Issues in Political Economy, Vol 22, 2013, 127-147

The Pursuit of Increased Human Capital: Who's Going to College, and Why?

Jack Nuland & Justin Ingram, St. Lawrence University

While an outlying few have enjoyed enormous success regardless of their choice to forgo a higher, formal education, empirical observations serve testament to the advantages of attending college. Increased starting salaries, higher earning trajectories, and the intangible benefits of living and studying in a socially robust and educationally stimulating environment are among the fruits enjoyed by those who decide to attend college. Here it is worth noting that the above diction has been specified carefully, as—despite predispositions resulting from uncontrollable factors such as the environment in which one is raised—the decision to attend college is just that: a decision. This begs the question of which individuals are making the decision to pursue a college education, and of those who are what incentives are among the primary contributors. In what follows, the answers to both these questions will be unveiled.

I. LITERATURE REVIEW

Recent literature regarding the decision to attend college has been focused on the socio-economic factors that predispose some individuals to be more likely to have the opportunity of a higher education than others. Determining these factors, as well as the magnitude of their respective influences, has been the task of economists, sociologists, and academics alike. Specifically, there has been an observable increase in female enrollment, and studies by Susan L. Averett, Mark L. Burton, and Alison Aughinbaugh suggest that women might be attending college for different reasons than men.

Averett and Burton (1996) examined the decision to go to college, and how that decision might be influenced by different incentives for men and women. They assert that this decision is a function of family background characteristics, such as: financial constraints (number of siblings, race, ethnicity), tastes (parental occupation; access to library cards, magazines, and newspapers when growing up; religion; and region of residence in adolescence), as well as the expected future earnings differential between college and high school graduates (the college wage premium). They also utilize AFQT test scores from the NLSY97 as an indicator of ability. They observed that men's decision to attend college is motivated by the college wage premium, while the college wage premium is an insignificant factor in the decision for women. In addition, they found support for the comparative advantage hypothesis, which suggests that individuals self-select themselves into the education that maximizes their utility.

Their results show that both men and women who attend college—as opposed to just high school—have better educated parents, fewer and better-educated siblings, AFQT scores that are nearly twice that of those who do not attend college, higher labor force participation rates, and are far more likely to have parents who held professional jobs.

Averett and Burton (1996) also highlight the phenomenon that since men have traditionally participated in the labor force, regardless of education, there is not a significant correlation between men attending college and joining the labor force. Women, on the other hand, do show a correlation between the probability of attending college and participating in the labor force. This has been an emerging trend, and suggests that women's decision to attend college is more oriented around the intent to join the labor force than that of men.

In a more recent study, Alison Aughinbaugh (2008) used data from the NLSY97 in order to identify the factors that influence two decisions: first, the decision of whether or not to attend college, and second, of those who do attend, who decides to stay in college. Aughinbaugh begins the article by conveying the motivation of her study, using the work of Day and Newberger (2002) to show that those who attend college earn significantly more money throughout their lives than those who do not. More specifically, someone with a bachelor's degree earns about a third more than someone who attended college but did not earn a degree. Those with a bachelor's degree earn about twice as much as those who never attended college. Evidently, the decision to attend college—and the subsequent decision to stay in college—significantly impacts the course of one's professional life, and as such Aughinbaugh sought to determine what influences these decisions.

Aughinbaugh's analysis uses data collected up through the 8th round of the NLSY97. Her sample consists of 6,580 respondents who were interviewed at age 21 or older, ensuring that the respondents were observed once they have been in college for at least 12 months. The independent variables considered were gender, race/ethnicity (Black, Hispanic, or mixed), age at last interview, family background characteristics (such as: family income in 1996, mother's education, father's education, mother's age at first birth, and whether the respondent lived with both parents at age 12), and educational characteristics (such as: high school grades, ASVAB scores, and a dummy variable for whether or not one took the SATs or ACTs).

The data showed that approximately 49 percent of the sample attended college by age 20; within that 49 percent, 40 percent started at a 2-year college while 60 percent started at a 4-year college. Those who did attend college by age 20 had parents who attained more schooling, had higher levels of family income, had mothers who were older at the birth of their first child, and were more likely to have lived with both parents at age 12 than their non-college-attending counterparts. Additionally, those who did attend college by age 20 had higher high school grades and higher ASVAB scores than those who did not.

Aughinbaugh goes on to demonstrate that both the background characteristics and the high school achievement of students who initially entered a 4-year college differ from those who initially enter a 2-year college. Specifically, respondents who first attend 4-year colleges were more likely to be female and less likely to be black or Hispanic, compared to those who initially entered a 2-year college. Furthermore, as opposed to the respondents who initially enrolled in 2-year colleges, those in 4-year colleges had higher family incomes in 1996, had better educated parents, had mothers who were older at the time of their first birth, were more likely to have lived with both parents at age 12, and performed better in high school and on the ASVAB test.

Aughinbaugh used a series of logit equations of the form:

$$(1)\ C_i = X_i\alpha_1 + X_{fi}\alpha_2 + X_{hsi}\alpha_3 + \varepsilon_I$$

in order to further examine how the respondent's characteristics are related to his/her decision to go to and remain in college. C_i is a dependent variable indicating the respondent's decision about college. X_i is a set of exogenous individual characteristics, X_{fi} is a set of family characteristics, X_{hsi} is a vector describing high school outcomes, ε is the individual error term, and the α's are the parameters to be estimated.

Regression results show that—with a full set of controls included in the estimation—blacks and hispanics are 11 and 8 percent, respectively, more likely to attend college by age 20 than their white counterparts. Furthermore, controlling for the full set of regressors suggests that men are 7 percentage points less likely than women to go to college. Not surprisingly, the results also show that youths with more advantaged family backgrounds (high family income, better-educated parents, older mothers at age of their first birth, and higher likelihood of living with both parents are age 12) and higher levels of high school achievement are more likely to attend college.

Averett, Burton (1996), and Aughinbaugh (2008) all found that possessing the logically advantageous individual and family characteristics included in their studies increased the likelihood of college attendance. Specifically, having higher ability—as measured by the AFQT and ASVAB—, better educated parents, and greater financial stability were among the studies' parallel factors that demonstrated positive impacts on the likelihood of college attendance. Furthermore, the studies mentioned also found evidence supporting the continued divergence of college enrollment rates for men and women, with women enrolling at an increasingly higher rate than men.

In summary, our review of the existing literature offers three key insights that will drive the course of the remainder of this paper: (a) confirmation of the benefits associated with pursuing a higher education in regards to post-college employment opportunities and earning trajectories; (b) a look at the significant determinants of college attendance in terms of personal demographic information, family circumstances and high school achievement; and (c) confirmation of the existence and persistence of the trend of increased college enrollment for females. These three findings, combined with the fact that we are conducting our research as male college students at an institution that we chose to attend, inclined us to delve deeper into this topic. Accordingly, our contribution to the current literature will be as follows: Given that the data from our study comes 20 years after that of Averett and Burton (1996), our results will illuminate the behavior of the variables included throughout that time period, and in turn will offer potentially valuable predictive information concerning which factors will continue to contribute to the likelihood of college attendance. Furthermore, the results of our Chow-test will provide an updated look at the differences in the factors that contribute to the college attendance decision between men and women. In doing so, we observe and demonstrate the persistence of the trend of increased female college enrollment.

II. DATA DESCRIPTION

We chose to extract our data from the National Longitudinal Surveys of Youths 1997. The NLSY97 consists of a nationally representative sample of approximately 9,000 youths who were 12 to 16 years old as of December 31, 1996. Round 1 of the survey took place in 1997. In that round, both the eligible youth and one of that youth's parents received hour-long personal interviews. Individuals continue to be interviewed on an annual basis (NLSY97). Here are our descriptive statistics of the variables in our model.

Table 1

Variable	Obs	Mean	Std. Dev.	Min	Max
college	8984	.2209484	.4149088	0	1
urban	7344	.7734205	.4186468	0	1
white	8984	.5192565	.4996569	0	1
female	8984	.4880899	.499886	0	1
feduc	5707	12.92238	3.800507	1	95
meduc	8010	12.58152	3.600808	1	95
faminc07	6235	57384.02	54579.71	0	279104
SATmath	2466	3.779805	1.211077	1	6

As our intent is to examine the determinants of the respondents' decisions as to whether or not to attend college, we chose a binary dependent variable that indicated college attendance. To find our dependent variable we used a variable from the NLSY97 called highest grade completed ever; this consists of a variety of categories such as: if the individual never completed high school, the individual completed high school, the individual earned his GED, the individual went to a 2-year college, the individual attended a 4-year college, the individual earned a BA, the individual earned a Masters, and the individual earned a doctorate. We split this variable into two categories; *COLLEGE* = 1 if the individual attended *COLLEGE* for 4 years or greater and *COLLEGE* = 0 for everything below attending college for four years. Just over 22 percent of individuals attended college for four years or more.

Our second variable is a dummy variable indicating whether or not the respondent grew up in an urban environment. This was created from a variable in the NLSY97 called *REGION*. *REGION* consisted of *URBAN*, *RURAL* and unknown region. Just over 77 percent of individuals live in an urban environment. For our next dummy variable—an indicator of race—we split a race variable from the NLSY97 to four subcategories; *WHITE*, *BLACK*, *HISPANIC* and *NON-HISPANIC MIXED RACE*. We chose to have our *RACE* variable take a value of 1 if the respondent was white, and zero otherwise. Just fewer than 52 percent of the individuals are white, 26 percent of the individuals are black, 21 percent of the individuals are hispanic and just less than 1 percent of the individuals are non-hispanic mixed race. Our final dummy variable was an indicator of gender, holding a value of 1 if the respondent is *FEMALE* and 0 if *MALE*. 51 percent of the individuals were male and 49 percent of the individuals were female.

Our next set of variables was included to encapsulate the respondents' family situations. Specifically, we chose the educational achievement of both one's mother (*MEDUC)* and father (*FEDUC)*, as we suspect that better-educated parents are more inclined to invest in their children's educations. On average, the individuals' fathers had approximately 13 years of education and on average the individuals' mothers had just over 12.5 years of education. Our next variable was a measure of one's family income (*FAMINC07)* as of 2007. The mean annual family income at this time was $57,384. With that said, it is important to note that the standard deviation of $54,579 suggests that our sample included a range of both low-income and high-income families.

The final variable, SAT math score (*SATMATH)*, was included as a measure of one's cognitive ability. We initially included respondents' SAT verbal (*SATVERBAL)* scores as well, however—given their high correlation with SAT math scores and their statistical insignificance in our regression—we dropped the variable. The NLSY97 standardizes SAT score variables to a range of 1-6, where a value of 1 indicates that the respondent received 200-300, 2 if the individual received 301-400, 3 if the individual received 401-500, 4 if the individual received 501-600, 5 if the individual received 601-700 and lastly a 6 if the individual received 701-800. The mean of the SAT math variable, 3.78, thus indicates that the average SAT math score among respondents was in the latter half of 401-500.

III. CONCEPTUAL FRAMEWORK

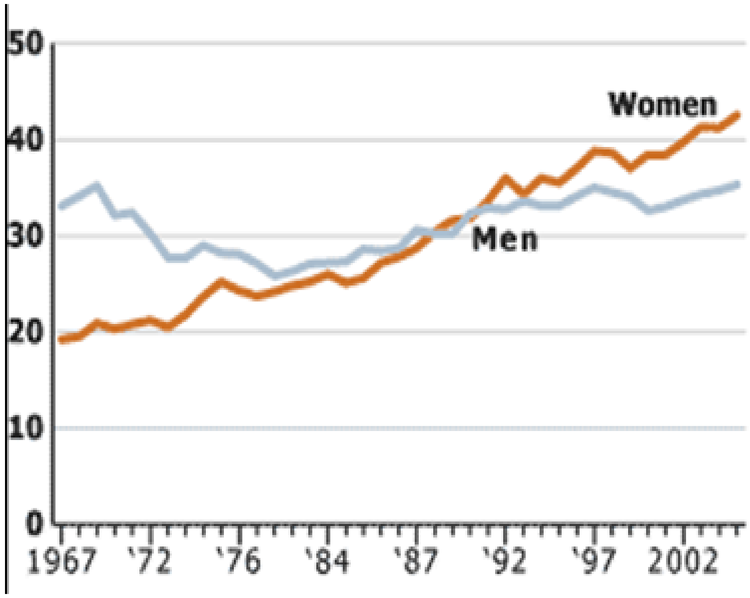
In the pursuit to increase one's human capital, attending college seems to be a logical step. Essentially, people feel that the costs associated with college enrollment are outweighed by the benefits of higher earnings down the road. In terms of net present value, the costs of paying for college—as well as the income foregone by attending college instead of joining the labor force immediately—are a worthwhile investment due to differences in starting salaries and earnings trajectories. In fact, work done by Jennifer Cheeseman Day and Eric C. Newburger shows that those who with a bachelor's degree earn almost twice as much as individuals with just a high school diploma. Evidently, attending (and completing) college proves advantageous for those looking to maximize their worth.

Yet, the decision to pursue a college education is one that is influenced by a variety of factors, and some people may be more likely than others to have that opportunity. To begin, college tuition is expensive, and the law of demand suggests that the higher the price of an education, the lower the demand will be. Accordingly, family income might be a significant determinant in the decision to attend college. Another family characteristic that may influence ones decision to attend college includes the education levels of one's parents. Becker's theory of human capital leads us to suspect that parents who themselves attended college might be more inclined to send their children to college, as the decision to invest in human capital hinges on a rational cost-benefit analysis of the return to investment, as well as cultural influences (Becker 1975). In other words, parents view investing in their children's educations partially as means to acquire financial stability for themselves at some point down the road.

Some individual exogenous variables that are worth examining in regards to one's likelihood of attending college include gender, race, and the region in which one was raised. Finally, it is reasonable to believe that one's level of high school achievement might be indicative of the decision to attend college.

In 2008, Alison Aughbaugh tested a similar set of variables to determine what factors influence the decisions to both attend college and to stay enrolled for more than one year. She regressed a dummy variable indicating one's decision regarding college on a set of exogenous individual characteristics, a set of family characteristics, and a vector conveying high school outcomes. The results show that, after controlling for a full set of variables, blacks, Hispanics, and females are more likely to attend college than their white male counterparts. Furthermore, those from more privileged family backgrounds (high family incomes and better educated parents) and those who performed well in high school are more likely to go to college. We are looking to focus on the gender variable and hope to find significantly different intercepts and slopes for the regression equations of males and females. Aughbaughs study provides evidence supporting an emerging trend that women are becoming increasingly more likely than men to attend college, and if the Chow test described above produces significant results, it may suggest that men and women are choosing to attend college for different reasons. Figure 1, below, displays the time series of college enrollment rates for men and women. The graph indicates that more women were enrolling in college than men by the late 1980's, and also conveys the continued divergence.

Figure 1



(US Bureau of Labor Statistics)

Our conceptualized prediction equation is as follows:

$$Pr(COLLEGE) = f(MOTHER'S\ EDUCATION(+), FATHER'S\ EDUCATION(+), FAMILY\ INCOME(+), SATMATHSCORES(+),$$

$$SATVERBALSCORES(+). ACTSCORE(+), RURAL(+), URBAN(-), BLACK(+), WHITE(-), HISPANIC(+) NON-HISPANIC(+), MIXEDRACE(+), FEMALE(+), MALE(-))$$

IV. EMPIRICAL SPECIFICATION

Our dependent variable reflects the decision of whether or not to attend college. This will hold values between 0 and 1, with 0 indicating no college and 1 indicating college enrollment. We felt that one's family situation would be a useful indicator of the decision to attend college. A 1996 study conducted by Susan L. Averett, Mark L. Burton shows that those whose parents have attended college are more likely to attend college themselves (with nuanced results for the correlations between mothers and fathers and sons and daughters), and a 2008 study by Alison Aughinbaugh has confirmed the persistence of this seemingly reasonable trend. As such, our first independent variable will be mother's education and our second will be father's education. For our third independent variable associated with one's family situation, we chose family income. This was for both the obvious reason that college can be expensive—and thus higher income families are more able to afford the costs of college than low income families—and also because a large family income might be associated with a well-paying job, which itself may be a result of attending college.

In order to attempt to account for the difficulty in capturing one's ability, we initially included independent variables for SAT Math, SAT Verbal, and ACT scores, which have been generally regarded as good indicators of cognitive ability. Typically, higher standardized test scores have been associated with a higher likelihood of college enrollment, *ceteris paribus*. Our next two independent variables are for race and gender. We chose to include race because both Averett and Burton, and Aughinbaugh's studies have shown that certain races are more likely to attend college than others. Finally, we chose to include a gender variable due to the emerging trend that females are now more likely to attend college than males. The existing literature suggests that males and females might have different incentives motivating their decision of whether or not to attend college. As such, we are going to split our data set based on the gender variable in order to see if the regressions have significantly different slopes and intercepts for males and females. If the results of the aforementioned Chow-test convey significance, then such a split would be appropriate.

Prior to such a split, our pooled prediction equation is as follows:

$$Pr(COLLEGE) = \beta_0 + \beta_1 MEDUC + \beta_2 FEDUC + \beta_3 FAMINCOME07 \\ + \beta_4 URBAN + \beta_5 WHITE + \mu$$

The fully interacted model for the Chow test is as follows:

$$College = \beta_0 + \beta_1 MEDUC + \beta_2 FEDUC + \beta_3 FAMINCOME07 + \beta_4 URBAN \\ + \beta_5 WHITE + \beta_6 FEMALE + \beta_7 FEMALE * MEDUC \\ + \beta_8 FEMALE * FEDUC + \beta_9 FEMALE * FAMINCOME07 \\ + \beta_{10} FEMALE * URBAN + \beta_{11} FEMALE * WHITE + \mu$$

The two models after the split are as follows:

$$COLLEGE_{female} = \alpha_0 + \alpha_1 MEDUC + \alpha_2 FEDUC + \alpha_3 FAMINCOME07 \\ + \alpha_4 URBAN + \alpha_5 WHITE + \mu$$

$$COLLEGE_{male} = \beta_0 + \beta_1 MEDUC + \beta_2 FEDUC + \beta_3 FAMINCOME07 \\ + \beta_4 URBAN + \beta_5 WHITE + \mu$$

If these two models have significantly different coefficients, then it is worthwhile to split the model by gender as the decision to attend college might be influenced by different factors between men and women.

V. RESULTS

Table 2 provides estimation results for linear and probit models that regress urban, white, and female dummy variables, as well as a set of family characteristics and standardized test scores, on a dichotic college variable. The following interpretations are derived from our four linear models.

Column (2) depicts the OLS estimates for our main regression model, which includes the aforementioned dummy variables, as well as measures of the individual's father's education, mother's education, and family income in the year 2007. We chose this model as our main regression due to the facts that it had the highest number of observations (3602) and all of the regressors are statistically significant at the 0.01 level. The model explains 15.2 percent of the variability in the decision to attend college.

Table 2

VARIABLES	**(1)Linear** *COLLEGE*	**(2)Linear** *COLLEGE*	**(3)Linear Robust** *COLLEGE*	**(4)Probit** *COLLEGE*	**(5)Probit** *COLLEGE*	**(6)Probit Robust** *COLLEGE*
URBAN	0.023	0.067***	0.0265	0.075	0.199***	0.199***
	(0.033)	(0.017)	(0.033)	(0.096)	(0.054)	(0.054)
WHITE	0.120***	0.102***	0.113***	0.331***	0.378***	0.378***
	(0.030)	(0.016)	(0.025)	(0.087)	(0.050)	(0.064)
FEMALE	0.119***	0.101***	0.119***	0.366***	0.327***	0.327***
	(0.026)	(0.014)	(0.025)	(0.078)	(0.046)	(0.045)
FEDUC	0.024***	0.022***	0.024***	0.075***	0.064***	0.064***
	(0.005)	(0.002)	(0.005)	(0.016)	(0.006)	(0.023)
MEDUC	0.0151**	0.026***	0.026**	0.047**	0.085***	0.085***
	(0.006)	(0.003)	(0.006)	(0.018)	(0.008)	(0.030)
*FAMINC*07	-8.48e-08	4.47e-07***	-3.97e-08	-2.28e-07	1.33e-06***	1.33e-06***
	(1.92e-07)	(0.000)	(1.91e-07)	(5.87e-07)	(0.000)	(0.000)
SATMATH	0.099***		0.089***	0.273***		
	(0.011)		(0.142)	(0.044)		
Constant	-0.505***	-0.491***	-0.522***	-3.10***	-3.093***	-3.093***
	(0.076)	(0.034)	(0.071)	(0.266)	(0.122)	(0.300)
Observations	1242	3692	1221	1242	3692	3692
R-squared	0.19	0.152	0.19			

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

The results indicate that growing up in an urban environment increases the likelihood of attending college by 6.7 percentage points, *ceteris paribus.* Being white increases the likelihood of attending college by 10.2 percentage points, while being female increases the likelihood by 10.1 percentage points, *ceteris paribus.*

The results for the set of family characteristic variables indicate that each additional year of an individual's father's education increases his or her likelihood of attending college by 2.2 percentage points, while an extra year of one's mother's education increases the likelihood by 2.6 percentage points, *ceteris paribus.* Furthermore, an additional $10,000 of annual income increases the likelihood of college attendance by 0.422 percentage points, *ceteris paribus.*

The constant term of -0.491 implies that a non-white male, growing up in a non-urban environment, whose parents have no education and no family income, has zero years of schooling. However, it is unlikely that someone has zero years of schooling so the intercept does not have an intuitive explanation.

While the previous model does demonstrate statistical significance, it does not include an indicator of one's ability. In order to account for the bias that results from this omitted variable, a regressor representing respondents' SAT Math scores were added to the main regression.

Doing so increased the R^2 from 0.152 to 0.19, although the number of observations dropped from 3692 to 1242.

As stated earlier, the *SATMATH* variable ranges from 1-6, with 1 being a score of 200-300, 2 being a score of 301-400, 3 being a score of 401-500, 4 being a score of 501-600, 5 being a score of 601-700 and lastly a 6 being a score of 701-800. Accordingly, the coefficient off 0.099 indicates that an additional 100 points on the SAT math section increases the likelihood of attending college by 9.9 percentage points.

The results from the second regression, displayed in Column (1), demonstrate that adding the *SATMATH* variable rendered the *URBAN* dummy variable insignificant at the 0.1 percent significance level. That being said, the *WHITE* dummy variable remained significant at the 0.01 level, and suggests that being white increases the likelihood of attending college by 12 percentage points, *ceteris paribus.* The *FEMALE* dummy variable also remained significant at the 0.01 level, and shows that being female increases the likelihood of attending college by 11.9 percentage points, *ceteris paribus.*

With the full set of regressors controlled for, an additional year one's father's education increases the likelihood of college attendance by 2.4 percentage points, while the equivalent statistic for one's mother suggests a 1.51 percentage point increase, *ceteris paribus.* Interestingly, with the *SATMATH* variable included, family income is no longer a significant estimator of the decision to attend college.

VI. SPECIFICATION TESTS

A. Multicollinearity

One of the assumptions associated with running a classical linear regression (CLRM) is that there are no exact linear relationships among the regressors. If there are one or more such relationships among the resgressors, it is referred to as multicollinearity. While the issue of multicollinearity impacts most empirical studies, the following specification tests demonstrate the low extent to which our model is affected by this issue.

In order to detect for multicollinearity, we first checked for high, pair-wise correlations amongst our regressors. Table 3 displays the correlations between our regressors. The only significant correlation is between father's education and mother's education, with a correlation of 0.5072. This relationship is reasonable, as educational achievement may be an influential factor in spousal selection. That being said, there were no other significant correlations between our regressors.

Our next test to detect for multicollinearity was to look for high partial correlations between coefficients (refer to Table 4). Once again, there is a significant partial correlation between the estimators for mother's education and father's education. There were no other significant partial correlations.

The next test for multicollinearity was to check whether or not the F-statistics for our auxiliary regressions were greater than 10. An auxiliary regression is a regression of each regressor on the remaining regressors (refer to Table 5). Our first auxiliary regression regressed the urban dummy variable on all of our other right-hand side variables (white, female, father's education, mother's education, and family income as of 2007). This regression yielded an F-statistic of 14.41. Since 14.41 is greater than ten, this test indicates collinearity issues.

Our next auxiliary regression regressed the white dummy variable on the remaining regressors. This model had an F-statistic of 91.89. Being greater than 10, this F-stat is also indicative of collinearity. The auxiliary regression with father's education as the dependent variable yielded an F-statistic of 289.81, indicating high collinearity. Similarly, the auxiliary regression for mother's education has an F-statistic of 282.32, which also indicates high collinearity. None of the remaining auxiliary regressions demonstrated significant collinearity.

Our final test for the detection of multicollinearity was an analysis of the variance inflation factors, or VIF's (refer to Table 6). The results of this test indicate significant collinearity if the VIF is greater than two. None of our regressors had VIF's greater than two, thus indicating a relative lack of multicollinearity amongst our regressors.

B. Heteroskedasticity

The next specification tests required was to detect for heteroskedasticity, or non-constant variance of the error term. In order to test for this, we first graphed a histogram of our squared residuals and checked for constant variance. To the contrary, our histogram in Figure 2 demonstrated non-constant variance, which is a sign of heteroskedasticity. Next, we plotted our fitted values against our squared residuals—displayed in Figure 3 below—and once again saw evidence of heteroskedasticity.

Figure 2

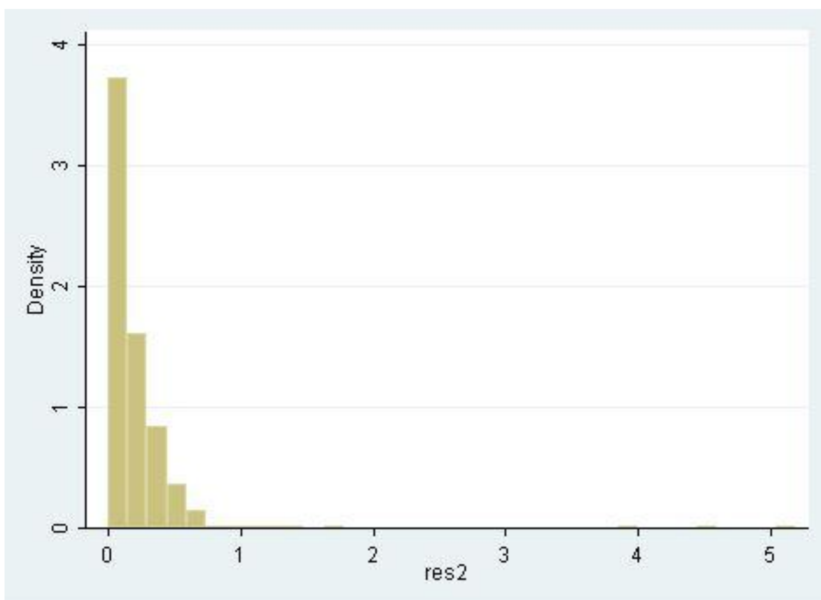
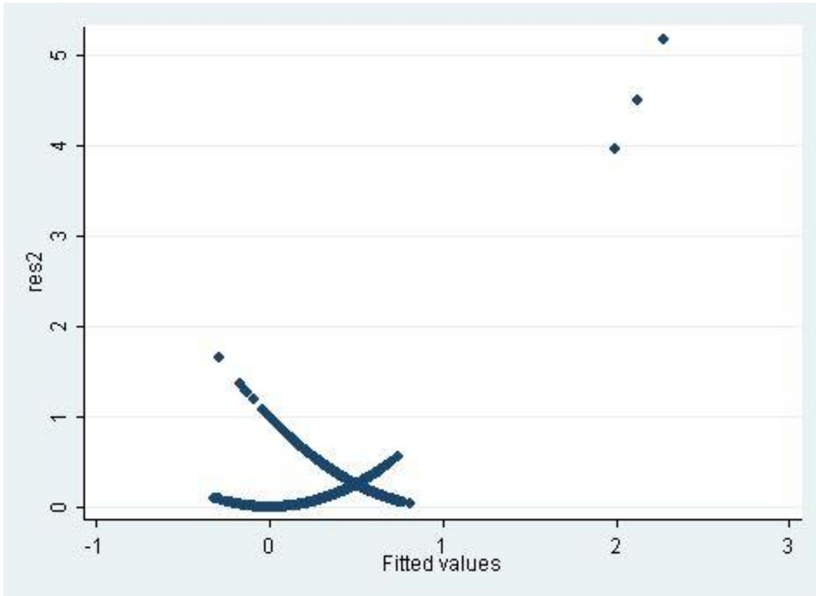


Figure 3



From here, we conducted a Breusch-Pagan test (refer to Table 7). In such a test, the null hypothesis is one of homoskedasticity, while the alternative implies heteroskedasticity. Looking at the results from STATA we concluded to reject the null hypothesis due to the chi-square p-value of 0.00, further indicating our model suffers from heteroskedasticity. (Refer to STATA specification test output)

Our final specification test for the detection of heteroskedasticity is the White test (refer to Table 8). In such a test, we check the null hypothesis of homoskedasticity against an alternative hypothesis of heteroskedasticity. From the p-value of the Chi-square test of 0.00, we can reject the null hypothesis and infer an issue of heteroskedasticity.

Ultimately, our model does suffer from heteroskedasticity. Potential sources of this issue include the presence of outliers in our data and the mixing of observations with different measures of scale. An example of the latter source is our family income variable, as there is a mix of high and low-income households. The heteroskedasticity present in our model does not affect the bias or consistency of our OLS estimators, yet it does render our estimators less efficient and they are no longer the best linear unbiased estimators (BLUE)—they are now simply LUE. To compensate for heteroskedasticity, we added the robustness test to our regression in Table 2.

C. Chow-Test

Due to the emergence and persistence of a recent trend of increased female college enrollment, we decided to conduct a Chow-test in order to determine the value of splitting our sample by gender (output from Chow-test can be seen in Table 9). A Chow-test is a useful way to compare two regressions, as it tests for significantly different coefficients and/or slopes between two samples of data. In order to conduct such a test, we first created a fully interacted model (withholding the *SATMATH* variable to represent a larger sample size):

$$COLLEGE = \beta_0 + \beta_1 MEDUC + \beta_2 FEDUC + \beta_3 FAMINCOME07 + \beta_4 URBAN + \beta_5 WHITE + \beta_6 FEMALE + \beta_7 FEMALE * MEDUC + \beta_8 FEMALE * FEDUC + \beta_9 FEMALE * FAMINCOME07 + \beta_{10} FEMALE * URBAN + \beta_{11} FEMALE * WHITE + \mu$$

The five-step chow procedure is below:

1) $H_o: \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$
   $H_o: \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} \neq 0$
2) $\propto = 0.05$
3) $F = \frac{SSR_{pooled} - (SSR_1 + SSR_2)/k+1}{(SSR_1 + SSR_2)/[n-2(k+1)]} \sim F_{k+1,[n-2(k+1)]}$
   $F_{6,3680} = F^* = 2.10$
   $SSR_{pooled} = 693.456133$
   $SSR_1(female = 1) = 351.072642$
   $SSR_2 = 331.399005$
4) $$F_c = \frac{693.456133(351.072642+331.399005)/6}{(331.399005+351.072642)/3680}$$
   $$F_c = \frac{1.83075}{0.18545} = 9.8719$$
5) Decision rule: $F_c > F^*$, $9.8719 > 2.10$

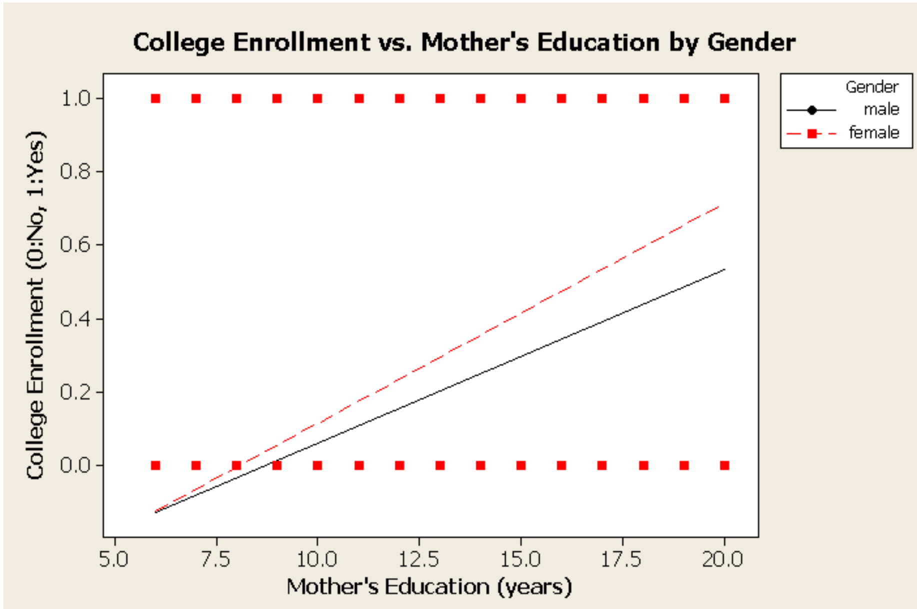
Since the critical value is greater than the test statistic, the results of our Chow-test indicate that it is worthwhile to split the sample, as the regression equation for females has significantly different slopes and intercepts than that of males. See Table 3.

VII. CONCLUSION

The results of our regressions and specification tests encapsulate the factors that influence the decision to attend college, and also convey how some of these factors hold different weights for men and women. For both men and women, being white, raised in an urban environment, and coming from a privileged family background (having educated parents in a high-income family)all increased the likelihood of attending college, *ceteris paribus.* Holding all else constant, based on the differences between the split models from our Chow-test, women are 18.77 percentage points more likely than men to attend college. These results confirm the persistence of the trend that women are increasingly more likely to attend college than men.

Interestingly, as derived from the Chow-test results, the marginal effect of mother's education is greater for daughters than sons, thus mothers' education is more influential in the likelihood of attending college for daughters than for sons. This implies that better-educated mothers are more inclined to afford their daughters the opportunity of higher education. This effect can be observed in Figure 4 below:

Figure 4



The fact that the best-fit line depicting females is steeper than that of males conveys that an additional year of one's mother's education has a greater impact on the likelihood of attending college for females.

Similarly, the marginal effect of family income is greater for daughters than sons, thus a larger family income is more influential in the likelihood of attending college for daughters than sons. This seems to suggest that higher income families are more willing to invest in their daughters' educations than low income families.

While our study does demonstrate that women are more likely to attend college than men, it also leaves room for future investigation. There has been much speculation as to how the differences in the incentives to gain an education between men and women have manifested tangibly as starkly different college enrollment rates for the two groups. Specifically, the impact of the college wage premium might be an area of further analysis, as the decline in uncertainty for future earnings in female college graduates (Cho 2007) might be among the reasons that more women are electing to attend college.

Finally, a potentially interesting implication of the marginal effect of mothers' education on daughters is that as the members of the increasing population of female graduates become mothers of daughters themselves, it follows that the impact of this effect might be accelerated. Yet, such a conclusion hinges on the assumption that this increasing population of female college graduates will go on to mother daughters of their own, when perhaps the initial decision to pursue higher education is a reflection of one's prioritization of her role as a professional over that of a mother. Accordingly, this divergence in college enrollment rates for men and women will continue if the increasing population of well-educated females creates a proportionally large population of college-bound daughters. Alternatively, this divergence will slow down if the women who are pursuing higher education prefer to cultivate their professional careers as opposed to raising a large family.

VIII. APPENDIX: STATA RESULTS

TABLE 3: Correlation among variables.

```
. corr urban white female feduc meduc faminc07
(obs=3692)

             |    urban    white   female    feduc    meduc faminc07
-------------+------------------------------------------------------
       urban |   1.0000
       white |  -0.1010   1.0000
      female |   0.0038  -0.0313   1.0000
       feduc |   0.0621   0.2709  -0.0182   1.0000
       meduc |   0.0296   0.2675  -0.0321   0.5072   1.0000
    faminc07 |   0.0067   0.0435  -0.0318   0.0753   0.0711   1.0000
```

TABLE 4: Partial correlation among explanatory variables.

```
. pcorr urban white female feduc meduc faminc07
(obs=3692)

Partial correlation of urban with

    Variable |    Corr.     Sig.
-------------+------------------
       white |  -0.1238    0.000
      female |   0.0021    0.896
       feduc |   0.0744    0.000
       meduc |   0.0172    0.297
    faminc07 |   0.0045    0.786

. pcorr white urban female feduc meduc faminc07
(obs=3692)

Partial correlation of white with

    Variable |    Corr.     Sig.
-------------+------------------
       urban |  -0.1238    0.000
      female |  -0.0226    0.170
       feduc |   0.1695    0.000
       meduc |   0.1564    0.000
    faminc07 |   0.0181    0.273

. pcorr female urban white feduc meduc faminc07
(obs=3692)

Partial correlation of female with

    Variable |    Corr.     Sig.
-------------+------------------
       urban |   0.0021    0.896
       white |  -0.0226    0.170
       feduc |   0.0026    0.873
       meduc |  -0.0215    0.192
    faminc07 |  -0.0291    0.077
```

```
. pcorr feduc urban white female meduc faminc07
(obs=3692)

Partial correlation of feduc with

    Variable |    Corr.     Sig.
-------------+------------------
       urban |   0.0744    0.000
       white |   0.1695    0.000
      female |   0.0026    0.873
       meduc |   0.4638    0.000
    faminc07 |   0.0416    0.011

. pcorr meduc urban white female feduc faminc07
(obs=3692)

Partial correlation of meduc with

    Variable |    Corr.     Sig.
-------------+------------------
       urban |   0.0172    0.297
       white |   0.1564    0.000
      female |  -0.0215    0.192
       feduc |   0.4638    0.000
    faminc07 |   0.0343    0.037

. pcorr faminc07 urban white female feduc meduc
(obs=3692)

Partial correlation of faminc07 with

    Variable |    Corr.     Sig.
-------------+------------------
       urban |   0.0045    0.786
       white |   0.0181    0.273
      female |  -0.0291    0.077
       feduc |   0.0416    0.011
       meduc |   0.0343    0.037
```

TABLE 5: F-stats, if greater than 10, sign of multicollinearity.

```
. reg urban white female feduc meduc faminc07

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =   14.41
       Model |  13.1235403     5  2.62470806           Prob > F      =  0.0000
    Residual |  671.565517  3686  .182193575           R-squared     =  0.0192
-------------+------------------------------           Adj R-squared =  0.0178
       Total |  684.689057  3691  .185502318           Root MSE      =  .42684

------------------------------------------------------------------------------
       urban |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |  -.1161067   .0153255    -7.58   0.000    -.1461541   -.0860593
      female |   .0018345   .0140778     0.13   0.896    -.0257666    .0294355
       feduc |   .0098487   .0021757     4.53   0.000      .005583    .0141145
       meduc |   .0026034   .0024943     1.04   0.297     -.002287    .0074938
    faminc07 |   3.26e-08   1.20e-07     0.27   0.786    -2.03e-07    2.68e-07
       _cons |   .6631977   .0319299    20.77   0.000     .6005957    .7257997
------------------------------------------------------------------------------

. reg white urban female feduc meduc faminc07

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =   91.89
       Model |  95.2048784     5  19.0409757           Prob > F      =  0.0000
    Residual |   763.82004  3686  .207221932           R-squared     =  0.1108
-------------+------------------------------           Adj R-squared =  0.1096
       Total |  859.024919  3691  .232735009           Root MSE      =  .45522

------------------------------------------------------------------------------
       white |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urban |  -.1320566   .0174308    -7.58   0.000    -.1662316   -.0978815
      female |  -.0206084   .0150098    -1.37   0.170    -.0500369      .00882
       feduc |   .0239393   .0022932    10.44   0.000     .0194433    .0284353
       meduc |   .0252667   .0026278     9.62   0.000     .0201146    .0304187
    faminc07 |   1.40e-07   1.28e-07     1.10   0.273    -1.11e-07    3.91e-07
       _cons |   .0969579   .0359547     2.70   0.007      .026465    .1674509
------------------------------------------------------------------------------
```

```
. reg female urban white feduc meduc faminc07

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =    1.80
       Model |  2.24833882     5  .449667763           Prob > F      =  0.1088
    Residual |   919.30827  3686  .249405391           R-squared     =  0.0024
-------------+------------------------------           Adj R-squared =  0.0011
       Total |  921.556609  3691  .249676675           Root MSE      =  .49941

------------------------------------------------------------------------------
      female |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urban |   .0025112   .0192712     0.13   0.896    -.0352719    .0402944
       white |  -.0248036   .0180653    -1.37   0.170    -.0602227    .0106154
       feduc |    .000408   .0025527     0.16   0.873    -.0045968    .0054128
       meduc |  -.0038072   .0029181    -1.30   0.192    -.0095285    .0019141
    faminc07 |  -2.48e-07   1.40e-07    -1.77   0.077    -5.23e-07    2.70e-08
       _cons |   .5536237   .0384163    14.41   0.000     .4783044     .628943
------------------------------------------------------------------------------

. reg feduc urban white female meduc faminc07

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =  289.81
       Model |  15046.5511     5  3009.31022           Prob > F      =  0.0000
    Residual |  38274.9527  3686  10.3838721           R-squared     =  0.2822
-------------+------------------------------           Adj R-squared =  0.2812
       Total |  53321.5038  3691   14.446357           Root MSE      =  3.2224

------------------------------------------------------------------------------
       feduc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urban |   .5613146   .1240029     4.53   0.000     .3181936    .8044355
       white |   1.199596   .1149098    10.44   0.000     .9743033    1.424889
      female |   .0169878    .106279     0.16   0.873    -.1913837    .2253593
       meduc |   .5303107   .0166855    31.78   0.000      .497597    .5630244
    faminc07 |   2.29e-06   9.05e-07     2.53   0.011     5.16e-07    4.06e-06
       _cons |   4.826125   .2420492    19.94   0.000     4.351561    5.300688
------------------------------------------------------------------------------
```

```
. reg meduc urban white female feduc faminc07

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =  282.32
       Model |  11211.307     5  2242.2614           Prob > F      =  0.0000
    Residual |  29274.891  3686  7.94218421           R-squared     =  0.2769
-------------+------------------------------           Adj R-squared =  0.2759
       Total |  40486.198  3691  10.9688968           Root MSE      =  2.8182

------------------------------------------------------------------------------
       meduc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urban |   .1134877   .1087331     1.04   0.297    -.0996952    .3266706
       white |   .9683936   .1007152     9.62   0.000     .7709305    1.165857
      female |  -.1212394   .0929264    -1.30   0.192    -.3034316    .0609529
       feduc |   .4056122    .012762    31.78   0.000     .3805909    .4306336
    faminc07 |   1.65e-06   7.92e-07     2.08   0.037     9.54e-08    3.20e-06
       _cons |    6.83534    .192273    35.55   0.000     6.458368    7.212312
------------------------------------------------------------------------------

. reg faminc07 urban white female feduc meduc

      Source |       SS       df       MS              Number of obs =    3692
-------------+------------------------------           F(  5,  3686) =    6.18
       Model | 1.0616e+11     5  2.1233e+10           Prob > F      =  0.0000
    Residual | 1.2655e+13  3686  3.4333e+09           R-squared     =  0.0083
-------------+------------------------------           Adj R-squared =  0.0070
       Total | 1.2761e+13  3691  3.4574e+09           Root MSE      =   58595

------------------------------------------------------------------------------
    faminc07 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       urban |   615.1785   2261.041     0.27   0.786    -3817.836    5048.193
       white |   2324.974   2119.779     1.10   0.273    -1831.081     6481.03
      female |  -3415.161   1931.712    -1.77   0.077    -7202.491    372.1682
       feduc |   757.1423   299.2422     2.53   0.011     170.4456    1343.839
       meduc |    712.275   342.2586     2.08   0.037     41.24009     1383.31
       _cons |   45309.35   4572.064     9.91   0.000     36345.33    54273.38
------------------------------------------------------------------------------
```

TABLE 6: Variance Inflation Factor test, if greater than 2, sign of multicollinearity.

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
       feduc |      1.39    0.717815
       meduc |      1.38    0.723083
       white |      1.12    0.889171
       urban |      1.02    0.980833
    faminc07 |      1.01    0.991681
      female |      1.00    0.997560
-------------+----------------------
    Mean VIF |      1.16
```

TABLE 7: Heteroskedasticity test—Breusch Pagan test

```
. ivhettest
OLS heteroskedasticity test(s) using levels of IVs only
Ho: Disturbance is homoskedastic
    White/Koenker nR2 test statistic    : 605.281  Chi-sq(6) P-value = 0.0000
```

TABLE 8: Heteroskedasticity test--White test

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(24)     =   1479.58
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1479.58     24    0.0000
            Skewness |     808.56      6    0.0000
            Kurtosis |       8.43      1    0.0037
---------------------+-----------------------------
               Total |    2296.57     31    0.0000
---------------------------------------------------

. *** We reject the null hypothesis once again due to the chi-square p-value of
>  0.00, further indicating our model suffers from heteroskedasticity. ***
```

TABLE 9: Chow test

VARIABLES	**(1)** ***COLLEGE*** **Pooled**	**(2)** ***COLLEGE*** **Split** ***FEMALE***	**(3)** ***COLLEGE*** **Split** ***MALE***
URBAN	0.067***	0.067***	0.065***
	(0.017)	(0.025)	(0.022)
WHITE	0.100***	0.088***	0.112***
	(0.016)	(0.023)	(0.021)
FEDUC	0.022***	0.019***	0.023***
	(0.002)	(0.003)	(0.003)
MEDUC	0.026***	0.035***	0.021***
	(0.003)	(0.004)	(0.003)
*FAMINC*07	4.22e-07***	6.14e-07***	2.97e-07*
	(0.000)	(0.000)	(0.000)
Constant	-0.436***	-0.464***	-0.438***
	(0.033)	(0.051)	(0.044)
Observations	3692	1773	1919
R-squared	0.140	0.148	0.144
SSR	693.456133	351.072642	331.399005

IX. REFERENCES

Aughinbaugh, A. (2008). Who goes to college? Evidence from the NLSY97. *Monthly Labor Review*

Averett, S., & Burton, M. (1996). College attendence and the college wage premium: Differences by gender. *Economics of Education Review*, *15*(1), 37-49.

Becker, Gary S. *Human Capital: A Theoretical and Empirical Analysis, with special references to Education.* 2d ed. New York: Columbia University Press for NBER, 1975.

Cho, D. (2007). The role of high school performance in explaining women's rising college enrollment. *Economics of Education Review, 26*(4), 450-462. doi:10.1016/j.econedurev.2006.03.001

Mather, M., & Adams, D. (2007). The crossover in female-male college enrollment rates. In *Population Reference Bureau* Retrieved from http//:www.prb.org/articles/2007/crossoverinfemalemalecollegeenrollmentrates.htm

US Bureau of Labor Statistics, *National Longitudinal Survey Program 97*. http://www.bls.gov/nls/nlsy97.htm