Do Professional Rankings Steer Students of Economics the Right Way?

(A study incorporating student rankings)

Ryan A. Meister

Applied Research for ECON 5670:

Applied Econometrics

University of North Texas

Denton, Texas

November 2023

Contents

1 Introduction **2**

2 Background **4**

3 The Meister Rankology **7**

3.1 Meister Professional Ranking 7

3.2 Meister Student Ranking 8

4 The Data **10**

5 Theoretical Details of the Probit Model **14**

6 Empirical Results **19**

7 Heteroskedasticity **28**

7.1 *Theory* 28

7.2 *Applied* 30

8 Conclusion **31**

9 References **34**

10 Appendix **35**

1 Introduction

Picture yourself at the ripe young age of 18, just getting ready to apply to college. How do you pick the universities to which you are going to apply? Some students are limited by their GPA and/or standardized test scores (such as the ACT or SAT) or financial resources, and some are not. Regardless, it is not unreasonable to assume that students would like to go to an excellent school. After all, going to college in the modern era is expensive, frequently tallying over a hundred thousand dollars or more over four years. As a rational consumer of education and given that an 18-year-old is not an independent unbiased researcher, they typically turn to those who are perceived to be. We will refer to these individuals as the "Rankings Professionals," or "Professionals" more broadly. The purpose of this research is to provide information about rankings of university economics departments from the perspective of the students, relative to professionals.

The bright young student, who has no collegiate experience his or herself, sometimes turns to professionals to find the "right" universities in which to apply. Again, some students may have academic and financial limitations to get into the best universities, however we will assume that this is not the case. That is, we will assume that the student has the GPA, standardized test scores, finances etc., to get into whatever university he or she wishes, they simply want to be aware of all the school's merit, then choose.

Why this research is important is that it challenges these professional rankings, something that has not been pursued in an academic endeavor. To put it simply, would you rather trust someone's recommendation who has never attended the university they are recommending to you, or would you rather hear from the students themselves? As a student, one would typically want to hear from a fellow student, an individual who has attended the university which they are considering attending and taking the same courses in the major they are thinking about declaring. This research aims to challenge the professional rankings by matching them up against student rankings. Since there are no aggregate student rankings, one is constructed using The Meister Rankology.

For economics specifically, there are more than a couple of professional ranking lists available online. To accurately determine what the overall ranking of a school is, all readily used, popular and/or respected rankings lists are used in this analysis. The five ranking lists are: US News, Quacquarelli Symonds (QS) Top Universities, Times Higher Education, Research Papers in Economics (RePEc), and the list by the University of North Texas' Department of Economics, which ranks economics departments based on the number of pages published (McPherson 2012). To get an aggregate view of the average rankings of each university, each list is matched up against one another and a new rankings list is calculated. This new ranking list contains the top 50 economics departments in the country, on the average, according to the rankings lists described above. This list will be referred to as the "Meister Professional Rankings," more on the specifics of this list in section three. After we have the top 50 economics departments, we analyze the department again, but this time from the students' point of view.

To obtain an aggregate measure of how students, as an aggregate, rank their university's economics department, Rate My Professor data is utilized. A total of over 70,000 professor reviews across more than 3,500 professors in the top 50 institutions are utilized for this list. There are many aspects that a student can review a professor on, however only the quantitative data points are considered, namely quality and difficulty. To generalize, after all the data is collected, aggregated then averaged, a student rankings list can be constructed, specifics of how this process is performed is detailed in section three. Given the assumption that students want excellent instruction, and their professors to be quality instructors and not make the course unnecessarily difficult, we assume that students want a professor, or namely, a department that has high quality and low difficulty instruction. To put this into perspective, the number one school on the students' rankings list will be the college that has the highest quality instruction with the lowest difficulty. On the other side, the last ranked school (rank of 50), will have the lowest quality instruction with the highest level of difficulty of the schools considered.

The purpose of obtaining both the professional and student rankings is to compare them, identify where they differ and analyze why they are different. Do professionals and students disagree, or agree in most cases? and, how much the chosen independent regressors contribute to this phenomena?

Why should anyone care about the findings of this paper? It is paramount in a progressive society to question authority and scrutinize it by striving for the same goal through different means, then compare the results. If these professional rankings are not ever put under a microscope and compared to other rankings derived from different methodologies, they would become the dominant source of information, and may steer students who consult them astray, leading them to amass debt and confront a hostile learning environment. Although this research focuses on comparing professional to student rankings in economics, The Meister Rankology can be applied to any major or discipline.

Those who this research can help and can use it to their benefit are high school seniors who know they are going to declare economics as their major. Another group is current undergraduate seniors, or any student who is applying to graduate schools and is looking for an excellent program to attend which could include current working professionals. These groups of individuals are typically only aware of the professional rankings. This project does not intend to replace the professional rankings entirely, it simply investigates the opinions of both professionals and students about

2 Background

Some baseline information is needed to understand the concept of this paper and its objective. To begin, RateMyProfessor.com is a website on which students can do exactly what is described in its name, rate their professors. Once a student visits the website, they are prompted to type in their school's name. After this, they can click to view all the professors who have been reviewed by their students. For instance, at the University of North Texas, as of October 30, 2023, there are 4749 professors who have been reviewed across over 50 different disciplines.

Students can search for the professor in which they wish to review via a search bar after they have correctly identified their school. After locating the professor and clicks on him/her to write a review, they are asked a few questions. The first of which is the course in which they took under the professor. Next students rate their professor's quality of instruction on a scale from one to five, one being awful and five being awesome. Next, students are asked to rate the difficulty of the coursework as presented by the professor, once again on a scale from one to five. In this case, a one is very easy and a five is very difficult. Next, students are asked a simple yes or no question; would you take this professor again? In terms of submitting a review, these are the three metrics that are explicitly required by RateMyProfessor.com to submit a review in addition to a written review containing up to 350 characters. However, there are other metrics that students can choose to communicate more in their ranking, which contain more yes or no questions such as "was this class taken for credit?", "did this professor use textbooks?", and "was attendance mandatory?" There are two more optional questions that are not simple yes or no questions, such as stating the grade received in the course and "tags." Tags are perceived characteristics of the professor. To name a few of the tags that students can choose from: tough grader, online savvy, hilarious, inspirational, clear grading criteria, lecture heavy, get ready to read, etc. For the purposes of this paper, vast availability of metrics and given a time constraint, only the variables quality and difficulty of professors are collected and utilized to form student rankings.

Let us now turn our attention to each of the professional rankings. First is US News and its Best Economics Schools rankings list. US News sends out surveys, which listed 139 economics programs in 2023, to department heads. The universities are chosen based on the criteria of having at least 5 doctoral degrees awarded in the past 5 academic years, the survey writers use Integrated Postsecondary Education Data System (IPEDS) to obtain this information. After department heads receive a survey, they are tasked with rating each program on a 5-point scale, with 5 being "outstanding" and 1 "marginal." However, if a given department head is unfamiliar or has not heard of a particular school's economics program, they have the option to select "don't know" as listed on the survey. Once all the surveys are collected, their scores for each school are determined by dropping the top and bottom two responses, then an average score is calculated and the departments are sorted in from greatest to least (Morse and Hines 2023). An important piece of information about this professional ranking list is that it relies entirely on opinions. The justifications for why a certain professional rated a university a certain way may vary widely between survey takers.

The next professional rankings list that is utilized is QS Top University's Rankings by Subject 2023: Economics & Econometrics. There are four categories on which QS gathers information to rank departments, namely: research reputation (40% of the overall score), the learning and teaching environment (30%), research impact (20%) and internationalization (10%). In terms of research reputation, QS asks academics to give their opinion on which institutions are the most respectable, of the ones they are familiar with. They spoke with nearly 8,000 academics during seminars that share the same focus of discovering the top programs in the world. Their justification for asking academics to participate in their survey relative to any other group is that academics are the best group of individuals to ask these questions on research and academic excellence because they collaborate across universities, attend conferences assist other academics in the revision process and sit on advisory boards, so if anyone has the knowledge to be able to give accurate answers, it is this group of individuals (QS 2021). As far as the learning and teaching environment, QS considers the student-to-faculty ratio of the specific university at hand, making the claim that the more faculty that are available the more seriously the university takes their teaching commitments (QS 2021). In terms of further analysis for the teaching and learning environment, no other metric, or form of measurement is considered or applied in the methodology. Research impact is measured by calculating the volume of citations being achieved on the average, by the given university. Internationalization is measured in two metrics, the International Student Ratio (ISR) and the International Faculty Ratio (IFR). The ISR is measured by taking the total number of international students and dividing it by the total number of students. The IFR is calculated the same way, but by using the faculty population at the institution. It is important to reiterate that, in terms of looking at academia from the students' point of view, only the student-to-faculty-ratio is considered. Although QS attempts to measure the learning and teaching environment, and weighs it at 30%, there is more to this environment than the simple student-to-faculty ratio of a university at large.

The third out of the five professional rankings is Times Higher Education's World University Rankings 2022 by subject: business and economics. Although "world" is included in its name, only universities in the United States were curated and considered for the purposes of this research. Times Higher Education (THE) breaks down its ranking into five categories: teaching (30% of the overall score), research (30%), citations (30%), international outlook (7.5%), and industry income (2.5%). Most of these categories are broken down into subsections. Teaching (the learning environment)'s category is made up of 5 subsections: reputation survey (15%), awarded-to-academic-staff ratio (6%), student to faculty ratio (4.5%), doctorate-to-bachelor's ratio (2.25%), and institutional income (2.25%). The reputation survey was given to academics who ranked the programs they are familiar with the respective university's teaching. Next is Research (volume, income and reputation) which is broken into three subsections: reputation survey (18%), research income (6%) and research productivity (6%). The reputation survey is was given to academics that are familiar with the respective university's research output and quality. This rankings institution measures productivity by counting the number of works published in academic journals, that are indexed by Elsevier's Scopus database (Times Higher Education 2021).

The fourth out of five rankings list that is included in this analysis is Research Papers in Economics (RePEc)'s Top 25% Economics Departments, all authors, all publication years. First, it is important to note that RePEc only considers authors who have registered with the site. The two main metrics that this ranking considers is the number of authors registered with RePEc as well as "Author shares," which measures how many times, on average, an author at a university is cited, this information is measured by the CitEc project (RePEc/IDEAS 2023). After this information is collected, an average ranking is calculated then departments are ordered from top to bottom.

The fifth and final professional ranking list that is used in this project is University of North

Teaxas' Ranking U.S. Economics Programs by Faculty and Graduate Publications. Michael McPherson gathered data on and ranks U.S. economics departments based on the number of pages published in top economics journals during the time frame of 2002 to 2009. For instance, Northwestern University is ranked seventh because it published 5007.5 pages within the time frame above, which was more than University of Pennsylvania (4948.2 pages) and less than New York University. As far as what this paper considers as a "top" journal here are a few of the top journals listed: American Economic Review; American Economic Review Papers and Proceedings; Econometrica; Economic Inquiry; Economic Journal; Economic Theory; Economica; Journal of Law, Economics, and Organization; Journal of Mathematical Economics; Journal of Monetary Economics and the Journal of Money, Credit, and Banking (McPherson 2012).

3 The Meister Rankology

As touched upon briefly in prior sections, there are two ranking lists that are calculated in this work. "The Meister Ranking System" refers to the collection of all relevant and reliable existing quantitative information, averaging them to obtain a unique rank, then ranking them in ascending order. A ranking of 1, under the Meister Ranking System, indicates the "best" of something. The term "best" is defined clearly before the data collection process begins. The Meister Ranking System is utilized in this research to rank the top 50 economics departments ranked by professionals, this rankings list is coined the "Meister Professional Rank." It is also used to figure out the "best" economics department from the students' perspective. The fundamental essence of the Meister Ranking System is to incorporate students' opinions into account, which is hardly ever done by mainstream rankings.

3.1 Meister Professional Ranking

For the "Meister Professional Rank"[1] list, the term "best" is defined as having the highest average rank across all the rankings lists from professionals. This is done by first collecting the data on the top 50 universities from the five reputable rankings lists (US News, Times Higher Education, QS Top Universities, McPherson's and RePEc).

One issue that calculating an average value is what to do if a university was not included on each of the five rankings lists. To handle this problem, a "exclusion penalty" is imposed, and is a numeric penalty taken to account when calculating the average rank of the department. This

[1] For reference, the Meister Professional Rank list is included in the appendix

penalty is imposed every time a university is excluded from a rankings list. This penalty is only imposed if the university at hand was simply not deemed worthy enough to include on a list. This means that not only was the department not in the top 50, it was not contained within the list at all. The numeric penalty imposed on a university for being excluded from a list is 75. This value is chosen to penalize it by the number of ranking spots on the Meister Professional Rankings List (50), multiplied by 1.5. Furthermore, not being included on the list is worse than being the 50th ranked school on a given list, how much worse? 150% worse.

After the correct penalties are applied, an average rank can be taken. The average rank is taken by summing the respective economics department's ranking spots and exclusion penalties, if there are any. For instance, Princeton University is not included in the Times Higher Education rankings list. The sum of its ranking spots on the other four lists is 27. After the exclusion penalty is imposed, Princeton's average value is (27+75)/5, which is 20.4. Princeton ranks above Boston University and below Cornell University in the 16th ranking position.

One issue arises from the Meister Ranking System, take the example of the University of Texas at Austin (UT) and the University of Maryland at College Park (UM). Both schools were included on all 5 lists, and have the same average rank of 25. In this case, McPherson's list is referenced and the university with the highest ranking spot on that list takes the higher ranking spot. In this case, UM takes the higher ranking spot because it was ranked by McPherson at 16, where UT was ranked at 23. There is only one other case in which an issue like this exists, which is between George Washington and Iowa State University.

3.2 Meister Student Ranking

For the "Meister Student Rank"[2], the term "best" is defined as a department having the highest quality and lowest difficulty of instruction. The data that comprises the Meister Student Rank is collected from RateMyProfessor.com (RMP) for the variables of quality and difficulty. All the available data is used in this analysis. What this means is that for a given economics department, all of the available student ratings on all the available professors are used. This method is chosen because economics departments are made up of professors, so to get the best picture of the department in the aggregate, we must look at all professors across all years. After these values are retrieved from RMP, calculation can begin

There were two variables, that are taken into consideration when determining the "best"

[2]For reference, the Meister Student Rank list is included in the appendix

economics department from the point of view of students, so calculation becomes a bit more complicated than prior. The average quality score is calculated by taking all of the quality scores from all of the student ratings of their economics professor at a specific university, adding them together, then dividing by the total number of ratings the department at hand has. For instance, the University of California at Berkeley (UCB) had 1659 student ratings, and when summed together the quality value is 5401. This indicates that the average quality of UCB is approximately 3.26. This process is done for all 49 other universities. Recall from the definition of "best" a high quality score is preferred.

The second variable collected and manipulated from RMP is difficulty of instruction. No different from quality, average difficulty score is calculated by dividing the sum of all difficulty ratings across a department by the total number of ratings that it has. For UCB, the difficulty is 5864. This means that UCB's average difficulty is about 3.53. Once again, this process is performed on all the remaining 49 universities. Recall from the definition of "best," low difficulty s preferred.

Since the heavy lifting of the calculation is finished, we can not proceed to discover how the universities match up against each other in terms of the definition of "best," being of highest quality and lowest difficulty. To do this, we must rank each university's quality from highest quality to lowest quality, and difficulty from lowest difficulty to highest difficulty by using the values calculated in the prior two paragraphs independent of one another. Using this definition, the school that is ranked 1 on the Meister Student Rank list is of the highest quality and lowest difficulty.

After quality and difficulty are ranked by each school independent of one another, it is time to combine them. The ranking of quality and the rank of difficulty are added together. After they are summed, we obtain the specific department's rank points. It is from here that we can now rank each school from "best" to not the best by listing each economics department by their rank points, from lowest to largest. For instance, the University of California at Los Angeles (UCLA), of quality rank 1 and difficulty rank 1, is ranked higher in the Meister Student rankings than the University of Minnesota at Twin Cities, of quality rank 9 and difficulty rank 2. The reason UCLA is ranked higher is because its rank points of 2 (1+1) is lower than the University of Minnesota's 11 (9+2).

Similar to the Meister Professional Rankings, the Meister Student Rankings run into the same issue. How does one discern which university is better than the other when they have the same number of rank points? This problem is handled by ordering the universities, from highest to lowest quality rating. Take the University of California - San Diego (UCSD) and Ohio State University (OSU), they are both tied at 48 rank points, and are competing for the 24th ranking position. Since UCSD has a higher quality rating (3.34) than OSU (3.14), it takes the 24th position, leaving the 25th for OSU. This issue presents itself a few more times, and is handled as described above (see the Meister Student Rankings in the appendix for more detail.

4 The Data

The dependent variable of study is $overrated_i$, which measures whether or not the economics department at university "i" is overrated. $overrated_i$=1 if the department at hand is overrated, which means that the professionals ranked it one or more spots above their student counterparts. On the other hand, if $overrated_i$=0, the economics department at hand is not overrated. This dependent variable, $overrated_i$, is calculated by comparing both the professional and student ranking lists and is derived by the researcher.

The first regressor that is utilized to explain whether or not an economics department is overrated is $Depfund_i$ which measures the average amount of funding, in millions of dollars, that the economics department at university "i" received over the past 10 years, 2012 to 2021 (National Science Foundation). This regressor is expected to have a positive sign, because the more funding that is brought into the department the more is usually required of its faculty, outside of the classroom, leaving educating students as a second priority, thus causing professionals to rank it higher because of the output that comes from the funding and students lower because they may feel abandoned by their educator. The next independent variable is $Pages_i$, which measures how many pages, in thousands, that the economics department at university "i" published from the years of 1994-2009, which is the last time the data was available (McPherson 2012). For the same justifications stated above, the expected sign on this regressor is positive as well. The next regressor is $Qual_i$ which measures the perceived quality of all professor instruction in economics courses at university "i" (measured as a percentage from 1 to 100). For instance, a department with a perceived quality of "90" would be interpreted as students perceiving the department of having better quality instruction than a department that has a perceived quality of "65" (Rate my Professor). This regressor is expected to have a negative sign, because there is a tradeoff between doing research and teaching in academia, where professionals pay attention to the output of a department, students pay more attention to the input of the economics professors such as coming to class prepared and making a lecture easy to understand and have clear expectations, traits that professors mainly concerned with research and publishing usually do not. The final regressor used to describe whether or not a university's economics department is overrated is $Diff_i$ which measures the perceived difficulty of all coursework taught by economics faculty at university "i," measured as a percentage from 1 to 100 (Rate My Professor). This sign is expected to be positive due to the fact that when professors do research and publish, a trait that professionals weigh heavily, they are trading off time they could be devoting to being a better educator. Perhaps it is the case that when professors mainly conduct research, they leave their students to teach themselves by reading the book, and maybe they do not set clear expectations which makes it difficult for students to pass the course.

The independent variables described above ($Depfund_i$, $Pages_i$, $Qual_i$ and $Diff_i$) and the dependent variable that measures whether or not a university's economics department is overrated ($Overrated_i$) are listed and defined in the table below along with their expected signs:

Table 1: Definitions of Variables and their Expected Signs

Variable	**Defintion**	
Overrated	=1 if the economics department at university ”i” is overrated	**Expected Sign**
Depfund	An average of the amount of funding that economics department ”i” in the past 10 years (measured in millions)	+
Pages	The number of pages published by economics faculty at University ”i” in the years of 1994 to 2009 (measured in thousands)	+
Qual	Perceived quality of all professor instruction in economics courses at university ”i” (measured as a percentage)	-
Diff	Perceived difficulty of all coursework in economics courses at university ”i” (measured as a percentage)	+

It is important to note that a rational person may conclude that the regressors of $Depfund_i$ and $Pages_i$ may be correlated. The degree of correlation between them is equal to 0.57, which is relatively weak. However, if we use $Depfund_i$ to explain $Pages_i$, on the average, for every million dollar increase in the department's funding, the number of pages published by the given department increases by 1,450, ceteris paribus. Another way of stating this is that it takes about $700 for every 1 published page by an economics department in this sample, all else equal. Although interesting, the correlation is not strong enough to pose as a problem within the model.

With regard to summary statistics we will consider three groups, of observations. The first group we will consider is all of the economics departments contained in the sample. We will then consider the summary statistics for the regressors for both values of the dependent variable. We will do this by first looking only at the non-overrated departments (where $Overrated_i$=0). Then, we will consider the group that contains all the overrated economics departments in the sample (where $Overrated_i$=1).

Let us first examine the summary statistics for all of the economics departments contained in the sample:

Table 2: Summary Statistics of all Economics Departments in the Sample					
Variable	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
Overrated	50	0.48	0.50	0.00	1.00
Depfund		4.73	6.56	0.01	32.23
Pages		2.23	2.21	0.02	9.80
Qual		67.31	6.03	57.00	86.40
Diff		66.65	6.72	37.80	76.00

From Table 2, we can observe that about half of the departments are overrated, while the other half are non-overrated. As far as the mean of $Overrated_i$, it does fall within the acceptable range of 0.2 and 0.8, thus confirming that the variation requirement is met, and we can proceed with confidence in the model. This requirement must be met to ensure that no single group is over represented, as that would bias the results. For analysis, it makes sense to consider each of these groups separately.

The summary statistics for each group are detailed in Table 3 and Table 4 below:

Table 3: Summary Statistics for Underrated Economics Departments						
Overrated	**Variable**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
0	Depfund	26	2.79	2.82	0.04	12.51
	Pages		1.64	1.66	0.02	5.01
	Qual		68.95	6.12	57.00	86.40
	Diff		62.58	5.62	37.80	67.40

Table 4: Summary Statistics for Underrated Economics Departments						
Overrated	**Variable**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
1	Depfund	24	6.83	8.62	0.01	32.23
	Pages		2.28	2.57	0.03	9.80
	Qual		65.53	5.00	58.20	78.00
	Diff		68.97	3.92	59.00	76.00

For each value of the dependent variable, we see noticeably different summary statistic values. On average, overrated economics departments are funded roughly 2.5 times more, publish 150% more pages, and have lower quality and higher difficulty instruction. Looking at it from the other perspective, non-overrated schools are funded less, published less, have higher quality and lower difficulty instruction. Additionally, we can observe that there are more non-overrated economics departments than there are overrated ones indicating that professional's may have an incentive to inflate their rankings.

To be sure we can move forward with this research confidently, the "Largest Smallest" rules need to be checked. To satisfy the first part of this rule, the largest value of a regressor, when the economics department is non-overrated ($y_i = 0$), must be bigger than the smallest value of the regressor when the economics department is overrated ($y_i = 1$), this is referred to as the "Largest Smallest Inside Rule." To simplify this rule, the value highlighted in blue for a given regressor in table 5 must be greater than the corresponding value of that regressor in green in table 6. The tables are as follows:

Table 5: Largest Smallest Inside Rule (Overrated = 0)

Overrated	Variable	Minimum	Maximum
0	Depfund	0.04	12.51
	Pages	0.02	5.01
	Qual	57.00	86.40
	Diff	37.80	67.40

Table 6: Largest Smallest Inside Rule (Overrated = 1)

Overrated	Variable	Minimum	Maximum
1	Depfund	0.01	32.23
	Pages	0.03	9.80
	Qual	58.20	78.00
	Diff	59.00	76.00

As we can observe, this data passes the "Largest Smallest Inside Rule." Because this rule is passed, we can move on the second part of it.

"The Largest Smallest Outside Rule" requires the largest value of the regressor, when the department is overrated to be bigger than the smallest value of the regressor when the department is non-overrated. If the data do not conform to this rule, we must take additional steps in order to proceed with confidence. Similarly, tables 7 and 8 display this information for analysis.

Table 7: Largest Smallest Outside Rule (Overrated = 0)			
Overrated	**Variable**	**Minimum**	**Maximum**
0	Depfund	0.00	11.47
	Pages	0.59	4.93
	Qual	64.04	86.38
	Diff	37.73	75.00

Table 8: Largest Smallest Outside Rule (Overrated = 1)			
Overrated	**Variable**	**Minimum**	**Maximum**
1	Depfund	0.04	31.22
	Pages	0.74	9.80
	Qual	58.13	70.27
	Diff	58.91	75.94

This time, the value highlighted in blue for a given regressor must be smaller than the corresponding value of that regressor highlighted in green in order to pass this rule. As we can observe from above, the data set passes the rule and we can proceed with confidence.

5 Theoretical Details of the Probit Model

The model used in this analysis is called the probit model. The term "probit" is derived from the term "probability unit" (Bliss 1934). Basically, what this means is that for the existence of some independent variable that influences the dependent variable, its effect is measured in probability units, or probits. Positive probits indicate a positive probability that an event occurs, or a trait exists, and negative probits have the opposite effect. The probit model's dependent variable must be binary, either 0 or 1. In this research, which uses the dependent variable $Overrated_i$, is 1 if an economics department is overrated or 0 if it is non-overrated. The probit model estimates the probability that a specific university's economics department will be overrated, which is a value between 0 and 1, so it makes sense why the dependent variable must be either 0 or 1.

The Ordinary Least Squares (OLS) regression model is not the best model to use in this case, or any case in which a researcher is calculating the probability of something occurring. The OLS model for calculating probability is called the linear probability model (LPM). There are many problems in using the LPM, such as the error term not being normally distributed, this is devastating because it does not allow for reliable hypothesis test results. Also in the LPM model, there is heteroskedasticity, which means that the variance across the regressors is not constant. This is a foundational issue because it violates the Gauss Markov (GM) Assumption 3 which indicates that there must be homoskedasticity, or constant variance across the regressors. Another reason the LPM should not be used in this case is that it can predict a probability that is outside the value of 0 and 1. This is problematic as the conclusion for a predicted probability of 1.25 is that "the probability of an economics department being overrated is 125%" which does not make logical sense, as probabilities must be within the range of 0 and 1. The third reason to not use LPM in this case is that it assumes constant partial effects. Basically, it lumps everyone together and has a blanket partial effect for everyone, when there is likely to be different partial effects for each economics department in the sample.

To measure the dependent variable, we use an indicator variable, $Overrated*_i$ which measures whether or not an economics department has the desire or ability to "ride" on the prestige that is brought about by the school's name. By "ride" it is assumed that the department at hand is not motivated to work harder and more effective with students because they have no incentive to improve the prestige of their department, as the university is already ranked so high. We assume that if an economics department has the desire or ability to "ride" on the prestige of the school's name, it is overrated. That is the same as stating that if $Overrated*_i > 0$, then $Overrated_i = 1$. The same pattern holds for economics departments who do not have the desire to "ride" on the prestige of their university's name; if they do not, then they are non-overrated. That is the same as stating that if $Overrated*_i \leq 0$, then $Overrated_i = 0$. Using the indicator variable is important because of the non-linear nature of the data. It is important to note that the estimation method used in the probit model is the maximum likelihood estimation (MLE) method, which indicates that the results from the model are as accurate and reliable as they can be.

The slope coefficients on regressors in the probit model are not the marginal effects. We are estimating the probability that an economics department is overrated, each department has a unique set of values for the regressors. Because of this fact, the probit model's marginal effects are estimated specifically for every single observation. It is because of this that it would not make sense to use the slope coefficients as the marginal effects, every observation has distinct independent variable values, what is true for one observation is not likely to be true for another. The actual formula for the estimated marginal effects in probit are as follows:

$$\frac{\partial F(\widehat{Overrated_i})}{\partial x_{ij}} = \frac{\partial F(\widehat{Overrated_i})}{\partial(\widehat{Overrated_i})} \times \frac{\partial F(\widehat{Overrated_i})}{\partial x_{ij}} = f(Overrated_i) \times \frac{\partial F(\widehat{Overrated_i})}{\partial x_{ij}} \quad (1)$$

Where $F(\widehat{Overrated_i})$ is the cumulative density function (CDF) of a standard normal distribution evaluated at $(Overrated_i)$ unique to observation "i" with its specific regressor values, and $f(Overrated_i)$ is the probability density function (PDF) of a standard normal distribution evaluated at the dependent variable unique to observation "i".

The probit model also "predicts" the value the dependent variable will take on. It essentially does this after calculating the estimated marginal effects for each observation. It then plugs the values for each of the regressors and multiplies them by the estimated marginal values to either obtain the predicted probability. Or, much simpler and in equation form:

$$\widehat{Prob}(Overrated_i = 1|X) = 1 - F(\widehat{-Overrated_i}) = F(\widehat{Overrated_i}) \quad (2)$$

where the predicted probability depends on one minus the probability that the dependent variable does not take the value of one, which leaves us with the probability that it does take on the value of one, or in this case, an economics department is overrated.

Another important aspect of the probit model is evaluating how good of a fit a model is in absolute terms. There are four absolute measures of fit that can be used to evaluate a probit model. The four are the: pseudo general F-test, Wald/t-test, Hosmer-Lemeshow test, and the percentage of correct predictions.

For reference to the statistical tests considered below, here is the population regression equation:

$$Overrated_i^* = \beta_0 + \beta_1 Depfund_i + \beta_2 Pages_i + \beta_3 Qual_i + \beta_4 Diff_i + \epsilon \quad (3)$$

recall that $Overrated_i^*$ represents the economics department at university "i"'s desire, or ability to ride on the prestige of the school's name or reputation, rather than making it an adequate academic environment. However, desire and ability can not be directly measured, so we turn to sample regression equation:

$$\widehat{Overrated_i} = \widehat{\beta_0} + \widehat{\beta_1} Depfund_i + \widehat{\beta_2} Pages_i + \widehat{\beta_3} Qual_i + \widehat{\beta_4} Diff_i \quad (4)$$

The pseudo general F-test is similar to the usual F-test in OLS, but because the probit model does not have OLS residuals, because we use MLE, we must use the pseudo general F-test. Furthermore, our regressors are non-linear, so it would be unwise to use a linear significance test to evaluate the joint significance of non-linear regressors. This test has a Chi-squared distribution. The point of running this test is to see if all the regressors are jointly statistically insignificant, which is represented by the null hypothesis as follows:

$$H_o : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \tag{5}$$

The alternative hypothesis is the opposite of the null hypothesis. This means that the regressors are not jointly significant in explaining the dependent variable. In equation form, the alternative hypothesis is:

$$H_a : not\ H_o \tag{6}$$

If the test statistic, which can be the Likelihood Ratio (LR) test statistic, is larger than the chi-squared critical value, we would reject the null. What this means is that the regressors are jointly significant when explaining the probability that something, such as an economics department being overrated, occurs.

The next test, is a Wald/t-test. This test is used to see if a single regressor is significant. It is similar to the prior test, but different in nature because it only test for significance of a single regressor and not all the regressors jointly. Here is the null hypothesis for this test:

$$H_o : \beta_j = 0 \qquad\qquad j = 1, 2, ... \tag{7}$$

which tests to see if the regressor, or parameter that goes with it is statistically significant when estimating the dependent variable. Here is the alternative:

$$H_o : \beta_j \neq 0 \qquad\qquad j = 1, 2, ... \tag{8}$$

Again, in this case, we want to reject the null if the goal is to find a significant relationship between the parameter and its ability to describe the dependent variable. If the test statistic is larger than the critical value, we reject the null, and conclude that the regressor is significant.

The third test is the Hosmer-Lemeshow test. This test is different from the above tests because we want to fail to reject the null as it indicates a good fit of the model. By failing to reject the null, we are stating that the model fits the data well.

$$H_o : The\ fit\ is\ good \tag{9}$$

And, on the other hand here is the alternative hypothesis, which is always the opposite of the null:

$$H_a : The\ fit\ is\ not\ good \tag{10}$$

Again, we use a test statistic, specifically the Lagrange multiplier (LM) in this case, and if it is larger than the chi-squared critical value, we reject the null. So, if the fit is good the LM statistic will be smaller than the critical value, the fit is good.

The final test is the simplest to interpret. The percentage of correct predictions test, which puts our model to the test, evaluating and comparing its calculated predictions to the actual values of the dependent variable. A correct prediction for outcome $Overrated_i = 1$ occurs if the predicted probability, $\widehat{prob}(Overrated_i) \geq 0.50$ and the actual value of $Overrated_i$ is 1. The same interpretation holds for a correct prediction for $Overrated_i = 0$, but in this case, $\widehat{prob}(Overrated_i) < 0.50$ 50 and the actual value of $Overrated_i$ must be 0 to be counted as a correct prediction. Any other relationship between the predicted and actual value of $Overrated_i$ is considered to be a incorrect prediction and counts against the model's accuracy. The statistic "sensitivity" measures to the percentage of correct predictions for the case when the value of $Overrated_i = 1$. Additionally, "specificity" measures to the percentage of correct predictions when $Overrated_i = 0$. After the predicted probability and actual value of $Overrated_i^*$ is computed for all the economics departments in the sample, if the model is 80% or more accurate in both the sensitivity and specificity categories, then the model at hand can be classified as fitting the data well. A model that has below 80% accuracy in either of these categories, then it is not adequate to use for further analysis.

When evaluating which model is of best fit, there are a couple of metrics that we can use to help us choose between a couple given models. Namely, they are the Akaike Information Criterion (AIC), the Schwarz Criterion (SC)/Bayesian Information Criterion (BIC), and a few pseudo R^2 measures.

The AIC measures "badness of fit," rather than goodness of fit. This relative measure of fit measures total error in the model, so the smaller this value is, the better suited the model is to the data. This is a relative measure of fit, so a value of 34.76 supplies us no useful information as to if this model is the best fit. The value of 34.76 must be compared to a different model's AIC to supply information about which model fits the data best. For instance, say we calculated another model and that model's AIC is 20.25, the latter model indicates a better fit as 20.25 is less than 34.76. The AIC is calculated using the following equation:

$$AIC = -2Ln(L(\theta)) + 2K \tag{11}$$

where "K" is the number of parameters. The term $Ln(l(\theta)$ indicates the log-likelihood function which assesses how well a model explains the observed data. As we can observe, the AIC is essentially multiplying the log-likelihood by negative two, then adding to it the number of parameters multiplied by 2. This gives us a measurement of the total error in the model, where the lower relative value indicates the best fitting model.

The SC is similar to the AIC as it is also a measurement of "badness of fit." The same interpretation is applied, as this too is a relative measurement of fit and needs to be compared to other models' to supply useful information to the research as how to which one is the best. The equation used to compute the value for the SC is as follows:

$$SC = -2Ln(l(\theta)) + KLn(N) \tag{12}$$

where "K" is the number of parameters, and "N" represents the number of observations in the sample. Again, the term $Ln(l(\theta)$ indicates the log-likelihood function which assesses how well a model explains the observed data. Basically, the SC still multiplies the log likelihood by negative two, but then adds the number of parameters, multiplied by the natural log of the number of observations in the sample. So, if the number of observations is greater than e^2 (which is approximately 7.4) the SC will be greater than the AIC. The SC should be larger than the AIC for any reasonable research project, indicating it used over 7.4 observations which is a shockingly low number of observations. Again, this is a relative measure of fit, so a value of 27.88 tells us nothing about whether we should use that model. It should be compared to another model's value, say 40.86. In this case, the former model fits the data better due to 27.88 being less than 40.86.

This analysis utilizes three pseudo R^2 measurements, Adrich-Nelson, McFadden's and McKelvey-Zavoina. The interpretation for each of these measurements of relative fit is rather simple, the model with the highest value indicates the strongest fit. Additionally, the model largest value of the¬pseudo R^2 also has the strongest predictive power of the dependent variable.

6 Empirical Results

The classification of a university's economics department being overrated is anything but random chance. Many factors play a role in an economics department being classified as overrated. Of all of the economics departments in the world, those in the top 50 are in a unique subgroup, many of which are a branch of a university that has copious amounts of prestige, wealth and perceived success. There is a clear conflict between the criteria by which professionals and students evaluate an economics department. Where the professionals care more about activities outside the classroom, such as publications, students care more about different aspects, such as how the professors interact with them inside the classroom. There are many variables at play in the classification of whether a department will be classified as overrated.

As a reminder the population, regression equation in this research is as follows:

$$Overrated_i^* = \beta_0 + \beta_1 Depfund_i + \beta_2 Pages_i + \beta_3 Qual_i + \beta_4 Diff_i + \epsilon \tag{13}$$

recall again that $Overrated_i^*$ represents the economics department at university "i"'s desire, or ability to ride on the prestige of the school's name or reputation, rather than making it an adequate academic environment. However, desire and ability can not be directly measured, so we turn to sample regression equation:

$$\widehat{Overrated_i} = \widehat{\beta_0} + \widehat{\beta_1} Depfund_i + \widehat{\beta_2} Pages_i + \widehat{\beta_3} Qual_i + \widehat{\beta_4} Diff_i \tag{14}$$

As a reminder, we assume that if an economics department has the desire to "ride" on the prestige of the school's name, it is overrated. That is the same as stating that if $Overrated{*}_i > 0$, then $Overrated_i = 1$. The same pattern holds for economics departments who do not have the desire to "ride" on the prestige of their university's name; if they do not, then they are non-overrated. That is the same as stating that if $Overrated{*}_i \leq 0$, then $Overrated_i = 0$. After estimating the above sample regression equation using MLE applied to the probit model, we obtain:

$$\widehat{Overrated_i} = -29.67 + 0.19 Depfund_i + .73 Pages_i - 0.19 Qual_i + 0.61 Diff_i \tag{15}$$

Recall that the slope coefficients are not the estimated marginal effects. In the Probit model, the estimated marginal effects vary for each regressor based on the given school.

In terms of evaluating the significance of the regressors, please refer to Table 9 below:

Table 9: Regressors and P-Values	
Regressor	**P-Value***
Depfund	0.09
Pages	0.01
Qual	0.03
Diff	>0.00
*Ho: each regressor independently has no explanatory power over the dependent variable	

The null hypothesis which accompanies each of these regressors listed in Table 9 is that a regressor, such as quality of instruction ($Qual_i$) is statistically insignificant when estimating whether or not an economics department is overrated. A p-value less than 0.10 indicates that the respective regressor is statistically significant at the 90% confidence level. We can observe that all the above regressors are statistically significant. It is important to note that these are the P-values before heteroskedasticity is tested for.

Before testing for heteroskedasticity, which means that the variance error term is not constant, we must make sure that we are working with a good fitting model. When evaluating which probit model fits the data best we use the following statistical information: Akaike Information Criterion (AIC), Schwarz Criterion (SC), Aldrich-Nelson's R^2, McFadden's R^2, and McKelvey-Zavoina's R^2. Additionally, the Hosmer-Lemeshow test is used to determine the best fitting model. Another bar that the model must exceed is the percentage correct test, where the model must correctly predict the dependent variable, both when it is equal to zero and equal to one, at least 80% of the time.

To see which model is the best fitting, three models are considered. The first model is already known, as it was previously stated. Model 1 uses the regressors of Depfund, Qual, Pages and Diff to explain the dependent variable. Model 2 uses only Depfund, Qual and Pages to explain $Overrated_i$. Model 3 uses Depfund, Pages, and Diff to explain whether or not an economics department is overrated. For completeness, the 3 models are listed below.

$$Model\ 1 : \widehat{Overrated_i} = \widehat{\beta_0} + \widehat{\beta_1} Depfund_i + \widehat{\beta_2} Qual_i + \widehat{\beta_3} Pages + \widehat{\beta_4} Diff_i \tag{16}$$

$$Model\ 2 : \widehat{Overrated_i} = \widehat{\beta_0} + \widehat{\beta_1} Depfund_i + \widehat{\beta_2} Qual_i + \widehat{\beta_3} Pages \tag{17}$$

$$Model\ 3 : \widehat{Overrated_i} = \widehat{\beta_0} + \widehat{\beta_1} Depfund_i + \widehat{\beta_2} Pages + \widehat{\beta_3} Diff_i \tag{18}$$

As one can observe, the difference between the models is whether to include $Qual_i$, $Diff_i$ or both.

First, let us consider the differences in the AIC and SC between the three models. This information is illustrated in Table 10:

Table 10: Comparing AIC and SC		
	AIC	**SC**
Model 1	27.18	36.74
Model 2	61.81	69.46
Model 3	32.94	40.59

Recall that alone, these numbers mean nothing, however when there is one or more model to compare it to, they become meaningful. It is best to have the lowest of the AIC and SC out of the three models. With respect to having the lowest AIC and SC, Model 1 is the ideal model.

Table 11: Comparing Various Pseudo R-Squares			
	Aldrich-Nelson	**McFadden's**	**McKelvey-Zavoina**
Model 1	0.51	0.75	0.95
Model 2	0.24	0.22	0.51
Model 3	0.47	0.64	0.90

Once again, in every different measurement of pseudo R^2, Model 1, and its respective regressors have the largest explanatory power as depicted in Table 11.

The next necessary test that a model must pass is the Hosmer-Lemeshow test, where, as stated in equation (7), the null hypothesis is that the model fits the data well. In this respect, failing to reject the null means that it is a good fitting model. For the sake of the purposes of evaluation, we will use the 90% confidence interval. For the three models, the respective p-values are indicated in Table 12:

Table 12: P-values* for Hosmer-Lemeshow Test	
Model 1	0.94
Model 2	0.34
Model 3	0.85
*Ho: the fit of the model is good	

Note, all of the models fail to reject the null hypothesis, so with respect to only the Hosmer-

Lemeshow test, these models would all be good fitting models. However, if we want to choose the model that fits the data the best, perhaps we would look at the model in which has the largest p-value, which indicates that it is the most unlikely of the three to reject the null. Using this selection criterion, Model 1 is the best fitting model.

Next, we use the LR, Score and Wald test statistics to compare the joint significance of the regressors to observe whether or not they have notable power in describing the dependent variable. The p-values for the 3 test statistics for the 3 models are listed in Table 13 below:

Table 13: P-Values for Test Statistics

	LR	**Score**	**Wald**
Model 1	<.0001	<.0001	0.020
Model 2	0.0015	0.008	0.028
Model 3	<.0001	<.0001	0.001

Similar to the Hosmer-Lemeshow test, all we are looking for here is if we reject the null or not. The null and alternative hypothesis are stated in equations (3) ad (4) above. As we can observe, all of the models have regressors that are jointly significant. Furthermore, we have minimize the probability of a Type-2 error since the Wald test statistic is listed above. All Models work, however if we were to pick the model that has the lowest possible p-values, which indicates how strong we reject the null hypothesis, Model 3 should be used.

The last aspect of these models that should be used for analytical comparison is the percentage of correct predictions. Remember, that a model must be correct at least 80% of the time when predicting both values of the dependent variable. So, it must correctly predict when an economics department is overrated ($Overrated_i = 1$), which is depicted as the term "sensitivity" at least 80% of the time. In addition to this criterion, the model must also correctly predict when an economics department is non-overrated ($Overrated_i = 0$), which is expressed by the term "specificity." This information is stated below in Table 14:

Table 14: Percentage of Correct Predictions		
	Sensitivity	**Specificity**
Model 1	87.50%	92.30%
Model 2	70.00%	66.70%
Model 3	92.00%	87.50%

Immediately, we can eliminate Model 2 as a potential candidate of best fit because it does not meet the 80% correct prediction mark on "sensitivity," for this reason, we are left between choosing Model 1 and Model 3. If we were to pick the model with the highest percentage of correct predictions overall, Model 1 would be the best choice.

Considering all of the information in Tables 10-14, we can conclude that Model 2 is not an option, Model 3 is an acceptable model, but Model 1 is the best possible model we can use for analysis in this data set. This is because not only does it exceed the minimums of the percentage correct test, have statistically significant regressors, and fail to reject the Hosmer-Lemeshow test, but it has the lowest AIC and SC statistics.

Given the fact that the best fitting model is being used, we can proceed with confidence, especially with respect to the estimated marginal effects (EMEs) of the regressors. Recall, the parameter estimates are not the marginal effects. In the Probit model, EMEs for each regressor are unique to the specific economics department at hand.

When evaluating the EME of the regressors, it is important to look at both overrated and non-overrated economics departments, to observe a more complete picture of the data. We will consider 2 departments from each category. From the overrated departments, Stanford and University of California - San Diego will be considered, and Northwestern and Iowa State will be analyzed at from the non-overrated group.

To begin, let us consider the two chosen overrated departments. First, Stanford University, who has the following values for its regressors in Table 15 and the estimated marginal effects of its regressors in Table 16.

Table 15: Regressor Values for Stanford University				
Overrated	**Depfund**	**Pages**	**Qual**	**Diff**
1	4.06	5.42	59.00	59.00

The associated EMEs for Stanford University are as follows:

Table 16: Estimated Marginal Effect of Regressors for Stanford University			
Depfund	**Pages**	**Qual**	**Diff**
0.07	0.28	-0.07	0.24

The interpretation for these EMEs are as follows, if Stanford's Economics Department, or others similar to it, receive an additional one million dollars in funding, its probability of being overrated increases by 7%, on the average. If this department, or others like it publish an additional 1,000 pages, it is 28% more likely to be classified as overrated, all else constant. In terms of quality of instruction at this university, or others like it, if it increases by 1%, it is 7% less likely to be classified as overrated, on average. Also, with respect to difficulty, Stanford, or other colleges like it, if difficulty increases by 1%, it is 24% more likely to be considered as overrated, ceteris paribus.

Let us note that the EME of $Pages_{Stanford}$ and $Diff_{Stanford}$ have the largest magnitude of influencing the probability of its economics department being overrated, both increasing the chance.

Let us continue with another overrated economics department, University of California - San Diego (UCSD). Here are the regressor values for UCSD along with the respective EMEs:

Table 17: Regressor Values for University of California - San Diego				
Overrated	**Depfund**	**Pages**	**Qual**	**Diff**
1	2.54	2.44	66.80	64.80

Table 18: Estimated Marginal Effect of Regressors for University of California - San Diego			
Depfund	**Pages**	**Qual**	**Diff**
0.06	0.23	-0.06	0.19

The interpretation for these EMEs are as follows, if UCSD's Economics Department, or others similar to it, receive an additional one million dollars in funding, its probability of being overrated increases by 6%, on the average. If this department, or others like it publish an additional 1,000 pages, it is 23% more likely to be classified as overrated, all else constant. In terms of quality of instruction at this university, or others like it, if it increases by 1%, it is 6% less likely to be classified as overrated, on average. Also, with respect to difficulty, UCSD, or other universities like it, if difficulty increases by 1%, it is 19% more likely to be overrated, ceteris paribus.

Once again, let us note that the EME of $Pages_{UCSD}$ and $Diff_{UCSD}$ have the largest magnitude of influencing the probability of its economics department being considered as overrated, both increasing the chance.

Let us now analyze the non-overrated economics departments, starting with Northwestern University, whose regressor values and associated EMEs are depicted in Table 19 and 20 respectively.

Table 19: Regressor Values for Northwestern University

Overrated	Depfund	Pages	Qual	Diff
0	2.39	5.01	76.40	65.80

Table 20: Estimated Marginal Effect of Regressors for Northwestern University

Depfund	Pages	Qual	Diff
0.07	0.29	-0.08	0.24

The interpretation for these EMEs are as follows, if Northwestern's Economics Department, or others similar to it, receive an additional one million dollars in funding, its probability of being overrated increases by 7%, on the average. If this department, or others like it publish an additional 1,000 pages, it is 29% more likely to be classified as overrated, all else constant. In terms of quality of instruction at this university, or others like it, if it increases by 1%, it is 8% less likely to be classified as overrated, on average. Also, with respect to difficulty, Princeton, or other colleges like it, if difficulty increases by 1%, it is 24% more likely to be considered as overrated, ceteris paribus.

The same trend continues even with non-overrated economics departments. Note that the EME of $Pages_{Northwestern}$ and $Diff_{Northwestern}$ have the largest magnitude of influencing the probability of its economics department being considered as overrated, both increasing the probability.

The next non-overrated university is Iowa State University (ISU), where its regressor values and associated EMEs are listed below in Table 21 and 22, respectively.

Table 21: Regressor Values for Iowa State University

Overrated	Depfund	Pages	Qual	Diff
0	3.16	0.15	61.00	65.60

Table 22: Estimated Marginal Effect of Regressors for Iowa State University

Depfund	Pages	Qual	Diff
0.06	0.24	-0.06	0.20

The interpretation for these EMEs are as follows, if Iowa State's Economics Department, or others similar to it, receive an additional one million dollars in funding, its probability of being overrated increases by 6%, on the average. If this department, or others like it publish an additional 1,000 pages, it is 24% more likely to be classified as overrated, all else constant. In terms of quality of instruction at this university, or others like it, if it increases by 1%, it is 6% less likely to be classified as overrated, on average. Also, with respect to difficulty, Georgetown, or other colleges like it, if difficulty increases by 1%, it is 20% more likely to be considered as overrated, ceteris paribus.

Once again, we observe the same pattern with the EME of $Pages_{ISU}$ and $Diff_{ISU}$ having the largest magnitude of influencing the probability of its economics department being considered as overrated, both increasing the chance.

Now we will select 2 economics departments, one overrated and one non-overrated, and put the Probit model to the test by first displaying the values of the regressors, then comparing the model's predicted probability to the actual value of the dependent variable. From the overrated category, Texas A&M (T-A&M) is selected. University of Minnesota - Twin Cities is the department that is selected to the non-overrated group.

Here are T-A&M's economics department's values for the regressors:

Table 23: Regressor Values for Texas A&M University

Overrated	**Depfund**	**Pages**	**Qual**	**Diff**
1	5.19	0.18	59.20	71.60

With these regressors, the Probit model predicts the probability that it will be overrated, this predicted probability is compared to its actual probability in Table 24 below:

Table 24: Predicted Probability and Actual Rating for Texas A&M University

Predicted	**Actual**
0.99	1

As we can observe, in this case not only is Texas A&M's economics department overrated, but the Probit model also predicts that it will be overrated, or extremely close to being 100% overrated. It is because of this fact that we can state that this Probit model correctly predicted Texas A&M University's economics department as overrated.

Next, let us turn to the non-overrated department, University of Minnesota - Twin Cities. Minnesota's values for the regressors are as follows:

Table 25: Regressor Values for University of Minnesota - Twin Cities				
Overrated	**Depfund**	**Pages**	**Qual**	**Diff**
0	7.09	1.93	72.40	58.80

Once again, the Probit model takes these values and computes a predicted probability that the department will be overrated, this information is listed below in Table 26:

Table 26: Predicted Probability and Actual Rating for University of Minnesota - Twin Cities	
Predicted	**Actual**
0.01	0

1% is quite close to 0%, so if one was to gamble on whether Minnesota economics department is overrated, a perceived 1% likelihood of being so would not be high enough for a rational individual to take that bet. Because of the closeness of the two values, we can state that this model has strong predictive power.

7 Heteroskedasticity

7.1 *Theory*

Attention must be paid to the population error term. If the population error term has a constant variance across the independent variables, then the data is said to have homoskedasticity. Homo means same and "skedasticity" fundamentally means variance, so together, homoskedasticity means same variance (in the population error term in this case). With homoskedasticity we obtain efficient parameter estimates. This is the ideal case, because we obtain the correct measurement of standard error, which assists us in calculating the p-values, which will tell us if a regressor is statistically significant in its predictive power of the dependent variable.

If the population error term does not have the same variance across the independent variables, the data is said to have heteroskedasticity. Hetero means different and we know that "skedasticity" means variance, so combining the two results in heteroskedasticity. This is not ideal, because now we would have biased and inconsistent parameter estimates. In turn, this issue may cause a false statistical significance or false insignificance in the model. By this, it means that we will not obtain the true values of standard error, which will directly influence the p-values. It may supply to us a false estimate of the parameter's p-value that is less than 0.10, leading us to interpret that the regressor is statistically significant when it is in fact not.

Given the fact that the heteroskedasticity pollutes the entire model to the point of being unreliable, the next question is; how does check to make sure that our model does not have heteroskedasticity? The most common test for this matter is the Breusch-Pagan (BP) test. Let us remind ourselves what we would be testing for here. We are testing to see if there is non-constant variance of the population error term. The existence of this non-constant variance can be caused by a single, all or a combination of the regressors. Because of this, there is some legwork to be done by the researcher to correctly check all the possible combinations of regressors to see if they cause heteroskedasticity.

The null hypothesis for the BP test is that the model has homoskedasticity. So, if we obtain a BP value that is less than 0.10, we reject the null hypothesis and conclude that the model in fact has heteroskedasticity. If heteroskedasticity is found to exist, it must be corrected using generalized least squares (GLS), and GLS-corrected standard errors are used. It is important to note that in order to obtain the true value of the GLS standard errors, the researcher must know the exact form of heteroskedasticity in the population error term, which is impossible to obtain since the population error terms is unobservable. A better term for an obtainable GLS method is feasible generalized least squares (FLGS). The reason for this is that although it is impossible to know the population error term, it is possible to estimate it, and that is what the "GLS" or FGLS method does. If we discover that we do not need to worry about heteroskedasticity, we can disregard the GLS estimates in their entirety. However, if we obtain a BP p-value that is less than 0.10, we must use the utilize the GLS estimates for how ever many regressors can be attributed to be causing heteroskedasticity in the model. This information begs the question; what are the steps involved in the BP test?

After finding the correctly specified regression model, the next step is to test all the regressors, one-by-one, to see if there is a single variable causes heteroskedasticity. If the p-values are all greater than 0.10, we fail to reject the null hypothesis, and can confidently conclude that the model at hand does not have heteroskedasticity. This is the ideal scenario, however let us consider the case in which we find ourselves having a model with heteroskedasticity.

In the case where we are testing the regressors, one-by-one, and we obtain a p-value that is less than 0.10 for a single regressor, there are further steps to be taken. We reject the null hypothesis, and conclude that the model has heteroskedasticity, but this is just one of the many variables that may be causing it. Next, we test two regressors at a time. Heteroskedasticity may be caused by not only $Depfund_i$, but also $Qual_i$. In any case, after all the variables that are causing heteroskedasticity have been identified, the next step is to obtain the FGLS estimates. After we obtain the FGLS estimates, we have the correct estimated standard errors. Now is when we can consider the p-values for the regressors and make conclusions about statistical significance, as heteroskedasticity has been accounted for.

In the case that the statistical software used does not converge when the BP test is run, it is problematic as heteroskedasticity must be checked for whether software can compile or not. There is no need to worry, because another option exists and can be used as a "back up." This option involves using what are called "robust" standard errors (RSE) to correct for the existence of heteroskedasticity, the RSE supply consistent estimates of the parameters.

How does a researcher know whether to use the "traditional" standard errors, and conclude that the model does not have heteroskedasticity, or to use the RSE and conclude that the model does have heteroskedasticity? This involves comparing both estimated values of standard errors, as well as looking to see if there is any change in significance of the regressors.

If there is little difference between the traditional and RSE, and there are no changes in the significance of any of the regressors, then it is best to use the traditional standard errors and conclude that there is no heteroskedasticity in this model. In the opposite case, it is best to use the RSE for evaluation of statistical significance because there is heteroskedasticity in the model.

7.2 *Applied*

Now that we know that heteroskedasticity can make our model useless due to potentially supplying us with inefficient results that affect the statistical significance of the regressors, we must test for it in our model.

To start, the Breusch-Pagan (BP) test is first implemented. After attempting to run the BP test on the model via statistical software SAS, the program would not converge, because of this, we can not test for heteroskedasticity in this domain using the BP test. Consequently, we are unable to obtain the FGLS estimates and interpret the BP test statistic to determine if the model does have heteroskedasticity. However, this is not a dead end as we can use robust standard errors (RSE) as a backup. This process is executed in the statistical software STATA.

Recall that the way to obtain the answer of whether our data has heteroskedasticity is to generate the RSE and then compare the standard errors as well as the statistical significance of the regressors. Below are two tables, table 27 shows each regressor's standard error and respective p-value using the “traditional” standard errors and table 28 shows the same information using RSE:

Table 27: Estimations using ”Traditional” Standard Errors		
Regressor	**”Traditional” Standard Errors**	**P-value**
Depfund	0.1203	0.12
Qual	0.0819	0.02
Pages	0.2923	0.01
Diff	0.1814	0.00

Table 28: Estimations using Robust Standard Errors		
Regressor	**Robust Standard Errors**	**P-value**
Depfund	0.0675	0.01
Qual	0.0999	0.06
Pages	0.2528	0.00
Diff	0.1383	0.00

From table 27, we can observe that $Depfund_i$ is not statistically significant. Also, when compared to the robust standard errors in table 28, we notice that the “traditional” standard errors are quite different, enough to bring $Depfund_i$ into the statistically significant range. Because the standard errors are noticeable different and that when RSE are used, it changes the significance of a regressor, we conclude that the data has heteroskedasticity and we should use the RSE.

Note that even though the estimated marginal effects (EME) are reported before heteroskedasticity is checked for, there is no issue. In terms of working with data that has heteroskedasticity, it does not affect the EME. So, the respective EME for each of the departments included in the sample, there is no need to re-estimate them because they would not change. Simply put, the EME are the same whether the data have heteroskedasticity or not.

8 Conclusion

There are clear differences in the rankings between students in professionals. They also use different criteria, but does this mean that a researcher should not match them up against each other? This is an excellent question, as it claims that the premise of this project is comparing a banana to broccoli. The banana is a fruit and broccoli is a vegetable, however they are both edible. In this case, student rankings are the result of aggregating the quality of instruction and difficulty of coursework and professionals typically look at the output of a university's economics department, or namely research, however they are both rankings. Matching them up against each other is vitally important.

Where the professional rankings are trusted, there does not seem to be a rival rankings list to serve as another point of reference for students considering attending a specific university. The purpose of this research is not to replace the professional rankings, rather it is to point out the illegitimacy of professionals to take in-the-classroom experience into consideration, because if they did, the two groups would be more in agreement.

As attending a university in the modern age is becoming increasingly expensive, students may want to take extra care when considering what university, they wish to attend. A student must ask them self this single salient question when in the consideration phase; do I value the professional's opinions, or the opinions of the students who attended this specific university, and took the same classes I am likely to take, more? For some, specifically those considering attending graduate school, they may choose to weigh the professional rankings more. Other groups of students, such as high school seniors, are likely to resonate more with student rankings because they are likely to care more about the in-class experience of a university rather than the out-of-class academic activity in which their professors take part in.

In most cases, the absolute value of the regressors of $Pages_i$ and $Diff_i$ were the largest of all the regressors. This has an important implication for schools that are looking to change their classification, or who are looking to strike a more equal balance between in and out-of-class academic activities. The sign on $Pages_i$ is positive and about the same, on the average, as $Diff_i$ indicating that there may be an intertwined relationship between these two variables.

Let us assume that a university's economics department devotes less time to publications, let us say publishes 1,000 less pages, and more time to lessening the difficulty of their course, let us say an adjustment is made that corresponds to a 1% decrease in difficulty, a university such as Stanford University would reduce its probability of being overrated by over 50%.

These implications are noteworthy, as perhaps some universities are ranked highly by the professionals and low by students because they spend too much time on out-of-class academic activities and too little time on in-class academic activities, so in terms of creating a more well rounded program, an overrated university may find it beneficial to strike more of a balance in this regard. Additionally, programs that are non-overrated are perhaps spending too much time on in-class academic activities and too little on out-of-class academic activities. Again, departments in this category could scale their resources to be more balanced and provide a well-rounded feel.

9 References

Bliss, Chester (1934), “The Method of Probits,” Science, New Series, Vol. 79, No. 2037, pp. 38-39.

McPherson, M. (2012), “Ranking U.S. Economics Programs by Faculty and Graduate Publications: An Update Using 1994-2009 Data,” Southern Economic Journal, 2012, 79(1), 71-89.

Morse, R., Hines, K. (2023), “Methodology: Best Social Sciences and Humanities School Rankings,” U.S. News & World Report, April 24, 2023.

National Science Foundation. (2021), https://ncsesdata.nsf.gov/profiles/site?method=rankingBySource&ds=herd , various pages.

QS Top Universities’ Staff. (2021), “Understanding the Methodology: QS World University Rankings,” QS Top Universities, June 8, 2021.

QS Top Universities. (2023), $https://www.topuniversities.com/university-rankings/university-subject-rankings/2023/economics-econometrics?tab = indicators\&sort_by = indicator_70\&order_by = des$, various pages.

RePEc/IDEAS. (2023), “RePEc/Ideas rankings,” RePEc 2023

RePEc/Ideas. (2023), ¡https://ideas.repec.org/top/top.econdept.html¿, various pages.

Times Higher Education’s Staff. (2021), “World University Rankings 2022: methodology,” Times Higher Education, August 26, 2021.

Times Higher Education. (2023, ¡https://www.timeshighereducation.com/world-university-rankings/2023/subject-ranking/business-and-economics¿, various pages.

Timmerman, T. (2008). On the Validity of RateMyProfessors.com. Journal of Education for Business, 84(1), 55–61. https://doi.org/10.3200/joeb.84.1.55-61

U.S. News. 2023, https://www.usnews.com/best-graduate-schools/top-humanities-schools/economics-rankings, various pages.

10 Appendix

Table 3A. The Meister Professional Ranking List							
Meister Professional Rank:	University Name:	Number of Lists Included in:	Exclusion Penalty:	Sum of Rank by list:	Rank + Penalty:	Average Rank:	Tie Breaker (McPerson's list)
1	Harvard University	5	0	7	7	1.4	-
2	Massachusetts Institute of Technology	5	0	11	11	2.2	-
3	Stanford University	5	0	18	18	3.6	-
4	University of Chicago	5	0	20	20	4	-
5	University of California - Berkeley	5	0	22	22	4.4	-
6	Yale University	5	0	38	38	7.6	-
7	Columbia University	5	0	40	40	8	-
8	New York University	5	0	46	46	9.2	-
9	University of Pennsylvania	5	0	48	48	9.6	-
10	Northwestern University	5	0	51	51	10.2	-
11	University of California - Los Angeles	5	0	67	67	13.4	-
12	University of Michigan - Ann Arbor	5	0	70	70	14	-
13	Duke University	5	0	77	77	15.4	-
14	Princeton University	4	51	27	78	15.6	-
15	University of California - San Diego	5	0	78	78	15.6	-
16	Cornell University	5	0	85	85	17	-
17	Boston University	5	0	103	103	20.6	-
18	University of Wisconsin - Madison	5	0	111	111	22.2	-
19	University of Minnesota - Twin Cities	5	0	112	112	22.4	-
20	University of California - Davis	5	0	119	119	23.8	-
21	University of Texas - Austin	5	0	125	125	25	23
22	University of Maryland - College Park	5	0	125	125	25	16
23	Universsity of Southern California	5	0	76	76	15.2	-
24	Michigan State University	5	0	134	134	26.8	-
25	Brown University	4	51	83	134	26.8	-
26	Carnegie Mellon University	4	51	84	135	27	-
27	Dartmouth College	4	51	88	139	27.8	-
28	Boston College	5	0	150	150	30	-
29	Pennsylvania State University - University Park	5	0	160	160	32	-
30	Ohio State University	5	0	164	164	32.8	-
31	Johns Hopkins University	5	0	166	166	33.2	-
32	Washington University in St. Louis	5	0	178	178	35.6	-
33	University of Illinois - Urbana/Champaign	5	0	180	180	36	-
34	University of Virginia	5	0	182	182	36.4	-
35	Texas A&M University	5	0	190	190	38	-
36	University of Rochester	5	0	192	192	38.4	-
37	Arizona State University	5	0	194	194	38.8	-
38	Georgetown University	5	0	198	198	39.6	-
39	University of Washington - Seattle	5	0	199	199	39.8	-
40	Vanderbilt University	4	51	155	206	41.2	-
41	Indiana University - Bloomington	5	0	207	207	41.4	-
42	Purdue - West Lafayette	5	0	209	209	41.8	-
43	University of California - Irvine	5	0	213	213	42.6	-
44	University of California - Santa Barbara	4	51	170	221	44.2	-
45	University of Arizona	5	0	232	232	46.4	-
46	University of Pittsburgh	5	0	252	252	50.4	-
47	Rutgers University	5	0	253	253	50.6	-
48	George Washington University	5	0	258	258	51.6	71
49	Iowa State University	5	0	258	258	51.6	31
50	Rice University	5	0	262	262	52.4	-

Table 3B. The Meister Student Ranking List						
Meister Student Rankings:	University Name:	Quality:	Difficulty:	Quality Rank:	Difficulty Rank:	Rank Points:
1	University of California - Los Angeles	4.32	1.89	1	1	2
2	University of Minnesota - Twin Cities	3.62	2.94	9	2	11
3	Columbia University	3.84	3.15	4	13	17
4	University of Arizona	3.76	3.19	7	15	22
5	Massachusetts Institute of Technology	3.87	3.24	3	21	24
6	Princeton University	3.49	3.11	15	10	25
7	Purdue - West Lafayette	3.55	3.17	13	14	27
8	Brown University	3.59	3.23	10	19	29
9	University of Illinois - Urbana/Champaign	3.35	2.98	24	5	29
10	Northwestern University	3.82	3.29	5	25	30
11	Pennsylvania State University - University Park	3.46	3.14	18	12	30
12	University of Pennsylvania	3.38	3.06	22	8	30
13	University of Pittsburgh	3.35	2.95	26	4	30
14	Cornell University	3.35	3.02	25	6	31
15	Boston College	3.81	3.31	6	26	32
16	University of Chicago	3.90	3.36	2	33	35
17	Rutgers University	3.32	3.06	29	7	36
18	Boston University	3.41	3.22	19	18	37
19	University of Southern California	3.51	3.28	14	24	38
20	Duke University	3.63	3.41	8	37	45
21	Yale University	3.55	3.40	11	35	46
22	University of Virginia	3.48	3.32	17	29	46
23	Georgetown University	3.40	3.32	20	27	47
24	University of California - San Diego	3.34	3.24	28	20	48
25	Ohio State University	3.14	3.09	39	9	48
26	Washington University in St. Louis	3.25	3.20	32	17	49
27	University of California - Davis	3.39	3.34	21	30	51
28	Stanford University	2.95	2.95	48	3	51
29	University of Washington - Seattle	3.24	3.25	33	22	55
30	Harvard University	2.97	3.12	44	11	55
31	George Washington University	3.34	3.34	27	31	58
32	Dartmouth College	3.55	3.75	12	49	61
33	University of Rochester	3.48	3.59	16	46	62
34	Michigan State University	3.22	3.32	35	28	63
35	University of California - Irvine	2.96	3.20	47	16	63
36	University of Michigan - Ann Arbor	3.27	3.40	30	36	66
37	Iowa State University	3.05	3.28	43	23	66
38	Vanderbilt University	3.20	3.36	36	32	68
39	University of Wisconsin - Madison	3.36	3.60	23	47	70
40	University of California - Berkeley	3.26	3.53	31	41	72
41	Arizona State University	3.22	3.43	34	38	72
42	New York University	3.20	3.45	37	39	76
43	University of California - Santa Barbara	3.08	3.52	41	40	81
44	Carnegie Mellon University	3.17	3.56	38	44	82
45	University of Maryland - College Park	3.12	3.55	40	43	83
46	Rice University	2.85	3.37	50	34	84
47	Johns Hopkins University	2.97	3.54	45	42	87
48	Texas A&M University	2.96	3.58	46	45	91
49	University of Texas - Austin	3.06	3.80	42	50	92
50	Indiana University - Bloomington	2.91	3.72	49	48	97