Investigating the Effect of Educational Attainment on Marital Status

Michael Iacobelli, Max Pawlowski, Henry Price

Department of Economics, Boston University

EC204: Empirical Economics 2

Dr. Ishita Dey

May 3, 2023

Abstract

Using IPUMS-CPS's extensive library of data, this report aims to investigate the impact of educational attainment on marriage in the United States. This paper utilizes OLS regression methods to analyze and predict a best-fit model to illustrate a possible correlation between education and marriage.

Drawing on a comprehensive literature review, education and marital status affects social and economic welfare. When reviewing the regression results, our findings suggest that for every additional year of education attained, the probability of marriage increases by 2.37%. Furthermore, out of the eight regression models, the multiple quadratic regression model provided the best fit for explaining our relationship.

In addition to our findings highlighting the importance and social relevance of education on marriage, our paper provides practical implications for policymakers. For instance, subsidizing education to make it more affordable and available can improve relationships and the chances of a healthy marriage.

I Introduction

The United States (U.S.) has the highest rate of children living in single-parent households, and the rate continues to rise. Almost a quarter (23%) of U.S. children under the age of 18 live in a single-parent family (Kramer, 2019). Moreover, the percentage in the U.S. is projected to increase further, making it worth investigating because children from single-parent families are more likely to develop high-risk behavioral health challenges (Casey, 2022).

According to the Census Bureau, the U.S. high school completion rate of the population 25 and older equated to 91.1% in 2021. Although high educational attainment, the U.S. continues to see an increasing trend of single-parent families. Conversely, according to the World Bank, Sub-Saharan Africa has the highest rate of educational exclusion (Klapper Panchamia, 2023). However, with poor educational infrastructure, countries in Sub-Saharan Africa, such as Mali (1%), are far below the global single-parent rate average of 6.8% (Kramer, 2019).

Formal research on educational attainment outcomes finds that increasing years of education and being married positively correlate with health outcomes such as suicide risk (Øien-Ødegaard et al., 2021). Furthermore, research finds that college-educated adults marry later in life when they are more financially secure, and higher education increases their chances of a long-term marriage (Wang, 2015).

We hypothesize that the relationship between an individual's educational attainment and the probability of being involved in a marriage is positively related. Before conducting the study at a comprehensive level and based upon the information that we are privy to, we have determined through preliminary analysis via a simple linear regression that a one-year increase in educational attainment increases the probability of someone being married by 3.46%. Using this information this can be seen as early confirmation of our hypothesis.

Summarizing our major findings, we found that the analysis shows a positive correlation between educational attainment and the probability of being married at the mean of the data set. There is a 3.46% increase in the probability of being married for every additional year of education in the simple linear model. However, using a multiple regression quadratic model, which was a better fit, we found that a one-year increase in educational attainment results in a 2.37% increase in the likelihood of marriage.

Our multiple regression model uses sample data entirely downloaded from IPUMS-CPS for 2019, 2020, 2021, and 2022, including controls for regional and demographic factors to create a more realistic model. Analyzing the impact of marriage on educational attainment, the primary x and y variables chosen for this analysis are educ_rev, which is a cleaned version of the IPUMS-CPS educ variable that represents a person's highest year of education, and married, which is a dummy variable made from the IPUMS-CPS marst variable that has a value of 1 if a person is married and a value of 0 otherwise.

The progression of this empirical paper will be as follows: Section II includes the literature review of major research surrounding the respected variables in our analysis, which further assisted in our paper. Section III explains the methodology of our econometric model utilized in our regression analysis and how we manipulated our data set for the desired outcomes. Section IV includes our data set information and an explanation of our descriptive statistics tables. Our description of results from experimenting with various functional forms in our regression analysis and how we controlled for endogeneity in our results can be found in Section V. Based on our results, conclusions of our results and final thoughts are in Section VI. In the final Section VII of our paper, you will find sources cited as well as the appendix that contains detailed tables, summary statistics, and other aspects of this material surrounding our paper.

II Literature Review

The research that we are conducting takes into consideration prior studies that have been published and seeks to create another suggestion into the relationship between education and the probability of marriage for individuals. This section focuses on studies that have involved education making an impact on variables relating to marriage and the stability of a marriage.

Education can affect numerous aspects of life beyond the probability of marriage. Abdullah, Doucouliagos, and Manning (2013) explain how education can impact income inequality. Utilizing a technique known as meta-regression analysis, the team reviewed over 1,200 studies with numerous different outcomes in order to come to their conclusions (Abdullah, et al. 2013). The result of this study proved to be that education is, in fact, significantly associated with lower income inequality, indicating an inverse relationship. The study also suggests that the strength of the relationship can vary with different factors, such as region and cultural standards. This important delineation is relevant to our paper as education's impacts on other dependent variables can certainly be influenced by other instances of culture and region, which is a primary reason for our inclusion of a regional variable in order to control for any omitted variable bias.

When it comes to the relationship between education and marriage, Berna Torr (2011) provides insight into the evolution of this dynamic between the years 1940 and 2000 (Torr 2011). The study takes data from the United States Census Bureau over this sixty-year period and applies logistic regression techniques in order to analyze the likelihood of marriage by level of education, gender, and race. The study finds that over this period of time, the relationship between education and marriage has, in fact, changed. For women, this change has been that there is a significant increase in the relationship between their education level and their probability of marriage. For men, the opposite is the case, with the relationship actually decreasing over the sixty-year time frame. Potential reasoning for this is the fact that more women have had the opportunity to go to college and gain higher education than what was the case in 1940. This, in turn, makes their earnings less reliant on a husband, decreasing the necessary relationship between the level of education and marriage for a man. As a result, Torr suggests that these changes are linked to broader societal changes in the United States. A difference in the probability of marriage between men and women identified in this regression paper has led us to create an interaction term between education and a gender dummy variable in order to explain this potential difference in our results.

At a similar time as this study, Köppen (2011) released her thesis and research into how education can affect marriage and cohabitation in Germany as well as France (Köppen 2011). Her findings in Germany happen to bring key insights into the research we are performing in this paper. Köppen used data from the German Socio-Economic Panel Study (SOEP) in order to supply her sample for her research into the relationship between education and marriage stability. She ran a logistic regression analysis while also controlling for other factors such as age, gender, and income. What she found was that when individuals had a greater education level, marital stability was much higher and more likely. Marital stability is defined in the study as a marriage with no "disruptions," such as separations, divorces, or anything of the sort. So, while not exactly the married variable that we are employing, this study also points toward a direct relationship between education and the probability of marriage. In the study, however, there is a caveat. Köppen notes that there is a distinct difference between the impact of this relationship in Western Germany and Eastern Germany, where Western Germany sees a much greater effect of this relationship than Eastern Germany. She points to the possibility that cultural and economic differences play a role in this regional difference, which is very important to our approach to our study. In order to determine if there is a regional difference in the impact in the United States, we have employed regional variables that can help us discover differing effects of this relationship in our study.

Hayford, et al. (2016) gives slightly more recent information regarding the connection between education and family formation in the United States (Hayford and Guzzo, 2016). The team used data from the National Survey of Family Growth (NSFG) in order to complete their research. The study employs discrete-time event history models to analyze. This modeling is when the probability of an event occurring given specific time periods, discretely analyzing those larger moments as opposed to a specific point in time. More specifically, the dependent variable in this study was discussing childbearing while not being married. Notably, Hayford and her team explore if educational levels and the length of one's time in school affects this number and gives a piece of a larger picture when it comes to modern family life in the United States. Given their data and analysis, the team concluded that women and men who have attained a higher level of education are less likely to marry at a young age and have children, while there is a higher number of people who are less educated who are having children at a younger age either out of wedlock or marrying as a result. This survey and analysis can give us insight into the likelihood of respondents in our data set being married, as age can play a role in the likelihood someone is married as well as the level of education someone may be able to obtain is somewhat dependent on age. A caveat to this study, however, is that this could perhaps just say more about when people are marrying, as opposed to the likelihood of the event of marriage in general. Our paper aims to shed more light on the likelihood of marriage, given there are many factors that can contribute to it.

Due to extensive research and analysis of the preceding literature that we have taken into account, we understand the nature of the study that we are conducting. We can see that education certainly plays an integral role in the likelihood of marriage not just in the United States of America but also around the world. The following sections will build upon this literature and provide more insight into this very complicated question.

III Econometric Model

(1) Simple Linear Model

$$married = \beta_1 + \beta_2 educ\_rev + e$$

To begin, we will examine the simple linear regression model. This model only contains one independent variable and thus is the most simple form that a regression model can take. Additionally, as the dependent variable is a dummy variable, this is a linear probability model and serves to illustrate the effect of an individual's years of schooling on their probability of being married[1]. The constant parameter, β_1, represents the probability of an individual being married having completed 0 years of schooling, and the coefficient parameter, β_2, represents the change in probability of an individual being married given a one-year increase in their years of school completed. As this model's independent variable is in linear form, the slope of the model is fixed for all values of x at β_2. The elasticity of this model at the means can be calculated using the equation $\beta_2 \frac{\overline{educ\_rev}}{\overline{married}}$, representing the percentage increase in the probability of an individual being married due to a one percent increase in their years of school completed. The semi-elasticity of this model at the mean can be calculated using the formula $\beta_2 * \frac{100}{\overline{married}}$, representing the percentage increase in the probability of an individual being married due to a one-year increase in their years of schooling. The error term, e, accounts for any error within the model due to things such as but not limited to omitted variables, functional form, and measurement.[2]

(2) Simple Quadratic Model

$$married = \beta_1 + \beta_2 educ\_rev^2 + e$$

This model is almost exactly identical in form to the simple linear model, the only difference being that in the quadratic model, the independent variable is squared. As such, the interpretation of the constant parameter, β_1, remains the same,[3] and the coefficient parameter, β_2, is interpreted as the increase in the probability that an individual is married due to a 1 unit increase in years of schooling squared. Due to the nonlinear functional form of this model, the slope is not constant for different values of educ_rev. The slope at the mean can be found using the formula $2\beta_2\overline{educ\_rev}$, the elasticity at the means can be found using $2\beta_2\overline{educ\_rev} * \frac{\overline{educ\_rev}}{\overline{married}}$, and the semi-elasticity at the mean can be found using $2\beta_2\overline{educ\_rev} * \frac{100}{\overline{married}}$. The interpretations of slope, elasticity, and semi-elasticity are consistent with the linear model.

(3) Simple Cubic Model

$$married = \beta_1 + \beta_2 educ\_rev^3 + e$$

The last simple regression model that we will examine is the cubic model. Once again, this model is almost identical in form to the simple linear model, the only difference being that the independent variable is cubed. Similar to the simple quadratic model, the slope

---

[1] This interpretation is true for all population regression models discussed in this paper

[2] This interpretation of the error term applies to all population regression models discussed in this paper

[3] This is because $educ\_rev^2 = 0$ when $educ\_rev = 0$

is different at different values of educ_rev. Thus, the slope, elasticity, and semi-elasticity must be calculated at a specific value of educ_rev. At the means, the slope can be calculated using the formula $3\beta_2 * \overline{educ\_rev}^2$, the elasticity using $3\beta_2 * \overline{educ\_rev}^2 * \frac{\overline{educ\_rev}}{\overline{married}}$, and the semi elasticity using $3\beta_2 * \overline{educ\_rev}^2 * \frac{100}{\overline{married}}$. Once calculated, the interpretations are consistent with the linear and quadratic models.

(4) Multiple Linear Model

$$married = \beta_1 + \beta_2 educ\_rev + \beta_3 age + \beta_4 white + \beta_5 taxinc + \beta_6 female + \beta_7 north + \beta_8 east + \beta_9 city + e$$

In the multiple regression model, we include variables that could be a source of omitted variables bias[4] in hopes controlling for endogeneity. For the purposes of this analysis, we decided to introduce 7 additional x variables: age, white, taxinc, female, north, east, and city.[5] When these variables are introduced, the interpretation of the main population parameters must now account for the other variables within the model. Thus, the interpretation of the constant, β_1, becomes the probability that an individual is married given they have completed 0 years of schooling, are 0 years old, are not white, have a taxable income of 0$, are a male, live in the south, live in the west, and do not live in a metropolitan area. The interpretation of the main x variable's coefficient parameter, β_2, becomes the change in the probability that an individual is married due to a one-year increase in their level of schooling, holding age, white, taxinc, female, north, east, and city constant. The dummy variables' coefficient parameters ($\beta_4, \beta_6 - \beta_9$) are interpreted as the change in probability of an individual being married due to the respective dummy variable being positive as opposed to in its base category, holding all other variables constant. The parameters for the additional quantitative independent variables (β_3, β_5) are interpreted as the change in the probability that an individual is married due to a one-unit increase in the respective variable, holding all other variables constant.

The demographic control variables included in this model are age, white, taxinc, and female. As age is a limiting factor in the level of education that someone can complete, we expect there to be a strong correlation between age and education; an individual who is 18 cannot possibly have completed 18 years of school. Additionally, there is likely a very strong correlation between age and married as individuals are likely to postpone getting married until they are older and in a more stable position to do so. The white variable is included because historically, the white population has greater educational attainment than the non white population and thus, the two variables must be correlated. This was observed in the 2021 United States Census survey, where the percentage of adults over 25 who had completed high school was 95.1% for the white population, 90.3% for the Black population, 92.9% for the Asian population, and 74.2% for the Hispanic population. The taxinc variable was included for two reasons. First, there is known to be a strong correlation between educational attainment and income, and second, there is likely a correlation between income

[4] Variables that could be correlated with the main x or y variable; $corr(OV, x) \neq 0$, $corr(OV, y) \neq 0$, control for x

[5] see table 1 for descriptions of variables

and married as people are more likely to get married when they are more financially stable. The final demographic control variable, female, was included as a historical relationship is observed between sex and educational attainment. This is consistent with the 2021 US Census survey which found that "29.4% of men age 25 and older had completed a high school diploma or GED as their highest level of educational attainment, compared with 26.5% of women age 25 and older," along with "of adults age 25 and older who had completed a bachelor's degree or more, 53.1% were women and 46.9% were men."

In the multiple regression model, the slopes, elasticities, and semi-elasticities must be calculated with respect to a specific independent variable. In the multiple linear model, the formulas are very similar to the simple linear model, the only difference being that we must specify which x variable is being examined. The formulas for slope, elasticity, and semi-elasticity, respectively, are $\frac{\Delta x_i}{\Delta married} = \beta_i$, $\varepsilon_i = \beta_i \frac{x_i}{married}$, and $\eta_i = \beta_i \frac{100}{married}$, calculated at the means. Respectively, these are interpreted as a change in the likelihood of an individual being married due to a one unit increase in x, a percentage change in the likelihood of being married due to a one percent increase in x, and a percentage change in the likelihood of being married due to a one unit increase in x, calculated at the means, holding all other variables constant.[6]

(5) Multiple Quadratic Model

$$married = \beta_1 + \beta_2 educ\_rev^2 + \beta_3 age + \beta_4 white + \beta_5 taxinc + \beta_6 female + \beta_7 north + \beta_8 east + \beta_9 city + e$$

The multiple quadratic model is identical in form to the multiple linear model, the only difference being that the main independent variable, educ_rev, is replaced with educ_rev^2 in order to fit the data better. The constant coefficient is interpreted the same as in the previous model, with the exception that instead of years of schooling being equal to zero, years of schooling squared are equal to zero. The slope, elasticity, and semi-elasticity formulas remain the same in this model as in the previous model, the only difference is that for the main x variable, these are now calculated using the same formulas as in the simple quadratic model.

(6) Multiple Quadratic Model With educ_female Interaction Term

$$married = \beta_1 + \beta_2 educ\_rev^2 + \beta_3 age + \beta_4 white + \beta_5 taxinc + \beta_6 female + \beta_7 north + \beta_8 east + \beta_9 city + \beta_{10} educ\_rev * female + e$$

Now that the fundamental regression models have been examined and interpreted, we can start introducing interaction terms to better examine the marginal effect that the main x variable has on the y variable, given different control variables. In this paper, we will examine the marginal effects of the female dummy variable, the white dummy variable, and the city dummy variable.

To begin, we will examine the model containing the educ_female interaction term. The goal of introducing this interaction term is to illustrate the difference in the marginal effect of education on the probability of being married between females and males. Given that in

[6] For dummy variables, this equates to the difference between the base case and the positive case

recent years, females in the US have higher educational attainment on average, some kind of marginal difference is expected(US Census Bureau). Empirically, this marginal effect is calculated by taking the partial derivative of married with respect to educ_rev using the formula, $\frac{\partial married}{\partial educ\_rev}$. This will result in the sample regression line, $\widehat{married} = 2b_2 educ\_rev + b_{10} female$. For all three models containing interaction terms, the marginal effect will be calculated at the mean level of educational attainment, $\overline{educ\_rev}$

(7) Multiple Quadratic Model With educ_white Interaction Term

$$married = \beta_1 + \beta_2 educ\_rev^2 + \beta_3 age + \beta_4 white + \beta_5 taxinc + \beta_6 female + \beta_7 north + \beta_8 east + \beta_9 city + \beta_{10} educ\_rev * white + e$$

Now we will introduce the interaction term educ_white into the main regression model to illustrate any difference in the marginal effect of educational attainment on the probability of being married between whites and non whites. Seeing as there are differences in educational attainment within different race categories, it is likely that there will be an observed difference in marginal effects. This is calculated the same as in the previous model and will result in the sample regression line, $\widehat{married} = 2b_2 educ\_rev + b_{10} white$.

(8) Multiple Quadratic Model With educ_city Interaction Term

$$married = \beta_1 + \beta_2 educ\_rev^2 + \beta_3 age + \beta_4 white + \beta_5 taxinc + \beta_6 female + \beta_7 north + \beta_8 east + \beta_9 city + \beta_{10} educ\_rev * city + e$$

Finally, the educ_city interaction term is introduced into the model to display any difference in the marginal effect of education on the likelihood of being married between those that live in a metropolitan area and those that do not. Once again, this is calculated using the same method as the last two models and will result in the sample regression line, $\widehat{married} = 2b_2 educ\_rev + b_{10} city$.

Hypothesis Testing

Once these regression models have been processed and results have been obtained, we can use statistical methods to test hypotheses relating to this analysis. For the purposes of this analysis, we will be testing two hypotheses that we believe to be the most relevant to this topic. The first test is a right-tailed t-test of β_2 in the simple and multiple linear models (models 1 and 4). The second and final test is a left-tailed t-test that will conclude if there is statistically significant evidence to suggest that the effect of one additional year of education on marital status is less than 0.0375. This hypothesis was formed based on a 2015 report that found that of adults 25 and older, 65% of those with a completed college degree, compared to 50% those with nothing more than a high school diploma(Parker & Stepler, 2017). This 15% difference converted to probability and divided over 4 years equates to a one-year change of 0.0375. Refer to Table 6 for formal definitions of these tests.

IV Data and Descriptive Statistics

To analyze our research question, the paper uses repeated cross-sectional data retrieved from IPUMS-CPS. Using the annual (ASEC) supplement from 2019, 2020, 2021, and 2022, the precisely chosen years are pre and post-COVID-19 pandemic in order to introduce a degree of variability. The two main variables chosen for our investigation are, respectively, the independent variable educ and marst as the dependent variable. The independent variable educ indicates the respondents' educational attainment by the highest school year or degree completed. Furthermore, the dependent variable, marst, provides each respondent's current marital status. This includes six categories: married spouse present, married spouse absent, separated, divorced, windowed, and single/never married.

Before running the simple regression of marst on educ, the data had to be cleaned of all observations containing not-in-universe (NIU) codes and prepared for analysis. Following, a revised version of the variable educ was generated as educ_rev as an exact copy of the educ variable. The new variable educ_rev is then correctly re-coded to the corresponding observation of educ to represent the highest level of school completed.

Marst is a categorical value taking six different categories. After dropping the category widowed, a dummy variable, MARRIED, was generated. The variable classifies married as married, spouse present or married, spouse absent. Not married is defined as separated, divorced, or never married/single. The married variable takes the value of 1 if an individual is married and 0 if the respondent is not married. After cleaning the data set, we were left with 487,931 observations.

For our multiple regression analysis, to refine and develop a more accurate model, we used the following additional variables: age, sex, race, taxinc, metro, and region. As seen in Table 1, categorical variables were cleaned by replacing and dropping any missing variables and recoded as dummy variables taking a value of either 1 or 0. For the race variable, we excluded any uncertain observations that were not clear and subsequently dropped races that were listed as unspecified or respondents who were listed as two or more races.

Collected and available in appendix table 2, summary statistics from these revised variables were created to understand the samples' properties better. The average age of a person in our sample was 44.52 years old, having completed a mean of 13.54 years of schooling with a standard deviation of 3.017. The proportion of the married sample, as indicated by the mean, is 56.79%, with a standard deviation of 0.4954. The sample proportion of females is 50.6% and 49.4% of males, close to equal representation. The proportion of white individuals in the sample is 75.7% (any other identified race equals 24.35%). The average likelihood of one of the surveyed individuals being from the north region of the United States is 64.1%. Furthermore, the average likelihood of someone from the east region is 50.7%. Lastly, the mean of the sample proportion for city, a dummy variable created from metro, indicates that 79.6% live in a metropolitan area.

When sorted by being married or not married (Refer to Table 3(A)), we can observe that those who were classified as married, on average, have 14.16 years of schooling, which is 1.3 years of additional schooling than those who were classified as not married and received on average 12.86 years of schooling. Moreover, married individuals had an average age of 51.03 years, a 14.49-year age gap compared to the average of those not married. There is an 80.8%

higher chance that a person not married lives in a city. An interesting observation is that married people produce an income of \$46,642, more than double the amount of \$21,545 for a person not married, representing a sizeable difference. Another category that produces an intriguing finding is that of WHITE. Those surveyed on race indicate that 80% are married, whereas those who are not married people who identified as white decreased by 9.6

A few notable observations are found by organizing the summary statistics with our dummy variable white (Refer to Table 3(B)). Firstly, those identified as white obtained 13.56 years of schooling, only a slight difference of -0.05 years compared to those who do not identify as white and received, on average, 13.61 years of schooling. Although receiving similar years of education, the gap in taxable income is considerably large. Secondly, on average, a person who identifies as white produces \$35,947 of taxable income, whereas a person who does not report a taxable income of \$29,999 on average.

There is an intriguing finding when examining our data by summary statistics sorted by whether the respondent was female or not female (Refer to Table 3(C)). The dummy variable, female = 0, indicates male, highlighting a gender gap. When observing the taxable income, in our sample, those who identified as female, on average, produced \$28,914 compared to men who produced a taxable income of \$41,746. The corresponding taxable income provides insight into the potential disparities and inequalities in the U.S. workforce. Although noted that taxable income and average pay are not interchangeable, our sample yields that the average taxable income for women is 69.26% of the average taxable income of males. 12.74% less than what Pew Research Center reported as the U.S. gender pay gap (Aragão, 2023).

When looking at the data sorted according to the north variable (Refer to Table 3(D)), on average, those who identified as from a northern region of the U.S. have completed 13.60 years of schooling, and 55.2% on average are married. On the contrary, those who do not identify as part of the northern region reported, on average, 13.52 years of schooling, slightly -0.08 years than those from the north. Furthermore, on average, 53.6% are married from a region, not from the north. This information provides insight into the demographic and regional differences in the United States and suggests a positive correlation between educational attainment and marriage.

Regarding demographic and regional differences in the United States, when sorting the data according to the east variable (Refer to Table 3(E)), we do not see the same positive correlation as in the northern region of the U.S. Those who identified to be from the eastern region of the U.S. have completed an average 13.75 years of schooling and 53.9% on average are married. Whereas the respondents who identified as not being from the East completed an average of 13.39% years of schooling and, on average, were 55.3% married. Observing the data from the east variable and comparing it to the north variable previously, there is a difference between the relationship between being married and educational attainment in regions of the U.S. Referring to the past literature review, the report by Katja Koppen concluded a similar finding in the difference of education and marriage stability between West and East Germany.

Lastly, an interesting relationship can be drawn by categorizing our data in summary statistics of whether respondents live in a city (Refer to Table 3(F)). Respondents who live in a city tend to have completed an average of 13.71 years of schooling at a mean age of 44.05 years. Note that those who responded that living in a city earned a taxable income significantly larger than those not living in a city. The average taxable income for those who live in a city is 37,935, and those who do not is 24,788. Although, on average, years of schooling is 13.06, a slightly older age of 46 years a lower taxable income, those who do not live in a city responded being married 57.4% compared to those who live in a city reported being married 53.9%—a 3.5% difference.

V Estimation Results

(1) Simple linear Model

$$\widehat{married} = \underset{(0.00348)***}{0.0760} + \underset{(0.000250)***}{0.0346}\ educ\_rev, R^2 = 0.045$$

In order for a regression model to be BLUE for a population, several assumptions must be met. To begin, the expected value of the error term must be equal to zero, $E(\hat{e}) = 0$. This assumption can be verified by evaluating the summary statistics of the error term for this model. This results in $E(\hat{e}) = -0.00000000118 \approx 0$, which satisfies this assumption. Secondly, the variance of $\hat{e}$ must be constant for all values of x, that is, $var(\hat{e}) = \sigma^2$, for all x. This assumption can be tested Breusch–Pagan/Cook–Weisberg test for heteroskedasticity. In this case, we reject the null hypothesis that the variance of $\hat{e}$ is constant for all x in favor of the null that it is not constant at a 0.01 level of significance. The third assumption is that there is no clustering nor autocorrelation within the data set, that is, $corr(e_i, e_j) = 0$. As this sample data is personal level data, we can assume that this assumption is met as it is unlikely that two individual's survey answers are impacted by others. The final assumption, which is not required for a regression line to be BLUE if the sample is sufficiently large but is required for hypothesis testing, is normality in the error term, $\hat{e} = N(0, \sigma^2)$. Referring back to the summary statistics table, the skewness of $\hat{e}$ is -.1570051, and the kurtosis is 1.213517, indicating a non normal distribution that is slightly left skewed.

No we can begin interpreting the results of the regression analysis.[7] Firstly, the constant, b_1, is interpreted as the probability that an individual is married having completed 0 years of schooling. In this model, this probability is 0.076, or very low. This is in line with what we would expect for a couple of reasons. Namely, school is both vital to one's social development, as well as one of the main sources of networking throughout one's early life and into adulthood. So, forgoing this entirely should undoubtedly have a negative impact on one's social development and thus their likelihood of being married. Secondly, people who have not completed any school are likely very young and inherently have a lower probability of being married. The coefficient is 0.0346, indicating that the probability of an individual being married increases by 0.0346 due to a one-year increase in their years of education. This indicates a positive relationship and is, once again, in line with our hypothesis. The elasticity of this model, calculated using the equation mentioned in the econometric model section, is

[7] Regression results for all models can be found in table 4

0.86, indicating that the probability of an individual being married increases by 0.86% due to a 1% increase in their level of education, evaluated at the means. The semi-elasticity is 6.337, indicating that a one-year increase in an individual's level of educational attainment results in a 6.337% increase in their probability of being married, evaluated at the means. The R^2 of this model is 0.045, indicating that 4.5% of the variance of the married variable is explained by the educ_rev variable.

(2) Simple Quadratic Model

In the simple quadratic model, educ_rev is squared in order to account for any functional form error. In this model, $E(\hat{e}) = -0.00000000118 \approx 0$, satisfying the first assumption. The assumption of homoskedasticity is violated as this depends on the data which we have previously concluded is heteroskedastic. The third assumption of $Corr(e_i, e_j) = 0$, is assumed to be satisfied as this is highly unlikely in personal-level data. The fourth and final assumption of normality is violated as it depends on the data which we have previously concluded is not normal.

The constant term, $b_1 = 0.290$, indicates that the probability of an individual being married is 0.29, having completed 0 years of schooling squared. This number is higher than in the previous model because of the difference in functional form. The coefficient term, $b_2 = 0.00132$, thus a one unit increase in $educ\_rev^2$ results in a 0.00132 increase in the probability of being married. The slope at the mean level of education is 0.036, illustrating that a one-year increase in education results in the likelihood of being married to increase by 0.036. The elasticity is 0.8904, indicating that an individual's probability of being married increases by 0.8904% due to a one-year increase in their level of schooling, evaluated at the means. The semi-elasticity is 6.5613, indicating that a one-year increase in an individual's level of educational attainment results in a 6.5613% increase in their probability of being married, evaluated at the means. This model has an R^2 of 0.049, while the interpretation of this does not change from that of the previous model, this higher value indicates a slightly better fit than the linear model. These results are all very similar to those of the previous model, which is consistent with what was expected. The one key difference is that the results from this model are slightly more accurate as this model fits the data better.

(3) Simple Cubic Model

The results of the simple cubic model were briefly examined and due to the R^2 being less than that of the simple quadratic model, indicating a worse fit for the data, will not be examined further.

(4) Multiple Linear Model

$$\widehat{married} =$$
$$\underset{(0.00414)***}{-0.278} + \underset{(0.000242)***}{0.0229}\, educ\_rev + \underset{(3.93e-05)***}{0.0102}\, age + \underset{(0.00164)***}{0.111}\, white + \underset{(8.59e-09)***}{4.50e-07} taxinc +$$
$$\underset{(0.00140)***}{-0.00531} female + \underset{(0.00151)***}{-0.00583} north + \underset{(0.00145)***}{-0.0325}\, east + \underset{(0.00175)***}{-0.0215}\, city, R^2 = 0.201$$

As explained in the econometric model section, in the multiple linear model we introduce additional x variables to control for endogeneity. That is variables that could be correlated with the main x or y variables. The constant term in this model, $b_1 = -.278$, indicating that the probability of someone being married given that they have 0 years of schooling, are 0 years old, are not white, have a taxable income of \$0, are a male, live in the south, live in the west, and do not live in a city, is -.278. This result is obviously not helpful nor realistic for two reasons. Namely, because the probability of someone being married cannot be below 0 but also, it is not possible for an individual to be married at zero years of age. This model also shows that holding all control variables constant, a one-year increase in the level of schooling results in a 0.0229 increase in the probability of being married. This result is useful in that it shows a positive correlation, but it is highly unlikely that this change in probability is constant for all levels of education. For example, the change in probability of being married from someone completing their 8^{th} year of education is likely much lower than the increase from completing their 22^{nd} year. A surprising conclusion is that (holding all else constant) age is not as important of a factor in one's probability of being married as one would expect. This model predicts that a one-year increase in one's age causes a 0.0102 increase in their probability of being married; a fairly insignificant change. Moreover, this model estimates that the probability of being married is 0.111 higher for white people compared to non white people. This is a very interesting result as we expected age to be one of if not the most influential factors on one's probability of being married. But, in fact, the race of an individual has the greatest impact by a significant margin. The coefficient of taxinc displays the increase in one's likelihood of being married given a \$1 increase in taxable income. In order to interpret this coefficient more realistically, the coefficient is multiplied by 10,000, which gives us that a person's likelihood of being married increases by 0.0045 due to a \$10,000 increase in taxable income, holding all else constant; showing that income has a negligible effect on marriage likelihood. The effect of being a female on the likelihood of being married is also quite low, only -.00531 holding all else constant. This result is expected as the population is known to be relatively evenly divided between sexes. The R^2 of this model is 0.201, indicating that 20.1% of the total variation in the likelihood of being married is explained by the independent variables included in this model. This value cannot be compared with those of the simple models as R^2 increases substantially for each additional x variable added to the model.

(5) Multiple Quadratic model

In the multiple quadratic model, the main y variable is squared to adjust for any functional form error. In this case, as shown by the R^2 of 0.202, this model fits the data slightly better than model 4. Due to this, the estimates from this model are inherently more accurate than in the previous model. As the only variable that changed between this model and model 4 is the main x, the only estimates that must be reinterpreted are the constant, $b_1 = -0.132$, and the coefficient of educ_rev, $b_2 = 0.000875$. The constant term, once again, is not useful for the same exact reasons as in model 4. The coefficient is not useful as it relates to changes in years of education squared, which is not a useful measure of education. In order to make this value more intuitive, we must convert it to years. This results in a slope at the means of 0.0237, illustrating that a one-year increase in educational attainment results in the likelihood of a person being married to increase by 0.0237. The elasticity at the means is 0.59, showing that a one percent increase in years of schooling results in a 0.59%. The semi-elasticity at the means is 4.35, indicating that a one-year increase in educational attainment results in a 4.35% increase in the likelihood of being married.

(6) Multiple Quadratic Model With educ_female Interaction Term

In this model, an interaction term between educ_rev and female is introduced to examine the difference in educational attainment's impact on the likelihood of being married between the two sexes.[8] Using the results from this regression, the marginal impact is calculated using the formula mentioned in the econometric model section. This results in $\frac{\partial \widehat{married}}{\partial educ\_rev} = 0.001742 educ\_rev + 0.000229 female$, concluding that the effect of each additional year of education on a female's likelihood of being married is 0.000229 higher than that of males. At the mean level of education, one additional year of education will result in a female's likelihood of being married to increase by approximately 0.023, while this increase would be approximately 0.0228 for a male. This result does not show a significant difference in the effect of education on the likelihood of being married between the two genders.

(7) Multiple Quadratic Model With educ_white Interaction Term

In this model, we introduce the educ_white interaction term in order to examine the marginal differences in the effect of educational attainment on the likelihood of being married between white people and non white people. By taking the partial derivative with respect to educ_rev, we can better examine this relationship. This results in $\frac{\partial \widehat{married}}{\partial educ\_rev} = 0.00208 educ\_rev - 0.00618 white$. From this, we can conclude that for all levels of education, the effect of an additional year of education on one's likelihood of being married is 0.00618 less for white people compared to that of non white people. At the mean level of education, this comes out to approximately 0.0272 for non white people and 0.021 for white people. Once again, this is a relatively insignificant difference.

(8) Multiple Quadratic Model With educ_city Interaction Term

The final interaction term that will be introduced into the multiple quadratic model is educ_city. The purpose of this variable is to examine the marginal change in the effect of educational attainment on the likelihood of being married for those who live in a city versus those who do not. Taking the partial derivative of this model with respect to educ_rev results in $\frac{\partial \widehat{married}}{\partial educ\_rev} = 0.00208 educ\_rev - 0.00533 city$. From this, we can deduce that for all years of education, the change in probability of being married resulting from a one-year increase in education is 0.00533 lower for those that live in cities versus those that do not, holding all else constant. At the mean level of education, this comes out to a 0.0272 increase for those that live in cities versus a 0.0218 for those that do not.

Hypothesis Testing

To begin, we will test the hypothesis that the $\beta_2 \leq 0$, or the change in the likelihood of being married due to a one-year increase in education, is less than or equal to 0. In model 1, this results in a t statistic of 138.44 and a corresponding p-value of 0. In model 4, this

[8] Refer to table 5 for regression results

results in a t statistic of 94.56 and a p-value of 0. Thus in both models, we reject the null hypothesis at a 0.01 level of significance. With this, we can conclude that the model provides statistically significant evidence to suggest that the effect of one year of education on the probability of being married is greater than 0. The second and final hypothesis that we will be testing is that the effect of educational attainment is greater than or equal to 0.0375. In model one, this results in the t statistic -116 and the corresponding p-value of 0. In model 4, this results in a t statistic of -603.31 and a p-value of 0. In both cases, we reject the null hypothesis in favor of the alternative at a 0.01 of significance, indicating that the model provides statistically significant evidence to suggest that the effect of one year of education on the likelihood of being married is less than 0.0375.

VI Conclusions

The goal of this research and study was to determine if education plays a significant role as it pertains to the probability that an individual is married while controlling for other relevant factors. We did this in order to help identify if the growing rate of education as identified by the United States Census Bureau is having any sort of effect on the marriage rates in the United States, and to perhaps show an explanation for this that could be modeled on a global scale. Our findings do in fact support a generally positive relationship between the educational attainment an individual has and their probability of being married.

When examining the simple linear regression model as our first method of study, we saw that the correlation was positive, meaning that with more education, the probability of marriage grew. We suspected that education would have some sort of impact, given the findings by Abdullah, et al. (2013) that education had a significant effect on income inequalities while controlling for other variables to protect against error. This led us to add variables that would protect against biases, such as regional differences as identified by Köppen (2011), age as identified by Hayford, et al. (2016), and race as shown by Berna Torr (2011). We also used Torr's literature as our reasoning to include interaction terms between education and gender, education and race, and education and region. All of this literature helped provide us with the justification to develop and advance our research into a more comprehensive nature of the study.

As we evolved our study into including other independent variables and thus creating a multiple regression model shows that our thinking on the general direction of the relationship is indeed true. Interestingly, however, the correlation coefficient of the educ_rev variable is less in the multiple linear regression model (0.0229) as compared to the simple linear regression model (0.0346). This goes to show that the other factors that we included in our study, including race, gender, region, income, and age, also had a strong impact on our regression model. Upon more review, we found that the multiple quadratic regression model had an R^2 value that was the highest, indicating the best possible fit. This model had a slightly higher slope of education, being 0.0237. This means that for every additional year of education attained, the probability of marriage increases by 2.37%. This is not an insignificant number when looking at the amount of time people spend in schooling.

While the above number certainly matters, we acknowledge that the slope is relatively low.

We attribute this to the numerous additional variables, both endogenous and exogenous, that also play an impact on the likelihood of marriage. While some of these are challenging to control for, we have attempted to include each control factor to the best of our ability. It is important to note that we experimented with models including logarithmic-linear regressions and logarithmic-logarithmic regressions, however the R^2 values of these functions were less than that of the multiple quadratic regression model, therefore we determined that they were not of value to the impacts of our study.

We introduced three different interaction terms to determine differences in the effect based on race, gender, and region. Two of these variables showed minor differences based on these factors. For the gender term, we found that women have a slightly more positive relationship between education and marriage than that of men by 0.000229 in terms of the probability. For the race term, we found that people who are not white had a more positive relationship between education and marriage than that of white individuals by 0.00618. Lastly, we found that the city interaction term yielded that people who live in cities see their change in marriage probability increase by 0.00553 less than that of those who don't live in cities.

While we discovered many interesting aspects of this relationship, there are certain limitations we discovered within our research. For one, the model we used was not able to satisfy all of the OLS assumptions as the data we used was heteroskedastic and as a result led to the distribution of e not being normal. Additionally, the impact of race on this study cannot be understated. As indicated in Table 4, race has a significantly greater impact than any other variable we have chosen to include in this study. It would be interesting and worthwhile to evaluate this further in later studies on the matter. Of course, the model itself had its limitations, one being the lack of logical understanding of the constant value in our best-fit model. The value of the constant was -0.132, which does not make sense as a negative probability value is impossible. Lastly, many aspects contribute to the likelihood of someone being married, some of which are not quantifiable. For example, an individual's mental health cannot be quantified in any sort of data set or sample, and thus we are unable to factor this important variable into someone's ability to complete schooling or be able to sustain a happy spousal partnership.

Despite the previously explained limitations of our study, the results and subsequent implications of our paper suggest that more research should be conducted on this particular topic. For one, policymakers may wish to study this further, as marriage is beneficial in the scope of economic and population development. Perhaps political officials would wish to subsidize education in an effort to promote marriage and socio-economic growth, if they felt that the relationship is strong enough that action would be beneficial. Further studies may also seek to provide more subgroupings based on age, racial groups, or region in more detail, as that can provide further explanations into the impact of those variables, as well as that of education.

References

[1] Abdullah, A., Doucouliagos, H., & Manning, E. (2013). Does education reduce income inequality? A meta-regression analysis. *Journal of Economic Surveys*, 29(2), 301–316. https://doi.org/10.1111/joes.12056

[2] Hayford, S. R., & Guzzo, K. B. (2016). Fifty years of unintended births: Education gradients in unintended fertility in the US, 1960-2013. *Population and Development Review*, 42(2), 313–341. https://doi.org/10.1111/j.1728-4457.2016.00126.x

[3] Koppen Katja. (2011). *Marriage and cohabitation in Western Germany and France.*

[4] Torr, B. M. (2011). The changing relationship between education and marriage in the United States, 1940–2000. *Journal of Family History*, 36(4), 483–503. https://doi.org/10.1177/0363199011416760

[5] Kramer, S. (2019, December 12). *U.S. has world's highest rate of children living in single-parent households*. Pew Research Center; Pew Research Center. https://www.pewresearch.org/short-reads/2019/12/12/u-s-children-more-likely-than-children-in-other-countries-to-live-with-just-one-parent/

[6] Casey, A. E. (2022, August). *Child Well-Being in Single-Parent Families*. The Annie E. Casey Foundation. https://www.aecf.org/blog/child-well-being-in-single-parent-families#:~:text=Kids%20from%20single%2Dparent%20families,peers%20raised%20by%20married%20parents

[7] *The high price of education in Sub-Saharan Africa*. (2023, March 13). World Bank Blogs. https://blogs.worldbank.org/developmenttalk/high-price-education-sub-saharan-africa#:~:text=Sub%2DSaharan%20Africa%20has%20the,education%20for%20low%2Dincome%20households.

[8] Carine Øien-Ødegaard, Lars Johan Hauge, & Reneflot, A. (2021). Marital status, educational attainment, and suicide risk: a Norwegian register-based population study. *Population Health Metrics*, 19(1). https://doi.org/10.1186/s12963-021-00263-2

[9] Wang, W. (2015, December 4). *The link between a college education and a lasting marriage*. Pew Research Center; Pew Research Center. https://www.pewresearch.org/short-reads/2015/12/04/education-and-marriage/

[10] Aragão, C. (2023, March). *Gender pay gap in U.S. hasn't changed much in two decades*. Pew Research Center. https://www.pewresearch.org/short-reads/2023/03/01/gender-pay-gap-facts/#:~:text=The%20gender%20gap%20in%20pay,%2D%20and%20part%2Dtime%20workers.

[11] Parker, K., & Stepler, R. (2020, May 30). *As U.S. marriage rate hov-*

ers at 50%, education gap in marital status widens. Pew Research Center. https://www.pewresearch.org/short-reads/2017/09/14/as-u-s-marriage-rate-hovers-at-50-education-gap-in-marital-status-widens/

[12] U.S. Census Bureau. (2022, February 24). Census Bureau releases New Educational Attainment Data. Census.gov. https://www.census.gov/newsroom/press-releases/2022/educational-attainment.html#:~:text=the%20Hispanic%20population.-,From%202011%20to%202021%2C%20the%20percentage%20of%20adults%20age%2025,20.6%25%20for%20the%20Hispanic%20population.

Appendix

Table 1

Variable Name	**Description**
educ_rev	Number of years of completed schooling
married	Dummy Variable which is 1 if respondent is currently married and 0 if otherwise
female	Dummy Variable which is 1 if respondent identifies as female and 0 if otherwise
white	Dummy Variable which is 1 if respondent is white and 0 if otherwise
north	Dummy Variable which is 1 if respondent is from a northern part of the United States and 0 if otherwise
east	Dummy Variable which is 1 if respondent is from an eastern part of the United States and 0 if otherwise
city	Dummy Variable which is 1 if respondent is from a metropolitan area of the United States and 0 if otherwise
taxinc	Respondent's taxable income following deductions
age	How old the respondent is
educ_female	An interaction term introduced to determine if the effect of sex on marriage varies with education level
educ_white	An interaction term introduced to determine if the effect of race on marriage varies with education level
educ_city	An interaction term introduced to determine if the effect of living in a metropolitan area on marriage varies with education level

Table 2

VARIABLES	(1) mean	(2) sd	(3) min	(4) max
age	44.45	18.05	15	85
taxinc	35,253	85,106	0	3.213e+06
educ_rev	13.57	3.051	0	21
married	0.546	0.498	0	1
female	0.506	0.500	0	1
white	0.756	0.429	0	1
north	0.641	0.480	0	1
east	0.507	0.500	0	1
city	0.796	0.403	0	1

Table 3a

Summary Statistics by Married

VARIABLES	(1) married 0 mean	(2) sd	(3) N	(4) married 1 mean	(5) sd	(6) N
age	36.54	18.14	184,340	51.03	15.09	221,873
taxinc	21,545	51,497	184,340	46,642	103,787	221,873
educ_rev	12.86	2.799	184,340	14.16	3.125	221,873
female	0.512	0.500	184,340	0.501	0.500	221,873
white	0.704	0.457	184,340	0.800	0.400	221,873
north	0.634	0.482	184,340	0.648	0.478	221,873
east	0.514	0.500	184,340	0.500	0.500	221,873
city	0.808	0.394	184,340	0.786	0.410	221,873

Table 3b

Summary Statistics by White						
	(1)	(2)	(3)	(4)	(5)	(6)
	white 0			white 1		
VARIABLES	mean	sd	N	mean	sd	N
age	43.56	17.83	99,036	44.74	18.11	307,177
taxinc	29,999	75,656	99,036	36,947	87,870	307,177
educ_rev	13.61	3.017	99,036	13.56	3.062	307,177
married	0.449	0.497	99,036	0.578	0.494	307,177
female	0.528	0.499	99,036	0.499	0.500	307,177
north	0.570	0.495	99,036	0.664	0.472	307,177
east	0.539	0.499	99,036	0.496	0.500	307,177
city	0.873	0.333	99,036	0.771	0.420	307,177

Table 3c

Summary Statistics by Female						
	(1)	(2)	(3)	(4)	(5)	(6)
	female 0			female 1		
VARIABLES	mean	sd	N	mean	sd	N
age	44.58	18.34	200,664	44.33	17.76	205,549
taxinc	41,746	91,929	200,664	28,914	77,348	205,549
educ_rev	13.46	3.090	200,664	13.69	3.009	205,549
married	0.552	0.497	200,664	0.541	0.498	205,549
white	0.767	0.423	200,664	0.746	0.435	205,549
north	0.646	0.478	200,664	0.636	0.481	205,549
east	0.502	0.500	200,664	0.512	0.500	205,549
city	0.793	0.405	200,664	0.799	0.401	205,549

Table 3d

Summary Statistics by North						
	(1)	(2)	(3)	(4)	(5)	(6)
	north 0			north 1		
VARIABLES	mean	sd	N	mean	sd	N
age	44.46	17.98	145,754	44.45	18.09	260,459
taxinc	32,714	82,123	145,754	36,674	86,699	260,459
educ_rev	13.52	3.035	145,754	13.60	3.060	260,459
married	0.536	0.499	145,754	0.552	0.497	260,459
female	0.513	0.500	145,754	0.502	0.500	260,459
white	0.708	0.455	145,754	0.783	0.412	260,459
east	0.678	0.467	145,754	0.411	0.492	260,459
city	0.801	0.399	145,754	0.793	0.405	260,459

Table 3e

Summary Statistics by East						
	(1)	(2)	(3)	(4)	(5)	(6)
	east 0			east 1		
VARIABLES	mean	sd	N	mean	sd	N
age	43.86	17.98	200,415	45.03	18.10	205,798
taxinc	34,012	82,290	200,415	36,461	87,745	205,798
educ_rev	13.39	3.070	200,415	13.75	3.022	205,798
married	0.553	0.497	200,415	0.539	0.498	205,798
female	0.501	0.500	200,415	0.511	0.500	205,798
white	0.772	0.420	200,415	0.741	0.438	205,798
north	0.766	0.423	200,415	0.520	0.500	205,798
city	0.787	0.409	200,415	0.804	0.397	205,798

Table 3f

Summary Statistics by City

	(1)	(2)	(3)	(4)	(5)	(6)
	city 0			city 1		
VARIABLES	mean	sd	N	mean	sd	N
age	46.00	18.65	82,856	44.05	17.87	323,357
taxinc	24,788	61,423	82,856	37,935	89,984	323,357
educ_rev	13.06	2.663	82,856	13.71	3.129	323,357
married	0.574	0.495	82,856	0.539	0.498	323,357
female	0.499	0.500	82,856	0.508	0.500	323,357
white	0.849	0.358	82,856	0.732	0.443	323,357
north	0.650	0.477	82,856	0.639	0.480	323,357
east	0.486	0.500	82,856	0.512	0.500	323,357

Table 4

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5
educ_rev	0.0346***			0.0229***	
	(0.000250)			(0.000242)	
age				0.0102***	0.0102***
				(3.93e-05)	(3.93e-05)
white				0.111***	0.111***
				(0.00164)	(0.00164)
taxinc				4.50e-07***	4.30e-07***
				(8.59e-09)	(8.62e-09)
female				-0.00531***	-0.00534***
				(0.00140)	(0.00140)
north				-0.00583***	-0.00594***
				(0.00151)	(0.00151)
east				-0.0325***	-0.0325***
				(0.00145)	(0.00145)
city				-0.0215***	-0.0241***
				(0.00175)	(0.00175)
educ_rev_sq		0.00132***			0.000875***
		(9.13e-06)			(8.89e-06)
educ_rev_cu			5.78e-05***		
			(4.08e-07)		
Constant	0.0760***	0.290***	0.380***	-0.278***	-0.132***
	(0.00348)	(0.00192)	(0.00140)	(0.00414)	(0.00329)
Observations	406,213	406,213	406,213	406,213	406,213
R-squared	0.045	0.049	0.047	0.201	0.202

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

Table 5

VARIABLES	(1) Model 6	(2) Model 7	(3) Model 8
educ_rev_sq	0.000871***	0.00104***	0.00104***
	(1.21e-05)	(1.63e-05)	(1.96e-05)
age	0.0102***	0.0102***	0.0101***
	(3.93e-05)	(3.93e-05)	(3.93e-05)
white	0.111***	0.195***	0.111***
	(0.00164)	(0.00700)	(0.00164)
taxinc	4.30e-07***	4.29e-07***	4.31e-07***
	(8.63e-09)	(8.62e-09)	(8.62e-09)
female	-0.00845	-0.00524***	-0.00546***
	(0.00627)	(0.00140)	(0.00140)
north	-0.00594***	-0.00609***	-0.00632***
	(0.00151)	(0.00151)	(0.00151)
east	-0.0325***	-0.0318***	-0.0321***
	(0.00145)	(0.00145)	(0.00145)
city	-0.0241***	-0.0247***	0.0456***
	(0.00176)	(0.00176)	(0.00774)
educ_female	0.000229		
	(0.000450)		
educ_white		-0.00618***	
		(0.000501)	
educ_city			-0.00533***
			(0.000576)
Constant	-0.131***	-0.165***	-0.159***
	(0.00359)	(0.00422)	(0.00443)
Observations	406,213	406,213	406,213
R-squared	0.202	0.202	0.202

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

Table 6

	Model 1	Model 4
$H_0 : \beta_2 \leq 0$ $H_1 : \beta_2 > 0$	0.0	0.0
$H_0 : \beta_2 \geq 0.0375$ $H_1 : \beta_2 < 0.0375$	0.0	0.0