

Spring 4-2011

COLLEGE ATTENDANCE RATIONAL CHOICE: MODELING AND EMPIRICAL ESTIMATIONS

Sergiy Polyachenko

Southern Illinois University Carbondale, sergiy.polyachenko@gmail.com

Follow this and additional works at: http://opensiuc.lib.siu.edu/gs_rp

Recommended Citation

Polyachenko, Sergiy, "COLLEGE ATTENDANCE RATIONAL CHOICE: MODELING AND EMPIRICAL ESTIMATIONS" (2011). *Research Papers*. Paper 99.
http://opensiuc.lib.siu.edu/gs_rp/99

This Article is brought to you for free and open access by the Graduate School at OpenSIUC. It has been accepted for inclusion in Research Papers by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

COLLEGE ATTENDANCE RATIONAL CHOICE: MODELING AND
EMPIRICAL ESTIMATIONS

By

Sergiy Polyachenko

BS in finance
Sumy State University, 2005

A research paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science in Economics

Department of Economics
in the Graduate School
Southern Illinois University Carbondale
May 2011

RESEARCH PAPER APPROVAL

COLLEGE ATTENDANCE RATIONAL CHOICE: MODELING AND EMPIRICAL
ESTIMATIONS

By

Sergiy Polyachenko

A Research Paper Submitted in Partial
Fulfillment of the Requirements
for the Degree of
Master of Science
in the field of Economics

Approved by:

Scott Gilbert, Chair

Graduate School
Southern Illinois University Carbondale
April 14, 2011

AN ABSTRACT OF THE RESEARCH PAPER OF

SERGIY POLYACHENKO, for the Master of Science degree in ECONOMICS,
presented on April 13, 2011, at Southern Illinois University Carbondale.

**COLLEGE ATTENDANCE RATIONAL CHOICE: MODELING AND
EMPIRICAL ESTIMATIONS**

MAJOR PROFESSOR: Dr. Scott Gilbert

Previous findings of economic literature pointed out that there is significant correlation of family characteristics and family background with an individual's rational choice of education. However, variable of abilities was omitted.

This study develops and estimates the model of rational choice of college attendance with respect to agent's individual and family characteristics as well as abilities to meet college requirements.

Paper concludes significant impact of an individual's cognitive abilities on his rational educational choice. Additional finding is that gender of individual has dual effect on his educational choice. First is negative via income mechanism. And the second is positive via motivation mechanism to compensate an income gap.

ACKNOWLEDGMENTS

This research paper would not be possible without help and support of many people. I would like to thank Professor Dr. Scott Gilbert, department of Economics at Southern Illinois University at Carbondale, for his support and supervision on this research paper. His invaluable assistance and guidance on methodological framework of this project was extremely helpful.

Also, I would like to thank docent of department of economic theory at Sumy State University (Ukraine), Candidate of Science in economics Maxym V. Brykhanov for his great help in developing and expressing ideas of this research project as well as his invaluable moral support and inspiration.

I would like to express special thanks to Assistant Professor Dr. Dmytro Hryshko, department of Economics, University of Alberta (Canada) for his help and sharing his experience in collecting statistical data for this research paper.

TABLE OF CONTENTS

AN ABSTRACT OF THE RESEARCH PAPER OF	i
ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
LIST OF FIGURES	v
CHAPTERS	
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 REVIEW OF RELATED LITERATURE.....	3
CHAPTER 3 THE MODEL OF DISCRETE EDUCATIONAL CHOICE.....	7
CHAPTER 4 DATA DESCRIPTION	11
CHAPTER 5 METHODOLOGY OF ESTIMATIONS	14
CHAPTER 6 ESTIMATION OUTPUTS	18
CHAPTER 7 CONCLUSION.....	26
REFERENCES	27
APPENDECIES	
APPENDIX A	28
VITA.....	29

LIST OF TABLES

TABLE 1.....	19
TABLE 2.....	19
TABLE 3.....	21
TABLE 4.....	23

LIST OF FIGURES

FIGURE 1	20
----------------	----

CHAPTER 1

INTRODUCTION

Human capital studies took a solid part of economic literature over the past decades. Scientists were interested in the impact of education on different aspects of social wellbeing starting from wellbeing of individual and concluding by government policy directed at educational development stimulation.

The question, of what exactly enforces an individual to invest his time and other restricted resources in his education, is still open. Numerous scientists try to answer this question using different approaches. Manski (1992) uses simulation based on modeling of high school student's behavior based on family income. He concludes that school choice will not have significant impact under condition of equalizing choice opportunities for different income groups. Wilson (2001) builds and estimates high school graduation model based on family background and expected income based on observations of previous cohorts educational outcome. She concludes a significant impact of income expectations on the high school graduation choice.

Previous studies found significant correlation of family background characteristics, childhood environment, expected income with either schooling attainment or duration of education at different levels. But all these studies are missing such an important parameter as individual's ability to study at a certain academic level. This research is dedicated to fulfill this gap and give an answer the question how individual's ability define this educational choice.

This study is modeling individual's college attainment choice with respect to his individual characteristics, such as abilities to meet requirements at the college level, family characteristics, social factors (gender, race, etc.) and financial possibilities of his family to afford his education. This research is made to answer the following questions:

- 1) To build theoretical model of college attendance response to family background, income expectations and abilities to fulfill academic requirements at the college level;
- 2) To estimate the model on individual level data using probit regression;
- 3) To evaluate weight of individual's abilities to fulfill academic requirements in the model of rational choice of college attendance;
- 4) To make conclusions based on obtained results.

To answer all questions specified above I develop the following structure of the study. Chapter 2 provides review of related work. In this Chapter I want to reveal recent works in the educational choice literature and point to their advantages and disadvantages. Chapter 3 develops theoretical model of agent's college attendance choice response to family background, personal characteristics, income expectations and abilities. Chapter 4 describes dataset used to estimate theoretical model developed in the Chapter 2. Also this Chapter provides methodology of some variables construction and data sampling. Chapter 5 provides methodology of the model estimation, its restrictions and assumptions made. Chapter 6 provides estimation results and Chapter 7 concludes.

CHAPTER 2

REVIEW OF RELATED LITERATURE

Related literature on rational educational choice topic includes a huge bunch of publications. In this Chapter I would like to point out the most important ones, methodology of which was used in this study.

Lazear (1997) in his paper “Education: consumption or production” investigates causality between income and education. In his model he treats education as a consumption (investments in education) and as a factor of production (higher salary in future). He assumed that initially each individual has a non-zero endowment of input factors which could be transformed into education using production function and zero endowment of education. Each period the individual decides which share of his endowment of goods to transform into education. After, he estimates OLS model for his educational choice framework. He concludes that most individuals do not reach wealth maximizing level of education due to utility costs of education, if education enters into utility function as consumption good. The idea of bilateral nature of education in the utility function was used later by many researchers.

One of the main disadvantages of Lazear’s approach was that he considered education-income relationship as a priory given. As was shown in later studies employing income expectation approach would be more appropriate and would provide better understanding of individual’s motivation to study.

Willis (1985) in the Chapter 10 of “Handbook of Labor Economics” considers functional form specification of regression equations of wage with respect to education

and vector of other variables. He develops a theoretical model which says that there is non-linear dependence between income and duration of education. That is why he proposes to include both linear and quadratic form of educational variable for OLS estimation aimed to capture this non-linearity.

Manski (1992) in his study investigates the problem of schooling and social mobility among students with different income levels. He performs a quantitative study of employing of voucher system to low-income students to allow them equal choice opportunity. He concludes that voucher system does not have definite positive effect on school choice in favor of high quality schools, but decreases motivation of poor family students from public schools to make their academic efforts to get other sources of financing to study in private high quality schools. At the same time it gives opportunity to study in private schools by employing vouchers system to low-income students. The high income ones get better schooling quality based on other income parameters (for instance, ability to get a better computer, books, etc). His study is valuable for my research, because I used his idea of consideration of other income and family background characteristics as determinants of educational choice.

Palme and Wright (1998) performed empirical study of rate of return to education changes in Sweden. They used empirical model specification suggested by Willis (1985) and found that Sweden experienced significant decline in returns to education (mainly university level) in 1968-1991. However, they pointed out that results for returns to working experience differed among different samples. Thus, they concluded that an individual's decision on educational choice might be influenced by future income expectation based on the observations of older cohort's income.

Bills and Klenow (2000) used a macroeconomics approach to investigate relationship between education and wellbeing. They developed a macroeconomic model of dependence of economic growth on education. However, they got a controversial result, which brought them to idea of reverse causality between economic growth and education. In other words they considered a scenario that richer economy can afford more education opportunities. Disadvantages of their approach based on the cross-country data analysis is that, it is hard to capture the effect of schooling duration on the economic growth due to difference in systems of education and its quality. That is why it is important to use micro-level analysis within the same system of education.

Wilson (2001) provided educational choice model and its estimation based on individual choice of high school graduation. She developed discrete choice model, where rational agent's choice of high school graduation is based on his family background, neighborhood characteristics and expected income based on observation of job market performance of older individuals who made their choice. This research is closely based on her approach, but applied to college attendance choice. Also, she did not include individual's abilities to meet academic requirements for high school graduation. That is why in my study I decided to fulfill this gap and include the variable of abilities in the model of college attendance choice.

Hanushek (2007) in his study "The Role of Education Quality in Economic Growth" appealed to income differences between countries which provide the same "amount" of education. In the result of his research he admitted that the main characteristic, which affects wellbeing, is its quality. Even though measurement of educational quality is beyond the scope of my study, Hanushek's modeling of income and

education relationships was very helpful for understanding education-to-income transformation path.

In the following chapters I will try to model and estimate individual's educational choice using experience of previous studies.

CHAPTER 3

THE MODEL OF DISCRETE EDUCATIONAL CHOICE

This Chapter is dedicated to description of impact mechanism of different factors on agent's educational choice. All agents are assumed to have rational utility maximizing behavior based on their ability to observe and compare educational outcomes from previous generations.

During educational choice process agents face income and non-income factors which determine their choice. Family background, individual characteristics, abilities will be referred to as non-income factors. And expected income outcomes will be referred to as income factor of educational choice. It is assumed that the rational agent is able to observe job market performance of older generation agents who have the same family background and have already made their choice. Income expectations are based on agent's ability to observe income differences of older people who already made their educational choice. This approach was developed and applied by Wilson (2001) as an extension of Manski (1993) and Freeman (1971) framework. She applied this approach to modeling and estimation of educational choice of high school graduation.

In this Chapter Wilson's approach will be extended by adding variable of abilities to her inputs vector in the education production function as well as the scope of the investigation will be moved to college attendance choice.

Suppose agent's utility function is described by the following equation:

$$U_i = S_i + \ln(C_i) + \epsilon_i \quad (1)$$

Where, S_i is utility of education and C_i is utility of consumption. In this model education is considered as consumer good and study process assumed to be associated with some non-income benefits and costs, for example pleasure of college life and efforts input to maintain satisfactory academic standing. C_i represents all other goods that could be consumed by individual. S is defined as function of actual studying s :

$$S_i = g(s_i, x_i) + u_i \quad (2)$$

Where x_i is vector of input factors important for the educational process (such as books, computer, abilities etc), $g(*)$ is a transformation function of inputs and actual study to utility from school attendance.

Agent maximizes his utility subject to the following budget constraints:

$$Y_i = f(V_i)s_i + \varepsilon_i \quad (3)$$

$$Y_i \leq C_i \quad (4)$$

Where Y_i is agent's discounted income, $f(V_i)$ is schooling to income transformation function, ε_i is a set of other factors of income. There are no loans in this model, so it is assumed that agent cannot consume more than earn. V_i is a vector of factors which can affect returns to education. In my case I selected the following list of factors: family income, sex, race and family size. Choice of the specified set of factors was based on previous educational choice studies and basic economic intuition.

Impact of sex and race parameters on educational attendance were widely discussed in economic literature (Hanushek, 2007) as well as gender and political economy studies (Dryler, 1998). That is why presence of these parameters in schooling-to-income transformation function is very important.

Family income was pointed out by Wilson (2001) as a very important factor of agent's job market performance. Higher family income indirectly implies better family standing on job market and thus better connections and opportunities to support each other for successful performance (for example, it is hard to believe that Roman Abramovich's children will work as cashiers in McDonalds after college graduation).

Family size is assumed to be a negative factor of education-to-income transformation. This assumption is dictated by simple logic that having a big family an individual would pay much more attention to his family then to his job, despite having a degree from a top college.

Following Wilson's and others approach to agent's income expectation I specify the following equation:

$$E(Y_i|V_i, s_i = s_j) = f(V_j)s_j \quad (5)$$

Where, subscript j identifies agents from the older cohort (those who already made their choice). In other words, it means that observing older people's performance with and without college degree agents are able to "fit" their family backgrounds to themselves and to form income expectations of their future income if they will or will not attend college.

Substitution of equations (2) to (4) into (1) gives:

$$E(U_i) = g(s_i, x_i) + E[\ln(Y_i) + \epsilon_i] \quad (6)$$

Since agent assumed to be rational utility maximize, it is logical to suggest that he will chose college attendance unless:

$$E(U_{i,ca}) = g(s_{i,ca}, x_{i,ca}) + E[\ln(Y_{i,ca})] + \epsilon_{i,ca} > E(U_{i,nca}) = g(s_{i,nca}, x_{i,nca}) + E[\ln(Y_{i,nca})] + \epsilon_{i,nca} = E(U_{i,nca}) \quad (7)$$

Where, subscripts ca and nca mean college attending and non-college attending individuals respectively.

Rearranging elements in (7) we will obtain:

$$g(s_{i,ca}, x_{i,ca}) - g(s_{i,nca}, x_{i,nca}) + E[\ln(Y_{i,ca})] - E[\ln(Y_{i,nca})] > \epsilon_{i,nca} - +\epsilon_{i,ca} \quad (8)$$

Relaxing subscripts and substituting $g(s_{i,ca}, x_{i,ca}) - g(s_{i,nca}, x_{i,nca}) = Dx$ and $\epsilon_{i,nca} - +\epsilon_{i,ca} = \epsilon$. This brings us to the following inequality:

$$Dx + E[\ln(Y_{ca})] - E[\ln(Y_{nca})] > \epsilon \quad (9)$$

Rational agent will choose college attendance, while inequality (9) holds. Then, probability of college attendance is given by the following equation:

$$\Pr(CA) = \Pr(\epsilon < Dx + (E[\ln(Y_{ca})] - E[\ln(Y_{nca})])) \quad (10)$$

Equation (10) is probabilistic interpretation of agent's rational choice. Hence, having actual data on income of older cohort and family background of both cohorts we can estimate expected income for primary group and then probabilistic regression of educational choice.

CHAPTER 4

DATA DESCRIPTION

In my research paper I combined two sources of data to capture a wider set of determinants of educational choice. Datasets are Panel Study of Income Dynamics (PSID) and National Longitudinal Survey of Youth 97 (NLSY97). From these two sources I was able to capture family background and individual characteristics.

Following Wilson (2001), I distinguished two cohorts of individuals. The first (main) cohort includes individuals aged 19-20 in the year 2000. Since the typical age of high school graduation is 18, this cohort will capture the main share of college attendance decision makers. Thus we assume that this sample will represent college attendance decision making process.

Selected sample of NLSY97 dataset contains actual and constructed variables. Actual variables describing family background are family income in the previous period (in our case in 1999, since agent makes his decision in year 2000), family size in 2000, mom's education, dad's education, region of interview. Personal characteristics are given by gender and highest degree attended. To measure person's abilities I used Scholastic Aptitude Test (SAT) score in math and verbal as proxy. I had to construct some variables based on available data in order to make my estimation results more robust and capture all necessary effects.

Below I provide a list of constructed variables with short explanations.

College attendance takes values zero if highest grade attended in the year 2000 is less or equal to twelve (twelve is a time period normally required to graduate from a high

school in US) and equal to one if highest grade attended is more than twelve.

Construction of this variable allows us to capture actual college attendance decision.

Race takes value zero if respondent is black and 1 if white. This variable is constructed since I wanted to capture income difference between the two biggest ethnic groups in the US.

Abilities are measured by interaction term of both math and verbal SAT scores, since I do not have college specification, so generally it is logical to assume that interaction of these two indicators will give us measurement for ability any college regardless of specialization.

According to the model, decision maker in the main cohort will make his decision of college attendance based on his objective factors and personal perception of possible outcomes of college attendance. Representative agent can build his expectation of income upon college attendance based on experience of older people with the same background. That is why I distinguish the second cohort (reference group). To this group belong people who are 23-36 years old as of 1990. Selection of this age range was based on the assumption that people of this group potentially were able to finish baccalaureate and difference in their income can help me to capture income effect of college attendance for this group. Also, while the main cohort observes the income status of older people (reference group), they are able to make their own income expectation conditional on their family background and educational choice. Observing the income difference of older fellows with a similar to representative agent's family (family income and family size) and personal (race and sex) backgrounds, but different educational attendance we will be able to capture income effect of education on college attendance choice. My

sample of PSID captures constructed and actual variables by the same principle as the sample for NLSY97. Descriptive statistics for NLSY97 and PSID data sample are provided in Table A1 and Table A2 (see Appendix A).

If we take a look at the common variables from two samples we can notice significant growth in college attendance from average 0.4 to almost 0.88. The same thing is about average family income. All other variables stay about the same level. There are two possible sources of such differences. The first is economic growth and growth of share of rich families in economy. The second one is that this growth might be caused by sampling bias (as you can see NLSY97 contains only 534 observations, while PSID has over 4000). This fact might bring some disturbances in estimation results, but having no other source of data we have to assume that both samples are representative.

As I specified above I used two dataset for this research: NLSY97 for main group data and PSID for reference group data. There are two reasons for incorporation of these two data samples. The first reason is that 1990 is the latest period where PSID dataset contains all necessary variables. But to create a reference group we need some reasonable time interval between decision makers. That is why we have to consider people from the sample NLSY97 as the main cohort. It satisfies our need in time gap between decision makers, since we observe decision making process in 2000. The second reason is that PSID dataset does not contain any variable which allows us to capture person's abilities and at the same time NLSY97 provides this data (SAT test).

CHAPTER 5

METHODOLOGY OF ESTIMATIONS

In the Chapter 3 I described probabilistic model of rational choice of college attendance. This model describes the mechanism of income and non-income factors of educational choice. In this Chapter I will try to estimate parameters of this model employing probabilistic (probit) regression approach.

In order to estimate this model, we need to set up equation of educational choice response to specified factors for empirical estimation of their impact.

$$CA_i = DX_i + \beta(E[\ln(Y_{i,ca})] - E[\ln(Y_{i,nca})]) \quad (11)$$

Where, CA is a binary variable of college attendance; β is a coefficient of income difference impact on the probability of college attendance; X_i is a vector of non-income parameters; D is a vector of coefficients for non-income parameters.

Into the vector of non-income variables and their product combination I include the following ones:

- 1) Family income is a parameter which measures individual's family possibility to afford college education;
- 2) SAT math score is a proxy for individual's analytical abilities. It takes discrete values from 1 to 10, where 10 is the best score;
- 3) SAT verbal score is a proxy for individual's verbal abilities. It takes discrete values from 1 to 10, where 10 is the best score;
- 4) Sex is a dummy binary variable to identify agent's gender. It takes values 1 for male and 1 for female

- 5) Race is dummy binary variable to identify agent's race. It takes values 1 for white and 0 for black;
- 6) Parents' education appears to be a significant determinant of educational attainment in the economics of education literature. This variable takes discrete values, which measure number of full years of educational attainment.
- 7) Region of residence is a discrete variable, which takes values 1 for North-East; 2 for North-Central; 3 for South; 4 for West. This variable allows us to capture regional income differences as well as regional factors magnitude differences.

In the Chapter 6 of model estimation I will use different sets of selected variables as well as their product combinations do to get explanatory variables set with the best fitness to the specified empirical model.

In order to estimate equation 11, first of all, we need to know $E[\ln(Y_{i,ca})]$ and $E[\ln(Y_{i,nca})]$. Under assumption of the theoretical model that agent can observe job market performance of previous generation of decision makers; we can estimate income equations for two subsamples for those who attended college and those who did not from the reference group. As described in Chapter 4, we can use data of PSID individual dataset for the year 1990 and treat this cohort as a reference group.

From equation (3), we can specify the following equation for estimation:

$$Y_i = \alpha_{1,i}c_i + \alpha_{2,i}fam_income + \alpha_{3,i}Sex + \alpha_{4,i}race + \alpha_{5,i}fam_size \quad (12)$$

Where c is a constant; fam_income is individual's family income; sex – individual's gender, dummy variable which takes values 1 for male and 0 for female; $race$ – race of an individual, dummy variables which takes values 1 for white and 0 for

black; *fam_size* is individual's family size. Also, I include *region* as a classificatory variable to capture regional income differences. This variable takes discrete integer values 1 for North-East; 2 for North-Central; 3 for South; 4 for West.

Substituting (12) into (11) will bring us to full probit model specification of college attendance choice.

Having estimated values of $\alpha_{t,i}$ for two subsamples of the reference group we can generate two series of expected income if agent would attend and if he would not attend college. In order to do that I use actual values of family background and individual characteristics into estimated regression equation (12). Taking difference of two generated series will give us expected income gap variable with respect to college attendance choice and increase in this gap should give more income motivation of college attendance to the agent.

A trick with usage of two datasets was a necessary step due to lack of data on agents abilities in PSID dataset as well as absence of individual income parameter after year 1990. Assumption of relevance of expected income estimation is based on the assumption that income equation (12) has unbiased estimators which will be tested in the next Chapter.

Reader can notice that probit model specification includes family income, gender and race twice. That is why these variables are considered to have direct and indirect (income) effect on the educational choice. Direct effect appears when family and individual characteristics affect educational choice in the straight forward way. For instance, family income will define whether or not individual's family is able to afford his education. Indirect effect appears in the function of education to income

transformation. Hence, the same variables will affect educational choice in the alternative way via this functional mechanism.

Logarithmic specification of the income effect¹ allows us to avoid perfect colinearity of family background and personal characteristics and capture magnitude of both direct and indirect effects of selected variables on agent's rational educational choice.

In the next Chapter I present estimation output results along with some standard tests of models and estimators relevance.

¹ Wilson, 2001 used this approach to avoid perfect colinearity of factors.

CHAPTER 6

ESTIMATION OUTPUTS

This Chapter provides estimation results and econometric test results for estimated models. I would like to start with the general income equation specification estimated on general sample but adding college attendance dummy variable. For estimation of this model I used simple OLS method. Estimation output is provided in Table 1.

As reader can see from the estimation result we got well fitted estimation of income equation. All variables are significant at 10% level or less (except intercept) which says that selected set of depended variables is able to explain changes in individual's income. Heteroskedasity test (see Table 2) fails to reject H_0 of absence of heteroskedasity at even at 1% level. Normality test, (see Figure 1) also gives us evidences of residuals' normality and unbiasedness of our estimators for basic income equation estimation.

Signs of coefficients are also correct with respect to our model framework. College attendance has a positive sign which means that people who attended college get higher salary. Higher level of education is assumed to be associated with higher income. Otherwise, agent would not have any income motivation to put forth his efforts to seek a college degree.

Family income is also positively associated with individual income which corroborates Wilson (2001) in the framework of my model specification.

Table 1

Basic income equation estimation

Dependent Variable: INCOME
 Method: Least Squares
 Date: 04/06/11 Time: 01:11
 Sample: 1 4483
 Included observations: 4483

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	457.9116	697.7321	0.656286	0.5117
CA	3663.006	383.2054	9.558858	0.0000
FAMILY_INCOME	0.316986	0.006909	45.88032	0.0000
RACE	718.8970	409.4634	1.755705	0.0792
SEX	12514.43	344.4338	36.33334	0.0000
FAM_SIZE	-1139.029	123.3836	-9.231608	0.0000
R-squared	0.494813	Mean dependent var		16939.78
Adjusted R-squared	0.494249	S.D. dependent var		16169.01
S.E. of regression	11498.78	Akaike info criterion		21.53921
Sum squared resid	5.92E+11	Schwarz criterion		21.54778
Log likelihood	-48274.13	Hannan-Quinn criter.		21.54223
F-statistic	877.0128	Durbin-Watson stat		2.153387
Prob(F-statistic)	0.000000			

Table 2

Basic income equation, Heteroskedasticity test

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	358.2539	Prob. F(5,4477)		0.0000
Obs*R-squared	1281.097	Prob. Chi-Square(5)		0.0000
Scaled explained SS	7168.116	Prob. Chi-Square(5)		0.0000
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.34E+08	22697478	-14.69346	0.0000
CA	-29015817	12465812	-2.327632	0.0200
FAMILY_INCOME	8910.516	224.7512	39.64613	0.0000
RACE	5628117.	13319992	0.422532	0.6727
SEX	96639430	11204558	8.625011	0.0000
FAM_SIZE	16977271	4013714.	4.229816	0.0000
R-squared	0.285768	Mean dependent var		1.32E+08
Adjusted R-squared	0.284970	S.D. dependent var		4.42E+08
S.E. of regression	3.74E+08	Akaike info criterion		42.31907
Sum squared resid	6.26E+20	Schwarz criterion		42.32764
Log likelihood	-94852.19	Hannan-Quinn criter.		42.32209
F-statistic	358.2539	Durbin-Watson stat		1.944634
Prob(F-statistic)	0.000000			

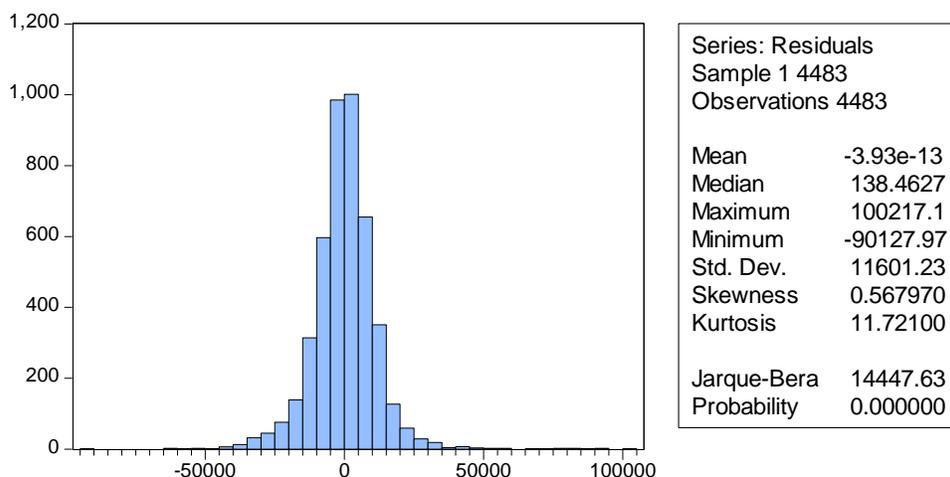


Figure 1. Normality Test for Basic Income Equation Estimation.

Positive sign of the dummy variables of race and sex one more time prove presence of income difference across gender and racial groups. To put it simply, I can summarize this finding as that it is objectively not profitable to be a women or a black on the US job market. Low significance of the coefficient (about 7%) of race points to the presence of some “disturbances” in its impact on the individual income. This disturbance will be revealed in the data analysis below.

Family size has a negative impact on individual’s impact. As I suggested above, this is most likely because an individual spends much more time on the family dues and relative’s interaction then on his job.

Such a good fitness of the model and robustness of results allows me to assume that this model estimation would give pretty decent income prediction based on family background characteristics.

Further estimations of income (see Table3) on two subsamples (people with and without college attendance) gives pretty similar results. All variables have this same sign as the basic model and all are significant at 10% level or less. Exception is variable of race. I noticed pretty low significance of this variable in basic model. Further analysis has shown its significance varies from region to region having the highest significance for population with college degree in the North-Central and South regions. This result is rather interesting, because from the historical point of view these two regions were famous for their intolerance to black population. Thus, from obtained results I can conclude that these regions still have this sort problem. This result is very interesting by itself and deserve separate study framework, but unfortunately it is behind the scope of my research.

Now we can generate series of expected income and its difference conditional on college attendance and for main cohort based on estimated coefficients of income equation for reference group. Descriptive statistics for this series are presented in Table 4. As we can see from the Table 4, median expected income of people who would attend college is significantly higher. That means that our simulated income expectation takes into account that higher level of education should be associated with higher income. Therefore, this fact proves the assumption of our model that individual will have income effect in his educational choice.

Table 3

Income equation estimation by region

Variable	General Sample	Significance	College attendance, 1	Significance	College Attendance, 0	Significance
All regions						
c	457.911600	0.511700	1181.426000	0.254400	3189.905000	0.000000
ca	3663.006000	0.000000	-	-	-	-
family_income	0.316986	0.000000	0.345152	0.000000	0.293205	0.000000
sex	12514.4300	0.000000	13619.4200	0.000000	10405.8400	0.000000
race	718.897000	0.079200	2105.266000	0.000400	-71.089090	0.844800
fam_size	-1139.0290	0.000000	-1303.7760	0.000000	-1085.51700	0.000000
Region 1						
c	2024.056000	0.062700	9246.321000	0.021300	3650.803000	0.163300
ca	4305.751000	0.000000	-	-	-	-
family_income	0.292374	0.000000	0.296601	0.000000	0.278080	0.000000
sex	11820.36000	0.000000	17792.88000	0.000000	13217.37000	0.000000
race	989.948100	0.094800	608.642400	0.799800	-480.956800	0.739500
fam_size	-1231.01400	0.000000	-2847.98000	0.000100	-1196.50800	0.007800
Region 2						
c	1044.832000	0.291200	-472.358700	0.857700	5279.218000	0.002500
ca	4150.893000	0.000000	-	-	-	-
family_income	0.288696	0.000000	0.317839	0.000000	0.218551	0.000000
sex	10759.250000	0.000000	14491.360000	0.000000	11026.770000	0.000000
race	839.573100	0.107000	3470.744000	0.031500	587.840000	0.563800
fam_size	-856.935400	0.000000	-968.163100	0.049600	-1225.977000	0.000100
Region 3						
c	-136.703600	0.874700	1464.909000	0.230000	3271.832000	0.000200
ca	3325.350000	0.000000	-	-	-	-
family_income	0.337791	0.000000	0.371111	0.000000	0.317270	0.000000
sex	10660.570000	0.000000	11047.850000	0.000000	8528.356000	0.000000
race	965.956200	0.029900	1386.287000	0.049200	-262.9877	0.557400
fam_size	-926.576300	0.000000	-1274.1110	0.000000	-1061.1930	0.000000
Region 4						
c	889.955900	0.298400	-4630.517000	0.143200	-1717.335000	0.362400
ca	3486.056000	0.000000	-	-	-	-
family_income	0.325977	0.000000	0.388033	0.000000	0.360259	0.000000
sex	11174.380000	0.000000	16541.670000	0.000000	12812.170000	0.000000
race	690.249100	0.148000	2952.540000	0.158300	1600.447000	0.256200
fam_size	-1109.221000	0.000000	-879.281200	0.128900	-1078.441000	0.000000

TABLE 4

Descriptive statistics for expected income

Statistics	Expected Income difference	Expected income if notattend college	Expected income if attend college
Mean	4988.336	29386.68	34375.02
Median	4465.228	25660.43	29835.80
Maximum	16780.54	89592.74	106373.3
Minimum	-3300.316	-3746.793	-6523.532
Std. Dev.	3998.283	20130.91	24015.79
Skewness	0.966852	1.298513	1.264310
Kurtosis	4.100673	4.631613	4.577118
Jarque-Bera	110.1530	209.2992	197.6072
Probability	0.000000	0.000000	0.000000
Sum	2663771.	15692490	18356261
Sum Sq. Dev.	8.52E+09	2.16E+11	3.07E+11
Observations	534	534	534

Probit model estimation is performed on the general sample of main cohort without division on regional subsamples. This decision was based on two reasons. First, above we obtained pretty similar result for all 4 regions in terms of income equation estimation. That is why it should not make a big difference in terms of significance and magnitude of coefficients. The second reason is that, main cohort data sample is very restricted; it includes only about 500 observations. That is why dividing it on subsamples will reduce number of our observation to roughly 125 per sample, but in this case we will increase risk of estimator biasedness.

Table 5 contains probit estimation outputs employing different sets of educational choice determinants. General model column includes all specified variables in the Chapter 4, but as you can see significance of estimate coefficients is very low as well as signs cannot be explained in the framework of the model specified in the Chapter 2.

TABLE 5

Probit estimation of the model of educational choice

Variable Name	General model	Significance	Selected Variables	Significance
log(d_income)	-0.092436	0.4562	0.07248	0.0271
test_math*test_verbal	-0.083206	0.0844	0.03069	0.0024
test_math	0.404249	0.0515	-	-
test_verbal	0.420745	0.0362	-	-
fam_incomes	5.65E-06	3.02E-02	4.41E-06	0.0215
sex	-0.437049	0.341	-0.3939	0.0209
dad_edu	-0.011244	0.8832	-	-
mom_edu	-0.085264	0.279	-	-
fam_size	-0.029794	0.6413	-	-
dad_edu*mom_edu	0.00646	0.2498	-	-
race	-0.042262	0.9158	-	-
sex*race	0.238698	0.6034	-	-
loglikelihood	-155.1599	-	-162.83	-

This result of general model estimation might be caused by presence of correlation between explanatory variables. Further, playing with the set of determinants of college attendance I attempted to remove variables which could have correlation (for example, an interaction term of parental education is highly correlated with family income), In the result of such model tuning, I defined set of selected variables, which gave me significant result which could be explained and which is logical in the framework of this study. As you can see from the selected variable column of Table 5 there are 4 variables which have a significance level of 3%. Difference of logarithms of income has positive sign and significant coefficients which defines the magnitude of income effect on the educational choice.

Interaction term of SAT scores has a positive effect on the college attendance choice. This follows straightforward from the fact that SAT exam is taken before

graduation and at the same time is considered by most colleges in the US as a standardized test for college entrance. Higher interaction term of these scores means higher probability of acceptance to good quality college. Also, this parameter might be considered as a proxy to self-confidence of the individual during application campaign. Hence, my hypothesis of significant impact of personal abilities on his or her rational educational choice is proven on the real data.

Family income has a positive sign. This is logical, since family income identifies an individual's ability to afford college level education.

Gender variable has a negative sign. This means that girls tend to attend college more than guys. Why? The answer, to this question is little more complicated. Reader should review estimation output for income equation. For that equation we got positive sign for the variable of gender which implies that being a female individual leads to lower return to education. But, ambition to have an appropriate level of income gives female individuals a higher motivation to attend college. For this reason, dummy variable of sex has a negative sign in my probit model.

CHAPTER 7

CONCLUSION

This paper answers three main questions addressed in the beginning of this study. In the Chapter 2 I built a theoretical model of an individual's rational choice of college attendance, which shows that rational agent will choose to attend college while joint utility of studying and consumption will be higher than consumption alone.

Empirical estimation of the income equation and educational choice provided us with three main findings. Firstly, there is a significant impact of abilities in educational choice. That's means that agent take his abilities to meet college program requirements into account during decision making process.

Secondly, our empirical study shows that there is bidirectional effect of gender on the educational choice. On the one hand, female individuals get a priory lower education-to-income transformation efficiency. But on the other hand, they are encouraged to get more education in order to compensate this inefficiency. It may be interesting topic for my future research to estimate magnitudes of positive and negative gender effect on the educational choice.

Thirdly, my empirical analysis shows, that there is non-systematic racial income discrimination of highly qualified people. Particularly, this discrimination appears to be significant in the samples of Southern and North-Central regions of the United States.

REFERENCES

- Dryler, H. (1998). Parental role models, gender and educational choice. *The British journal of sociology*.
- Eric A. Hanushek, Ludger Wößmann. (February, 2007). *The Role of Education Quality in Economic Growth*. World Bank Policy Research Working Paper 4122,.
- Lazear, E. (June, 1977). Education: Consumption or Production? *The Journal of Political Economy*, 569-598.
- Manski, C. F. (1992). Educational choice (vouchers) and social mobility. *Economics of Education Review*, 351-369.
- Mark Bilz, P. J. (2000). Does Schooling Cause Growth? *American Economic Review*, 1160-1183.
- Marten O. Palme, Robert E. Wright. (1998). Changes in the rate of return to education in Sweden 1968-1991. *Applied Economics*, 1653-1663.
- Willis, R. J. (1985). *Wage determinants: a survey and interpretation of human capital earnings functions*. In *Handbook of Labor Economics*, O. Ashenfelter and R. Layard . Amsterdam: North-Holland.
- Wilson, K. (2001). Determinants of educational attainment: Modeling and Estimating the Human Capital Model and Education Production Function. *Southern Economic Journal*, 518-551.

APPENDICIES

APPENDIX A

TABLE A1

NLSY97 sample descriptive statistics

Variable	Mean	Median	Maximum	Minimum	Std.Dev.	Skewness	Kurtosis	Sum	SumSq.Dev.	Observations
age	19.11049	19	20	18	0.758295	-0.186518	1.760155	10205	306.4813	534
ca.	0.883895	1	1	0	0.320651	-2.396718	6.744259	472	54.8015	534
dad_edu	14.56742	14	20	6	2.907283	0.036736	2.545746	7779	4505.073	534
fam_income	87726.54	76025	270543	250	66820.9	1.38018	4.765202	46845973	2.38E+12	534
fam_size	3.941948	4	9	1	1.298867	0.261994	3.910948	2105	899.2004	534
mom_edu	14.04869	14	20	1	2.572307	-0.4247	4.998634	7502	3526.734	534
race	0.833333	1	1	0	0.373027	-1.788854	4.2	445	74.16667	534
sex	0.462547	0	1	0	0.499063	-0.150235	1.02257	247	132.7509	534
test_math	3.990637	4	6	1	1.138573	-0.126712	2.768234	2131	690.9532	534
test_verbal	4.022472	4	6	1	1.07481	-0.207988	2.910681	2148	615.7303	534
ysch_2000	13.48502	14	16	3	1.090106	-1.692554	18.22623	7201	633.3801	534

TABLE A2

PSID sample descriptive statistics

Variable	Mean	Median	Maximum	Minimum	Std.Dev.	Skewness	Kurtosis	Sum	SumSq.Dev.	Observations
age	29.98861	30	36	23	3.904394	-0.13504	1.878084	126402	64239.45	4215
ca	0.400474	0	1	0	0.490053	0.406231	1.165024	1688	1011.999	4215
fam_size	3.807117	4	11	2	1.403111	0.947825	4.847246	16047	8296.186	4215
family_income	40973.94	36000	280500	1	26768.79	2.347991	13.76343	1.73E+08	3.02E+12	4215
ysch	12.46833	12	17	0	2.947956	-1.58172	7.841626	52554	36621.52	4215
race	0.752313	1	1	0	0.43172	-1.16901	2.36659	3171	785.4149	4215
income	17304.29	15000	195000	-9850	16371.47	2.388739	16.65031	72937568	1.13E+12	4215
sex	0.484223	0	1	0	0.49981	0.063139	1.003987	2041	1052.701	4215

VITA**SERGIY POLYACHENKO**

March 3, 1988

Current address:

1575 Logan Dr. Carbondale, IL, 62901

Permanent address

1-y prov. Budivelynykind bld, 2 apt, 20, Konotop 41631, Sumska oblast, Ukraine

sergiy.polyachenko@gmail.com

Sumy State University

Bachelor of Science, Finance, June, 2005

Special Honors and Awards:

Fulbright Graduate Exchange Student scholarship

Research Paper Title:

College Attendance Rational Choice: Modeling and Empirical Estimations

Major Professor: Scott Gilbert