferences for grades (described in the next section). We cannot separate out μ_1, ψ_1, and the normalization on γ for one of the departments. We make note of how this lack of identification does and does not affect the interpretation of our results and counterfactuals later in the paper.

Utility parameters

We now turn to estimation of the parameters of the utility function. Given our estimates of the grade equation, equation (3.12), we can calculate expected grades in each of the courses given optimal study choices:

$$\widehat{E[g_{ij}]} = \hat{\theta}_{0j} + \hat{\gamma}^*_j \left(w_i \hat{\theta}_{1k(j)} + X_i \hat{\theta}_{2k(j)} \right) \tag{3.24}$$

Given the estimates of the unnormalized returns to study and ability, $\hat{\gamma}$, we can express the utility i receives from choosing course j and studying optimally as:

$$U_{ij} = \delta_{0j} + w_i\delta_{1k(j)} + Z_{ij}\delta_{2k(j)} + \left(\widehat{E[g_{ij}]} - \hat{\gamma}_j\right)(\phi_0 + w_i\phi_1) + \epsilon_{ij} \tag{3.25}$$

The goal is then to recover the course fixed effects, δ_{0j}, the value women place on courses in particular departments, $\delta_{1k(j)}$, other department-specific preferences as well as preferences over instructor characteristics, $\delta_{2k(j)}$, and preferences over grades, ϕ.

We assume that ϵ_{ij} is distributed Type 1 extreme value. If individuals were choosing one course, estimation of the parameters in (3.25) would follow a multinomial logit. Students, however, choose bundles of courses. Even though the structure of the model is such that there are no complementarities for choosing particular combinations of courses, the probability of choosing a particular bundle does not reduce to the probabilities of choosing each of the courses separately.

We use simulated maximum likelihood to estimate the choice parameters. To illustrate the approach, 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 $\{1, 2, 3\}$. The probability of choosing $\{1,2,3\}$ could then be expressed as:

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

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

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:

$$\max_{\phi,\delta} \sum_i \ln \left(\left[\sum_{r=1}^{R} \prod_{j=1}^{J} \left(1 - G(M_{ir} - \overline{U}_{ij})\right)^{d_{ij}} \right] / R \right) \tag{3.26}$$

Given $\hat{\phi}_0$ and $\hat{\phi}_1$, we can calculate $\hat{\psi}_1$ using (3.20) where $\hat{\psi}_1 = \ln(\hat{\phi}_0 + \hat{\phi}_1) - \ln(\hat{\phi}_0) - \hat{\kappa}_1$ from the study equation. We can then recover $\hat{\alpha}_{1k(j)}$ from (3.14) where $\hat{\alpha}_{1k(j)} = \hat{\theta}_{1k(j)} - \ln(\hat{\phi}_0 + \hat{\phi}_1) + \ln(\hat{\phi}_0) + \hat{\psi}_1$ from the grade equation.

3.4.3 Estimation with Unobserved Heterogeneity

We now consider the case when one of the components of X_i is unknown to take into account correlation across outcomes for the same individual. We assume that this missing component takes on S values where π_s is the unconditional probability of the sth value. Let X_{is} be the set of covariates under the assumption that individual i is of type s. The components of the unobserved heterogeneity are identified through the correlation of grades in each of the courses as well as the probabilities of choosing different course combinations.

Integrating out over this missing component destroys the additive separability of the log likelihood function suggesting that the estimation of the three sets of parameters (grades, course choices, and study time) can no longer be estimated in stages. However, using the insights of Arcidiacono and Jones (2003) and Arcidiacono and Miller (2011), it is possible to estimate some of the parameters in a first stage.

In particular, note that the selection problem occurs because students select into courses. By focusing just on the grade estimation as well as a reduced form of the choice problem, we can greatly simplify estimation, recovering the grade parameters as well as the conditional probabilities of being each of the types. These conditional type probabilities can then be used as weights in the estimation of the choice and study parameters.

First consider the parameters of the grade process and the course choices. With unobserved heterogeneity, we now need to make an assumption on the distribution of η_{ij}, the residual in the grade equation. We assume the error is distributed $N(0, \sigma_\eta)$. We then specify a flexible choice process over courses that depends on an parameter vector φ. The integrated log likelihood is:

$$\sum_i \ln\left(\sum_{s=1}^{S} \pi_s \mathcal{L}_{igs}\left(\theta, \gamma\right) \mathcal{L}_{ics}\left(\varphi\right)\right) \tag{3.27}$$

where $\mathcal{L}_{igs}\left(\theta, \gamma\right)$ and $\mathcal{L}_{ics}\left(\varphi\right)$ are the grade and choice (of courses) likelihoods respectively conditional on i being of type s.

We apply the EM algorithm to then estimate the grade parameters and course choice parameters in stages. We iterate on the following steps until convergence, where the mth step follows:

1. Given the parameters of the grade equation and choice process at step $m-1$, $\{\theta^{(m-1)}, \gamma^{(m-1)}\}$ and $\{\varphi\}$ and the estimate of $\pi^{(m-1)}$, calculate the conditional probability of i being of type s using Bayes rule:

$$q_{is}^{(m)} = \frac{\pi_s^{(m)} \mathcal{L}_{igs}\left(\theta^{(m-1)}, \gamma^{(m-1)}\right) \mathcal{L}_{ics}\left(\varphi^{(m-1)}\right)}{\sum_{s'} \pi_{s'}^{(m)} \mathcal{L}_{igs'}\left(\theta^{(m-1)}, \gamma^{(m-1)}\right) \mathcal{L}_{ics'}\left(\varphi^{(m-1)}\right)} \tag{3.28}$$

2. Update $\pi_s^{(m)}$ using $\left(\sum_{i=1}^N q_{is}^{(m)}\right)/N$.

3. Using the $q_{is}^{(m)}$'s as weights, obtain $\{\theta^{(m)}, \gamma^{(m)}\}$ by maximizing:

$$\sum_i \sum_s q_{is}^{(m)} \ln\left[\mathcal{L}_{igs}\left(\theta, \gamma\right)\right] \tag{3.29}$$

4. Using the $q_{is}^{(m)}$'s as weights, obtain $\varphi^{(m)}$ by maximizing:

$$\sum_i \sum_s q_{is}^{(m)} \ln\left[\mathcal{L}_{ics}\left(\varphi\right)\right] \tag{3.30}$$

Once the algorithm has converged, we have consistent estimates of $\{\theta, \gamma, \varphi\}$ as well as the conditional probabilities of being in each type. We can use the estimates of q_{is} as weights to form the average type probabilities of students of year in school l in class j to then estimate the parameters in (3.23). Finally, we use the estimates of q_{is} as weights in estimating the structural choice parameters using (3.26).

3.4.4 Implications from the Demand-Side Estimation

Even without estimating professor preferences, much can be learned from the demand-side estimates. First, we can explain some of the persistent gender gap in STEM majors. 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 (δ_{ij}),

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

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

The differences in preferences can also be linked to characteristics of the instructor. For example, we can link courses to gender of the instructor and see the extent to which female students prefer female professors by regressing δ_{1j} on indicators for whether the professor was female and departmental fixed effects. We can then use these estimates to forecast how the course choices would change if each department had a larger (or smaller) representation of female professors, holding fixed grading standards.

The estimates of the model can also be used to see how enrollment in STEM courses by both men and women would be affected by changes in grading practices. First, we can adjust the intercepts in the grading equation such that the average student's expected grade is the same across courses, isolating the role of the level of the grade from the differences in the slopes, and therefore return to effort. Second, we can forecast course choices if all professors were to have the same grading practices.

3.5 Equilibrium Grading Policies

Examining the effects of grading policies like those at Wellesley where the fraction of A's are capped are difficult to analyze because professors can respond to constraints on the number of A's given by changing the returns to effort as well as the intercept. For example, if certain departments are more generous in handing out A's than what the policy mandates, a way to keep their courses attractive is to require less work. This is equivalent to lowering γ_j. Similarly, if a policy were to mandate that some departments increase their grades, they can deter some of the increases in enrollment by requiring more work by increasing γ_j. In this section we specify the objective function of the professor and, supposing that the professor is restricted to linear grading policies, describe its solution. We then discuss the sorts of counterfactuals that can be conducted once the professor's preferences are recovered.

3.5.1 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: $c(\beta, \gamma)$.

Learning for student i is what is rewarded in the grade equation, the term hit by γ_j. Learning for individual i in course j is given by:

$$\begin{aligned} L_{ij}(\gamma_j) &= A_i + \ln(s^*_{ij}) \\ &= A_i + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \end{aligned} \tag{3.31}$$

Student study time is included as a way of capturing student complaints as workload increases.[12] Note that this negative component to the professor's utility is *separate and distinct* from the utility gained from students' learning. The professor's objective function is then assumed to follow:

$$\begin{aligned} V_j &= \lambda_{0j} a(\beta,\gamma) - \lambda_{1j} b(\beta,\gamma) - \lambda_{2j} c(\beta,\gamma) \\ &= \lambda_{0j} \left[\sum_i P_{ij}(\beta,\gamma) \left(A_i + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \right) \right] \\ &\quad -\lambda_{1j} \left[\sum_i P_{ij}(\beta,\gamma) \right]^2 - \lambda_{2j} \left[\sum_i P_{ij}(\beta,\gamma) \frac{\phi_i \gamma_j}{\psi_i} \right] \end{aligned} \tag{3.32}$$

The term P_{ij} is the probability of student i choosing course j, defined in the student's problem. Utility functions are only identified up to scale. Hence we normalize λ_{0j} to one.

Professors are assumed to know the preferences of the other professors and all professors simultaneously set their grading standards in a non-cooperative fashion. In a pure strategy equilibrium, the first order conditions of the professors problem must be satisfied at the grading standards observed in the data. The choice of β_j and γ_j satisfy the two first order conditions:

$$\frac{\partial V_j}{\partial \beta_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \beta_j} - \lambda_{1j} \frac{\partial b(\beta,\gamma)}{\partial \beta_j} - \lambda_{2j} \frac{\partial c(\beta,\gamma)}{\partial \beta_j} \tag{3.33}$$

$$\frac{\partial V_j}{\partial \gamma_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} - \lambda_{1j} \frac{\partial b(\beta,\gamma)}{\partial \gamma_j} - \lambda_{2j} \frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \tag{3.34}$$

[12] Another way to interpret this term is to regard it as a proxy for professor's own time lost due to increased need to help students learn via extra classes, more preparation, or additional office hours.

The solution to this system is then:

$$\begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \end{bmatrix} = \begin{bmatrix} \frac{\partial b(\beta,\gamma)}{\partial \beta_j} & \frac{\partial c(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial b(\beta,\gamma)}{\partial \gamma_j} & \frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial a(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix} \tag{3.35}$$

Note that, 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 derivatives on the right hand side (3.35) can be calculated, allowing us to recover the professor preferences.

3.5.2 Implications

Once professor preferences are recovered, we can see how characteristics of these vary by department as well as with other characteristics of the professor such as gender or professor rank.[13] Counterfactual policies, however, would require solving for the new equilibrium. Given the professor preferences, we could then see how the solution to the grading standards game would change given different environments. Note that this would entail solving a fixed point problem among all the professors.

Conditional on the feasibility of solving for new equilibria, a number of counterfactuals can be conducted. First, we can see how the equilibrium grading policies would change if all professors had the same preferences and correspondingly how enrollment in STEM courses by males and females would change as well. Second, we can examine what the equilibrium grading policies would look like if preferences for courses, the δ_j's, were the same. This would then give a sense for how much of the differences we see in grading practices are driven by demand versus professor

[13] The data also categorizes teachers into: student, staff, post-doc, faculty, and retired faculty. Rank of faculty is easily discernible from publicly available data.

preferences. Finally, we can investigate how placing caps on the mean grade given would affect the equilibrium.

3.6 Results

3.6.1 Preference estimates

Table 3.4 presents the preference parameters with the exception of the study costs, the class-specific intercepts, and the coefficients on year in school cross department cross level of the course. Recall that the parameter on expected grades is identified from variation in how abilities are rewarded in different classes. Both men and women value grades, with the estimates suggesting that women value grades a little over 20% more than men. The estimate of female preferences for female professors is positive, with the estimate suggesting that women be indifferent between a class that had a female professor and one that had a male professor who gave grades that were 0.3 points higher. This coefficient is likely biased upward due to the aggregation of departments. To the extent that female professors are more likely to be in departments that females have a preference for and there is variation within our aggregated groups, we may be picking up within-group preferences for departments.

The second set of rows of Table 3.4 shows female preferences (relative to male preferences) for different departments. The omitted category is Agriculture. The largest difference in preferences is between Engineering and English: 1.57 points, which translates into over 3 grade points. Engineering, however, is an outlier with all the other gaps smaller that one point (or 2 grade points).

Table 3.4: Estimates of Preference Parameters

Preference for:	Coeff.	Std. Error
Expected grades (ϕ)	0.4124	(0.0197)
Female × expected grades	0.0800	(0.0131)
Female × female professor	0.1460	(0.0186)
Female preferences for Departments		
Regional Studies	0.1981	(0.0695)
Communication	-0.1714	(0.0535)
Education & Health	0.3266	(0.0582)
Engineering	-1.1963	(0.0664)
Languages	-0.1241	(0.0580)
English	0.3663	(0.0760)
Biology	0.2277	(0.0631)
Mathematics	-0.1212	(0.0598)
Chemistry & Physics	-0.1558	(0.0603)
Psychology	0.3440	(0.0651)
Social Sciences	-0.3067	(0.0542)
Mgmt. & Mkting.	0.1220	(0.0662)
Econ., Fin., Acct.	-0.5278	(0.0589)

3.6.2 Study effort estimates

Estimates of the study effort parameters are presented in Table 3.5.[14] Lower study costs result in women studying a little over 7% more than men conditional on taking the same class. However, women also study more because they value grades more, with this effect at over 19.5%,[15] again conditional on taking the same class resulting in an overall effect of over 26.5%.

The results also suggest lower study costs for higher ACT scores but, surprisingly,

[14] Because of measurement error in the γ's that is compounded by it entering as a log in the study effort equation, we drop classes in the bottom 40% of the γ distribution. Parameters of the study effort equation stabilize after this point.

[15] This number comes from the difference in the log of the preferences for grades: $\ln(\phi_0 + \phi_1) - \ln(\phi_0) = \ln(.4963) - \ln(.4083)$.

higher study costs for higher high school grades. While blacks and first generation students have higher study costs, Hispanics and miscellaneous minorities have lower study costs. Those who are the second unobserved type have lower study costs and, as we will see in subsequent tables, also are more able.[16]

The second set of columns shows how the returns to study effort vary across classes, taking the median γ class for each course grouping. The heterogeneity is quite large. A ten percent increase in study effort would translate into almost a quarter of a grade increase in mathematics but would translate into less than a tenth of a grade point in agriculture, management and marketing, and education.

Table 3.5: Estimates of Study Effort and Departmental Returns to Studying

	Study Effort Coeff. ($-\psi$)	Std. Error	Department	Median γ Coeff.
Female	-0.0228	(0.0651)	Agriculture	0.9618
ACT read	-0.0071	(0.0088)	Regional Studies	1.1799
ACT math	0.0297	(0.0095)	Communication	1.1333
HS GPA	-0.0777	(0.0840)	Education & Health	0.8072
Black	0.0152	(0.1437)	Engineering	1.3908
Hispanic	0.2186	(0.2432)	Languages	1.0656
Other Min.	0.3634	(0.3152)	English	1.1132
First Generation	0.0296	(0.1016)	Biology	1.6240
Unobs. Type	0.1091	(0.0554)	Mathematics	2.0694
			Chemistry & Phsyics	1.6304
			Psychology	1.2581
			Social Sciences	0.9741
			Mgmt. & Mkting.	0.9056
			Econ., Fin., Acct.	1.2465

[16] The population probability of being the second unobserved type is 0.213.

3.6.3 Grade estimates

With the estimates of the grading equation, we can reported expected grades for an average student. We do this for freshmen, separately by gender, both unconditionally and conditional on taking courses in that department in the semester we study. Results are presented in Table 3.7. Three patterns stand out. First, there is positive selection into STEM courses: generally those who take STEM classes are expected to perform better than the average student. This is the not the case for many departments. Indeed, the second pattern is that negative selection is more likely to occur in departments with higher grades. Finally, women are disproportionately represented in departments that give higher grades for the average student. Of the seven departments that give the highest grades for the average student, only one has a smaller fraction female than the overall population. In contrast, of the seven departments that give the lowest grades, only two have a fraction female that is larger than the overall population.

3.6.4 Professor preference estimates

We now turn to the estimates of the professor preference parameters. Recall that professor utility was a function of (i) total learning (the sum of student ability in the class plus the amount of studying), (ii) squared enrollment (so professors may value the first enrolled student more than the second), (iii) a distaste for assigning more work. With the payoff to learning to normalized to one, Table 3.8 shows the mean and standard deviation of the disutility of squared enrollment and assigning more work by upper and lower classes and STEM/non-STEM.

For lower-level classes, the marginal student is more costly in non-STEM classes

Table 3.6: Expected Freshmen GPA for Median Classes By Department, Unconditional and Conditional on Taking Courses in that Department

	Female Uncond.	Female Cond.	Male Uncond.	Male Cond.
Ed & Health	3.52	3.39	3.37	3.10
Communication	3.35	3.35	3.10	3.09
Agriculture	3.31	3.22	3.21	3.06
Languages	3.25	3.22	3.11	3.08
Mgmt & Mkting	3.10	2.78	2.97	2.72
Regional Studies	2.99	3.09	2.78	2.94
Biology	2.97	3.01	2.88	2.85
English	2.95	2.94	2.83	2.82
Social Sciences	2.94	2.93	2.79	2.69
Engineering	2.89	3.04	2.77	2.88
Mathematics	2.77	2.83	2.66	2.74
Econ., Fin., Acct.	2.77	2.66	2.77	2.68
Psychology	2.75	2.69	2.53	2.44
Chem & Physics	1.99	2.18	2.05	2.28

than in STEM classes. This actually translates into higher grades in STEM classes as higher grade intercepts increases enrollment with no effect on study time. At the same time, assigning work is less costly in STEM classes (though not significantly so) implying STEM courses will set higher returns to studying, which in turn may result in lower grading intercepts. Large classes are more costly for upper-level STEM courses than lower-level STEM courses, though the pattern is surprisingly reversed for non-STEM courses.

3.7 Counterfactuals

Given the estimates of the student's choices over classes and effort and given the estimates of the grading process, we now turn to examining the sources of the male-

Table 3.7: Professor Preferences Over Enrollment and Study Time by Course Level and Type

		Disutility of:			
Level	Category	Enrollment[2] (λ_{1j})		Study Time (λ_{1j})	
		Mean	Std. Dev.	Mean	Std. Dev.
Lower	non-STEM	0.060	0.036	1.832	0.784
Level	STEM	0.038*	0.041	1.752	0.489
Upper	non-STEM	0.049†	0.080	1.942	1.016
Level	STEM	0.069*,†	0.036	1.922	1.062

Note: * indicates STEM mean statistically different at 95% level from non-STEM mean at the same course level. † indicates mean upper-level course statistically different at the 95% level from mean lower-level course of the same category (STEM/non-STEM).

female gap in choice of STEM classes. Table 3.9 shows share of STEM classes taken for males and females as well as how that share changes for women as we change different characteristics. The baseline share of STEM classes for men and women is 0.400 and 0.284, respectively. The first counterfactual changes female preferences for grades to be the same as male preferences for grades. This increases the share of STEM course for women to 0.293, closing the gender gap by eight percent. Turning off observed ability differences such as differences in ACT scores and high school grades (row 3) and study costs (row 4) have smaller effects on the gap, though larger effects are found for unexplained gender differences in ability (row 2). Note that these effects are not driven by women being weaker academically per se, but in part due to women being relatively stronger in non-STEM courses. The next three counterfactuals, which equalize ability differences and costs of studying, all have smaller effects.

Counterfactuals (5) through (7) look at differences in tastes. Counterfactual (5) turns off taste differences for departments, which increases the share of women to 0.31, closing the STEM gap by 22 percent. These taste differences may be a mixture of pre-college experiences and the culture of different departments. Hence anything the university can do to close the STEM gap on this end is likely bounded above by this number. Counterfactual (6) turns off female preferences for female professors. One way of closing the gender gap in STEM would be to hire more female professors. However, even representation across fields would only close the gap by a little over three percent.

Table 3.8: Decomposing the Gender STEM Gap

	Share	Pct of Gap	Gender Ratio
Baseline Male	0.400		
Baseline Female	0.284		0.710
(1) No diff in grade prefs ($\phi_1 = 0$)	0.293	0.080	0.733
(2) No gender ability diff ($\alpha_1 = 0$)	0.301	0.142	0.751
(3) No observed ability diff ($\overline{X}_f = \overline{X}_m$)	0.289	0.043	0.722
(4) No study effort diff ($\psi_1 = 0$)	0.286	0.013	0.713
(5) No taste diff	0.310	0.220	0.773
(6) No female professor pref	0.288	0.034	0.720
(7) Both (5) and (6)	0.314	0.324	0.784

Our next set of counterfactuals focus on grading policies. Results are presented in Table 3.10. We consider two counterfactuals: (i) adjusting grading intercepts for each course such that the expected grade for the average student is the same across courses and (ii) changing preferences of STEM professors to on average match those of humanities professors.

The first grading policy counterfactual equalizes expected grades across courses for the average student by increasing (or decreasing) the course-specific intercepts until expected grades are the same for the average student. However, there is still heterogeneity in grades due to the relative difference in γ's and α's, the former being especially important as it dictates the returns to studying. This counterfactual raises the share of STEM courses taken by females to .334, higher than any of the counterfactuals in Table 3.9. The gender ratio also tilts significantly towards females but not as much as in counterfactuals (5) and (7) because men too see their probabilities of taking STEM courses increase. The reason the effects are larger here on the gender ratio than in the first counterfactual is that the returns to studying are much higher in STEM courses and women are willing to study more than men, due both to valuing grades more.

While the patterns here suggest a potentially cheap way of closing the gender gap is to equalize average grades across fields, professors are likely to respond to restrictions on grading policies. However, the response may further reduce the gender gap. The reason is that, if STEM courses are forced to give higher grades, they are likely to assign more work to deter entry. More work translates into higher γ's which make STEM courses relatively more attractive to women. The reverse holds for departments that are forced to lower their grades: in order to attract more students, they must lower workloads, implying lower values of γ which makes these courses relatively less attractive to women.

The final counterfactual changes STEM professor preferences by the average difference between STEM and non-STEM professor preferences so that the means are the same. Because professors in lower-level STEM courses have lower disutility from increased class size than their non-STEM counterparts, equalizing preferences exacerbates the gender gap. Lower shares for STEM classes are seen for both men and women but the effects are larger for women.

Table 3.9: Supply-Side Counterfactuals

	Male Share	Female Share	Gender Ratio
Baseline	0.400	0.284	0.710
(8) Equalize exp grades for avg student	0.440	0.334	0.760
(9) Change STEM prof prefs to non-STEM	0.398	0.278	0.700

3.8 Conclusion

The lack of graduates in STEM majors–particularly among under-represented groups–has been of some policy concern. We show that there is a potentially cheap way to change the number and composition of STEM majors. Namely, grading policies have a substantial effect on sorting into STEM classes. We show that a substantial portion of the gender STEM gap can be removed by having STEM classes give grades that are on average similar to those in non-STEM classes.

These grading policies, however, are in part choices by professors. Hence administrative policies designed to change how professors grade will elicit responses by professors on other dimensions such as workload. These responses by professors may result in an even further closing of the STEM gender gap. Namely, if classes across departments are forced to give similar grades on average, then STEM (non-STEM) classes will employ alternative means to deter (encourage) enrollment in their courses by changing workloads. This will result in STEM classes assigning even more work and non-STEM classes assigning less work. Since women are willing to study more, the increased STEM workload works as less of a deterrent to women taking STEM courses.

Appendix A

EM Algorithm Estimation Details

This appendix contains explicit solutions to $\text{argmax}_{\Theta} Q_{het}(\Theta \mid \Theta^m)$ and expressions for the distribution variables $q_{i\tau}^m$, $\xi_{i\tau}^m$, and $\Upsilon_{i\tau}^m$ used to form the expected likelihood function in the iteration m E step.[1] Formulas are presented in the general case in which there is type specific unobserved heterogeneity in the distribution parameters γ_τ and Δ_τ.

Distribution variables used to form expected likelihood

Unobserved Abilities $\xi_{i\tau}^{m+1}$ and $\Upsilon_{i\tau}^{m+1}$ characterize the distribution of each individual's ability vector α_i conditional on student type, iteration m parameter estimates, and all observed grades. These are derived by updating type specific initial prior distributions with all grade signals received by student i. Degroot (1970) shows the

[1] For clarity, I present formulas which assume the observed sample is randomly drawn from the population. The formulas actually used in my estimation routine include sampling weights to correct for the stratified sampling of the CLL and to partially eliminate non-response bias.

resulting posterior distribution is $N\left(\xi_{i\tau}^{m+1}, \Upsilon_{i\tau}^{m+1}\right)$ where:[2]

$$\xi_{i\tau}^{m+1} = \left(\left(\Delta_\tau^m\right)^{-1} + \sum_{l=1}^{L} D_{il}\left(\Phi_l^m\right)^{-1}\right)^{-1}\left(\left(\Delta_\tau^m\right)^{-1}\gamma_\tau^m + \sum_{l=1}^{L} D_{il}\left(\Phi_l^m\right)^{-1}\bar{z}_i\right) \quad \text{(A.1)}$$

$$\Upsilon_{i\tau}^{m+1} = \left(\left(\Delta_\tau^m\right)^{-1} + \sum_{l=1}^{L} D_{il}\left(\Phi_l^m\right)^{-1}\right)^{-1} \quad \text{(A.2)}$$

Unobserved Types $q_{i\tau}^{m+1}$ is the probability student i is type τ conditional on iteration m parameter estimates and observed grades and choices. These conditional probabilities can be expressed as the joint likelihood that individual i is type τ and of observing individual i's grades and choices divided by the marginal likelihood of observing individual i's grades and choices.

$$q_{i\tau}^{m+1} = \frac{\pi_\tau^m \mathscr{L}_i^c\left(\mathbf{d_i} \mid \tau, \gamma_\tau^m, \theta_k^m, \sigma_{kl}^m, \Delta_\tau^m, \omega^m\right)\left[\int_\alpha \mathscr{L}_i^g\left(\mathbf{g_i} \mid \alpha_i, \theta_k^m, \sigma_{kl}^m\right) f\left(\alpha_i \mid \gamma_\tau^m, \Delta_\tau^m\right)\right]}{\sum_{\tau=1}^{\mathscr{T}} \pi_\tau^m \mathscr{L}_i^c\left(\mathbf{d_i} \mid \tau, \gamma_\tau^m, \theta_k^m, \sigma_{kl}^m, \Delta_\tau^m, \omega^m\right)\left[\int_\alpha \mathscr{L}_i^g\left(\mathbf{g_i} \mid \alpha_i, \theta_k^m, \sigma_{kl}^m\right) f\left(\alpha_i \mid \gamma_\tau^m, \Delta_\tau^m\right)\right]} \quad \text{(A.3)}$$

[2] $\bar{z}_i$ is a vector of average grade signals received by individual i given by:

$$\bar{z}_i(k) = \frac{\sum_{t=1}^{T}\sum_{c=1}^{C_{it}}\left(d_{ict} = k\right)\left(g_{itklc} - X_{itk}\theta_k^m\right)}{\sum_{t=1}^{T}\sum_{c=1}^{C_{it}}\left(d_{ict} = k\right)}$$

Φ_l is a $K \times K$ matrix defined as,

$$\Phi_l\left(k, k'\right) = \begin{cases} \sigma_{kl}^2 & k = k' \\ 0 & k \neq k' \end{cases}$$

and D_{il} is a $K \times K$ matrix defined as,

$$D_{il}\left(k, k'\right) = \begin{cases} n_{ikl} & k = k' \\ 0 & k \neq k' \end{cases}$$

where n_{ikl} is the total number of field k level l courses taken by individual i.

Unconditional Type Probabilities Unconditional type probabilities π_τ^m are updated by averaging conditional type probabilities over the population of students:

$$\pi_\tau^{m+1} = \frac{1}{N}\sum_{i=1}^{N} q_{i\tau}^{m+1} \tag{A.4}$$

Solutions to $\text{argmax}_\Theta Q_{het}\left(\Theta \mid \Theta^m\right)$

Type Specific Ability Distributions Ability vectors are drawn from the type specific distributions $N\left(\gamma_\tau, \Delta_\tau\right)$. These parameters can be directly recovered from individual-type specific distribution parameters $\xi_{i\tau}^m$ and $\Delta_{i\tau}^m$ and individual specific type probabilities $q_{i\tau}^m$. The closed form expressions are:

$$\gamma_\tau^m = \frac{\sum_{i=1}^N q_{i\tau}^m \xi_{i\tau}^m}{\sum_{i=1}^N q_{i\tau}^m} \tag{A.5}$$

$$\Delta_\tau^m = \frac{\sum_{i=1}^N q_{i\tau}^m \left(\Upsilon_{i\tau}^m + \left(\xi_{i\tau}^m\right)\left(\xi_{i\tau}^m\right)'\right)}{\sum_{i=1}^N q_{i\tau}^m} - \frac{\sum_{i=1}^N q_{i\tau}^m \left(\gamma_\tau^m\right)\left(\gamma_\tau^m\right)'}{\sum_{i=1}^N q_{i\tau}^m} \tag{A.6}$$

Grade Production Parameters If α_i were observed, grade production parameters θ_k and σ_{kl} could be estimated by regressing $g_{itklc} - \alpha_{ik}$ on covariates X_{itkl}. Although α_i is not observed, individual-type specific ability means $\xi_{i\tau}^m$ and individual specific type probabilities $q_{i\tau}^m$ can be used to form best guesses of this variable given by $\tilde{\alpha}_i^m = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \xi_{i\tau}^m$. Denote the errors in these best guesses by $\upsilon_i^m = \alpha_i - \tilde{\alpha}_i^m$. By construction, the information set used to construct α_i^m contains student i's information set at the beginning of term t. As such, the residual υ_i^m is orthogonal to student information sets at the beginning of term t. This suggests the following regression can be used to estimate θ_k^m:

$$g_{itklc} - \tilde{\alpha}_{ik}^m = X_{itkl}\theta_k + \tilde{\eta}_{itklc} \tag{A.7}$$

where the composite error term is given by $\tilde{\eta}_{itklc} = \upsilon_{ik}^m + \eta_{itklc}$.

Standard deviation in grading noise can be estimated using residuals from this regression with an adjustment for the extra noise introduced by υ_{ik}^m.

Denote residuals from iteration m estimates of grading parameters as:

$$z_{itklc}^m = g_{itklc} - X_{itkl}\theta_k \tag{A.8}$$

The closed form expression is then: [3]

$$\left(\sigma_{kl}^{m+1}\right)^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T \sum_{c=1}^{C_{it}} \left(d_{ict} = \{k,l\}\right) \left\{ \left(z_{itklc}^{m+1}\right)^2 - 2 z_{itklc} \mathbb{E}_m\left[\alpha_{ik}\right] + \mathbb{E}_m\left[\alpha_{ik}^2\right] \right\}}{\sum_{i=1}^N \sum_{t=1}^T \sum_{c=1}^{C_{it}} \left(d_{ict} = \{k,l\}\right)} \tag{A.9}$$

Choice Parameters Choice parameters are estimated by numerically maximizing the choice contribution to the expected log likelihood taking iteration m estimates of γ_τ, Δ_τ, θ_k, σ_{kl} as given. Formally,

[3] Where iteration m expectations are given by:

$$\mathbb{E}_m\left[\alpha_i\right] = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \xi_{i\tau}^m$$

$$\mathbb{E}_m\left[\alpha_i\right]^2 = \sum_{\tau=1}^{\mathscr{T}} q_{i\tau}^m \left(\Upsilon_{i\tau}^m\left(k,\,k\right) + \left(\xi_{i\tau}^m\left(k\right)\right)^2\right)$$

$$\omega^{m+1} = \text{argmax}_{\omega} \left\{ \sum_{i=1}^{N} \sum_{\tau=1}^{\mathscr{T}} \sum_{t=1}^{T} \sum_{c=1}^{C_{it}} q_{i\tau}^{m} \ln \left[\frac{\exp\left(u_{itkl}\right)}{\sum_{k'=1}^{K} \sum_{l'=1}^{L} \exp\left(u_{itk'l'}\right)} \right]^{(d_{itc}=\{k,l\})} \right\} \quad \text{(A.10)}$$

Appendix B

Uncertain Admissions and Class Composition

This appendix presents a simple theoretical model demonstrating how uncertainty in admissions can generate a negative relationship between the means and variances of ability distributions. Suppose all students are endowed with a unidimensional ability vector α_i which is drawn from an individual specific distribution as: $\alpha_i \sim N(\gamma_i, \Delta_i)$. When universities evaluate applications for admissions, they do not observe α_i but instead observe an independent application draw a_i which is also drawn as $a_i \sim N(\gamma_i, \Delta_i)$. This is equivalent to assuming universities observe noisy ability $\alpha_i + \eta_i$ where $\eta_i \sim N(0, \Delta_i)$.

Given the set of application signals $\{a_i\}$, universities choose an admission threshold $\underline{a}$ such that students with $a_i \geqslant \underline{a}$ are admitted and students with $a_i < \underline{a}$ are rejected. The probability of admission given γ_i and Δ_i is then:

$$\Pr(a_i \geqslant \underline{a} \,|\, \gamma_i, \Delta_i) = 1 - \Phi\left(\frac{\underline{a} - \gamma_i}{\sqrt{\Delta_i}}\right) \tag{B.1}$$

where $\Phi(\cdot)$ is the standard normal CDF.

This probability is increasing in γ_i; additionally, the probability is increasing in Δ_i for students with $\gamma_i < \underline{a}$ and decreasing in Δ_i for students with $\gamma_i > \underline{a}$. Intuitively, this second property occurs because lower ability students only gain admission if they produce an unusually impressive application which occurs more often when there is greater uncertainty. Conversely, higher ability students will only fail to gain admission if they produce an unusually poor application which occurs less often when there is less uncertainty.

As a result, admitted students with $\gamma_i < \underline{a}$ have selectively high Δ_i (and this selection strengthens as γ_i gets further below $\underline{a}$). Additionally, admitted students with $\gamma_i > \underline{a}$ will have selectively low Δ_i (and this selection strengthens as γ_i gets further above $\underline{a}$). This implies that if γ_i and Δ_i are independent or inversely correlated in the general population then they will be inversely correlated in the selected population which gains admission. This is not necessarily true if γ_i and Δ_i are positively correlated in the general population but it seems unlikely that more confident students are less certain.

If enrollment decisions conditional on admission are independent of γ_i and Δ_i then the above statement applies to the selected population of enrolled students as well. This is consistent with my empirical finding: type 1 students have low expectations about their unobserved component of grade production and high uncertainty in these expectations while type 2 students have high expectations about their unobserved component of grade production and low uncertainty in these expectations.

Appendix C

Expressions for effects of spending on choice probabilities and class sizes in general equilibrium

In this appendix, I derive expressions for the effects of spending on instruction in one course on enrollments in all courses in a general equilibrium setting in which class size affects choice utility. These effects are complicated because spending on instruction has direct effects on own course enrollment through changes in instructor quality but also indirect effects on all course enrollments through changes in class sizes. In this appendix only, I assume idiosyncratic preferences ϵ_{ij} are drawn independently from a type 1 extreme value distribution for tractability.

The effects of excess spending on instruction in course j on expected enrollment in course j' can be written in terms of effects on individual choice probabilities:

$$\frac{d\tilde{n}_{j'}}{de_j} = \sum_{i=1}^{N} \frac{dP_{ij'}}{de_j} \tag{C.1}$$

The effects of excess spending in course j on probabilities of choosing course j' can be written as the effects of excess spending in course j on choice utility for all courses multiplied by the effects of these choice utilities on the probabilities of choosing course j':

$$\frac{d\tilde{n}_{j'}}{de_j} = \sum_{i=1}^{N} \sum_{k \in \mathbf{d}} \frac{\partial P_{ij'}}{\partial u_{ik}} \frac{du_{ik}}{de_j} \tag{C.2}$$

where

$$\frac{du_{ik}}{de_j} = \begin{cases} \frac{\partial u_{ij}}{\partial I_j} \frac{\partial \phi_j}{\partial e_j} + \frac{\partial u_{ij}}{\partial \tilde{n}_j} \frac{d\tilde{n}_j}{de_j} & j = k \\ \frac{\partial u_{ik}}{\partial \tilde{n}_k} \frac{d\tilde{n}_k}{de_j} & j \neq k \end{cases} \tag{C.3}$$

and the type 1 extreme value assumption imples that:

$$\frac{\partial P_{ij'}}{\partial u_{ik}} = \begin{cases} P_{ij'} \left(1 - P_{ij'}\right) & j' = k \\ -P_{ij'} P_{ik} & j' \neq k \end{cases} \tag{C.4}$$

Equation (C.2) holds for any multinomial choice model; however, without class size effects the terms $\frac{du_{ik}}{de_j} = 0$ for $j \neq k$. In this setting, these cross course effects are non-zero because spending in course j affects the desirability of course j which affects expected class sizes in all courses.

Combining (C.2), (C.3), and (C.4) yields own spending effects of:

$$\begin{aligned} \frac{d\tilde{n}_j}{de_j} \quad = \quad & \sum_{i=1}^{N} \left[P_{ij} \left(1 - P_{ij}\right) \left[\frac{\partial u_{ij}}{\partial I_j} \frac{\partial \phi_j}{\partial e_j} + \frac{\partial u_{ij}}{\partial \tilde{n}_j} \frac{d\tilde{n}_j}{de_j} \right] \right. \\ & \left. - \sum_{k \in \mathbf{d} \setminus j} P_{ij} P_{ik} \left[\frac{\partial u_{ik}}{\partial \tilde{n}_k} \frac{d\tilde{n}_k}{de_j} \right] \right] \end{aligned} \tag{C.5}$$

and cross course effects of:

$$\frac{d\tilde{n}_{j'}}{de_j} = \sum_{i=1}^{N}\Bigg[P_{ij'}\left(1-P_{ij'}\right)\frac{\partial u_{ij'}}{\partial \tilde{n}_{j'}}\frac{d\tilde{n}_{j'}}{de_j} - P_{ij'}P_{ij}\left[\frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j} + \frac{\partial u_{ij}}{\partial \tilde{n}_j}\frac{d\tilde{n}_j}{de_j}\right] - \sum_{k\in\mathbf{d}\setminus\{j,j'\}} P_{ij'}P_{ik}\left[\frac{\partial u_{ik}}{\partial \tilde{n}_k}\frac{d\tilde{n}_k}{de_j}\right]\Bigg] \tag{C.6}$$

where $j \neq j$. In both cases, effects of spending on enrollment depend on the effects of spending on enrollment in all courses. To convert these implicit definitions into explicit solutions for $\frac{d\tilde{n}_{j'}}{de_j}$, rearrange Equation (C.5) to obtain:

$$0 = \left[\sum_{i=1}^{N} P_{ij}\left(1-P_{ij}\right)\frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j}\right] + \frac{d\tilde{n}_j}{de_j}\left[\left(\sum_{i=1}^{N} P_{ij}\left(1-P_{ij}\right)\frac{\partial u_{ij}}{\partial \tilde{n}_j}\right) - 1\right] - \sum_{k\in\mathbf{d}\setminus j}\frac{d\tilde{n}_k}{de_j}\left[\sum_{i=1}^{N} P_{ij}P_{ik}\frac{\partial u_{ik}}{\partial \tilde{n}_k}\right] \tag{C.7}$$

and rearrange Equation (C.6) to obtain:

$$0 = \frac{d\tilde{n}_{j'}}{de_j}\left[\left(\sum_{i=1}^{N} P_{ij'}\left(1-P_{ij'}\right)\frac{\partial u_{ij'}}{\partial \tilde{n}_{j'}}\right) - 1\right] - \left[\sum_{i=1}^{N} P_{ij'}P_{ij}\frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j}\right] - \frac{d\tilde{n}_j}{de_j}\left[\sum_{i=1}^{N} P_{ij'}P_{ij}\frac{\partial u_{ij}}{\partial \tilde{n}_j}\right] - \sum_{k\in\mathbf{d}\setminus\{j,j'\}}\frac{d\tilde{n}_k}{de_j}\left[\sum_{i=1}^{N} P_{ij'}P_{ik}\frac{\partial u_{ik}}{\partial \tilde{n}_k}\right] \tag{C.8}$$

where $j \neq j'$.

For ease of notation, define:

$$
\begin{aligned}
A_j &= -\sum_{i=1}^{N} P_{ij}\left(1 - P_{ij}\right) \frac{\partial u_{ij}}{\partial I_j} \frac{\partial \phi_j}{\partial e_j} \\
B_j &= \left(\sum_{i=1}^{N} P_{ij}\left(1 - P_{ij}\right) \frac{\partial u_{ij}}{\partial \tilde{n}_j}\right) - 1 \\
C_{kj} &= -\sum_{i=1}^{N} P_{ij} P_{ik} \frac{\partial u_{ik}}{\partial \tilde{n}_k} \\
E_{jj'} &= \sum_{i=1}^{N} P_{ij'} P_{ij} \frac{\partial u_{ij}}{\partial I_j} \frac{\partial \phi_j}{\partial e_j} \\
F_{jj'} &= -\sum_{i=1}^{N} P_{ij'} P_{ij} \frac{\partial u_{ij}}{\partial \tilde{n}_j}
\end{aligned}
$$

Index courses in **d** from 1 to J. For any $j \in \mathbf{d}$, the system can then be written as:

$$
\begin{bmatrix}
B_1 & \cdots & F_{j1} & \cdots & C_{J1} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
C_{1j} & \cdots & B_j & \cdots & C_{Jj} \\
\vdots & \ddots & & \ddots & \vdots \\
C_{1J} & \cdots & F_{jJ} & \cdots & B_J
\end{bmatrix}
\begin{bmatrix}
\frac{d\tilde{n}_1}{de_j} \\ \vdots \\ \frac{d\tilde{n}_j}{de_j} \\ \vdots \\ \frac{d\tilde{n}_J}{de_j}
\end{bmatrix}
=
\begin{bmatrix}
E_{j1} \\ \vdots \\ A_j \\ \vdots \\ E_{jJ}
\end{bmatrix}
$$

The first row of the matrix generates Equation (C.8) where $j' = 1$; the middle row of the matrix generates Equation (C.7) where $j' = j$; and the bottom row of the matrix generates Equation (C.8) where $j' = J$. The explicit solution is then given by:

$$\begin{bmatrix} \frac{d\tilde{n}_1}{de_j} \\ \vdots \\ \frac{d\tilde{n}_j}{de_j} \\ \vdots \\ \frac{d\tilde{n}_J}{de_j} \end{bmatrix} = \begin{bmatrix} B_1 & \cdots & F_{j1} & \cdots & C_{J1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{1j} & \cdots & B_j & \cdots & C_{Jj} \\ \vdots & \ddots & & \ddots & \vdots \\ C_{1J} & \cdots & F_{jJ} & \cdots & B_J \end{bmatrix}^{-1} \begin{bmatrix} E_{j1} \\ \vdots \\ A_j \\ \vdots \\ E_{jJ} \end{bmatrix} \tag{C.9}$$

With functional form assumptions on the structure of u_{ij}, researchers can estimate a general equilibirum choice model to obtain choice probabilities P_{ij} and marginal effects $\frac{\partial u_{ij}}{\partial \tilde{n}_j}$ and $\frac{\partial u_{ij}}{\partial I_j}\frac{\partial \phi_j}{\partial e_j}$ for all offered courses $j \in \mathbf{d}$. These can be used to construct A_j, B_j, C_{kj}, $E_{jj'}$, and $F_{jj'}$ which in turn can be used to construct $\frac{d\tilde{n}_{j'}}{de_j}$ using Equation (C.9).

Appendix D

Identification of γ_f from tangency conditions

In this appendix, I show how the system of tangency conditions given by Equation (2.23) can be used to identify university preferences for students choosing courses in field f .[1]

Because the total number of enrolled students is fixed, only relative parameters $\{\gamma_f - \gamma_F\}_{f=1}^{F-1}$ are identified. To see this more concretely, combine university payoffs given by Equation (2.19) with the constraint $\sum_{j=1}^{J} \tilde{n}_j(\mathbf{e}, \mathbf{d}) = N$ to obtain:

$$\mathbb{E}[\Pi \,|\, \mathbf{e}, \mathbf{d}] = \sum_{i=1}^{N} \mathbb{E}[V_i \,|\, \mathbf{e}, \mathbf{d}] + \sum_{f=1}^{F-1} \gamma_f \left[\sum_{j \in f} \tilde{n}_j(\mathbf{e}, \mathbf{d}) \right] + \gamma_F \left(N - \sum_{f=1}^{F-1} \sum_{j \in f} \tilde{n}_j(\mathbf{e}, \mathbf{d}) \right)$$

[1] I present an identification argument for the general equilibrium setting in which class size affects choice utility. The baseline setting with no class size effects is a specific case of the general equilibrium setting. As such, identification in the general equilibrium setting implies identification in the baseline setting.

Rearranging yields:

$$\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right] = \sum_{i=1}^{N} \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] + \sum_{f=1}^{F-1} \left(\gamma_f - \gamma_F\right) \left[\sum_{j \in f} \tilde{n}_j \left(\mathbf{e}, \mathbf{d}\right)\right] + \gamma_F N$$

The last term $\gamma_F N$ is uninformative because N is fixed across course offerings $\mathbf{d}$ and excess spending vectors $\mathbf{e}$. This demonstrates that relative parameters $\{\gamma_f - \gamma_F\}_{f=1}^{F-1}$ fully characterize university preferences for the courses students choose. As such, without loss of generality I normalize $\gamma_F = 0$.

To identify relative γ_f using the university's tangency conditions, I begin by selecting one non-contract offered course with positive excess spending for each academic field. The tangency conditions imply the marginal effects of excess spending on university payoffs must be the same across these selected courses. I show how this system of tangency conditions can be inverted to solve for the unique values of γ_f which make this true.

For ease of notation, define:

$$\begin{aligned} dW_f &= \sum_{i=1}^{N} \sum_{k \in \mathbf{d}} \left(\frac{du_{ik}}{de_j}\right) P_{ik} \\ dn_{fk} &= \frac{d\tilde{n}_k}{de_j} \end{aligned}$$

where course j is the selected course in academic field f. dW_f represents the marginal effect of excess spending on total student welfare for the selected course in field f while dn_{fk} represents the marginal effect of excess spending on the number of students choosing course k for the selected course in field f .

The general equilibrium tangency conditions given by Equations (2.21) and (2.23) relating these selected courses can then be represented with the following system of $(F-1)$ equations and $(F-1)$ unknown parameters:

$$\begin{aligned}
dW_1 + \sum_{f=1}^{F-1} \gamma_f \left[\sum_{k \in f} dn_{1k} \right] &= dW_F + \sum_{f=1}^{F-1} \gamma_f \left[\sum_{k \in f} dn_{Fk} \right] \\
&\vdots \\
dW_{F-1} + \sum_{f=1}^{F-1} \gamma_f \left[\sum_{k \in f} dn_{F-1k} \right] &= dW_F + \sum_{f=1}^{F-1} \gamma_f \left[\sum_{k \in f} dn_{Fk} \right]
\end{aligned}$$

To solve this system, define the matrices and vectors:

$$dn = \begin{bmatrix} \sum_{k \in 1} (dn_{1k} - d_{Fk}) & \cdots & \sum_{k \in F-1} (dn_{1k} - d_{Fk}) \\ \vdots & \ddots & \vdots \\ \sum_{k \in 1} (dn_{F-1k} - d_{Fk}) & \cdots & \sum_{k \in F-1} (dn_{F-1k} - d_{Fk}) \end{bmatrix}$$

$$dW = \begin{bmatrix} dW_F - dW_1 \\ \vdots \\ dW_F - dW_{F-1} \end{bmatrix}$$

$$\gamma = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{F-1} \end{bmatrix}$$

The system of tangency conditions can then be written as:

$$(dn)\,\gamma = (dW)$$

Assuming there is enough variation to invert dn, γ is identified using:

$$\gamma = (dn)^{-1}\,(dW)$$

This argument demonstrates γ is identified as long as there is at least one non-contract offered course with positive excess spending in each academic field. In all practical cases, there are multiple non-contract offered course with positive excess spending in each academic field. In these cases,γ is over-identified allowing for the over-identification test described in Subsection 2.4.2.

Appendix E

Optimal Excess Spending Decisions

In this appendix, I present methods for computing $\mathbf{e}(\mathbf{d})^{\star}$ for several alternative settings and utility structures. In most cases, it is infeasible to solve for $\mathbf{e}(\mathbf{d})^{\star}$ explicitly; however, it is often possible to define $\mathbf{e}(\mathbf{d})^{\star}$ implicitly and solve for a fixed point of these implicit definitions using a iterative algorithm.

Example 1: Welfare Maximizing University - no effects of class size

This example solves for $\mathbf{e}(\mathbf{d})^{\star}$ for a welfare maximizing university in the baseline setting where class size does not effect course utility. Suppose choice utility is given by:

$$U_{ij} = \theta_j \ln(e_j + 1) + \psi_{ij}(Z_j, X_i) + \epsilon_{ij} \tag{E.1}$$

With this structure, a welfare maximizing university's tangency conditions are given by:

$$\frac{\theta_j \tilde{\mathbf{n}}_j(\mathbf{e})}{e_j + 1} = \frac{\theta_{j'} \tilde{\mathbf{n}}_{j'}(\mathbf{e})}{e_{j'} + 1} \quad \forall j, j' \in \mathbf{d} \, s.t. \, c_j > m_j, \, c_{j'} > m_{j'} \tag{E.2}$$

and the binding budget constraint is given by:

$$\sum_{j \in \mathbf{d}} (m_j + e_j) = E \tag{E.3}$$

$\mathbf{e}(\mathbf{d})^\star$ can then be implicitly defined as:

$$e_j(\mathbf{d})^\star = \frac{\left[E + J - \sum_{j=1}^{J} m_j\right] \theta_j \tilde{\mathbf{n}}_j(\mathbf{e})}{\sum_{j'=1}^{J} \theta_{j'} \tilde{\mathbf{n}}_{j'}(\mathbf{e})} - 1 \tag{E.4}$$

The following iterative algorithm can then be used to solve for a fixed point of this implicit definition:

1. Set initial excess spending values to be uniform across offered courses: $e_j^1 = \frac{E - \sum_{j=1}^{J} m_j}{J}$
2. Compute expected class sizes given initial excess spending values: $\tilde{\mathbf{n}}_j^1 = \tilde{\mathbf{n}}_j(\mathbf{e}^1)$
3. Use Equation (E.4) to compute new excess spending values: e_j^2
4. Repeat until sequential values of $\mathbf{e}$ become arbitrarily close.

Example 2: General Objective Structure - no effects of class size

This example solves for $\mathbf{e}(\mathbf{d})^\star$ for the general university objective structure in the baseline setting where class size does not effect course utility. As before, suppose choice utility is given by:

$$U_{ij} = \theta_j \ln\left(e_j + 1\right) + \psi_{ij}\left(Z_j,\, X_i\right) + \epsilon_{ij} \tag{E.5}$$

where ϵ_{ij} follows a type 1 extreme value distribution. With this structure, the university's tangency conditions are given by:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} = \frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_{j'}} \quad \forall j, j' \in \mathbf{d}\, s.t.\, c_j > m_j,\, c_{j'} > m_{j'} \tag{E.6}$$

where

$$\begin{aligned}
\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} &= \left\{\sum_{i=1}^{N}\left(\frac{\theta_j}{e_j+1}\right)P_{ij}\right\} + \sum_{k\in\mathbf{d}}\gamma_k\left(\frac{\partial\tilde{n}_k}{\partial e_j}\right) \\
&= \left(\frac{\theta_j}{e_j+1}\right)\left[\sum_{i=1}^{N}P_{ij} + \gamma_k\left(\sum_{i=1}^{N}P_{ij}\left(1-P_{ij}\right)\right)\right] \\
&\quad -\sum_{k\neq j}\left(\frac{\theta_k}{e_k+1}\right)\gamma_k\left(\sum_{i=1}^{N}P_{ik}P_{ij}\right)
\end{aligned} \tag{E.7}$$

and the binding budget constraint is given by:

$$\sum_{j\in\mathbf{d}}\left(m_j + e_j\right) = E \tag{E.8}$$

To simplify notation, I make the following substitutions:

$$\alpha_j\left(\mathbf{e}\right) = \sum_{i=1}^{N}P_{ij} + \gamma_j\left(\sum_{i=1}^{N}P_{ij}\left(1-P_{ij}\right)\right) \tag{E.9}$$

$$\kappa_j\left(\mathbf{e}\right) = \left(e_j+1\right)\left[\sum_{k\neq j}\left(\frac{\theta_k}{e_k+1}\right)\gamma_k\left(\sum_{i=1}^{N}P_{ik}P_{ij}\right)\right] \tag{E.10}$$

This simplifies the first order conditions to:

$$\frac{\theta_j \alpha_j (\mathbf{e}) - \kappa_j (\mathbf{e})}{e_j + 1} = \frac{\theta_{j'} \alpha_{j'} (\mathbf{e}) - \kappa_{j'} (\mathbf{e})}{e_{j'} + 1} \tag{E.11}$$

$\mathbf{e}(\mathbf{d})^\star$ can then be implicitly defined as:

$$e_j (\mathbf{d})^\star = \frac{\left(\theta_j \alpha_j (\mathbf{e}) - \kappa_j (\mathbf{e})\right) \left(E + J - \sum_{j=1}^{J} m_j\right)}{\sum_{j'=1}^{J} \left(\theta_{j'} \alpha_{j'} (\mathbf{e}) - \kappa_{j'} (\mathbf{e})\right)} - 1 \tag{E.12}$$

Appendix F

Marginal Improvement Algorithm for U-SWM Course Offerings

In this appendix, I describe a Marginal Improvement Algorithm (MIA) for solving for the optimal course offerings and excess spending decisions of a university. The algorithm requires that the university's objective as a function of offering vector $\mathbf{d}$ and excess spending vector $\mathbf{e}$ is known. Furthermore, the algorithm requires that the set of feasible courses J and the budget endowment E are observed.

Let $\Pi(\mathbf{d}, \mathbf{e})$ represent the university's course offerings $\mathbf{d}$ and excess spending vector $\mathbf{e}$ and let $\mathbf{e}(\mathbf{d})^\star$ represent the university's optimal excess spending vector given course offerings $\mathbf{d}$. Algorithms for deriving $\mathbf{e}(\mathbf{d})^\star$ for various structures of student utility and university objectives are presented in Methodological Appendix D. Finally, let $\mathbf{v_j}$ represent the elementary $J \times 1$ vector which contains 1 in entry j and zeros in all other entries. The MIA proceeds as follows:

1. Solve for the best single course to offer alongside contracted courses:

$$j_1^\star = \text{argmax}_{j \in J} \left\{ \Pi \left(\mathbf{v_j}, \mathbf{e} \left(\mathbf{v_j} \right)^\star \right) \right\} \quad \text{s.t.}\, m_j \leqslant E \tag{F.1}$$

2. Solve for the best course to offer alongside $j_1^\star$ and contracted courses:

$$j_2^\star = \text{argmax}_{j \in J \setminus j_1^\star} \left\{ \Pi \left(\mathbf{v_j} + \mathbf{v_{j_1^\star}}, \mathbf{e} \left(\mathbf{v_j} + \mathbf{v_{j_1^\star}} \right)^\star \right) \right\} \quad \text{s.t.}\, m_j + m_{j_1^\star} \leqslant E \tag{F.2}$$

3. In general, solve for the best $k+1$ courses to offer alongside previously chosen k courses and contracted courses:

$$\begin{aligned} j_{k+1}^\star &= \text{argmax}_{j \in J \setminus \cup_{k'=1}^{k} j_{k'}^\star} \left\{ \Pi \left(\mathbf{v_j} + \sum_{k'=1}^{k} \mathbf{v_{j_{k'}^\star}}, \mathbf{e} \left(\mathbf{v_j} + \sum_{k'=1}^{k} \mathbf{v_{j_{k'}^\star}} \right)^\star \right) \right\} \\ & \quad \text{s.t.}\, m_j + \sum_{k'=1}^{k} m_{j_{k'}^\star} \leqslant E \end{aligned} \tag{F.3}$$

The algorithm terminates when either the best additional course decreases the university's objective or when no additional courses can be added without violating the budget constraint. Formally, the algorithm terminates if:

$$\Pi \left(\sum_{k'=1}^{k+1} \mathbf{v_{j_{k'}^\star}}, \mathbf{e} \left(\sum_{k'=1}^{k+1} \mathbf{v_{j_{k'}^\star}} \right)^\star \right) < \Pi \left(\sum_{k'=1}^{k} \mathbf{v_{j_{k'}^\star}}, \mathbf{e} \left(\sum_{k'=1}^{k} \mathbf{v_{j_{k'}^\star}} \right)^\star \right) \tag{F.4}$$

or if

$$\min_{j \in J \setminus \cup_{k'=1}^{k} j_{k'}^\star} \left\{ m_j + \sum_{k'=1}^{k} m_{j_{k'}^\star} \right\} > E \tag{F.5}$$

Appendix G

Non-Utilitarian Welfare Weights

In this appendix, I discuss extending my theoretical framework and inference methods to include non-utilitarian weights on student welfare. This extension can be used to estimate weighting parameters which best explain observed classroom spending decisions. Non-unitary weights reveal institutional preferences for the welfare of some students relative to others.

With welfare weights, the university's payoff given by Equation (2.19) is modified to:

$$\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right] = \sum_{i=1}^{N} \omega_i \mathbb{E}\left[V_i \,|\, \mathbf{e}, \mathbf{d}\right] + \sum_{j=1}^{J} \gamma_j \tilde{n}_j \left(\mathbf{e}, \mathbf{d}\right) \tag{G.1}$$

where $\sum_{i=1}^{N} \omega_i = 1$.

For the baseline setting where class size does not affect utility, marginal effects given by Equation (2.22) are modified to:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} = \sum_{i=1}^{N} \omega_i \left(\frac{\partial u_{ij}}{\partial I_j}\right)\left(\frac{\partial \phi_j}{\partial e_j}\right) P_{ij}\left(\mathbf{e}, \mathbf{d}\right) + \sum_{k=1}^{J} \gamma_k \left(\frac{\partial \tilde{n_k}}{\partial e_j}\right) \tag{G.2}$$

For the general equilibrium setting where class size affects choice utility, marginal effects given by Equation (2.23) are modified to:

$$\frac{d\mathbb{E}\left[\Pi \,|\, \mathbf{e}, \mathbf{d}\right]}{de_j} = \left\{\sum_{i=1}^{N} \sum_{k\in\mathbf{d}} \omega_i \left(\frac{du_{ik}}{de_j}\right) P_{ik}\right\} + \sum_{k=1}^{J} \gamma_k \left(\frac{d\tilde{n_k}}{de_j}\right) \tag{G.3}$$

The variance minimization estimation problem given by Equation (2.24) then jointly estimates the parameter values for γ_j and ω_i which minimize the variance in $\frac{d\mathbb{E}[\Pi \,|\, \mathbf{e},\mathbf{d}]}{de_j}$ across non-contract offered courses for which spending on instruction exceeds minimum costs. Formally,

$$(\hat{\gamma}, \hat{\omega}) = \operatorname{argmin}_{(\gamma,\omega)} \left\{ \operatorname{Var}_{j\in\mathbf{d}\setminus K \text{ s.t. } c_j > m_j} \left(\frac{d\mathbb{E}\left[\Pi \,|\, \gamma, \omega, \tilde{\mathbf{e}}, \tilde{\mathbf{d}}\right]}{de_j} \right) \right\} \tag{G.4}$$

For identification and estimation, it is necessary to group students into types indexed by τ and restrict welfare weights to be the same for all student types $\omega_i = \omega_{\tau(i)}$. With this restriction, relative values of ω_τ and γ_f are over-identified as long as the number of student types plus the number of course types is less than the total number of non-contract courses for which spending on instruction exceeds minimum costs. Intuitively, identification of ω_τ comes from students of type τ concentrating in classes where the university is under (over) investing relative to a U-SWM university. This implies that the university puts less (more) weight on the welfare of type τ students relative to the general student population. Identification of γ_g comes from residual differences in marginal returns on spending which cannot be explained by student weights. For example, if course groups g and group g' have identical student compositions but the marginal returns on spending are higher (lower) in group g it must be that the university values enrollment in group g courses less (more) than enrollment in group g' courses.

The maximum likelihood estimator given by (2.31) then jointly estimates the parameter values for γ_j and ω_i which maximize the likelihood that observed course offerings yield a greater payoff for the university than all feasible alternatives. Formally,

$$(\hat{\gamma}, \hat{\omega}) = \text{argmax}_{(\gamma,\omega)} \left\{ \Pr\left(\hat{\Pi}\left(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \gamma, \omega\right) + \zeta_{\mathbf{d}} \geqslant \hat{\Pi}\left(\mathbf{e}\left(\mathbf{d}\right)^{\star}, \mathbf{d}, \gamma, \omega\right) + \zeta_{\mathbf{d}} \, \forall \mathbf{d} \in \mathbf{D}\left(E\right) \right) \right\} \tag{G.5}$$

where $\hat{\Pi}\left(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \gamma, \omega\right)$ use the modified payoff structure in (G.1).

Once again, it is necessary to group students into types indexed by τ and restrict welfare weights to be the same for all student types $\omega_i = \omega_{\tau(i)}$. Intuitively, identification of ω_τ comes from differences across student types in preferences for observed offerings $\hat{\mathbf{d}}$—if type τ students prefer $\hat{\mathbf{d}}$ to most alternative offerings but type τ' students do not particularly like $\hat{\mathbf{d}}$ this suggests the university values the welfare of type τ students more than the welfare of type τ' students. Identification of γ_g comes from the composition of courses offered in $\hat{\mathbf{d}}$ relative to the composition of courses in alternative offerings. If $\hat{\mathbf{d}}$ and $\mathbf{d}$ are equivalent from the perspective of students but $\hat{\mathbf{d}}$ offers more type g courses this implies the university values enrollment in type g courses more than other courses.

Bibliography

Achen, A. C. and Courant, P. N. (2009), "What are grades made of?" *The Journal of Economic Perspectives*, 23, 77.

Ackerberg, D. A. (2003), "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review*, 44, 1007–1040.

Allen, J. and Van der Velden, R. (2001), "Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search," *Oxford Economic Papers*, 53, 434–452.

Altonji, J. G. (2005), "Employer learning, statistical discrimination and occupational attainment," *The American Economic Review*, 95, 112–117.

Altonji, J. G. and Pierret, C. R. (2001), "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

Altonji, J. G., Blom, E., and Meghir, C. (2012), "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, 4, 185–223.

Altonji, J. G., Kahn, L. B., and Speer, J. D. (2014), "Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs," *The American Economic Review*, 104, 387–393.

Altonji, J. G., Arcidiacono, P., and Maurel, A. (2015), *Handbook of Labor Economics*, vol. 5, chap. 7, Elsevier.

Antonovics, K. and Golan, L. (2012), "Experimentation and job choice," *Journal of Labor Economics*, 30, 333–366.

Arcidiacono, P. (2004), "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

Arcidiacono, P. and Jones, J. B. (2003), "Finite mixture distributions, sequential likelihood and the EM algorithm," *Econometrica*, 71, 933–946.

Arcidiacono, P. and Miller, R. A. (2011), "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 7, 1823–1868.

Arcidiacono, P., Aucejo, E. M., Fang, H., and Spenner, K. I. (2011), "Does affirmative action lead to mismatch? A new test and evidence," *Quantitative Economics*, 2, 303–333.

Arcidiacono, P., Aucejo, E. M., and Spenner, K. (2012), "What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice," *IZA Journal of Labor Economics*, 1, 1–24.

Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2014), "College Attrition and the Dynamics of Information Revelation," Duke University.

Arcidiacono, P., Aucejo, E. M., and Hotz, V. J. (2015), "University differences in the graduation of minorities in STEM fields: Evidence from California," *American Economic Review*.

Bayer, P. and Timmins, C. (2007), "Estimating Equilibrium Models Of Sorting Across Locations," *The Economic Journal*, 117, 353–374.

Bernal, R. and Keane, M. P. (2010), "Quasi-structural estimation of a model of childcare choices and child cognitive ability production," *Journal of Econometrics*, 156, 164–189.

Bettinger, E. P. and Long, B. T. (2010), "Does cheaper mean better? The impact of using adjunct instructors on student outcomes," *The Review of Economics and Statistics*, 92, 598–613.

Bhattacharya, D., Kanaya, S., and Stevens, M. (2014), "Are university admissions academically fair?" Working Paper.

Bordon, P. and Fu, C. (2015), "College-major choice to college-then-major choice," *The Review of Economic Studies*, 82, 1247–1288.

Bound, J., Lovenheim, M. F., and Turner, S. (2012), "Increasing Time to Baccalaureate Degree in the United States." *Education Finance and Policy*, 7, 375–424.

Brint, S., Cantwell, A. M., and Saxena, P. (2012), "Disciplinary categories, majors, and undergraduate academic experiences: Rethinking Bok's "Underachieving Colleges" thesis," *Research in Higher Education*, 53, 1–25.

Bronson, M. A. (2014), "Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women," Georgetown University.

Butcher, K. F., McEwan, P. J., and Weerapana, A. (2014), "The Effects of an Anti-Grade-Inflation Policy at Wellesley College," *The Journal of Economic Perspectives*, pp. 189–204.

Cai, H., Chen, Y., Fang, H., et al. (2009), "Observational Learning: Evidence from a Randomized Natural Field Experiment," *The American Economic Review*, 99, 864–82.

Carrell, S. E., Page, M. E., and West, J. E. (2010), "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 125, 1101–1144.

Chade, H. and Smith, L. (2006), "Simultaneous search," *Econometrica*, 74, 1293–1307.

Chan, W., Hao, L., and Suen, W. (2007), "A signaling theory of grade inflation," *International Economic Review*, 48, 1065–1090.

Crawford, G. S. and Shum, M. (2005), "Uncertainty and learning in pharmaceutical demand," *Econometrica*, 73, 1137–1173.

DeGroot, M. H. (1970), *Optimal statistical decisions*, McGraw Hill.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, Series B (methodological), 1–38.

Downs, J. S., Loewenstein, G., and Wisdom, J. (2009), "Strategies for promoting healthier food choices," *The American Economic Review*, 99, 159–164.

Dupas, P. (2011), "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya," *American Economic Journal: Applied Economics*, pp. 1–34.

Dupas, P. (2014), "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica*, 82, 197–228.

Farber, H. S. and Gibbons, R. (1996), "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

Figlio, D. N., Schapiro, M. O., and Soter, K. B. (2015), "Are tenure track professors better teachers?" *Review of Economics and Statistics*, 97, 715–724.

Gemici, A. and Wiswall, M. (2014), "Evolution Of Gender Differences In Post-Secondary Human Capital Investments: College Majors," *International Economic Review*, 55, 23–56.

Gibbons, R., Katz, L. F., Lemieux, T., and Parent, D. (2005), "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 23, 681–724.

Greenwald, C., Stiglitz, J. E., et al. (1990), "Asymmetric Information and the New Theory of the Firm: Financial Constraints and Risk Behavior," *The American Economic Review*, 80.

Heckman and Singer, B. (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52, 271–320.

Hoffmann, F. and Oreopoulos, P. (2009), "A professor like me the influence of instructor gender on college achievement," *Journal of Human Resources*, 44, 479–494.

Hoxby, C. and Avery, C. (2013), "The missing" one-offs": The hidden supply of high-achieving, low-income students," *Brookings Papers on Economic Activity*, 1, 1–50.

Jalan, J. and Somanathan, E. (2008), "The importance of being informed: Experimental evidence on demand for environmental quality," *Journal of Development Economics*, 87, 14–28.

James, J. (2012), "Learning and occupational sorting," Federal Reserve Bank of Cleveland Working Paper 12-25.

Johnson, V. E. (2003), *Grade inflation: A crisis in college education*, Springer.

Jovanovic, B. (1979), "Job matching and the theory of turnover," *The Journal of Political Economy*, pp. 972–990.

Jovanovic, B. and Nyarko, Y. (1997), "Stepping-stone mobility," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 46, pp. 289–325, Elsevier.

Malamud, O. (2010), "Breadth versus Depth: The Timing of Specialization in Higher Education," *Labour*, 24, 359–390.

Malamud, O. (2011), "Discovering one's talent: learning from academic specialization," *Industrial & Labor Relations Review*, 64, 375–405.

Mansfield, H. C. (2001), "Grade inflation: It's time to face the facts," *Chronicle of Higher Education*, 47, B24.

Manski, C. F. (1992), *Parental income and college opportunity*, Democratic Study Center.

McCall, B. P. (1990), "Occupational matching: A test of sorts," *The Journal of Political Economy*, pp. 45–69.

McFadden, D. (1978), "Modeling the Choice of Residential Location," *Spatial Interaction Theory and Planning Models*, pp. 75–96.

Miller, R. A. (1984), "Job matching and occupational choice," *The Journal of Political Economy*, 92, 1086–1120.

Neal, D. (1999), "The Complexity of Job Mobility among Young Men," *Journal of Labor Economics*, 17, 237–261.

Nordin, M., Persson, I., and Rooth, D.-O. (2010), "Education–occupation mismatch: Is there an income penalty?" *Economics of Education Review*, 29, 1047–1059.

OECD (2014), *Education at a Glance 2014: OECD Indicators*, OECD Publishing.

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

Papageorgiou, T. (2014), "Learning your comparative advantages," *The Review of Economic Studies*, 81, 1263–1295.

Pavan, R. (2011), "Career choice and wage growth," *Journal of Labor Economics*, 29, 549–587.

Rask, K. and Tiefenthaler, J. (2008), "The role of grade sensitivity in explaining the gender imbalance in undergraduate economics," *Economics of Education Review*, 27, 676–687.

Rask, K. N. and Bailey, E. M. (2002), “Are faculty role models? Evidence from major choice in an undergraduate institution,” *The Journal of Economic Education*, 33, 99–124.

Sabot, R. and Wakeman-Linn, J. (1991), “Grade inflation and course choice,” *The Journal of Economic Perspectives*, pp. 159–170.

Shaw, K. L. (1987), “Occupational change, employer change, and the transferability of skills,” *Southern Economic Journal*, 53, 702–719.

Stinebrickner, R. and Stinebrickner, T. (2014a), “Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model,” *Journal of Labor Economics*, 32, 601–644.

Stinebrickner, R. and Stinebrickner, T. (2014b), “A major in science? Initial beliefs and final outcomes for college major and dropout,” *The Review of Economic Studies*, 81, 426–472.

Stinebrickner, T. and Stinebrickner, R. (2012), “Learning about Academic Ability and the College Dropout Decision,” *Journal of Labor Economics*, 30, 707–748.

Turner, L. J. (2014), “The Road to Pell is Paved with Good Intentions: The Economic Incidence of Federal Student Grant Aid,” Working Paper.

Wikström, C. and Wikström, M. (2005), “Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools,” *Economics of Education Review*, 24, 309–322.

Zafar, B. (2011), “How do college students form expectations?” *Journal of Labor Economics*, 29, 301–348.

Biography

James Thomas was born in Schenectady, NY on September 9, 1986. His family lived in Schenectady and Tokyo, Japan before settling in Palo Alto, CA where James graduated from Henry M. Gunn Sr. High School in 2005. James initially enrolled at the US Naval Academy but left voluntarily in 2007. James went on to graduate *summa cum laude* from the University of Pennsylvania in 2010 with Bachelor's degrees in Economics and Mathematics. Subsequently, James enrolled at Duke University where he earned a Master's degree in Economics in 2011 and a Ph.D. in Economics under the supervision of Peter Arcidiacono in 2016. James is an avid outdoorsmen with particularly strong interests in rock climbing and backcountry camping.