

**Capturing a College Education's Impact on Industry  
Wages Across Time:  
An Analysis of Academic Factors that Affect Earnings**

**Ian Low**

*Professor Kent Kimbrough, Thesis Instructor*

*Professor Peter Arcidiacono, Faculty Advisor*

---

Honors thesis submitted in partial fulfillment of the requirements for Graduation with  
Distinction in Economics in Trinity College of Duke University.

Duke University  
Durham, North Carolina  
2012

## **Acknowledgements**

First, I would like to thank my advisor, Professor Peter Arcidiacono. Not only has been tremendously helpful this past year with my thesis, he has also been a great mentor, counselor, and teacher for the past 4 years here at Duke University. His guidance throughout this process was an integral part of my enthusiasm for writing this thesis.

In addition, I would like to thank Professor Kent Kimbrough and the rest of the Econ 198S/199S class for providing insightful commentary and assistance throughout the year. They truly made my thesis an engaging and exhilarating experience.

## Abstract

Studying how a college education can impact one's wages has always been an area of interest amongst labor and education economists. While previous studies have stressed using single academic factors (i.e. college major choice, performance, or college prestige) to determine the effect on wages, there has not been a focus on predicting wages given industries and a combination of these academic factors across time. Therefore, the crux of my thesis seeks to provide a new model which incorporates college major choice, GPA, industry selection across time, college type (private or public), natural ability (standardized test scores), and several demographic variables in order to predict percent increase/decrease in wages. My results show that college major choice, academic performance, natural ability, and industry selection (together) do have a significant impact on earnings, and they are appropriate measures to predict post-graduation wages.

*JEL Classification:* A2; A22; J3; J31

**Key Words:** College, Wages, Industry

## I. Introduction

The choice between working now and receiving a higher education is a common decision graduating high school students often make. At a higher education level, participating college students expect certain gains from attending universities as opposed to working immediately out of high school. Specifically, college students forgo current earnings in order to achieve higher future earnings. Once in college, students must decide an area of study to focus on. However, there are potentially many factors that these students consider when choosing a major. This can range from pure interest to monetary reasons.

Media sites, such as CNN, The Wall Street Journal, and Yahoo! Finance, report returns on college majors selections. For example, Larson (2010) from AOL's DailyFinance website reports the highest and lowest paying majors. Her findings are reported in Table 1 below. She concludes that the highest paying major (using first year salaries) is engineering (of all sorts), followed by computer science, and economics. This information is meant to display desirable majors that students can choose in order to gain monetary satisfaction; this could supplement/influence major choices and decisions. Thus, based on reports above, it is common for college students to choose majors that are associated with higher earnings.

However, most major/earnings reports only take into consideration the "major choice" variable and apply them to their respective salaries. There are, however, several other important factors that may help to explain the effect of a college education on earnings. Larson's (2010) findings may not account for college prestige and academic performance in her analysis. These factors may be critical in terms of future earnings. Therefore, it is important to understand all aspects of a college education and its correlation to wages.

Table 1: Highest and Lowest Paying Majors

<b>College Major</b>	<b>Starting Median Salary</b>
Engineering	\$59,000
Economics	\$50,200
Physics	\$51,100
Computer Science	\$56,400
Statistics	\$48,600
Biochemistry	\$41,700
Mathematics	\$47,000
Construction Management	\$53,400
Information Systems	\$51,400
Geology	\$45,000
Social Work	\$33,400
Elementary Education	\$33,000
Theology	\$34,800
Music	\$34,000
Spanish	\$35,600
Horticulture	\$37,200
Education	\$36,200
Hospitality and Tourism	\$37,000
Fine Arts	\$35,800
Drama	\$35,600

For instance, the majors/earnings reports never take into account the performance of the students when they take some of the higher paying jobs. Students who outperform their peers in their respective fields of study might have a strong correlation with the industry they go into. Specifically, students who do well in their major studies are likely to enter an industry that is very similar to what they studied in college. With some limitations to how many people can enter an industry at one time, underperforming students may not be qualified enough to enter the most competitive industries. These low performing majors go into lower paying sectors than their high performing counterparts thus pulling down the average salary for their particular major. To this end, this might skew these best paying majors reports. This leads to an inaccurate depiction of

highest/lowest paid majors. At the same time, there might be a correlation with how well a student does in school, but it may not take into account the difficulty of his/her major.

By accounting for the effects of major selection, academic performance, and college prestige, a more accurate picture of the effects of a college education on earnings is likely to emerge. This will also display an accurate correlation to how well a student performs in his/her major and how much he/she can expect to earn in his/her desired industry selection. Thus, in order to have a better analysis of college education on wages, there must be college major choice, academic performance, and college prestige as the primary variables to measure against the return on wages.

In an area that already has vast contributions from distinguished individuals, my thesis argument still remains unique. Previous literature (expanded in Section II) has only focused on one academic factor and various demographic covariates (such as gender and age) and its effect on wages. Past studies have regressed wages on isolated single college education variables, such as majors, performance, ability, and college prestige. However, it is important to consider simultaneously all of the academic factors that may affect a college graduate's wages. By measuring the impact of those variables (for a certain industry) on earnings, there is a unique method of explaining an overall college education's impact on future salaries. This will be a new model that incorporates the above variables (major, performance, ability, college type) and apply them to different industries and measure their impacts on wages across time. Thus, I am hypothesizing that (1) using the stated variables as measurements against wages, these variables are statistically significant in the time series model (meaning all variables are meant to be measured to understand a college education's impact on wages) and (2) that if we hold everything else constant, lower performance results in lower wages in any given industry.

To fully convey my argument, I structured my paper as follows. Section II of the paper discusses the existing literature that has been done on this topic and I will explain how I can use these discussions to help draw a foundation for my thesis. I will also go in depth about my contribution to the education and labor economic fields. I will also discuss the theoretical framework, including using previous regressions/equations, for my thesis. Section III is the data set that I will be using (National Longitudinal Survey of Youth 1997) and a basic analysis/description of what my data set looks like and the motivation for choosing the variables in my regression model. Section V and VI will include the discussion of results from my regressions and what impact these results provide to the labor and education economic fields. The results in Section VI illustrate the importance of measuring academic variables together when predicting wages. Specifically, there is a significant change in wages across time if a college student selected a certain college major, attained a higher GPA, entered a particular industry, and has a certain level of ability. Lastly, Section VII is the conclusion of my study and a baseline for future work in this area.

## II. Literature Review & Theoretical Framework

In order to fully grasp the theory behind returns to a college education, an analysis of investment in human capital and its returns is necessary. Becker (1962) hypothesizes the following: (1) with the increase in age, there will also be an increase in earnings at a decreasing rate and (2) unemployment rates are negatively correlated with skill – meaning that people who are employed are more likely to have less “skill” (Becker 1962). The above arguments Becker contends are essential to understanding how “skill” affects the returns on wages. The term, “skill,” is used to describe additional training in school, on the job, and/or from work experience.

This provides a great foundation for analyzing how college majors, academic performance, and college selectivity play roles in industry salaries. Using Becker's analysis, students who believe that certain college majors are able to provide more skillsets and knowledge as opposed to other majors are more inclined to choose those majors with the expectation of higher earnings.

Becker concludes that "since earnings are gross of the return on human capital, some persons may earn more than others simply because they invest more in themselves" (Becker 1962). Therefore, students who invest more in their education may expect to earn more in the industry they choose to go into. The theoretical analysis of Becker's paper provides a baseline for arguing that the investment in skill (for the purposes of this paper, skill is related to major selection/college education) is positively correlated to higher earnings. This foundation is necessary to understand college major selection and the expected returns to specific industry salaries. To this end, the theory of investment in human capital is necessary to understand the relationship between a college education and industry wages.

Other studies also measure the impact of a college education as an investment for future earnings. For instance, Schultz (1960) proposed "the task of determining whether there are some economic benefits from education that may appropriately be treated as capital that can be identified and estimated" (Schultz 1960). He also analyzed that the "increase in national income is attributable to increase in the stock of human capital and what the 'rate of return' on investment in education has been" (Schultz 1960). Similar to Becker's work, Schultz focused his research on analyzing the economic tradeoff of a college education instead of working immediately. His study is crucial because it suggests that students enter college expecting a higher payoff than going straight into the workforce after high school. Treating a college

education as an investment is similar to treating a college degree as an investment; students expect to pursue a certain degree knowing that there are higher returns.

There have been many empirical models that have measured a component of a college education to wages (or percent increases/decreases in wages). All of the studies above followed a regression model where wage (or the natural log of wage) is the output and independent variables, such as average SAT score for a school, academic performance, or college majors, predict future earnings. In the following paragraphs, I will show the basics of these regressions and how they, combined, contribute to the theoretical background of my empirical method.

From Arcidiacono's et al (2010) he and his colleagues introduced a model that evaluates one's ability and covariate variables, which are race and gender, and how they affect the respective person's wages. The measurement of ability they use is a standardized test called AFQT (Armed Forces Qualification Test), which is "large for college graduates immediately upon entering the labor market, and do not significantly change with labor market experience" (Arcidiacono et al 2010). To compare the differences in ability between college graduates and high school graduates, they incorporated the following model (simplified), where  $X$  is a vector that represents the demographic/covariate variables Arcidiacono et al includes, and  $\gamma$  represents the vector of the respective coefficients:

$$w_i = \beta_0 + \beta_{AFQT}AFQT_i + \gamma X + \varepsilon_i \quad (1)$$

While I plan on using similar variables (AFQT Verbal/Quantitative and GPA) instead of the composite AFQT score to measure ability, the theory behind measuring ability (from education) and its effect on future earnings is consistent with Becker's hypotheses (1962). These ability measures are aspects that are incorporated in the proposed model. In theory, the output wage should be higher with higher AFQT scores. Wages also have a strong correlation with

ability and performance. The above equation provides the necessary theoretical basis for what factors affect wages positively and negatively. By measuring a single variable, such as GPA or another ability variable, we know whether that factor will explain how wages react in different scenarios.

In addition to understanding the theory of investing in human capital, showing and explaining the correlation between attending a more selective college and its payoff is equally important. For instance, Dale and Krueger (2002) sought to describe the relationship between students who elected to attend a more prestigious university and their expected future earnings. While the common belief is that by attending a more selective school, any student can expect to earn more, Dale and Krueger argue that students are more likely to maximize their payoffs by choosing schools that are better fits for their interests and “desired future field of employment” (Dale & Krueger 2002). This means that if a particular school is known for its engineering field, for example, but not as prestigious as other schools, students who are interested in pursuing a career in engineering could potentially be better off by going to that engineering focused school.

Dale and Krueger (2002) use a similar regression model (with different variables), but it analyzes the increase/decrease percentages in wages. The model used was to measure the impact of choosing a selective college and how that affects future income. To define “selectivity” for colleges, Dale and Krueger use the average SAT score at the school the observed subject attends, his/her SAT score, and dummy variables that could affect the log (earnings).  $X_{1i}$  and  $X_{2i}$  are two sets of demographic variables that the admissions committee use to determine admission (such as race, essay scores, etc.) use Below is the regression model that Dale and Krueger use to capture this effect:

$$\ln w_i = \beta_0 + \beta_1 SAT_{j*} + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_i \quad (2)$$

Dale et al (2002) conclude that with this regression model, it does not explain the true effects of percent increase/decrease in wages. There are several other factors (and covariates) that play a part in measuring future income – these include public or private colleges, amount of tuition, and undergraduate class percentile. In summary, the theory behind labor and education economics suggest that there needs to be other variables that we need to analyze to capture a more complete (or at least more applicable) model that incorporates the excluded variables that can be a factor in earnings. Specifically, within Dale and Krueger’s model, college prestige does not always result in higher earnings (although it never hurts). They conclude that students “maximize their payoffs by choosing schools that are better fits for their interests and desired future field of employment” (Dale et al 2002). The underlying education economic theory implies that people do well because they choose to pursue their interests (that interest can be completely monetary reasons). The above equation does suggest that college selectivity does have an impact on net increase/decrease in wages, but there needs to be other variables to fully explain a college education’s impact on future earnings.

While the previous papers discussed the development of student potential and its return on future earnings, there needs to be an analysis of in-school studies that illustrate how college majors and performance affect wages. Arcidiacono (2003) analyzed this in detail: he argues four main points: (1) there is a strong correlation between future income and major selection; (2) “ability sorting across majors occurs both before and during college” (Arcidiacono 2003); (3) more math oriented students typically select the more lucrative majors at each school; (4) there exists a correlation between underperforming students and dropping out/switching to a less lucrative major (Arcidiacono 2003). For example, people who majored in natural science and stuck with it had earnings of \$54,000, whereas the students who originally majored in natural

science, but dropped it had earnings of \$44,000. Thus, there is strong correlation to how well a student performs in his/her academic major and future earnings. Those who are strong performers in a lucrative major have the capability of potentially earning more than the weaker performers in a less lucrative major.

Arcidiacono concludes that “large earnings and ability differences exist across majors. Selection into majors depends upon the monetary returns to various abilities, preferences in the workplace, and preferences for studying particular majors in college” (Arcidiacono 2003). This helps show that the more math oriented majors are more attractive to students who expect high earnings as well as strong performers in these majors are more likely to earn more in the industry they want to pursue. This offers a phenomenal baseline to understanding the effects of underperforming students and their impact on industry wages. Arcidiacono’s work indirectly suggests that the students who underperform can correlate to lower industry wages because these students do not have the same opportunities as those who are high performers in more lucrative majors. This study illustrates how people who underperform in the more lucrative majors switch to less attractive majors, but they can indirectly affect the respective major wages. Therefore, there needs to be a further breakdown on how college majors, academic performance, and college selectivity affects individual industry salaries, so that we can eliminate the bias of having underperforming students switch majors and affecting average major salaries as a whole.

Moreover, analyzing major selection and academic performance only takes into account a student’s ability, but there also needs to an understanding of how a student selects their occupation of their choosing. James (2011) sought to show how (1) high ability workers are more inclined to change occupations simply because they have higher abilities (more flexible), (2) workers may switch occupations that are closely related to their current occupation, and (3)

workers may leave their occupation only to return several years later. This paper hints that people who gain more skillsets have the capability to switch occupations because they have that flexibility. Having that flexibility can indirectly affect future earnings because with different occupations, people can gain even higher skillsets/abilities, which make them more valuable. With that in mind, James offers analysis that with that flexibility (with high ability people), these people have the potential earn more, which could also skew average college major returns. Choosing certain occupations could directly result in higher future earnings, and it is logical to say that people who pursue more lucrative majors and do well in them have the ability to switch occupations.

James finds that “flexibility allowing information to enter the worker’s mobility decision in this manner broadens the type of career moves that occupational matching models can address” (James 2011). Furthermore, his work suggests that “workers can potentially realize large wage gains by finding occupations they are well suited for” (James 2011), and he shows “the importance of sorting on ability in occupational decisions and its effect on wage growth” (James 2011). This helps demonstrate that people with higher abilities, which could result from major studies, have the potential to earn more in certain industries. Upon choosing occupations, people expect certain payoffs and learning experiences that could potentially help them in their next occupation.

Another piece of literature that provides insight on how students choose majors is an analysis of how students perceive their expected returns on college majors. In Arcidiacono, Hotz, and Kang (2011) aimed to survey how Duke University students perceive each individual major’s expected payoff is (Arcidiacono, Hotz, Kang 2011). They find “evidence of sorting on the basis of both expected earnings and ability...[and] students in [their] study are much more

likely to state that they are more able (more competitive) in the major they choose relative to the ones they did not” (Arcidiacono et al 2011). This provides a background on the students’ perspectives on what they expect each major to earn, which strongly influences how students decide on what majors to pursue. With this expectation, students are able to draw certain conclusions on their future earnings potential and which industries they elect to go into.

This is important for my framework because I take into account the selection of college majors and what factors go into choosing majors. Arcidiacono et al (2011) show that people have certain expectations of what sector they want to go and their expected returns 10 years out. For instance, it is interesting to note that almost everyone views education as a sector that has low returns, and people think they have relatively low probabilities of entering the education market. From an overarching theoretical point view, the correlation between choosing a major that relates to higher earnings is high and prevalent, and there needs to be an incorporation of major selection in any college education – wage regression model.

Using the above past economic literature, there is a better understanding of the impact of a college education as a whole on future earnings. While each individual paper hit on specific points that correlate to earnings, I seek to combine all these studies and provide a more uniform model that explains the different factors in a college education that correlate to industry salaries. By separating out average major returns to different industry earnings, we can eliminate the skew in the data. The above studies offer a foundation to understanding the effects of major selection, academic performance, and college type on individual industry wages. Combined, their analysis, both theoretical and empirical, explains the motive to further research the phenomena of a college education’s impact on sector salaries. My contribution seeks to incorporate all the prevalent variables in a college education and its effect on industry wages. Not only will I be

expanding on several of the previous literature data sets and theories, I will use the previous studies as a foundation for a new model and build on the analyses and conclusions these works have done. Building upon the foundation that the above authors provided, my proposed model will take these previous theories and results and further cement their analyses by incorporating a holistic regression that takes into account industry wages and the above education variables. This will be an original contribution as no one has sought to combine college majors, academic performance, and college type into one regression model.

Together, using the above theoretical models, I will combine the variables listed above that are important (and have already been proven) to studying a college education's effect on wages across time. Moreover, it is also important to measure these impacts on percent increases/decreases of wages given a certain industry. The significance of this is to ensure that in a given industry, how wages react to different performance levels, major selections, and college type (public or private schools).

### III. Data

The data set that I collected is from the National Longitudinal Survey of Youth 1997 (NLSY97). This is an economic survey of 9000 people starting in the year 1997 when they were between the ages of 12-16 (they are about 26-30 years old now). I decided to use this particular survey because it is public information and it is easily accessible and they offer all the variables necessary for the proposed model. Arcidiacono et al (2010) uses the same data set for his measurement of AFQT scores, race, and gender and their impact on wages. However, the above previous studies use separate data sets because (1) they were focused on specific impacts within a school (the NLSY97 does not provide school data) or (2) several variables, like SAT scores,

are not available in the NLSY97, and, therefore, have to use different data sets. The NLSY97 surveys the participants on various economic areas. This can range from income to marriage to health surveys. It is important to keep in mind that this is all self-reported information, but we are going on the assumption that this is accurate or near-accurate data. The information that I will be extracting from the NLSY97 survey is the following: (1) the college GPA of the participants, (2) the subjects' wages per hour (the dependent variable), (3) the major/subject the participants elected to study in college, (4) the industries (and in the years following) the participant elected to work for post school, (5) college type (private or public), (6) AFQT verbal and quantitative scores, and (7) 2 demographic variables: gender and age. I included Age and Age<sup>2</sup> to measure whether experience will have a significant impact on wages and if experience is diminishing. The industry variable will be modeled so that in the regression equation, we will compare on an industry-by-industry basis. The dependent variable for this data set will be the log(wage).

The NLSY97 data set offers a year-by-year panel survey for my model. By having an industry, performance, major choice, and respective wages, the data set offers the empirical evidence needed to understand the effects of the above variables on their wages. Instead of a college prestige variable (that was not present in the panel data), there are several other college characteristic breakdowns that may have an effect on earnings, such as public vs. private colleges. Private colleges might have more resources/money to invest in their students – which might have a positive impact on their students' career options. Moreover, private colleges generally have lower faculty to student ratios, which encourages more counseling and focus for each student. This provides a crude proxy variable for prestige that the NLSY97 does not provide. Therefore, I have included an indicator variable for college type to measure the effect of

a student choosing to attend a private institution as opposed to a public school. This is recorded as either 1 (public school) or 2 (private school).

In making sure the data is consistent, I recognize that some individuals may not feel comfortable in responding to these specific surveys. Therefore, I decided not to consider people who did not report any one of the above independent variables and wages. This is a result of missing data that cannot be recovered and therefore must be taken out from my data analysis.

In choosing between the wages and salaries as dependent variables, several factors must be accounted for: (1) people who report salaries might only be reporting their base salaries while people reporting total wages include bonuses and incentives and (2) people may enter the job market much later on in the year, which may result in a lower salary; using wages accounts for the number of hours the subjects work, and using their compensation divided by the hours is fairer comparison. In previous literature, researchers also use wage as their main dependent variable. In order to stay consistent with past studies, using wages as opposed to salaries is a more even comparison, especially when I am using their models as the foundation for my analysis. Moreover, wages also accounts for the number of hours each individual works and is a better representation of how much the person is being compensated for each hour; whereas salary is only the end compensation (i.e. someone could work many hours and will get higher compensation because of that). The participants calculate their total compensation that year divided by the number of hours they worked that year (formatted: \$/hour), and then report their respective wages on the NLSY97 survey. The reasoning behind this is because end of year salaries include the total compensation starting from the participants' starting work dates (which may vary). Wages is a better standardization measurement for earnings because it accounts for

external factors that salaries cannot address. Therefore, wages will be a better dependent variable to measure against to eliminate the biases listed above.

The wages and industries that I have included in my dataset are on a year by year basis. In this panel survey, subjects report their wages from 1999 (or whenever the subjects enter the workforce) until 2009. Using wages eliminates that bias as it standardizes the compensation on an hourly basis. The majority of subjects start reporting their wages and industries starting 2005, which is the time I will start extracting from. Each subject reports his/her wage for that year and I will then measure the changes in wage from year to year as well as the changes in industries. After cleaning up the data so that I only have the people who responded to all of the surveys above (my variables), I have attained 1739 participants/students for the year 2009 as my preliminary group. This reduction (from the 9000 surveyed) is mainly due to the amount of **undergraduate** graduates in the data set. Not all of the original 9000 correspondents that participated in this annual survey attended college, and the purpose of this research is to conduct an undergraduate education's effect on earnings.

In this data, majors are assigned with numbers (ranging from 0-99). Numbers 0 and 99 are listed as people who have not chosen a major; thus, I removed these groups of students as they are not applicable to my regression analysis. The college majors list condenses into numbers from 1-33. Then, I segmented the listed majors into 3 major categories (which are used for my regression): (1) humanities, (2) science/engineering, and (3) business/economics. GPAs are also listed in the hundreds. I divided the GPAs list out by 100 so that they will be in a more readable format (ex. #.##). Lastly, the industry that the participants are working is also given in numbers with corresponding industry codes. I broke these codes down from industries 1-16. Both college

majors and industries are categorical variables, which means each major/industry will be treated as a dummy variable for whether the participant chose that major/industry or not.

The average wage in this group is roughly \$21/hour with a standard deviation of \$49. The average college GPA for this group is also relatively high: 3.22 with a standard deviation of 0.55. Overall, the most popular major is business administration, which represents 22% of the data. Other significant majors include: engineering (5%), biology (3%), education (10%), nursing (8%), and psychology (4%). Other significant trends include education majors having a higher overall GPA than most other majors with an average of 3.40; whereas home economic majors perform the worst overall with an average GPA of 2.94 (frequency greater than 30). The full summary statistics table for college majors can be found in Appendix under Table 6.

In terms of industries, the highest paying industry is the finance field with \$29.05/hour with a standard deviation of \$59.04. The industry with the highest GPA (with a frequency more than 30) is transportation with 3.30 and a standard deviation of 0.51. The worst paying industry is agriculture with \$17.97/hour and a standard deviation of \$12.70. The industry with the worst performance in school is also agriculture with a GPA of 2.95 and a standard deviation of 0.66. The full summary statistics table for these industries can be found in the Appendix (Table 7).

After running a preliminary one independent variable – one dependent variable model (I used overall GPA with wage), I found that there is a lot of uncertainty to this correlation as I only have an  $R^2$  of 0.64%. With more variables included, this correlation will change positively. Similarly, after running a regression for college majors and wages, the correlation is 11%, which is better, but this correlation will be higher as more variables are added to the regression.

Overall, from a preliminary summary statistical analysis, we see that one independent variable by itself is not enough to explain trends in wages. After combining several other

variables with academic performance, we see there is a higher correlation for those variables on wages. This is a step in the right direction as I argue that there needs to be an inclusion of other variables associated with a college education. However, I recognize that there might be a survey error and a limited response rate, since it is relatively sensitive data. To this end, with the data set that I have gathered, we can see that there is an increasing correlation when adding more direct variables into determining wages.

Together the summary statistics breakdown for both majors and industries provides an overall look of how college majors and industries interact with GPA and wages. This could suggest that there is an indirect effect between majors and industries; specifically, students who choose majors with a higher average wage might flock to industries that a higher average pay.

Table 2: Major Average GPA and Average Wage

Major	Average GPA	Average Wage
Biology	3.28	\$18.81/hour
Business	3.17	\$23.05/hour
Computer Science/Engineering	3.26	\$20.00/hour
English	3.15	\$19.78/hour
Physical Sciences	3.22	\$16.57/hour
Psychology	3.30	\$15.86/hour
History	3.22	\$14.69/hour

Table 3: Industry Average GPA and Average Wage

Industry	Average GPA	Average Wage
Agriculture	2.95	\$17.97/hour
Mining	3.16	\$18.93/hour
Manufacturing	3.26	\$28.91/hour
Finance/Insurance/Real Estate	3.22	\$29.05/hour
Professional & Related Services	3.27	\$18.20/hour
Education, health, and Social Sciences	3.21	\$18.75/hour
Entertainment	3.17	\$19.29/hour

For the purposes of my paper, I removed people who decided to go to professional schooling because they skew the wage data and their entry into the work force is much later than the average person who enters the workforce immediately after graduation of undergraduate schooling. The purposes of my data is to measure the effects of wages post-graduation, and going to graduate school provides a potential unfair advantage where they have the potential to earn higher wages and therefore skew the data. To equalize the playing field, there needs to be stratification between those who attended graduate school and those who just attended undergraduate institutions. Moreover, during 2005-2009, people might not have graduated from professional schools, and their compensation (for PhD students) is relatively low. In other words, this panel data is unable to observe them in a late enough stage to see whether they earn higher wages from their professional schooling. More importantly, I only seek to measure the impact of an undergraduate education, not graduate education, on wages.

To stay consistent with Arcidiacono's et al (2010) study, I also included an ability measurement by taking portions of the AFQT (Armed Forces Qualification Test) and standardizing them to see whether there that ability measure affects wages. These AFQT scores control for natural ability, which could affect the students' academic performances. Therefore, it is helpful to understand how natural ability can affect earnings after college. The two scores, verbal and quantitative, were calculating by summing up the Paragraph Comprehension and Word Knowledge sections for Verbal and Arithmetic Reasoning and Mathematics Knowledge sections for Quantitative. In order to standardize these scores, the participant's verbal and quantitative sections is subtracted by the means (of the respective sections) of the panel group divided by the standard deviations of the respective sections (Z-Score). By including a standardized test, it provides another means of calculating performance/ability: measuring

potential in a certain time period (similar to IQ testing). Similar to Arcidiacono's et al (2010) study, someone who has a higher ability level (higher test score) should correlate to higher earnings.

For my full dataset, I included people who reported wages and industries from 2005-2009. This is important for my time series regression (to be discussed in the next section). This gives me a group of 2013 participants (as people may have not reported their wages in 2009, but could have reported earlier). The panel time series dataset takes the year they recorded their industries and wages and transpose them so that participants who recorded information for all 5 years, will be recorded 5 times once in 2005, 2006, etc until 2009. This will measure the effect of post-graduation and how frequently the subjects change industries and how that reflects on wages. Again, I took out people who did not report an industry or wage in any of their 5 years. This provides a total of 8712 observations for my time series regression. The Appendix provides the full summary statistics (including average wage, GPA, and frequencies) results for the year 2009 (Table 6 – College Majors; Table 7 – Industries).

#### IV. Empirical Methodology and Results – Single Year Regression

From the empirical models that I mentioned in my theoretical framework and literature review sections, I am creating a new regression model that uses the previous frameworks as the basis of my empirical methodology and analysis. A combination of the analysis Arcidiacono et al (2010) and Dale & Krueger's (2002) performed showed that the variables of measuring ability/performance, major choice, and college prestige are statistically significant by themselves, but my contribution is to combine those variables in one model and measure the effect on wages given the industry the data subjects chose. Because previous studies (mentioned above) also have used a single year in their regression models, I ran a preliminary model for the

year 2009 to measure the impact when participants graduate and enter their respective industries and what they earn respectively. I used the year 2009 because this is the last year the NLSY97 surveyed its participants and this allows the younger people in the data set to be included in the analysis (as the younger individuals might not have graduated from college if we were to use earlier years). I illustrate this using the below model:

Regression 5 (Results on page 26):

$$\ln Wage_{09_i} = \beta_0 + \beta_1 GPA_i + \beta_2 Major_i + \beta_3 ColType_i + \beta_4 Gender_i + \beta_5 Age_i + \beta_6 Age^2_i + \beta_7 Industry_{09_i} + \varepsilon_{it} \quad (3)$$

- $Gender_i$  represents male or female (female – 0, male – 1)
- $Age_i$  and  $Age^2_i$  represent the ages and the square of the ages of the participants
- $GPA_i$  represents final college GPA once graduated
- $Major_i$  represents major choice (categorical variable: 1 – Humanities, 2 – Science/Engineering, 3 – Business/Economics)
- $ColType_i$  represents college type attended (public or private)
- $Industry_{09_i}$  represents industry selection for year 2009 (List of industry codes in Appendix)

Although the areas of education and labor are not novel, my model is different from previous methodologies as I seek to incorporate all of the other models that only measure a single variable (that is related to college education) and apply it across time. Regression 7 is a cross-section, time series regression, whereas Regression 5 is just a cross-section regression. With a combination of these previous variables, my main time series empirical regression model is below.

Regression 7 (Results on page 34):

$$\ln Wage_{it} = \beta_0 + \beta_1 GPA_i + \beta_2 Major_i + \beta_3 ColType_i + \beta_4 Gender_i + \beta_5 Age_{it} + \beta_6 Age^2_{it} + \beta_7 Industry_{it} + \beta_8 AFQT\_V_i + \beta_9 AFQT\_Q_i + \varepsilon_{it} \quad (4)$$

- $Industry_{it}$  represents industry selection for year t
- $AFQT\_V_i$  represents the standardized **verbal** test scores (participant verbal score minus mean verbal score, all divided by standard deviation of verbal scores)
- $AFQT\_Q_i$  represents the standardized **quantitative** test scores (participant quantitative score minus mean quantitative score, all divided by standard deviation of verbal scores)

$Major_i$ ,  $ColType_i$ ,  $Industry_{it}$ , and Gender variables are categorical/binary variables. Together I want to measure the percent change in wages given the above variables and see whether to play a significant role in determining salaries after graduating from college.

In Arcidiacono et al's (2010) study, economists' created regressions on salary and ability (test AFQT). With my data set, I ran similar regressions to see their individual effects. This one academic independent variable (with gender, age and age-squared as the demographic independent variables), one dependent variable regression instead includes GPA (academic performance) and wages in 2009. These results are reflected in Table 4, Regression 1 (Regressions 1-4 equations are provided in the Appendix) on page 26.

Similar to the conclusions of Arcidiacono et al (2010), attaining a higher performance in college results in having a higher income level. The p-value suggests that the coefficient is statistically significant at a 5% level, which means that there is a strong indication that having a higher GPA is highly correlated to earning a higher salary, while holding all the other covariates constant (Gender, Age, and Age<sup>2</sup>). A further examination of the GPA coefficient shows that a 1.00 increase in grade point average results in an 8.4% increase in salary. Economically

speaking, college graduates who find a higher paying job are highly correlated to having a higher GPA. However, it is pretty rare to find magnitudes of 1.0 GPA differences. A more appropriate transformation is for every 0.1 increase in GPA, salary increases by 0.84%. For example, a person in the finance industry who earns \$29.05/hour and he/she obtained a 3.00 GPA in college, another person in the finance industry who got a 3.10 GPA is predicted to earn \$29.29/hour wage (predicted linear model). To put this in perspective, if everyone works 40 hours a week (normal work week) and 52 weeks in a year, the person who got a 3.0 in college makes \$60,424 as his/her annual salary. With that \$0.24 increase in wage, the person who got a 3.1 in college, with the same work hours and weeks, will earn \$60,932 as his/her annual salary. In the long run, people who academically perform better in college will jump start their careers with higher salaries, regardless of major, industry selection, college type, or ability. This impacts the education economics field because it reinforces the theory that students who do well in college (performance-wise) are seeking larger paychecks out of their first few years of college.

Furthermore, using studies from Arcidiacono (2003) and James (2011) from my literature review section, they have also regressed college majors (as categorical variables) with wages. Likewise to my above section, I regressed the same group's major choice (categories are: 1 = humanities, 2 = sciences/engineering, and 3 = business/economics). The omitted variable is the humanities major and that will be used as the comparison for science/engineering and business/economics majors. These results are reflected in Regression 2 in Table 4 on page 26 (again using my data points for the year 2009).

Table 4: Regression Results for Year 2009

Log(Wage)	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5
GPA	0.0841*** (0.0285)	-	-	-	0.0839*** (0.0284)
Science & Engineering	-	0.0745 (0.0487)	-	-	0.0752 (0.0488)
Economics & Business	-	0.1501*** (0.0347)	-	-	0.1518*** (0.0345)
College Type (Private)	-	-	0.0024 (0.0330)	-	-0.0159 (0.0330)
Mining	-	-	-	0.0097 (0.2509)	-0.0176 (0.2357)
Utilities	-	-	-	0.4517*** (0.1674)	0.4139** (0.1639)
Construction	-	-	-	0.0150 (0.1517)	-0.0160 (0.1410)
Manufacturing	-	-	-	0.0997 (0.1560)	0.0728 (0.1462)
Wholesale Trade	-	-	-	0.2504 (0.1698)	0.2123 (0.1560)
Retail Trade	-	-	-	0.0510 (0.1538)	0.0185 (0.1439)
Transportation	-	-	-	0.1074 (0.1589)	0.0682 (0.1484)
Information & Communication	-	-	-	-0.0209 (0.1859)	-0.0391 (0.1786)
Finance/Insurance/Real Estate	-	-	-	0.2534 (0.1595)	0.2168 (0.1487)
Professional Services	-	-	-	0.0033 (0.1578)	-0.0302 (0.1475)
Education & Health & Social Services	-	-	-	0.0397 (0.1500)	0.0059 (0.1392)
Entertainment Food Services	-	-	-	0.0522 (0.1516)	0.0175 (0.1411)
Other Services	-	-	-	0.0833 (0.1541)	0.0456 (0.1435)
Public Administration	-	-	-	0.1418 (0.1620)	0.1133 (0.1510)
Active Duty Military	-	-	-	0.1606 (0.1896)	0.0928 (0.1887)
Gender (1=Male)	0.0969*** (0.0315)	0.0668** (0.0304)	0.0861*** (0.0311)	0.0886*** 0.0311	0.0801*** (0.0307)
Age	-0.1400 (0.4912)	-0.0180 (0.4858)	-0.0583 (0.4912)	-0.1137 (0.4929)	-0.1569 (0.4863)
Age <sup>2</sup>	0.0030 (0.0285)	0.0008 (0.0090)	0.0015 (0.0091)	0.0025 (0.0092)	0.0033 (0.0090)
Constant	4.0491 (6.5826)	2.6256 (6.5159)	3.2009 (6.5882)	3.8802 (6.6100)	4.2922 (6.5103)
R <sup>2</sup>	0.0123	0.0166	0.0069	0.0203	0.0352

Omitted Variables: Female, Humanities Major, Public Schools, Agriculture Industry

\*.,\*\*.,\*\*\* represents 10%, 5%, 1% significance levels respectively

Number of Observations for each regression = 1739

Regressions 1-4 Equations can be found in the Appendix

In this case, I omitted the humanities major to avoid multi-collinearity bias. In this case, both science/engineering and economics/business variables do have some strong impacts (statistically significant) of predicting the percent increase or decrease in salaries. It seems that both science/engineering type major and business/economics have a positive effect between earnings (compared to humanities). Specifically, choosing a science/engineering type major vs. humanities results in a 7.5% increase in salary versus choosing an economics/business major vs. humanities results in a 15% increase in wages. In translation, using a person who earns \$20/hour from majoring in English (humanities) as the baseline, a person majoring science/engineering can earn \$21.50/hour and a person majoring in business/economics can earn almost \$23/hour. Expanding this to an annual basis, science/engineering majors earn \$3200 more and business/economics majors earn over \$6500 more. In that scenario, the difference of a \$3200 and \$6200 increase in salary can definitely be influential for students to pursue those majors (especially if students have to pay back loans).

This reinforces Arcidiacono's (2003) and James (2011) conclusions that choosing a college major has some monetary impact that people consider. The majors with higher earnings, described as more "lucrative" financially, are more attractive for students to pursue, which explains the reason why science/engineering and business/economics majors have larger frequencies (average 42.12/major = humanities; 47.17 = science/engineering; 201.50 = business/economics). These results also indirectly imply that the larger the expectation of wages, the more people will pursue the majors that have higher wages. However, while business/economics major is statistically significant, science/engineering major is not. This suggests that science/engineering majors may not earn more than humanities majors. This

reinforces the fact that there needs to a model where all the college education independent variables are combined to predict increase/decrease in wages.

Regressing just college type and industries (separately) as the sole academic independent variable also have an insignificant impact on wages. Specifically, while going to a private school has a positive effect on wages, there is little evidence that suggests that this prediction is true. Similarly, most industries (except for information/communication) have positive impacts on wages when being compared to the farming industry, but the results are insignificant. This suggests that measuring the independent effects of these academic variables is insufficient; there needs to be a combined model to capture these effects along with other academic variables. The discussion in the following paragraphs will analyze the college type and industry variables in more detail.

The next step of my regression is to evaluate the dependent variable, wage, with my proposed independent variables: GPA, College Major, College Type, Industry, Gender, Age, and Age<sup>2</sup> (shown on equation 3). Results are reflected in Regression 5 in Table 4 on page 26. Similar to the GPA and college major discussion above, we find that the p-value for GPA is even more significant and college major choice is even more insignificant. The results from running a single variable regression and this model are also similar, suggesting that choosing a certain major and attaining a certain GPA will more than likely result in higher earnings. Moreover, while adding covariates, Gender and Age, we see that gender has a significant effect, but not age. In addition, whether your college is private or public does not seem to matter when evaluating percent increase/decrease in salaries.

More importantly, when measuring the categorical variables in my industry variable, there are some industries where income plays a significant role and other industries that do not.

This model reiterates the previous point that choosing a major that the media lists as “highest paying major” does not necessarily translate into automatically suggesting earning higher salaries. In fact, what seems to be more important is how well you do in college (in terms of academic performance) and which industry you select to go in. While the p-values of the variables may not be ideal, the small standard errors of each variable means that there is some significance of the academic factors, and it is more of a power issue. This is further exemplified in the high magnitudes of the coefficients. The high magnitudes imply that the academic variables still have somewhat, but not strong, significant impact on wages.

Interestingly, the more variables I added into the regression gives GPA an even higher significance level. Statistically, this results from adding variables that often go with GPA in a regression (such as major, earnings, and college type). This altogether gives a better explanation for GPA – increasing its p-value signals that the GPA variable is important to the model. This means that when evaluating starting salary from a college graduate, including all these variables helps explain why a 0.10 increase in grade point average results in a 0.84% increase in income. When adding in college major, college type, and industry variables, there is a higher significance level with GPA. Moreover, the coefficient in this regression is larger, meaning there is a larger effect on wages when a person selects a certain type of major and industry, and achieves a higher GPA. Altogether, the combination of these variables clearly proves one of my hypotheses: that having a higher GPA results in higher income. These results show that doing well in college and choosing the right industries that interest the students gives a higher percent increase in income. We concluded that just choosing a major and industry for its financial benefits (from media sites) is sufficient in obtaining a higher salary. Moreover, going to a private college does not imply higher income. This makes sense, as it is ultimately up to the student how he or she makes of the

resources that his or her school provides for its students. All these factors must be included to evaluate how a college education can impact a student's starting salary. With a higher GPA, choosing certain industries may result in lower income levels. For that reason, someone with a lower GPA who enters an industry that statistically drops income level (from my regression), may be set up not to earn as much as their peers. This may imply that people with higher GPAs might not want to pursue an area that does not result in a percent decrease in starting salary, which means that people with lower GPAs pursue those industries since there is an unmet demand. In conclusion, there is much more that to be studied to figure out the true implications as to why certain industries and majors may not automatically mean higher salaries.

However, the major variables become less statistically significant with the addition of my proposed variables above. While the coefficients for science/engineering and business/economics are very similar to running a regression with just wages as a dependent variable and majors as the independent variables, selecting certain types of college majors and industry with a higher GPA has less of an impact of just choosing a major and evaluating its individual impact on wages. The indirect effect in play (or lack thereof) shows that choosing certain majors will set one up to earn higher wages. Nonetheless, compared to the humanities majors, choosing a science/engineering or business/economics degree still have a percent increase on wages (about 7% and 15% increase respectively). Also, this is for the year 2009, where the job market was particularly harsher than previous years. These outside factors, such as job economy, may influence the way students go about searching for employment. Having a negative job economy can also push students who majored in areas that traditionally have higher rates of going into graduate school go to professional schools to delay their job market entry. Therefore, my time series model will capture that effect and measure whether changes in industry across time

(measured in years) will affect wages given the students' final college major and GPA. Thus, while the science/engineering major variable became less significant, there is still a strong indication that choosing certain majors is more financially "lucrative" than others.

However, the industry variables that people reported on the surveys in the year 2009 are more alarming. Of the 16 industries that were reported, only 1 had a significance level that achieved the 5% level. An explanation for this is that when measuring whether industries have an impact on wages, it is better to see whether changes between industries affect wages. Depicting a model where industries and wages are only represented for a single year may not be the best explanation for this regression. Therefore, these industry variables will be further analyzed in my time series results in the next few pages. Outside of significance levels, the industries that have the largest impacts on wages are utilities and finance (41% and 22% increase in wages respectively when compared to the agriculture industry). This means that while achieving a higher GPA and choosing certain majors are important, entering particular industries may influence wages in a very positive way. Using Becker's (1962) analysis, people with higher skill sets should, theoretically, achieve higher levels of pay. Likewise, industries that require more human capital and less working capital require people who have higher abilities, and therefore, should pay at a higher wage rate. This is represented in the regression model: industries that are usually high-skilled generally have a higher percent increase in wages.

To this end, combining these college education variables together in one regression for a single year is not the best prediction for increase/decrease of wages. While the statistical significance results in Regressions 1 and 2 are similar to the results from previous literature, measuring the combined effects is subject to many external factors that cannot be accounted for in a single year regression. Single academic independent variables provide the best resource for

predicting wages that year. However, the lack of significance with the industry variables (in both Regression 4 and 5) illustrates that there needs to a combined regression that evaluates college education and industry wages across time. Therefore, there needs to a combined time-series academic variables model, which is reflected in my model (equation 4 – this is the model where I will test my hypotheses).

#### IV. Empirical Methodology and Results – Time Series Regression

To see whether industries play a larger role in determining percent increases/decreases in wages, there needs to be an evaluation of this effect across time (as this will limit the single year bias stated in the previous section). In particular, whether someone, given a final GPA, college major, ability score, and college type, can earn more from different industries (if they switched industries from 2005-2009). This also provides a measure of which industries provide higher wages, while limiting outside biases such as the job economy discussed above). The results for without industry and with industry regressions (Regressions 6 and 7 respectively) are located in Table 5 on page 34 (Regression 6 equation is provided in the Appendix).

It is important to note the differences in coefficients between Regressions 6 and 7. Upon removing the industry variables, there is minimal change between the GPA coefficients in Regressions 6 and 7 while still maintaining high p-values. This illustrates that regardless of industry, GPA is likely to be a factor in how much one makes out of college no matter which industry he/she enters. In other words, having a higher GPA is likely to be more beneficial in attaining higher earnings. Similar to GPA, both standardized test sections also do not change much when removing the industry variables. However, both Science/Engineering and Business/Economics coefficients increased by a significant amount without the industry

variables. This jump of 1.5%, signifies that there is an indirect effect in play between college majors and industry selection. Even though this model predicts that one can earn more based on which majors he/she chooses (conditional on industries), it also means that certain majors can lead one to industries that pay better than others. Statistically, this makes sense as the inclusion of certain industry variables may decrease the weight of the major variables depending on a subject's choice of college major and industry. The indirect effect between majors and industries helps explain the trend of people choosing certain majors to get into certain industries. More importantly, it validates the proposed model as measuring the impact of industries and a college education influences wages significantly.

With the above results for my proposed model (Regression 7), the significance levels of the industry variables are better explained using a time series regression. This is shown in my  $R^2$  as it increases to almost 10% of the data being explained by this regression model. Compared to the 2009 regression results  $R^2$  of 3.5% of the data being explained by that regression, the time series model offers more explanation for my data and therefore, there is more confidence in using the time series model to predict wages ( $R^2$ 's reported in the bottom of Regression 5 on page 26 and Regression 7 on page 34). The coefficients of the industry and major variables also change drastically (by a magnitude of 1-2% increases/decreases) while making most other variables statistically significant. Logically, this makes sense as measuring wages across time and industries will provide more data points and a better explanation of whether GPA, college majors, and ability have lasting effects on wages. Moreover, this is a better prediction model as measuring one year may not account for various external factors, such as job economy. Taking industry wages across time minimizes that bias as fluctuations in the job market will be accounted for in these 5 years.

Table 5: Time Series Regression (Equation 4)

Log(Wage)	Regression 6	Regression 7
GPA	0.0849*** (0.0173)	0.0852*** (0.0169)
Science/Engineering	0.0961*** (0.0266)	0.0836*** (0.0261)
Economics/Business	0.1249*** (0.0230)	0.1082*** (0.0227)
College Type (Private)	-0.0237 (0.0212)	-0.0257 (0.0207)
Mining	-	0.3463*** (0.1220)
Utilities	-	0.4930*** (0.1491)
Construction	-	0.2024** (0.0849)
Manufacturing	-	0.2489*** (0.0835)
Wholesale Trade	-	0.1368 (0.0908)
Retail Trade	-	0.0894 (0.0812)
Transportation	-	0.2287*** (0.0881)
Information & Communication	-	0.1701* (0.0873)
Finance/Insurance/Real Estate	-	0.2463*** (0.0826)
Professional Services	-	0.1877** (0.0816)
Education & Health & Social Services	-	0.1393* (0.0808)
Entertainment	-	0.1292 (0.0812)
Other Services	-	0.1222 (0.0856)
Public Administration	-	0.2148** (0.0864)
Active Duty Military	-	-0.0659 (0.1338)
Standardized Verbal Score	-0.0670*** (0.0143)	-0.0659*** (0.0139)
Standardized Quant Score	0.0437*** (0.0137)	0.0425*** (0.0133)
Gender (1=Male)	0.0846*** (0.0195)	0.0769*** (0.0192)
Age	0.4826*** (0.0583)	0.4484*** (0.0587)
Age <sup>2</sup>	-0.0082*** (0.0012)	-0.0076*** (0.0012)
Constant	-4.6136*** (0.7228)	-4.3010*** (0.7308)
R <sup>2</sup>	0.0808	0.0981

Omitted Variables: Female, Humanities Major, Public Schools, Agriculture Industry  
 \*.,\*\*.,\*\*\* represents 10%, 5%, 1% significance levels respectively  
 Number of Observations for each regression = 2013 Groups or 8712 Observations

Measuring the impact of a college education's effect across time will offer a more holistic view of earnings post-graduation. Therefore, this model offers a better representation of whether a college education has a positive effect over a time span than running an one-year regression (Regression 5 on Page 26).

The demographic variables, Age and Age<sup>2</sup>, are consistent with previous economic theories that Becker (1962) made. Both coefficients are statistically significant at the 1% level, which means that experience is a strong predictor for wages (using Regression 7 on page 34). This makes sense intuitively as more experience generally translates to earning higher wages. In this case, the difference between being 28 and 29 years old is a 44.8% increase in wages. Having one extra year of experience in any given industry will result in higher wages, as the person might gain more skill, human capital, or knowledge in a specific area. Therefore, people who attain higher skill levels often make themselves more desirable (Becker 1962), and will be compensated at a higher rate.

However, there is also a diminishing increase in wages as one ages. The Age<sup>2</sup> variable captures that effect as there is a negative impact between Age<sup>2</sup> and wages. Specifically, the difference between someone who is 28 and 29 years old results in a 0.8% decrease in wages. Economically, as people age, there are diminishing returns to productivity as people past their prime are more focused on retirement and have less room for growth in ability. If there are increasing returns, then there is no incentive for people to leave work (outside of the desire to retire). It makes sense for employers to gradually increase wages, but at a decreasing rate, to properly compensate people that have extremely high skill, but have little room to improve. Therefore, people at a younger age, have higher potential and generally have higher energy than older people.

In the majors categories, both science/engineering and business/economics majors are extremely statistically significant, which differs from the results in Regressions 2 and 5 (page 26) where these majors were only somewhat significant. It is evident that by adding industries and wages across time affects majors in an extremely positive way. Moving to strong significance validates the proposal and contribution of adding the GPA, college type, and industries across time variables into a model. Similar to my discussion above, we can conclude that choosing certain types of majors do have a significant impact on wages, especially if one elects to move across certain industries. Specifically, choosing a science/engineering major or a business/economics major results in a 8% or 11% increase in wages (when being compared to choosing a humanities major). In numbers, if someone who majored in english humanities is earning \$20/hour, choosing a science/engineering degree can potentially increase his/her wage to \$21.60/hour; and choosing a business/economics degree can increase his/her wage to \$22.20/hour. A reason for these increases is because some industries may look for certain types of majors that are high skilled. For example, under a computer engineering job, the hiring manager may look for someone who has a more prominent programming background. This can potentially lead to the hiring manager searching through candidates who majored in computer science or electrical engineering. Because a computer engineering job requires more human capital (knowing different programming languages), it is understandable that people who majored in computer science/engineering can get paid more. However, in the previous regression models (Regressions 2 and 5), because the results were not significant, it cannot be concluded that choosing certain majors had a strong effect on wages. With the addition of measuring industries and wages across time, there is a stronger impact on majors selection and percent

increases/decreases in wages. To this end, the transformation into a time series model helped provide a clearer and stronger understanding of majors selection and its impact on wages.

Likewise to the majors variables, the final GPA variable in this regression not only remains significantly strong, but it also increased in statistical significance. While in Regressions 1 and 8, the GPA variable were already statistically significant, the time series regression model increased the z-score to an even higher level, suggesting that my proposed equation is even further validated. In this regression, a 0.1 increase in GPA represents a 0.84% increase in wages. With a near 1% increase in wages, it shows that academic performance is highly correlated with achieving higher wages, even with a 0.1 increment. Employers associate students who do better academically as people with higher skill. It comes to no surprise that employers often seek students who perform very well academically for any high demanding job. As per the discussion above, people who have higher GPAs are set up with more opportunities to enter more financially lucrative jobs. This is evident in the regression as GPA represents a variable with significant impact on wages. Moreover, this result confirms my second hypothesis: when adding college majors, college type, and industries to the regression, a higher performance in school results in higher wages (visa versa for lower performance and low wages). While previous studies from Arcidiacono et al (2010) show that performance and ability individually have significance, I have shown that even with the added variables listed above, GPA achieves a stronger significance, which, in the end, provides a better explanation for a college education's impact on future wages. Therefore, by earning a higher GPA, one can expect to earn more in the future in a given industry and college major.

Unlike my regression for equation 3 (results in Regression 5 on page 26), most of the industries in my proposed time series regression achieve statistical significance, which now

offers insight as to which industries can earn higher earnings across time. This suggests that people who switch industries for career, monetary, or personal reasons, can expect some sort of impact on their wages/income. Across time, these industry variables provide significant results to my regression (using equation 3 and Regression 5 as the basis for comparison). In the above regression, the omitted industry for comparison is agriculture again. This means choosing any industry code 2-16 (1 being agriculture) across time means the percent change in wage is represented as a comparison to the earnings potential in the agricultural industry. The industries that are statistically significant represent that electing to go to those industries are highly correlated with income level. To analyze this further, I selected to analyze several industries that show significant results: mining/utilities/construction, finance, professional services, and education.

The industries with the highest coefficients that are significant are mining, utilities, and construction (35%, 49%, and 20% increases in wages compared to agriculture respectively). This means that compared to a person working in agriculture and making \$18/hour, people in mining, utilities, or construction make \$24.30/hour, \$26.82/hour, and \$21.60/hour respectively. While these numbers seem obscure at first, further qualitative analysis shows that these results are not that far off. Because mining, utilities, and construction are industries with higher risks in terms of health and life, these industries are inclined reward individuals who do these jobs with higher pay. For example, West Virginia is infamous for its mining industry; however, about 2.4% of all miners experience some sort of injury due to accidents in the mine. This dangerous job has to pay well in order to entice people to join this industry; otherwise, people would not go into the mining caves. The average mining salary is approximately \$80,000 (Bureau of Labor Statistics 2012). In terms of wages (assuming 52 weeks and 40 hour work weeks), this is equivalent to

\$38/hour. While miners can work more than 8 hours a day, this helps explain the 35% increase in wages in comparison to the agriculture industry, which is less risky and dangerous compared to mining/utilities/construction. Although those industries are considered to be hard labor jobs, the degree of danger is at a much higher level than most industries. Therefore, comparing these industries for academic performance and college majors may not be the best reasons as to why people join these particular industries, as there is a level of risk involved as well. Regardless of this reasoning, people who join the mining/utilities/construction industries typically earn more than their agriculture counterparts.

Another statistically significant industry is the finance/real estate/insurance industry. Typically associated as a “lucrative” industry, people in this field achieve higher bonuses, while not as dangerous as the mining industry discussed above. In this model, the finance industry earn a 22% increase in wages compared to the agricultural industry. Again to break this down numerically, if a person in the agriculture industry is making \$18/hour, a person in the finance industry is projected to earn \$22.50/hour. This makes sense as people in finance require higher level of “skill” (i.e. financially modeling, investments, etc) than people in the agriculture industry. As per Becker’s (1962) discussion above, because this industry involves more skill in human capital, compensation will be calculated accordingly. In this dataset in particular, typically people who switch into the finance industry will achieve a higher income level. Compared to the results I found in Regression 5 on page 26, the finance industry variable here is much more significant, and it effects wages in a more meaningful way. Therefore, this result comes to no surprise as people in the finance industry expect to earn higher wages than people in the agriculture industry.

Lastly, the education market is unlike the other industries described above, because it does not achieve statistical significance, yet they do offer a higher wage when compared to agriculture. Because the borderline significance of this variable, I am unable to conclude whether the education industry truly provides an explanation for percent change in wages. This can be driven by the fact that education can range from public school salary to private school salary. Because education is a broad industry, it can vary in earnings, which explains the statistical insignificance. The vagueness in this education is captured in the standard deviation of the people in the education industry: \$25. As one of the industries with the larger standard deviations, it shows that this industry wage can really vary from job to job. This further emphasizes the inconclusive analysis of how being in the education market affects one's potential earnings. Nonetheless, in this model, choosing to enter into the education market results in a predicted 12% increase in wages compared to a person in the agriculture industry. Again, this makes sense because people in education require some level of knowledge to be able to teach. While nowadays being in the education industry may not seem as financially lucrative, there still requires some aptitude compared to being in the agricultural industry. Again using Becker's (1962) theoretical framework, it makes sense that people in education make more than individuals in agriculture. To this end, while the education industry can imply earning a higher wage than agriculture wages, the variable is still not significant, which does not help explain whether the education industry has a strong impact on wages.

In measuring ability, the two coefficients, standardized verbal and standardized quantitative scores, are both statistically significant to the regression. In line with academic performance, this ability measure depicts the amount of human capital and potential the participants have. Similar to Becker's (1962) and Arcidiacono's et al (2010) analyses, ability and

human capital are important measures for any job as well as determining wages. However, in this case, scoring well in the verbal section has a negative correlation with wage. Specifically, scoring 1 standard deviation above the mean will lead to a 6.6% decrease in wages. This is consistent with my findings because most humanities majors require a higher level of verbal skills as opposed to mathematical skills. Evidenced by the major coefficients, a humanities major individual is predicted to earn less than his/her science/engineering and business/economics counterparts. This is also reflected in terms of industry field selection. People who are stronger in their verbal skills (as opposed to quantitative) will more than likely choose fields that are more verbal heavy, such as the education or communication industries. These industries, on average, pay less than the more quantitative heavy industries, such as finance. This suggests a trickle-down effect where people who are stronger in their verbal skills, choose humanities majors, which lead to more verbal heavy jobs. Also, the interaction between scores, majors, and industries further emphasizes the indirect effect between majors and industries as people with lower paying majors might be pushed into lower paying industries. Therefore, people who perform well in their verbal scores (as opposed to quantitative) are likely to earn less than their quantitative oriented classmates.

On the other hand, there is a positive relationship between people who do well in their quantitative section and their wages. In particular, each standard deviation increase from the mean results in a 4.3% increase in wages. Again, this makes sense as industries or jobs that require more math heavy daily functions, usually pay higher. For example, in the finance field, where math is prevalent in day-to-day work, there is a 20% increase in wages compared to the agriculture industry (where math is not the main function of the job). Moreover, majors that are more intensive tend occur in the science/engineering and business/economics fields. The above

results exemplify that as people who study science/engineering or business/economics are predicted to earn more than people who study humanities. Similar to the explanation above, people who achieve higher quantitative scores are more likely to pursue more quantitative oriented majors in college and, therefore, are more likely to enter more quantitative focused industries, which, on average, pays higher than verbal focused industries. This intuition further stresses the interaction between majors and industries, which, in turn, makes it significant to study all these college education variables together. To this end, standardized testing provides a measurement of ability and human capital capacity needed to enter certain industries, and both the verbal and quantitative coefficients results economically make sense given my above trends.

## V. Conclusion

The previous literature has traditionally modeled a single academic variable with several demographic covariates to predict wages. Building on past theoretical work, such as Becker's (1962) *Investment in Human Capital*, these existing studies focus on evaluating single academic factors and their effects on wages. However, there needs to be a model that captures multiple academic variables (along with demographic covariables) and apply them on wages across time. Given student preferences in major and industry choices, it is pertinent to determine which educational variables have strong impacts on earnings. This proposed model distinguishes itself from past empirical work as it incorporates factors that labor and education economists have regressed separate into different models. Specifically, adding industry variables change the weights of the major variables coefficients. This difference in coefficients implies that there is an effect between college majors and industry selection. The effect implies that people who choose science/engineering or business/economics majors are more likely to select industries that are higher paying, which is something that has not been previously studied. To this end, there is a

new way to interpret how one's academic performance, ability, college major selection, college type, and industry selection can affect his/her future earnings.

Therefore, by gathering data from the NLSY97 panel data, I took the participants' responses across time (from 2005-2009) and regressed by both an individual year (2009) and across time (from 2005-2009). Both GPA (Regression 1) and college major (Regression 2) in their respective single academic variable models obtained similar results as the previous literature's conclusions. Similar to the results of Arcidiacono et al (2010) and Arcidiacono (2003)/James (2011), both GPA and college major are statistically significant. Moreover, the positive correlation between GPA and wages signals that my results in Regression 1 is consistent with Arcidiacono et al's (2010) findings. Likewise, choosing majors that are more business/science oriented tend to obtain higher wages, which is similar to Arcidiacono's (2003)/James' (2011) research. However, college type and industries variables (Regression 3 and 4 respectively) produced insignificant results. This validates combining these academic variables into one model; however, because this only takes into account a single year, there are various external factors (such as job economy and timing) that are not accounted for. Thus, there needs to be an evaluation of these variables across time to minimize these biases.

The results from the time series regressed better explained the data (higher  $R^2$ ), and qualitatively, it makes sense to regress industry wages across time as it limits external factors, such as job economy, from skewing the results. Tying in the foundations from previous literature, the time series model (Regression 7) successfully predicts percent increase or decrease in wages while obtaining statistically significant variables. GPA seems to be an important independent factor in determining wages, as no matter which industry and major one chooses, having a higher GPA predicts percent increases in wages. However, there is an indirect effect

between majors and industries as the removal industries increases significantly the coefficients of science/engineering and business/economics majors. This means that choosing the above majors correlate with selecting higher paying industries. Therefore, it is important to regress all these factors across time together to figure out which industries pay higher with certain majors.

Testing my hypotheses with these results show that (1) choosing certain industries and majors, pending on academic performance and ability level display significant impact on wages, and these results are best explained by the time series regression model, and (2) there is a positive effect between GPA and wages as performing better in school will lead to higher wages. These results offer significant findings that build on both labor and education economics. This model helps explain the trends in college education and which factors are integral in getting into higher paying industries. Moreover, it not only reaffirms previous conclusions, the proposed model also brings new insight as to the reasoning behind why certain students now often try to pursue certain majors, attain high GPAs, and gain higher ability levels. Nowadays, a combination between choosing majors, performing well in school, gaining ability, and selecting certain industries are often in the minds of college students. Therefore, it is appropriate to measure these variables together to predict future wages.

While I offer explanations that can help explain the correlation between college education and earnings, there is still much more room to expand in this field. This study is meant to provide new considerations as to how different academic factors impact wages across time. A suggestion to help improve the results of this study is to evaluate on a per university basis. The data set used contains a broad range of individuals that were selected to take the NLSY97 survey. However, because of this macro level analysis, it is hard to determine how specific universities (along with academic variables) play a role in industry wages. Offering a micro analysis will be helpful as

there will be an inclusion of a college prestige variable, which was lacking in this research. Being able to determine whether the prestigiousness of a college impacts wages across time will offer a better understanding of a college education's effect on earnings.

In addition, the last year that NLSY97 surveys its participants is 2009. A natural extension is to find more recent data and compare to the year 2009. Including more current data will also illustrate how the increasing competitiveness of college admissions has affected industry wages. Providing a micro analysis and obtaining more recent data are just two suggestions that will help enhance the findings in this paper. All in all, this study seeks to provide a foundation for future researchers to further improve the conclusions of previous literature. To this end, although my results contribute to an already vast field of labor and education economics studies by incorporating new methods of depicting post-graduation industry earnings, they can offer new bases for further research in the labor and education economics field.

Works Cited

- Arcidiacono, P (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1), 343-375. Retrieved from <http://econ.duke.edu/~psarcidi/arcidimetrics.pdf>
- Arcidiacono, P (2010). Beyond signaling and human capital: education and the revelation of ability. *American Economic Journal: Applied Economics* 2,2(4), 76-104. Retrieved from <http://pubs.aeaweb.org/doi/pdfplus/10.1257/app.2.4.76>
- Arcidiacono, P, Hotz, V, & Kang, S (2011). Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals. *National Bureau of Economic Research*. Working Paper. Retrieved from <http://www.nber.org/papers/w15729>
- Becker, G (1962). Investment in human capital: a theoretical analysis. *Journal of Political Economy*, 70(5), 9-49. Retrieved from <http://www.jstor.org/stable/pdfplus/1829103.pdf?acceptTC=true>
- Bureau of Labor Statistics (2011). Occupational Outlook Handbook, 2010-11 Edition. Retrieved from <http://www.bls.gov/oco/>
- Dale, S & Krueger, A (2002). Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117(4), 1491-1527. Retrieved from <http://qje.oxfordjournals.org/content/117/4/1491.full.pdf+html?sid=edab91de-6a9f-42e8-97da-db79418a015f>
- Ellis, B (2011). Best paying college major: engineering. *CNN Money*, Retrieved from [http://money.cnn.com/2011/04/08/pf/college/best\\_paying\\_college\\_majors/index.htm](http://money.cnn.com/2011/04/08/pf/college/best_paying_college_majors/index.htm)
- James, J (2011). Ability Matching and Occupational Choice. *Federal Reserve Bank of Cleveland*. Working Paper. Retrieved from [www.clevelandfed.org/research](http://www.clevelandfed.org/research)
- Larson, J (2010). 10 Lowest Paying College Majors. *Daily Finance: An AOL Money & Finance Site*. Retrieved from <http://www.dailyfinance.com/2010/05/13/10-lowest-paying-college-majors/>
- Larson, J (2010). 10 Most Profitable College Majors and Highest Paying College Degrees. *Daily Finance: An AOL Money & Finance Site*. Retrieved from <http://www.dailyfinance.com/2010/05/10/the-10-most-profitable-college-majors/>

National Longitudinal Survey of Youth 1997 (1997-2009). *Bureau of Labor Statistics*. Retrieved from <http://www.bls.gov/nls/nlsy97.htm>

Schultz, T (1960). Capital formation by education. *Journal of Political Economy*, 68(6), 571-583. Retrieved from <http://www.jstor.org/stable/pdfplus/1829945.pdf?acceptTC=true>

Appendix

Industry Codes:

- 1: AGRICULTURE, FORESTRY AND FISHERIES
- 2: MINING
- 3: UTILITIES
- 4: CONSTRUCTION
- 5: MANUFACTURING
- 6: WHOLESALE TRADE
- 7: RETAIL TRADE
- 8: TRANSPORTATION AND WAREHOUSING
- 9: INFORMATION AND COMMUNICATION
- 10: FINANCE, INSURANCE, AND REAL ESTATE
- 11: PROFESSIONAL AND RELATED SERVICES
- 12: EDUCATIONAL, HEALTH, AND SOCIAL SERVICES
- 13: ENTERTAINMENT, ACCOMODATIONS, AND FOOD SERVICES
- 14: OTHER SERVICES
- 15: PUBLIC ADMINISTRATION
- 16: ACTIVE DUTY MILITARY/ACS SPECIAL CODES

## REGRESSION MODELS:

$$\text{Regression 1: } \ln Wage_{09i} = \beta_0 + \beta_1 GPA_i + \beta_2 Gender_i + \beta_3 Age_i + \beta_4 Age_i^2 + \varepsilon_{it}$$

$$\text{Regression 2: } \ln Wage_{09i} = \beta_0 + \beta_1 Major_i + \beta_2 Gender_i + \beta_3 Age_i + \beta_4 Age_i^2 + \varepsilon_{it}$$

$$\text{Regression 3: } \ln Wage_{09i} = \beta_0 + \beta_1 ColType_i + \beta_2 Gender_i + \beta_3 Age_i + \beta_4 Age_i^2 + \varepsilon_{it}$$

$$\text{Regression 4: } \ln Wage_{09i} = \beta_0 + \beta_1 Gender_i + \beta_2 Age_i + \beta_3 Age_i^2 + \beta_4 Industry_{09i} + \varepsilon_{it}$$

$$\text{Regression 6: } \ln Wage_{it} = \beta_0 + \beta_1 GPA_i + \beta_2 Major_i + \beta_3 ColType_i + \beta_4 Gender_i + \beta_5 Age_{it} + \beta_6 Age_{it}^2 + \beta_7 AFQT_V_i + \beta_8 AFQT_Q_i + \varepsilon_{it}$$

Table 6: Final Summary Statistics for Major – GPA – Wage Year 2009

<b>Final Major</b>	<b>Mean GPA</b>	<b>Std. Dev. GPA</b>	<b>Mean Wage</b>	<b>Std. Dev. Wage</b>	<b>Frequency</b>
<b>Agriculture</b>	3.36	0.50	23.23	29.64	12
<b>Antropology</b>	3.14	0.50	22.51	32.02	10
<b>Archaeology</b>	2.70	0.00	10.00	0.00	1
<b>Architecture</b>	3.18	0.53	25.79	35.14	22
<b>Area Studies</b>	2.80	0.57	15.49	11.44	4
<b>Biology</b>	3.28	0.53	18.82	12.06	50
<b>Business</b>	3.17	0.54	23.06	48.08	387
<b>Communications</b>	3.14	0.56	18.47	8.43	82
<b>Comp. Sci.</b>	3.27	0.52	17.75	8.69	90
<b>Criminology</b>	3.09	0.58	18.36	9.20	88
<b>Economics</b>	3.21	0.52	41.46	57.54	16
<b>Education</b>	3.39	0.55	16.41	13.35	169
<b>Engineering</b>	3.26	0.56	23.13	12.21	84
<b>English</b>	3.15	0.82	19.78	17.36	30
<b>Ethnic Studies</b>	3.40	0.00	12.63	5.13	2
<b>Fine Arts</b>	3.24	0.55	18.95	37.45	54
<b>Foreign Lang.</b>	3.23	0.74	15.25	5.65	10
<b>History</b>	3.22	0.40	14.69	4.43	30
<b>Home Econ.</b>	2.95	0.39	13.35	4.35	10
<b>Interdisciplinary Studies</b>	3.18	0.57	16.23	8.40	26
<b>Mathematics</b>	3.10	0.73	60.73	141.76	22
<b>Nursing</b>	3.26	0.51	20.55	14.31	139
<b>Other Health</b>	3.18	0.58	27.77	136.29	119
<b>Philosophy</b>	3.54	0.54	14.55	5.39	17
<b>Physcial Sci.</b>	3.22	0.59	16.57	11.14	32
<b>Poli. Sci.</b>	3.25	0.57	33.17	86.93	40
<b>Pre-Dental</b>	3.00	0.00	17.00	12.73	2
<b>Pre-Law</b>	3.49	0.64	12.07	2.91	5
<b>Pre-Med</b>	3.32	0.54	11.67	3.75	5
<b>Pre-Vet</b>	3.04	0.65	16.74	9.52	2
<b>Psychology</b>	3.30	0.51	15.86	8.62	80
<b>Sociology</b>	3.12	0.47	15.01	7.60	56
<b>Theology</b>	3.16	0.74	13.21	6.09	13
<b>TOTAL:</b>	<b>3.22</b>	<b>0.55</b>	<b>21.01</b>	<b>49.20</b>	<b>1739</b>

Table 7: Final Summary Statistics for Industry – GPA – Wage Year 2009

<b>Industry 2009</b>	<b>Mean GPA</b>	<b>Std. Dev. GPA</b>	<b>Mean Wage</b>	<b>Std. Dev. Wage</b>	<b>Frequency</b>
<b>Agriculture</b>	2.95	0.66	17.97	12.70	18
<b>Mining</b>	3.16	0.43	18.93	13.19	13
<b>Utilities</b>	3.40	0.34	24.15	7.23	9
<b>Construction</b>	3.17	0.60	16.41	7.52	116
<b>Manufacturing</b>	3.26	0.56	28.91	126.86	138
<b>Wholesale Trade</b>	3.29	0.52	22.66	19.54	41
<b>Retail Trade</b>	3.21	0.51	24.41	67.26	248
<b>Transportation</b>	3.30	0.51	18.34	9.34	56
<b>Info. &amp; Communication</b>	3.09	0.61	20.01	28.75	33
<b>Finance</b>	3.22	0.53	29.05	59.04	113
<b>Professional</b>	3.27	0.57	18.20	11.45	197
<b>Educational</b>	3.21	0.58	18.75	24.67	356
<b>Entertainment</b>	3.17	0.55	19.29	24.44	240
<b>Other</b>	3.29	0.53	17.93	9.53	89
<b>Public Admin.</b>	3.23	0.54	19.98	16.62	59
<b>Active Duty</b>	3.35	0.57	19.06	8.53	13
<b>TOTAL:</b>	<b>3.22</b>	<b>0.55</b>	<b>21.01</b>	<b>49.20</b>	<b>1739</b>

Table 8: Summary Statistics for Above Regressions

<b>Table #</b>	<b>R<sup>2</sup></b>	<b>R</b>	<b>Number of Observations</b>	<b>Test Statistic (F, Z, Chi<sup>2</sup>)</b>
<b>6 (GPA)</b>	0.0046	0.0678	1739	F = 7.44
<b>7 (Majors)</b>	0.0117	0.1082	1739	F = 11.07
<b>8 (Combined 2009)</b>	0.0352	0.1876	1739	F = 3.77
<b>10 (Time Series)</b>	0.0981	0.3132	8712	Chi <sup>2</sup> = 875.57