Issues in Political Economy, Vol 24, 2015, 48-63

The Impact of Major Choice on Earnings: A Longitudinal Study

Michael Bassett, American University '15

Over the last 30 years, labor economists have looked extensively at the returns to education. This analysis began with comparisons between high school students and college graduates, with most data suggesting a significant increase in income following an undergraduate education. From there, more in-depth studies were conducted, collectively looking at differences in high school, community college, undergraduate, masters, and doctorate students. Without much dispute, the results remained the same—individuals with more education tended to earn substantially more than their counterparts.

Much more recently, scholars began looking at field of study as an income determinant, due to the vast array of majors that exist at the undergraduate level. It quickly became apparent that individuals with similar education levels had at time a rather wide income gap, resulting in a thorough analysis of different majors. Almost universally, the literature agreed that major field of study was statistically significant as well (Finnie and Frenette, 2003; Daymont and Andrisani, 1984; James Lavoie and Finnie, 1999; Montmarquette, 2002). Specifically, students studying fields such as engineering, computer science, and business often earned considerably more than peers in the fine arts and humanities. These results, more often than not, make intuitive sense, as the more professionally-oriented disciplines would naturally lead to higher payoffs down the road.

A major limitation in this literature, however, has been its seemingly exclusive look at short-term payoffs immediately following graduation. Most studies have analyzed graduates' income 1-3 years after completion of their major, leaving the statistical significance of field of study in the long-term up for debate. With this in mind, a statistical study that addresses the issue with concrete, longitudinal evidence would not only add to the literature as a whole, but provide insights for policymakers, economists, and students alike.

This paper aims to find out which major areas of study, if any, continue to impact earnings after several years in the workforce. Rather than looking at earnings immediately following graduation, the data analyzes changes in income from 1997-2011. A data set this extensive gives the participants in the sample several years on the job, providing us with a rich, longitudinal collection of observations.

The empirical model used is an ordinary least squares model, with yearly income being the dependent variable. In order to properly evaluate the effect field of study has on earnings, a wide variety of control variables were included in our model. For instance, prior studies have indicated that work experience plays a critical role in determining one's income, so controls pertaining to prior work were included in the model (Fuller and Schoenberger, 1991). Other independent variables, such as demographic indicators and job industry, were integrated as well. Collectively, this combination of earnings figures, field of study indicators, and control variables provides a well thought out, thorough, and complete model that displays almost all measurable determinants of income.

In the past, a recurring problem faced by economists when examining income determinants has been the inability to include certain key factors. Attributes such as personality, charisma, and networking ability are often deemed unmeasurable, and as a result are excluded from many

empirical models. To account for these factors, the model used includes panel data, which looks at the same individuals over time rather than random assortments each year. This drastically increases the validity of the model, making the results more reliable.

Collectively, the results of such a study are intriguing for a variety of reasons. First off, determining the long-term significance of field of study is an important issue in and of itself. If every field of study remains significant over a 15-year period, then high school students ought to strongly consider attending college, as its financial gains would be not just in the short-term, but in the long-term as well. In the contrary, if field of study is not significant in the long-term, then evidence would suggest that other income determinants, such as work experience and job industry, eventually become much more important than one's major.

From a current or prospective students' perspective, this analysis could potentially provide insight as to what fields to pursue, or which to avoid. Perhaps in the long-term, wages in the engineering field level off, while individuals in the business profession benefit from promotions and higher levels of income. Lastly, from a policymaker's perspective, the results of the study could provide hints as to which fields demand the most attention, and could perhaps influence future spending as well.

Going forward, to truly analyze the longitudinal effects of major choice on earnings, this paper seeks to do the following:

1. Summarize and synthesize the appropriate literature, and properly explain how this study builds off of and adds to the conversation
2. Develop and explain the empirical model, and how it can be used to effectively answer our questions of interest
3. Regress the model, provide the appropriate results, and determine which fields of study, if any, remain statistically significant

I. LITERATURE REVIEW

Before jumping into the empirical model and statistical analysis, understanding and synthesizing past literature in the field of labor economics is of utter importance. While the returns to education had long been an intriguing area of research for economists, it was not until the late 80's that focus began to shift away from high school and college comparisons and towards the analysis of field of study indicators.

One of the first studies that dealt exclusively on college major choice was conducted by Mark C. Berger in 1988, and still today remains an extremely influential article in the field. Berger, a former professor at the University of Kentucky, sought to discover how predicted earnings played a role in major choice. Using the National Longitudinal Surveys of Young Men, a colossal panel-data set that ran from 1966-1978, the scholar extracted men's expected earnings and field of study (which was, for simplicity, broken down into five categories). After the inclusion of crucial control variables, Berger determined that the men in the sample tended to study what led to the most income down the road. Interestingly enough, the men also gravitated towards fields that yielded higher lifetime earnings, as opposed to more tempting careers with high starting salaries (Berger, 1988).

Following the publication of this article, among others, interest in the topic drastically increased. With the clear increase in college and university attendance, the payoffs of higher education suddenly became a much more important and practical issue, and a plethora of studies were conducted as a result. Following suit with Berger, a variety of scholars began breaking down major choice into broad categories as to simplify their analysis. Using rich data sets, often with thousands of participants, labor economists attempted to determine which major areas led to the highest payoffs.

One such study, led by Canadian researchers Ross Finnie and Marc Frenette, was particularly thorough in examining this issue, and served as a key inspiration for this paper. Finnie and Frenette's experiment used the National Graduates Surveys to look at three small groups, or cohorts, of Canadian graduates both 2 years and 5 years after completion of their undergraduate degrees (Finnie and Frenette, 2003). Using 10 broad major categories, ranging from Education to Commerce, and over 30,000 survey participants, the scholars were able to answer their research question with ease.

A particularly unique aspect to the paper, however, was its usage of not one but three econometric models to analyze the situation. Finnie and Frenette's first equation served as a base model, only including the dependent variable of income and field of study indicators. From there, a second and third model was constructed, which include a myriad of control variables, ranging from work experience and self-employment indicators to industry and geographic location (Finnie and Frenette, 2003). This gradual evolution from simple to complex was very easy to understand, intuitive, and effective, and as such, a similar tactic was implemented in this paper's analysis as well.

After running their regressions, Finnie and Frenette's experiment suggested that the health, science, and technology fields led to the highest payoffs. Furthermore, students studying education, business, and economics were around the middle of the spectrum, and arts and humanities majors finished near the bottom (Finnie and Frenette, 2003). These conclusions were not uncommon—in fact, other studies in the United States and Canada indicated that students studying the hard sciences and engineering typically earned much higher than their counterparts, while individual's studying philosophy, literature, and similar majors earned much less (Lavoie and Finnie, 1999; Rumberger and Thomas, 1993; Rumberger 1994).

In addition to these types of studies, economists also sought out answers to different, but equally intriguing, questions. Using another longitudinal study, the scholar Estelle James, for instance, determined that not only a student's major mattered, but their entire college experience. In the scholars' sample, GPA, courses taken, and club involvement collectively explained more of the variance in earnings then family background and college characteristics combined (James 1989). In another study, Temple University professors Thomas Daymont and Paul Andrisani sought to further analyze the gender gap in earnings, which many assume is a result of gender discrimination. Interestingly enough, the scholars determined that job preferences explained a significant portion of the gender gap, and that the public's thoughts on gender discrimination were for the most part exaggerated (Daymont and Andrisani, 1984).

In large part, all of the aforementioned studies were able to analyze their questions of interest with success. However, the vast majority of the literature in this topic has one major limitation; an overwhelming emphasis on short-term earnings, with little to know analysis on the long-term.

One notable exception to this claim was a study conducted by Rex Fuller and Richard Schoenberger in 1991. This experiment, not unlike Daymont and Andrisani's, spends extensive time examining the gender salary gap and major, but in contrast, also looks at academic achievement and internship experience (Fuller and Schoenberger, 1991).

Somewhat expectedly, the data in the study suggests that women earn less than men upon entering the workforce. However, once people are granted an initial job opportunity, the researchers claimed that on-site performance, work experience, and even luck begin to play a much bigger role (Fuller and Schoenberger, 1991). This notion, that over time job performance eventually becomes more significant than college coursework, is quite interesting, and is an idea that will be addressed at length throughout the course of this paper.

Overall, it is obvious that an extensive amount of literature analyzes the returns to education as well as the importance of one's field of study. However, there are two limitations to the literature that are quite significant. First off, little to no literature exists that uses 21st century data. This study will address that problem directly, by using a data set with observations as recent as 2011 to determine changes in income in the 2000s. The second, and perhaps more important, limitation has been the aforementioned lack of long-term analysis, perhaps due to the limited amount of data sets available. This issue is tackled in this paper as well, with a large scale data set covering 15 years.

Moving forward, this paper's questions of interest remain quite clear. Using panel-data, this study aims to examine which fields of study, if any, remain significant over a large period of time. To do this, an elaborate ordinary least squares model with a variety of controls is used, with income serving as the dependent variable. Consequently, this analysis will not only test the importance of fields of study, but also inform us of the most significant income determinants over time.

II. DATA

For this study, the National Longitudinal Survey of Youth 1997 (NLS97) was the data set of interest, and was looked at and used extensively throughout this paper's analysis. This survey, which was conducted from 1997-2011, is extremely vast, with thousands of variables at play and a collection of over 130,000 observations. Roughly 9,000 participants aged 12 to 16 were involved in an in-depth screening and questionnaire at the start of the data collection. In 1997, parents were also questioned to determine family background characteristics, demographic indicators, and other variables of interest. From 1998 onwards, the original participants were interviewed once a year, with the process often not taking more than one or two hours (United States).

As mentioned, participants were asked a wide range of questions that span a myriad of disciplines. However, the NLS97 was primarily built to display the participant's transition from high school (and likely college) into the workforce, and from there into adulthood. Consequently, questions and variables in the data set typically dealt with labor market indicators, such as job industry, educational attainment, and the like. In addition to these indicators, the NLS97 includes supplementary information as well. These variables cover a range of topics, such as criminal behavior, drug use, and even dating history, making this data set truly massive and quite extraordinary (United States).

With regards to this study, the NLS97 was able to provide a variety of variables that pertain directly to the research question. The core of this paper is the goal to analyze the relationship between major choice and income, and both of those variables are present in the data set. For this paper, income generally refers to earnings the participant received during the last year. To simplify the interpretation of the model, the natural logarithm of income is used in all regressions, resulting in a log-linear model.

To simplify the analysis of participants' major choice, all majors selected were broken down into eight categories. These are: Arts and Humanities, Social Science, Hard Science, Medical and Natural Science, Education and Social Work, Business, Political Science & International Relations, and Other Majors. For further detail on this breakdown, Figure I conveniently lists which majors were included in each field.

Figure I: Field of Study and their Corresponding Majors

Field of Study	Corresponding Majors
humanities	English, Fine and Applied Arts, Foreign Languages, History, Philosophy, Theology, Liberal Arts
socsci	Anthropology, Archaeology, Area Studies, Psychology, Sociology, Ethnic Studies, Economics, Criminology, Home economies, Geography
sci	Computer Science, Engineering, Mathematics, Physical Science, Architecture
med	Other Health, Pre-dental, Pre-medical, Pre-veterinary, Nutrition, Agriculture, Biological Sciences
edu	Social Work, Nursing, Human Services, Education, Communications
business	Business Administration
psir	Political Science, Pre-law, International Relations
othermajor	Electrical/Repair, Automobile, Hotel/Hospitality, Interdisciplinary Studies, No major yet, Other, Uncodable

Besides the inclusion of these more obvious variables, it is of utter importance to include control variables in the analysis. These controls, of which there are several, can essentially be broken down into three broad categories. The first group of variables deals with basic demographic indicators; namely, age, age squared, gender, and race. These controls are commonly considered key determinants of income, and as such play a pivotal role in the upcoming empirical analysis. Age squared was included because the gap in earnings is often much different from year to year. For instance, the difference in income between a 22 and 25 year old may differ quite significantly compared to the difference between a 47 and 50 year old.

This study's second broad grouping deals with the participants' level of experience, both in education and the workforce. Hours spent in the workforce, time spent unemployed, and even weeks in self-employment are all included in this paper's model. Additionally, dummies that account for education level are included; for simplicity, this is broken into participants who have completed high school, an undergraduate degree, and postgraduate study. These education levels are of extreme importance in the regression, as a lack of education may result in lower incomes, and vice versa.

Lastly, the third broad area pertains to the participant's job industry. Ross Finnie and Marc Frenette, along with other scholars, determined the importance of industry, so these variables play a pivotal role in this paper's analysis as well (Finnie and Frenette, 2003; Fuller and Schoenberger, 1991). Due to the rich complexity of the NLS97, there were a staggering 18 industries for participants to choose from. To simplify this, this paper combined several of them, leaving us with just 5 industry dummies for the empirical model. Specifically, the four largest industries, which were Retail, Professional Services, Educational Services, and Entertainment, were left untouched, while the remaining 14 industries were combined into an Other Industries dummy. For more concrete definitions of this study's variables, please refer to Figure II.

In addition to the aforementioned tables, Figure III includes summary statistics for the variables of interest. Fortunately, every variable had 40,000 observations or more, and the total number of observations reached 134,000, providing us with a large assortment of figures and improving the validity of our upcoming analysis. To provide the reader with more basic statistics, each dependent and independent variables' mean and standard deviation are included in the chart as well.

Collectively, the inclusion of the aforementioned variables provides us with a comprehensive model that, in conjunction with panel-data, effectively includes a significant portion of income determinants. From here, the econometric model will be specified, and the statistical analysis and empirical findings follow.

Figure II: Definitions of Dependent and Independent Variables

Variable	Definition
lnincome	How much income the participant received from wages, salary, commissions, or tips from all jobs, before deductions for taxes or anything else during the last year (natural logarithm)
humanities	Dummy to account for humanities field of study
socsci	Dummy to account for socsci field of study
sci	Dummy to account for sci field of study
med	Dummy to account for med field of study
edu	Dummy to account for edu field of study
business	Dummy to account for business field of study
psir	Dummy to account for psir field of study
othermajor	Dummy to account for othermajor field of study
highschool	Dummy to account for participants' completion of high school
undergrad	Dummy to account for participant's completion of an undergraduate degree
postgrad	Dummy to account for participant's completion of a postgraduate degree
age	Participant's age as of interview date
age2	Participant's age as of interview date, squared
female	Dummy to account for female participants
black	Dummy to account for African American participants
workexp	Cumulative hours worked at all jobs from age 20 as of the interview date
self	Total number of weeks participant was self-employed since last interview date
unemployment	Cumulative number of hours participant was unemployed since last interview date
retail	Dummy to account for retail industry
proservices	Dummy to account for proservices industry
eduservices	Dummy to account for eduservices industry
entertainment	Dummy to account for entertainment industry
otherindustry	Dummy to account for otherindustry industry
recession	Dummy to account for years 2008 and beyond

Figure III: Summary Statistics of Independent and Dependent Variables (N: 134,760)

Variable	Mean	St. Dev.	Count
lnincome	1.995	1.720	68,236
humanities	0.087	0.282	64,830
socsci	0.111	0.314	64,830
sci	0.125	0.331	64,830
med	0.165	0.872	64,830
edu	0.165	0.871	64,830
business	0.173	0.378	64,830
psir	0.022	0.147	64,830
othermajor	0.152	0.359	64,830
highschool	0.385	0.487	134,760
undergrad	0.077	0.269	134,760
postgrad	0.011	0.105	134,760
age	21.661	4.625	116,958
age2	490.595	201.368	116,958
female	0.488	0.500	134,760
black	0.317	0.465	99,825
workexp	4129.972	4010.399	41,876
self	2.933	12.063	91,139
unemployment	2.077	5.179	129,030
retail	0.182	0.386	90,559
proservices	0.103	0.304	90,559
eduservices	0.16	0.866	90,559
entertainment	0.204	0.403	90,559
otherindustry	0.351	0.477	90,559
recession	0.267	0.442	134,760

III. EMPIRICAL APPROACH/MODEL

The empirical model implemented in this paper, despite containing over 20 variables of interest, is actually quite simple in nature. As mentioned previously, an ordinary least squares model is used, with the natural logarithm of income serving as the dependent variable. In order to account for unmeasurable factors, panel-data has also been implemented. This ensures that all observations are matched with the appropriate survey participant, resulting in a model that is more accurate and valid. Finally, a random-effect model was used, because the primary variable of interest, major field of study, is unchanged over time.

This approach was used in this paper for two primary reasons. First off, after examining the literature, a variety of studies tended to favor this method (Berger, 1998; Finnie and Frenette, 2003; Daymont and Andrisani, 1984; Rumberger and Thomas, 1993; Lavoie and Finnie, 1999; Fuller and Schoenberger, 1991). This is most likely because an OLS model is not only relatively easy to understand, but often succeeds in displaying key income determinants. Secondly, this

method was implemented because other models seen were quite complex in nature, and would require extensive economic research and intuition to carry out successfully.

Rather than creating just one model, this study splits up the data into four models, each of which builds on the equation before it. The first model (M1) serves as our baseline model, and only includes the natural logarithm of income, major fields, and education dummies. In order to analyze these variables, which take the form of dummies, one must be excluded from the regression equation. With this notion in mind, the social science field and high school education dummy are not included in the model. Consequently, all dummies in each category are compared against the variables excluded. As a simple example, because high school is missing from the regression, we would expect positive coefficients alongside the undergraduate and postgraduate dummy, indicating higher earnings for those participants.

This paper's second model, referred to as M2 henceforth, expands on this baseline model by including demographic indicators. These indicators are added first because the vast majority of literature suggests that they are statistically significant, and ought to be in the long-term as well. Model 3 (M3) continues this pattern, but this time adding in work experience indicators. The richest model and final equation, Model 4 (M4), adds in industry dummies, with the retail industry variable being excluded from our regression equation.

While most data was extracted from the NLS97 with relative ease, this studies empirical analysis was not without its issues. For instance, despite the data set literally containing thousands of variables, there were a few variables of interest that were not found in the data set, and as such were excluded from the regression. These variables include the quality of high school and college education, private and public school dummies, part-time and full-time dummies. In addition to this issue, observations in which there was no income were excluded from our analysis. This leads to an element of selection bias, because participants that are in school or not seeking work were not included in our models.

Fortunately, despite these and other shortcomings (which will be addressed at the conclusion of this paper), four models were successfully regressed and analyzed, with most yielding high R^2s as well. This clearly indicates that, despite the models imperfections, all still serve as valuable equations with accurate insights. Now that the empirical model has been established, we examine the results of the regressions.

(1) $Income = \beta_1 + \beta_2 * Humanities + \cdots + \beta_{10} * Postgraduate\ Dummy$
(2) $Income = \beta_1 + \beta_2 * Humanities + \cdots + \beta_{14} * Black$
(3) $Income = \beta_1 + \beta_2 * Humanities + \cdots + \beta_{17} * Unemployment$
(4) $Income = \beta_1 + \beta_2 * Humanities + \cdots + \beta_{21} * Other\ Industry$

IV. RESULTS

After analyzing and synthesizing the literature, explaining the origin and implementation of the data set, and thoroughly clarifying the econometric model, we can now dig into our statistical findings. After running the first regression, Model 1 appears to reach results not unlike those in prior studies. Namely, the majority of the fields of study, when ran with education level dummies, appear to be significant. For instance, a one unit increase in humanities leads to an income 26.3% less than participants majoring in social science, a result that, while admittedly a

bit extreme, aligns with past literature (Rumberger and Thomas, 1993; Finnie and Frenette, 2003). In contrast to the humanities field, M1 suggests that the hard science, medical, business, and other major fields earn higher than social science. Meanwhile, education and political science/international relations appear to lack statistical significance over a 15-year span.

Besides major areas of study, the education level dummies appear to have a drastic impact on earnings. For instance, M1 indicates that those with an undergraduate degree earn 205.2% more than individuals with high school diplomas. Similarly, participants with a postgraduate degree earn 258.0% more, according to the model. While these overall signs appear to agree with our intuition, it is important to note that M1 yields an R^2 of just 0.1498. This indicates that about 14.98% of the variation in income can be explained by the variation in the dependent variables; in other words, there are other control variables at work here. Fortunately, this study's future models effectively capture these controls.

In M2, for instance, the coefficient of determination (R^2) jumps from 0.1498 to a staggering 0.6833, suggesting that more than two thirds of the variation in the natural logarithm of income is explained by the dependent variables. As the data slowly becomes more accurate in determining income, it appears that field of study becomes less significant, at least in a longitudinal context. This time around, only three of the eight major categories appear to be statistically significant. A one unit increase in the humanities field, once again, leads to a decrease in income (21.0%) compared to social science. The hard science and business fields, in the meantime, benefit from earnings 13.2% and 15.0% higher, respectively.

With respect to the control variables, every single one is determined to be statistically significant. For instance, a one unit increase in age leads to an increase in earnings (of 124.4%), an intriguing result backed up by the literature (Lavoie and Finnie, 1999). Unfortunately, M2 appears to have negative coefficients for gender differences and racial differences, with coefficients of -0.222 and -0.166, respectively. This indicates that, not surprisingly, females and African-Americans earn substantially less than males and whites, the variables that each are compared to. With an R^2 reaching such a high level, it is safe to that this model is quite successful in determining the effects of income. Despite this, there are still even more controls that ought to be addressed, which leads to the third and fourth models.

Moving forward, M3 aims to build off of M2 by adding in work experience controls. Following an intriguing trend set by M2, major field of study is, for the most part, statistically insignificant over the long-term period of 15 years. Unfortunately for humanities majors, this field manages to remain statistically significant, with participants studying these types of majors earning 16.6% less than social science graduates. The Other Majors field, which includes miscellaneous topics such as Automobile and Electric repairs, is also deemed statistically significant.

While not directly related to this paper's research question, it is quite important to note that, similarly to M2, every control variable included in the model is statistically significant at the 1% level. This suggests that, like Fuller and Schoenberger's experiment, when looked at from a long-term perspective, other determinants of income appear to become more important than one's experiences in college (Fuller and Schoenberger, 1991). This model also has a respectable coefficient of determination (0.3316), providing statistical proof that the dependent variables successfully explain changes in income.

Figure IV: Regression Output

VARIABLES	Model 1 lnincome	Model 2 lnincome	Model 3 lnincome	Model 4 lnincome
humanities	-0.263***	-0.210***	-0.166***	-0.138**
	(0.0628)	(0.0480)	(0.0567)	(0.0542)
sci	0.363***	0.132***	0.103*	0.102**
	(0.0571)	(0.0454)	(0.0536)	(0.0516)
med	0.287***	0.0139	0.0511	0.0455
	(0.0543)	(0.0421)	(0.0498)	(0.0480)
edu	0.0455	0.00362	-0.00198	0.0250
	(0.0540)	(0.0417)	(0.0493)	(0.0472)
business	0.235***	0.150***	0.106**	0.0966**
	(0.0533)	(0.0417)	(0.0492)	(0.0472)
psir	-0.258**	-0.0666	-0.125	-0.145
	(0.102)	(0.0793)	(0.0945)	(0.0918)
othermajor	0.459***	0.0818*	0.165***	0.154***
	(0.0555)	(0.0438)	(0.0520)	(0.0500)
undergrad	2.052***	0.331***	0.334***	0.329***
	(0.0216)	(0.0165)	(0.0230)	(0.0230)
postgrad	2.580***	0.630***	0.649***	0.690***
	(0.0501)	(0.0348)	(0.0752)	(0.0754)
age		1.244***	0.934***	0.955***
		(0.0104)	(0.0777)	(0.0787)
age2		-0.0215***	-0.0195***	-0.0205***
		(0.000234)	(0.00170)	(0.00172)
female		-0.222***	-0.190***	-0.136***
		(0.0229)	(0.0270)	(0.0262)
black		-0.166***	-0.141***	-0.143***
		(0.0248)	(0.0296)	(0.0286)
workexp			0.000118***	0.000130***
			(3.84e-06)	(3.95e-06)
self			-0.00232***	-0.00240***
			(0.000806)	(0.000799)
unemployment			0.0122***	0.00987***
			(0.00260)	(0.00253)
proservices				0.0536*
				(0.0281)
eduservices				0.0141
				(0.0236)
otherindustry				0.140***
				(0.0204)
o.recession				-
Constant	1.330***	-14.73***	-9.296***	-9.382***
	(0.0421)	(0.118)	(0.890)	(0.903)
Observations	38,436	32,147	12,579	12,024
Number of participant	4,290	3,457	3,163	3,081

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

Finally, the fourth and most complete model (M4) reiterates the key points from the aforementioned models, and can be seen alongside M1-M3 in Figure IV. Oddly enough, only one field of study is statistically significant at this point in the regression, and it is the Other Major category. Specifically, in the long-term a one unit increase results in earnings 15.4% higher than participants that had studied social science. Not one other major area of study, when regressed over a 15 year period, appeared statistically significant, an intriguing result given the high number of observations (12,024), participants (3,081) and coefficient of determination (0.3629).

Nearly all of the other controls are statistically significant. As with M2 and M3, demographic indicators impact earnings. For example, female participants and black participants earned 13.6% and 14.3% less than males and white, respectively, suggesting an element of gender and racial discrimination over time. Additionally, job experience indicators were shown to be significant as well; a one unit increase in work experience resulted in an increase in earnings of 0.013%.

This paper's industry dummies, for the most part, were not statistically significant in this regression. The Other Industry category was significant and positive, but included a variety of different industries and thus does not serve any practical purpose. The other four industries were not found to be significant, a shocking result that does not make intuitive sense.

Collectively, our results seem to preach one common message: when analyzed and regressed over a long-term period of time, one's choice of major is not nearly as significant as it once was. When looking at the control variables included in the four models, this conclusion is only reaffirmed. In M2-M4, the vast majority of control variables, unlike field of study, were deemed significant. This suggests that external factors, such as job experience, age, hours away from the workforce, and even gender play a bigger role than major. These results, whether expected or not, are undoubtedly quite intriguing, and provide insight to not only policy makers and economists, but the general public alike.

V. CONCLUSIONS AND POLICY IMPLICATIONS

At the onset of this paper, the existing literature mentioned clearly indicated that in the years following graduation, there was an often substantial income gap amongst graduates. With little to no dispute, economist's concluded that one's major choice was extremely significant in determining that income variation. Generally speaking, engineering and hard science majors tended to earn more than their peers, while graduates in the humanities and fine arts fields consistently lagged behind the crowd. Most of the literature had dealt with short-term payoffs, however, leaving major choices' impact over the long-term up for debate.

In the early 1990s, scholars Rex Fuller and Richard Schoenberger teased that perhaps major field was not as significant as we all once thought. Using salary data, field of study indicators, and a variety of controls, they determined that over time, external factors such as job performance and work experience began to play much more significant roles, to the point where college factors began to not matter. This assertion, which was discussed briefly in the literature review, ended up aligning up perfectly with this paper's empirical analysis and results.

Indeed, over a 15 year period the models above hinted at one common truth—for better or for worse, field of study is not significant in the long-term. According to prior literature in the field,

major choice is extremely important 1-5 years after graduation, and this econometric study is not attempting to dispute that claim. However, it ought to be noted that using an extensive data set with over 9,000 participants and 130,000 observations, other determinants of income became much more important as time went on.

This claim is quite intriguing, and can be used to provide insights to today's youth, labor economists, and policymakers. From a potential or current student's perspective, this consequence could be viewed in either a positive or negative light. For those eager to slip past college and work as soon as possible, this paper's conclusions are perhaps a subtle, positive reminder that once in the workforce, with a little luck anything can happen. For bookworms and students with a high grade point average, our conclusions probably serve as a grim reminder that other factors outside of academics are always at play.

From a policymaker's perspective, it is a bit more difficult to draw concrete policy recommendations from the empirical analysis. But to list just one example, the study's emphasis on work experience and job industry indicates that perhaps more funding could be used towards internship preparation, resume critiques, and career seminars, with less of an emphasize on college enrollment. Similarly, more public spending could go towards the alleviation of gender and racial discrimination, which plays a critical role in this paper's regression output. Improvements in those two areas would do the country well both in the short-run and long-term.

Despite the overall success of this study, the econometric model, data analysis, and results were not without their shortcomings. As mentioned previously, capturing every measurable determinant of income is simply an impossible task, and an assortment of variables was not included as a result. A second drawback to the model was its exclusion of participant's whose income value was set to "missing." This means that, for instance, qualified individuals who were not seeking work or were studying at a university were excluded from our empirical analysis, leading to a potentially high degree of selection bias. At the appendix of this paper, a supplementary model that addresses this issue is discussed at length, providing us with one last model in our analysis.

Besides these issues, the extraction of major field was difficult as well. Due to a lack of questions in the NLS97, major field of study was calculated by finding the mode major choice of each participant. Consequently, there is a chance that a collection of participants are paired to the wrong field, which would obviously challenge the validity of our econometric analysis. Lastly, all observations were integrated into the study. Going forward, perhaps taking out certain types of people (say, participants without a degree) would yield significant insights.

Going forward, there are still ample opportunities to add to this field within labor economics. For one, a study that examines majors individually, rather than in groups, could be practical. A study that looks at both short-term data (1-3 years) and long-term data (5+ years) using the same data set would be interesting as well, as the past literature's assertions as well as my own could be tested at once. Lastly, more experiments could be conducted outside of the United States and Canada, in an effort to determine differences across countries and continents.

VI. APPENDIX A - IMPLEMENTATION OF HECKMAN-SELECTION MODEL

As mentioned previously, Models 1-4, while for the most part a success, were unable to capture participants that left income blank when questioned. This meant that any participants not seeking work or studying full-time were excluded from the econometric models, leading to a degree of selection bias when running our regressions. As a supplementary topic, this paper also developed a fifth model that addresses this problem.

This paper's fifth and final empirical model is a Heckman-Selection Model, which deals with selection bias by determining the probability that each participant is employed (and in turn, reports results for income). When constructing the model, there are essentially two key parts. The first is fairly standard, and includes all variables that influence earnings; in the context of this paper, all variables from Models 1-4 are included. As we would expect, the natural log of income is our dependent variable.

The second portion, in contrast, involves variables or factors that impact the chances of being employed, with a new variable, "select", serving as the dependent variable. With this in mind, our self-employment variable, unemployment variable, and industry dummies are taken out of this regression. New factors, marital status and number of children, are also added in, as they in all likelihood influence chances of employment.

After running the regression, it quickly becomes apparent that not much has changed when compared to Models1-4. The other major category remains the only statistically significant major category, once again suggesting that over a 15-year period, major is not as important. All demographic indicators also remain significant, although in this empirical model, black participants tended to earn more than their white counterparts. In fact, being black lead to an income 4.24% higher, perhaps indicating that there was a degree of selection bias previously that favored white Americans. Elsewhere, work experience and unemployment were shown to have a significant impact on earnings, while most industry variables were not. This for the most part aligns with our conclusions in previous models.

Elsewhere, the Heckman-Selection model also determines how different factors influence the likelihood of working. For instance, compared to high school graduates, participants with undergraduate degrees are 16.3% more likely to be employed. Meanwhile, postgrads are 40.2% more likely than their peers, an intriguing result that perhaps indicates a growing importance in education at the graduate level. In addition to these results, our model also showed that females and blacks were 9.54% and 22.0% less likely to be employed, respectively. The female result makes intuitive sense, as pregnancy, child rearing, and family gender roles often decrease the chances of working. Our result with regards to African Americans, however, is a bit surprising, and ought to be studied in the future.

In conclusion, the results of our Heckman-Selection model match up nearly identically with our previous empirical models. Thus, the five models presented in this study agree that in the long-term, one's major field of study does not play a statistically significant role in determining income. Meanwhile, most demographic indicators and work experience variables remain significant, suggesting that factors such as those remain impactful, and maybe even become more impactful, as time progresses. Our new model's overwhelming resemblance to prior models suggests that there may not have been a large degree of selection bias, despite the

exclusion of several observations. This result is quite beneficial, as it ensures the validity of our core models and as such, strengthens their claims as well.

Figure AI: Results – Heckman Selection

	(1)	(2)
VARIABLES	Heckman Ln(income)	select
humanities	-0.0667*	-0.0881**
	(0.0399)	(0.0433)
sci	0.0913**	0.0170
	(0.0378)	(0.0420)
med	0.0834**	-0.0373
	(0.0354)	(0.0384)
edu	0.0564	-0.0329
	(0.0347)	(0.0376)
business	0.0771**	0.0228
	(0.0345)	(0.0383)
psir	-0.108	0.00184
	(0.0683)	(0.0735)
othermajor	0.200***	-0.130***
	(0.0370)	(0.0399)
undergrad	0.0386	0.163***
	(0.0260)	(0.0303)
postgrad	0.213**	0.402***
	(0.0925)	(0.124)
age	0.865***	-0.372***
	(0.107)	(0.122)
age2	-0.0170***	0.00546**
	(0.00233)	(0.00267)
female	-0.0676***	-0.0954***
	(0.0193)	(0.0214)
black	0.0424**	-0.220***
	(0.0215)	(0.0224)
workexp	7.47e-05***	0.000113***
	(4.16e-06)	(4.94e-06)
self	-0.00121	
	(0.000738)	
unemployment	0.00617***	
	(0.00154)	
proservices	0.0553*	
	(0.0288)	
eduservices	-0.0588**	
	(0.0240)	
entertainment	-0.0486**	
	(0.0247)	
otherindustry	0.223***	
	(0.0219)	
marital		0.225***
		(0.0255)
child		-0.0754***
		(0.0150)
Constant	-8.484***	5.730***
	(1.224)	(1.388)
	athrho	lnsigma
	-1.405***	0.0665***
	(0.0368)	(0.00960)
Observations	17,453	17,453

*Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

VII. REFERENCES

Berger, Mark C. 1988. "Predicted Future Earnings and Choice of College Major." Industrial and Labor Relations Review 41 (3): 418-429.

Daymont, Thomas N. and Paul J. Andrisani. 1984. "Job Preferences, College Major, and the Gender Gap in Earnings." Journal of Human Resources 19 (3): 408-428.

Finnie, Ross and Marc Frenette. 2003. "Earning Differences by Major Field of Study: Evidence from Three Cohorts of Recent Canadian Graduates." Economics of Education Review 22 (2): 179-192.

Fuller, Rex and Richard Schoenberger. 1991. "The Gender Salary Gap: Do Academic Achievement, Internship Experience, and College Major make a Difference?" Social Science Quarterly 72 (4): 715-726.

James, Estelle. 1989. "College Quality and Future Earnings: Where should You Send Your Child to College?" American Economic Review 79 (2): 247-252.

Lavoie, Marie and Ross Finnie. 1999. "Is it Worth Doing a Science Or Technology Degree in Canada? Empirical Evidence and Policy Implications." Canadian Public Policy 25 (1): 101-121.

Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. 2002. "How do Young People Choose College Majors?" Economics of Education Review 21 (6): 543-556.

Rumberger, Russell W. 1984. "The Changing Economic Benefits of College Graduates." Economics of Education Review 3 (1): 3-11.

Rumberger, Russell W. and Scott L. Thomas. 1993. "The Economic Returns to College Major, Quality and Performance: A Multilevel Analysis of Recent Graduates." Economics of Education Review 12 (1): 1-19.

United States Department of Labor—Bureau of Labor Statistics. "National Longitudinal Surveys." BLS.gov (accessed October 1st, 2014).