Does Workers with Higher Education Level Work More or Less in Hours?

Boston University
Freddie Guo, Yukun Lan, Jiaxuan Han
December 12th, 2023

Abstract

While the return on education in terms of earnings has been widely studied in the field. However, the effect of education on people's working hours has drawn less attention. This paper hopes to examine the conventional wisdom that education could serve as a driving factor for shorter working hours, reflecting a contemporary view that prioritizes life quality over the volume of work. We used data from IPUMS CPS (Current Population Survey) and employed the OLS (Ordinary Least Squares) multiple linear regression model to answer the question empirically. Contrary to the conventional wisdom's expectation, we find that more years of education is associated with longer hours worked per week at all jobs after controlling for various confounding factors such as age, childcare duty, income, race, gender, and health status.

I. Introduction

Work plays a pivotal role in pursuing meaningful and fulfilling lives. As writer Annie Dillard (1989) eloquently stated, “How we spend our days is, of course, how we spend our lives.” For most people, a substantial fraction of these days is dedicated to the workplace, where they invest their time, energy, aspirations, and well-being. An astonishing statistic underscores the significance of this commitment: the average person will spend about 90,000 hours of their lifetime to work. (Pryce-Jones, 2010) This vast amount of time spent in the labor force raises an essential question that has captured our attention: How does an individual’s level of education impact the number of hours they work?

Our motivation for this study finds resonance in the observation of news that has come to our attention. The Dubai government released the 2019 results of its Government Employee Happiness Index, designed to measure government workers’ well-being across the emirate. (Thomas, n.d.) The survey findings suggest a positive correlation between employee well-being and enhanced productivity, greater creativity, reduced illness and absenteeism, and decreased presenteeism, which refers to employees spending more time at work than necessary. (Thomas, n.d.) The Dubai Government’s approach exemplifies the growing recognition of the importance of workplace happiness. This raises the consideration of whether individuals pursue higher education primarily to create a foundation for a life that offers a better balance between professional pursuits and personal well-being. (Thomas, n.d.) The hypothesis of this research paper is that attaining higher levels of education correlates with a reduction in working hours. This idea assumes that individuals seeking advanced degrees tend to favor work schedules that promote a balanced work-life dynamic. Pursuing higher education is commonly motivated by the desire for improved career prospects and personal development, with the anticipation that these benefits will ultimately contribute to greater life satisfaction. Therefore, this paper suggests that education could serve as a driving factor for shorter working hours, reflecting a contemporary view that prioritizes life quality over the volume of work.

We used OLS (Ordinary Least Squares) regression models to study the empirical relationship between education and working hours. We included controls in various confounding factors such as age, childcare duty, income, race, gender, and health status. Contrary to our expectations, all specifications of the model suggested that more years of education are associated with longer hours worked in a usual week.

The rest of the paper is organized as follows: we first review the past works that have been done to answer related questions in section II; then we explained the econometrics model used for our study and various specifications of the model we tested in section III; we discussed our data source and how we cleaned the data so that they are ready for regression in section IV; then we showed the regression results in section V and provided our interpretation; lastly we concluded our research and proposed some future work to be done in section VI.

II. Literature Review

Return to education is a common theme in the field of labor economics as it is central to making well-informed decisions on school and major choices for individual students, understanding the complex interplay between school choice and labor market outcomes for school officials, and developing proper public education policy that fuels the growth of our economy. A long work has been devoted to understanding the relationship between education attainment and labor income, health outcomes, and gender disparities. We listed three papers that are mostly related to our research question: what's the relationship between years of education and weekly working hours?

By analyzing the relationship between higher education levels and working hours, it is also important to consider individual earnings and economic growth. Unlike many studies that emphasize the years of schooling as a measure of human capital, Hanushek and Woessmann (2008) find that not only school attainment but also the cognitive skills of the population are powerfully positively related to individual earnings, emphasizing that quality of education (cognitive skills) is more significant than quantity (years of schooling). Hanushek and Woessmann (2008) begin with a very simple earnings model with individual earnings as the dependent variable and the labor market skills of the individual as the independent variable, allowing for a better understanding of the relationship between education quality and economic outcomes. Hanushek and Woessmann (2008) focus on the quality of education as reflected by cognitive skills. The paper uses international assessment data (like standardized tests in math, science, and reading), which provides a broader, more diverse perspective compared to studies with only region-specific datasets and can be better applied widely in the world. This paper also shares a focus on human capital, which is a common theme in education economics. Both our study and Hanushek and Woessmann used Ordinary Least Squares (OLS) multiple regression models. While Hanushek and Woessmann (2008) focus on the quality of education (cognitive skills) and its impact on economic growth and individual earnings, our paper looks at a different facet of education's economic impact – the relationship between the quantity of education (years of schooling) and working hours. Both studies fundamentally explore how education influences economic outcomes, albeit from different angles. Although the measurements vary, the basic concept of educational impact is the same. Insights from their findings on the importance of education quality could provide a nuanced perspective to your analysis of education quantity. Also, their international dataset offers a broader perspective compared to our studies with national datasets, so our analysis will be more limited by comparison.

Unlike our studies that use the number of years of education to measure educational attainment, Zajacova, Hummer, and Rogers (2012) analyzed by broadly categorized educational attainment, including six levels of postsecondary education, high school, GED, and years of education if did not finish high school(e.g., 0-8 years, 9 years, 10 years, 11 years, 12 years). They focused on working-age adults, using the National Health Interview Survey data from 1997 to 2009, which covers a significant timeframe and a large sample, including 178,103 individuals. They mainly focus on non-Hispanic U.S.-born white and black working-age adults (30-65 years). This demographic often receives less attention in the relationship between education and health. The study also includes a wide range of control variables such as age, sex, race, economic standing, and health behaviors, offering a more comprehensive understanding of the education-health relationship. Zajacova et al.(2012) found obvious health differentials across educational levels, with higher education correlating with better health outcomes. The study supports both the quantity and credential models of education's impact on health. For example, substantial improvements in health are observed between those with 11 and 12 years of schooling and between individuals with an academic AA and a BA degree. Working-age adults with lower educational levels face significantly higher odds of reporting worse health than those with higher education. Zajacova et al.(2012) focuses on Self-reported health (SRH) as a main independent variable, which might have subjective biases, the economic indicators and health behaviors explain about 40% of the education-health relationship. Our paper also has a control variable(x) of health status. We use data based on self-rated health status on a five-point scale; with the relationship between education and health, the use of regression models and focus on adult demographics will be similar, and we might find a similar relationship in the result.

Wirtz et al. (2012) studied the gender differences in the effects of weekly working hours on occupational injury risk. They used data from the U.S. National Health Interview Survey between 2004 and 2010, which contains 96915 employed individuals. Wirtz et al. (2012) used injury risk as the dependent variable, where injury in the sample was "an individual reporting at least one injury or poisoning episode that required medical attention during the three months prior to the interview." Wirtz et al. (2012) break the independent variable, weekly working hours, into four levels: less than or equal to 30 hours per week, between 31 and 40 hours per week, between 41 and 50 hours per week, and greater than 50 hours per week. Wirtz et al. (2012) added the interaction term between gender and working hours, signaling a strong gender disparity in the effect of working hours on injury risk. Therefore, we included the interaction term between gender and education years as well in our own study. Wirtz et al. (2012) using a multivariable weighted logistic regression model, the authors found that longer working hours may increase work-induced fatigue and thus increase the chance of work-related injury. Specifically, Wirtz et al. (2012) found that occupational injury risk is significantly higher for male than for female workers. In addition, they documented that female workers tend to have longer sleep and experience a higher level of psychological pressure. Those are additional control variables to be included. However, due to data limitations, we cannot access those variables in the IPMUS CPS data set for our analysis.

III. Econometric Model

First, our study considered the simple linear regression model to analyze the impact of educational levels on working hours. The dependent variable is the weekly working hours, while the primary independent variable is the years of education received. In the simple linear regression model, we first need to learn if years of education have a statistically significant impact on weekly working hours and whether the effect is positive or negative. If the coefficient of the years of education is positive and significant, this paper then recognizes that the more years of education, the more hours worked per week. Suppose the coefficient of the years of education is negative and significant. In that case, this paper recognizes that the more years of education, the fewer hours worked per week, which is our expected result.

To find a better fit, we tried six different functional forms to establish this foundation. The second model we used, the quadratic model, allows for the possibility that the effect of an additional year of education on weekly working hours might increase or decrease at a diminishing or accelerating rate rather than remain constant. In the third model, we used the cubic model, allowing for more complex non-linearities, such as an initial increase in the effect of education that then starts to diminish or even reverse after a certain point. We used the following models: log-linear, log-log, and linear-log. We perform a logarithmic transformation to correct for the skewness. We will choose the best-fit model based on the highest comparable R squared.

OLS Assumption Check

We checked OLS assumptions by plotting out the histogram of the dependent variable, weekly working hours in original form and in log forms in Graph 1 and Graph 2, respectively. The distribution is symmetric and roughly normal. Due to the large size of our sample, by the central limit theorem, it can be assumed that the distribution of weekly working hours is approximately normally distributed. We also provided the scatter lots between our main dependent variable and main independent variables in Graph 3 and Graph 4. Due to too many observations in the data, we plot out the mean of the weekly working hours for each independent variable value to better reflect the trends. We also performed a heteroskedasticity check for the simple linear regression and found that the homoscedasticity assumption was rejected. We included the plot of linear regression model prediction errors in Graph 7. Due to the large size of our data set, we cannot directly observe patterns from the graph. Based on the Stata test results, we concluded that there is a heteroskedasticity issue in our data and model. Therefore, we reported the heteroskedasticity standard errors for all future analyses. We also plotted the predicted errors from linear and quadratic models in Graph 5 and Graph 6. We see that the error terms are centered around 0 with a large mass and are roughly normally distributed. After we used squared years of education as our independent variable, we saw an increase in the percentage of predicted errors around 0 and less dispersion.

Simple Linear Regression Models

We considered six model specifications of the simple linear regression model:

Model 1: $\widehat{uhrswork}t = \beta_1 + \beta_2 * educ\_rev$

Model 2: $\widehat{uhrswork}t = \beta_1 + \beta_2 * (educ\_rev)^2$

Model 3: $\widehat{uhrswork}t = \beta_1 + \beta_2 * (educ\_rev)^3$

Model 4: $\log(\widehat{uhrswork}t) = \beta_1 + \beta_2 * educ\_rev$

Model 5: $\log(\widehat{uhrswork}t) = \beta_1 + \beta_2 * \log(educ\_rev)$

Model 6: $\widehat{uhrswork}t = \beta_1 + \beta_2 * \log(educ\_rev)$

We presented all regression results in Table 5. Based on the regression results, we chose the second specification because it gives us the highest R-squared statistics.

Multiple Regression Models

Due to the complex interplay between many endogenous variables and omitted variables, a simple linear regression of working hours on education years would be misleading. We have to include many independent variables in the regression model other than the key variable we are interested in. Therefore, we also consider the following multiple regression model:

$$\begin{aligned}\widehat{uhrswork}t = {} & \beta_1 + \beta_2 * age + \beta_3 * nchilt5 + \beta_4 * disabwrk + \beta_5 * health + \beta_6 * female \\ & + \beta_7 * colored + \beta_8 * (educ_{rev})^2 + \beta_9 * married + \beta_{10} \\ & * logPersonalInc + \beta_{11} * logHouseholdInc + \beta_{12} * y_{2003} + \beta_{13} * y_{2006} \\ & + \beta_{14} * y_{2009} + \beta_{15} * y_{2012} + \beta_{16} * y_{2015} + \beta_{17} * y_{2018}\end{aligned}$$

In our study, we have tried to avoid omitted variable bias by including additional control variables in our analysis. These controls, like gender, race, age, marital status, presence of young children in the household, health status, and personal and household income, ensure our results are more accurate. We expected that the effect of additional years of education on working hours would be different across different respondents. By including these variables, we're aiming to make our findings more robust.

Control Variables

In our multiple regression models, we introduced a number of demographic controls, such as gender, rsce, age, and marital status. We expected the effects of education on weekly working hours might be different on different demographic groups.

We then introfuce a couple more of control variables. The variable “nchilt5” can be used to study the impact of the presence of young children in the home on an individual’s working hours, as it often entails additional caring responsibilities, especially for women. The variable “health” can be used to study the impact of an individual’s health status on their ability to work and their hours. The variable “logPersonalInc” can reveal if individuals with higher incomes tend to work fewer hours, possibly due to financial security that allows for a more relaxed work schedule, or if they work more hours, potentially driven by greater responsibilities or career commitments associated with the higher-paying position. The variable “logHouseholdInc” is similar to personal income. We expected that households with higher total incomes could lead to more flexible work choices and a reduced need to work long hours for economic stability.

We also added year control variables (2000, 2003, 2006, 2009, 2012, 2015, and 2018) to our regression model to improve its accuracy further. These controls help account for changes over time that might affect our dependent variables, such as economic fluctuations, economic events, and policy changes.

Furthermore, we added four interaction terms to reveal the differential effects of additional working hours given different ages, gender, race, and marital status. We made this by multiplying years of education by the dummy variable in a certain sequence. Thus, our model would directly show the result affected by the fluctuation of related variables. Then, the full model with interaction terms becomes :

$$\begin{aligned}\widehat{\text{uhrsworkt}} = {} & \beta_1 + \beta_2 * \text{age} + \beta_3 * nchilt5 + \beta_4 * \text{disabwrk} + \beta_5 * \text{health} + \beta_6 * \text{female} \\ & + \beta_7 * \text{colored} + \beta_8 * \left(educ_{rev}\right)^2 + \beta_9 * married + \beta_{10} \\ & * \text{logPersonalInc} + \beta_{11} * \text{logHouseholdInc} + \beta_{12} * y_{2003} + \beta_{13} * y_{2006} \\ & + \beta_{14} * y_{2009} + \beta_{15} * y_{2012} + \beta_{16} * y_{2015} + \beta_{17} * y_{2018} + \beta_{18} \\ & * \left(educ_{rev}\right)^2 * female + \beta_{19} * \left(educ_{rev}\right)^2 * colored + \beta_{20} \\ & * \left(educ_{rev}\right)^2 * married + \beta_{21} * \left(educ_{rev}\right)^2 * age\end{aligned}$$

Years of Education Interaction with Being Male:

In the final regression model, we added the interaction terms between “educ_rev” and the gender dummy variable, “female,” to account for the fact that there might be a differing effect of education on working hours across genders. We do not have a clear expectation of the sign of this interaction term because males might be working less given the same increase in education because they enjoy a higher wage premium compared to females, or they might be working more because they tend to spend less time with family and the social expectation for male is to earn more. We included the test of the significance of this interaction term in the results section.

Years of Education Interaction with Being Non-white:

In the final regression model, we added the interaction terms between "educ_rev" and the race dummy variable, "colored," to account for the fact that there might be a differing effect of education on working hours across races. We expect colored people to work more given the same marginal change in years of education because there is a racial difference between the wages all else equal. We included the test of the significance of this interaction term in the results section.

Years of Education Interaction with Being Married:

In the final regression model, we added the interaction terms between "educ_rev" and the marital status dummy variable, "married," to account for the fact that there might be a differing effect of education on working hours between married and not married people. We do not have a clear expectation on the sign of this interaction effect because married people might work more because they need to feed two mouths, or they might work less because they have two income sources. We included the test of the significance of this interaction term in the results section.

Years of Education Interaction with Being Old:

In the final regression model, we added the interaction terms between "educ_rev" and the age variable, "age," to account for the fact that there might be a differing effect of education on working hours across age groups. We expect older people to work less given the same years of education. We also included the test of this interaction term's significance in the results section.

Hypothesis Testing

We performed various F-tests for the control variables we included. To start with, we tested if they were equal to 0. Then, we tested some closely related variables to see if they have an effect on weekly working hours given the education backgrounds and other controls fixed. We tested if personal income has the same effects as household income. We also tested if self-perceived health status has the same effects as an actual disability that hinders people from working. We included all the hypothesis testing results in Table 7.

IV. Data and Descriptive Statistics

This paper used the IPUMS CPS data from 2000, 2003, 2006, 2009, 2012, 2015 and 2018. IPUMS CPS provides census and survey data across time, so it is easy to study change, conduct research, and analyze. (Flood et al., 2023) For our analysis, we selected data at three-year intervals throughout the 21st century, up to the latest data available before the COVID period. By narrowing our dataset to these selected years, we can create a more manageable dataset for our research while still capturing a wide range of data over time to study changes and trends. We have 543,482 observations after cleaning the dataset.

All unrelated default variables are dropped, including variables with missing or irrelevant values. We clean all the NIU values in the weekly working hours variables. We clean all 999, 0, or 1 values in education level variables because these values indicate missing, invalid, or uninformative education levels. We remove all NIU values in disability to work variables. We clean all undefined or missing marital status values in the marital status variables. We cleaned all unspecified or missing race categories in race variables. We cleaned unspecified or missing gender information in the gender variable. These data-cleaning steps help us to ensure our dataset contains only valid, relevant, and complete information, which is important for running accurate regression analysis.

Next, we created three dummy variables for female, colored, and married. Then, we transformed the education levels recorded in the data into the years of education an individual has received. We dropped outliers for individuals with more than 100 hours of weekly working hours. We also dropped people over 80 years old to the working age to help reduce the potential impact of outliers in our analysis. We clean extreme values in personal and household income variables by dropping those with incomes greater than 1000000 or with negative incomes. This helps us to handle outliers. Then, we take the log transformation for both personal and household income variables. This transformation helps handle skewed income data, achieve a more symmetric distribution, and make statistical analysis more robust.

<u>Descriptive Statistics</u>

The average person in this sample is male, white, 41.15 years of age, single, with a working hour of 39.63 hours per week, having an average of 13.7 years of education, having 0.196 children under five years old in his family, having the value of disabled equals to 0.0202, which means people are likely to be not affecting work because of disability. People with an average self-rated health equal 2.036, which means people tend to have good but not perfectly good health.

As we can see in Table 4, for respondents that are males, they have a mean of 41.26 years old, have an average of 13.6 years of education, have approximately 0.216 children under five years old in their household, have the value of disabled equals to 0.0190, having an average of self-rated health equals to 2.017, and have 82.8% percent of them are white, 35.9% of them are single.

As we can see in Table 4, for respondents that are females, they have a mean of 41.03 years old, have an average of 13.82 years of education, have approximately 0.175 children that are under five years old in their household, have a value of disabled equals to 0.0214, having an average of self-rated health to 2.057, while 20.6% percentage of them are people of colored, 56.1% percentage of them are married. Male respondents, on average, work 5.19 hours more per week than the average female and have slightly fewer years of education compared to females. This implies the trends in educational attainment where women are increasingly surpassing men in higher education degrees. Additionally, the higher number of young children in male respondents' households could indicate changing family dynamics and possibly a shift in paternal involvement in early child-rearing.

As we can see in Table 3, white people have a mean of 41.19 years old, an average of 13.69 years of education, approximately 0.202 children under five years old in their household, having 39.68 average working hours per week, having the value of disabled equals to 0.0201, and having self-rated health to 2.007, 47.1% percent of them are female, and 62.8% of them are married. For people of color, they have a mean of 40.96 years old, an average of 13.78 years of education, approximately 0.172 children under five years old in their household, 39.42 average working hours per week, the value of disabled equals 0.0203, and having self-rated health to 2.161, 52.6% percent of them are female, and 49.5% of them are married. People of color in the sample show a slightly younger average age and higher educational attainment, trends that may reflect demographic shifts and changing educational landscapes.

According to the histogram of working hours plotted in Graph 1, we can conclude that the trend of its distribution is relatively symmetric and that over 50% of the observations centered around 40 hours per week, which matches our expectation that most jobs are 8 hours a day and five days a week. We also plotted the weekly working hours data after log transformation in Graph 2 and found that the data is distributed more like a Normal distribution in shapes.

We plotted a scatter plot of our main dependent and independent variable in Graph 3 and saw that, vertically speaking, dots lined up almost identically, except that those with fewer years in education have fewer large extreme values in weekly working hours. To learn more about the relationship between the two variables, we plotted a scatter plot between the average weekly working hours by years of education and years of education and found different patterns. We observe people with around ten years of education working the fewest hours on average, which corresponds to those who didn't finish senior high school education. Once one completes high school, there is a positive correlation between the years of education and weekly working hours. For those who didn't complete high school, there is an initially slightly increasing trend and then a sharp drop in working hours as we approach ten years of education. It seems to suggest that those who didn't start high school need to work more hours to make a living, while those high school dropouts are working much less than an average person.

V. Results

<u>Simple Linear Regression Model Interpretation</u>

We tested different specifications models for the simple linear regression model to see which one fits better and whether someone with a higher education level tends to work more or fewer hours.

As shown in Table 5, model 1(Linear Model) of the regression output, the constant is 31.15, implying that when the number of years of education is 0, the estimated weekly working hours is 31.15 hours. The coefficient is 0.619, which means that when the number of years of education increases by one year, the estimated weekly working hours increase by 0.619 hours. This means that the higher the level of education, the longer weekly working hours one will have. The slope is the same as the coefficient for the linear model. Therefore, the interpretation will be the same. The elasticity and semi-elasticity values, 0.21 and 1.56, respectively, indicate moderate responsiveness of working hours to changes in educational years. Specifically, the elasticity being 0.21 means that when the number of years of education increases by 1%, estimated weekly working hours increase by 0.21%, evaluated at the mean; and the semi-elasticity being 1.56 means that when the number of years of education increases by one year, estimated weekly working hours increase by 1.56% evaluated at the mean.

As shown in Table 5 Model 2 (Quadratic Model) of the regression output, with an initial estimated working hour count of 35.12 (when educational years squared equals zero), the model yields a coefficient of 0.0230, reflecting a subtle increase in work hours per unit increase in the square of education years. This translates into a slope of 0.6302 evaluated at the mean, meaning that when the number of years of education increases by one year, the estimated weekly working hours increase by 0.6302 hours. Elasticity and semi-elasticity values stand at 0.2179 and 1.59, respectively, both are very similar to the estimates obtained from Model 1.

As one can see in Table 5, Model 3 (Cubic Model) of the regression output, the constant is 36.71, which implies that when the cube of the number of years of education is 0 units, the estimated weekly working hours is 36.71 hours. The coefficient is 0.001, which means that when the cube of the number of years of education increases by 1 unit, the estimated weekly working hours increase by 0.001. The slope is 0.563, which means that when the number of years of education increases by one year, the estimated weekly working hours increase by 0.563, evaluated at the mean education level. Elasticity and semi-elasticity are calculated as 0.195 and 1.42, respectively. Both measures are smaller in magnitudes compared to those in the linear and quadratic models. To be exact, the elasticity being 0.195 means that when the number of years of education increases by 1%, estimated weekly working hours increase by 0.195%, evaluated at the mean, and the semi-elasticity being 1.42 meaning that when the number of years of education increases by one year, estimated weekly working hours increase by 1.42%, evaluated at the mean.

After examining all the linear specifications with higher-order independent variables, we then check if applying log transformation will obtain better fits of the model. Table 5 Model 4 (Log-lin Model) of the regression output shows that the constant is 3.362, which implies that when the number of years of education is 0 years, then log(estimated weekly working hours) is 3.362 units, which means that the estimated weekly working hours is 28.85 hours. The coefficient is 0.0187, meaning that a 1.87% increase in weekly work hours is observed for each additional year of education. The slope and elasticity evaluated at the mean equal 0.74 and 0.256, respectively, meaning that when the number of years of education increases by one year, estimated weekly working hours increase by 1.87% and that when the number of years of education increases by one year, then estimated weekly working hours increase by 0.74 hours evaluated at the mean of weekly working hours. We obtain a semi-elasticity estimate of 1.87, which means that when the number of years of education increases by 1%, estimated weekly working hours increase by 0.256% evaluated at the mean.

In Table 5 Model 5, we presented the result of the Log-log Model output. The constant is 3.039, meaning that when the log number of years of education is 0 units, which corresponds to 1 year of education, the log of estimated weekly working hours is 3.039 units, which corresponds to a weekly working hour of 20.88 hours. The coefficient is 0.223, which means that when the number of years of education increases by 1%, then estimated weekly working hours increase by 0.223%, and 0.223 is called elasticity. The slope is 0.65, which means that when the number of years of education increases by one year, then estimated weekly working hours increase by 0.65, evaluated at the mean of weekly working hours and education levels. Therefore, the slope is not constant, and it changes depending on both the y and x. The semi-elasticity is 1.64, which means that when the number of years of education increases by one year, then estimated weekly working hours increase by 1.64%.

Lastly, Model 6, the Linear-log Model, yields an initial estimate of 20.74 work hours and a coefficient of 7.278, meaning that when the number of years of education increases by 1%, the estimated weekly working hours will increase by 0.073. This model notes a slope of 0.53 and an elasticity of 0.18, with a semi-elasticity of 1.34, demonstrating a somewhat consistent but modest change in work hours with increased education. The slope is 0.53, which means that when the number of years of education increases by one year, the estimated weekly working hours will increase by 0.53, evaluated at the mean of education levels. Repeat our calculation one more time, we obtain an elasticity of 0.18, meaning that a 1% increase in the number of years leads to an estimated weekly working hours increase by 0.18% evaluated at the mean. While a semi-elasticity estimate of 1.34 means that an one year increase in education leads to a 1.34% increase in estimated weekly working hours increase, evaluated at the mean.

Model Specification Choice

We reported the R-squared for all six model specifications in Table 5. We adjusted the R-squared for model 4 and model 5 so that the statistics are comparable across models. We reported the transformed version in parentheses. Among the six models, we found that model 2 yields the highest R-squared of 0.0243, meaning that 2.43% of the variations in weekly working hours can be explained by "educ_rev." Therefore, we decided to use squared years of education as our primary independent variable in the multiple regression models. We reported the multiple regression results in Table 6.

Multiple Linear Regression Model without Interaction Terms Interpretation

Building on the insights from the simple regression model, we extend our analysis to a multiple regression model to allow for additional explanatory variables and to better understand the correlation between education levels and weekly working hours, as seen in Table 5, model 1 of the regression output. All the coefficients are significant at the 1% level. The constant is 14.28 in this model, but this number does not make sense because the minimum age in our data set is 15 years old, which is under the legal age to get a job. As the square of the number of years of education increases by 1 unit, estimated weekly working hours increase by 0.00502 hours, holding all other factors constant. The coefficient for the square of the number of years of education in the multiple regression model is 0.0198 smaller compared to the simple regression, suggesting that the impact of education on weekly working hours is less obvious when accounting for a broader range of other factors. As people become more educated, they still work more, which may be because they are more passionate about what they do.

As age increases by one year, estimated weekly working hours decrease by 0.0185 hours, holding all other factors constant. This means that as someone gets older, on average, they tend to work fewer hours, which makes sense because as someone gets to a certain age, one might just retire as well. As self-rated health status increases by 1 unit, estimated weekly working hours increase by 0.0607 hours, holding all other factors constant. This means that when people's health and body are getting worse, they work a little bit longer hours. This is a surprising result since longer working hours could be a factor that leads to worse health. If one is female, estimated weekly working hours decrease by 3.194 hours, holding all other factors constant. This means that women tend to work fewer hours than men, and female labor force participation and gender equality are always a concern in our society. If one is married, estimated weekly working hours increase by 1.507 hours, holding all other factors constant. Married individuals might have more financial responsibilities, such as supporting a spouse and children, this can lead to a greater desire to work more hours. If the number of children five years old or under in the household increases by 1, estimated weekly working hours decrease by 0.204 hours, holding all other factors constant. This means that if more children are five years old or under in the household, people tend to work fewer hours because they might need to take care of the children. If one is a person of color, the estimated weekly working hours increase by 0.213 hours, holding all other factors constant. This means that non-whites tend to work a little longer. If one is disabled to work, estimated weekly working hours decrease by 6.064 hours, holding all other factors constant. This means people with health problems that can prevent them from working will work fewer hours. As the log(personal income) increases by 1 unit, estimated weekly working hours increase by 4.837 hours, holding all other factors constant. This means someone with a higher wage works more hours compared to others, higher income jobs might come with greater responsibilities, and many high-paying jobs, such as doctors, lawyers, and managers, require extended working hours. As the log(household income) increases by 1 unit, estimated weekly working hours decrease by 2.046 hours, holding all other factors constant. Higher household income might discourage an individual within the household from working long hours, as their financial needs are already comfortably met. One might prefer more leisure time over additional work.

Weekly working hours were 0.956 less in 2006 than in all other years, holding all other factors constant. Weekly working hours are 1.913 less in 2009 than in all other years, holding all other factors constant. Weekly working hours were 2.005 less in 2012 than in all other years, holding all other factors constant. Weekly working hours were 1.817 less in 2015 than in all other years, holding all other factors constant. Weekly working hours are 2.105 fewer in 2018 than in all other years, holding all other factors constant. The results indicate a trend of decreasing weekly working hours over time compared to the base year. This could suggest changes in the labor market, economic, and work-life balance. The impact of technology and productivity improvements allows people to work fewer hours for the same output. The reduction in working hours is not constant. For example, the drop in hours was larger in 2009, 2012, and 2018 compared to 2003 and 2006. This could be due to specific economic events that occurred in those years. For example, 2009 was the year following the global financial crisis, which may have led to reduced working hours due to an economic downturn.

<u>Multiple Linear Regression Model with Interaction Terms Interpretation</u>

We have introduced four distinct interaction terms into our model. Adding in these interaction terms allows us to check whether the effect of an extra year of education on weekly working hours varies with gender, race, marital status, and age. As illustrated in Table 5, Model 2, the inclusion of these interaction terms has shown significant variations in these results. The constant changes from 14.28 to 7.950. The coefficient of age changes from -0.0185 to 0.0874 as age increases by one year, and estimated weekly working hours change from decrease to increase by 0.0874 hours, holding all other factors constant. The coefficient of colored people changes from 0.213 to 0.809, and the impact of being a person of color on weekly working hours is larger, which means they are working more hours. The coefficient of married people changes from 1.507 to 3.741, that the impact of getting married on weekly working hours is much larger, meaning they are working more than 2 hours per week. All other coefficients remain similar.

<u>Education interaction with gender(female)</u>

The coefficient of the interaction of education level and females is significant at the 5% level. The effect of an extra unit of the square of education is associated with 0.000769 fewer weekly working hours for women compared to men. This could be a result of females tend to shoulder a larger share of family responsibilities, including childcare. The effect of an extra unit of the square of education increases by 0.0388 weekly working hours for women and increases by 0.0396 hours for men, holding all other factors constant.

Education interaction with race(colored)

The coefficient of the interaction of education level and colored people is significant at the 1% level. The effect of an extra unit of the square of education is associated with 0.00328 fewer weekly working hours for people of color compared to white people. This could be due to the inequality in job opportunities. The effect of an extra unit of the square of education increases 0.806 weekly working hours for people of color and increases by 0.809 hours for white people, holding all other factors constant.

Education interaction with marital status(married)

The coefficient of the interaction of education level and married people is significant at the 1% level. The effect of an extra unit of the square of education is associated with 0.0122 less weekly working hours for married people compared to single ones. The effect of an extra unit of the square of education increases by 3.729 weekly working hours for married people and increases by 3.741 hours for single people, holding all other factors constant. This may reflect that married people make a conscious choice to seek a better work-life balance as their educational attainment increases, although the weekly hours of work for both married and single people increase with each unit increase in the square of educational level, but less so for married people.

Education interaction with age

The coefficient of the interaction of education level and age is significant at the 1% level. For a high school graduate(12 years of education), an extra year of age increases estimated weekly working hours by 0.0019 hours; and for a college graduate(16 years of education), an extra year of age decreases estimated weekly working hours by 0.0004 hours; and for master's graduate(18 years of education), an extra year of age, decrease estimated weekly working hours by 0.0015 hours; and for doctor graduate(24 years of education), an extra year of age, decrease estimated weekly working hours by 0.005 hours. The interaction between education level and age documented that higher education levels are associated with a decrease in weekly working hours beyond high school graduates. We see a trend where individuals with more education, such as those with master's or doctoral degrees, tend to reduce their working hours a lot more as they age, which may make it easier to achieve a work-life balance or retire earlier.

In addition, the interaction terms all have negative coefficients, meaning that compared to a single white male at age 0, being female, being colored, being married, or becoming old would all decrease the marginal increase in the number of hours worked in a usual week given a one year increase in education.

With all the interaction terms added, our final model gives an R-squared of 0.22, meaning that the variations in our independent variables and interaction terms can jointly explain 22% of the variation in weekly working hours, significantly better than what the single linear regression model yields.

Hypothesis Testing Results

We listed the hypothesis tests we conducted in Table 7. Specifically, we reported the null hypothesis, the associated F-statistics, and the p-value of the F-statistic calculated from the multiple linear regression model with interaction terms. We first tested if the coefficient of the control variables is equal to zero, meaning that the control variable does not help us predict weekly working hours. We found that the corresponding p-values are all 0.0000, meaning that they are extremely helpful in understanding people's weekly working hours. Then, we tested if the additional interaction terms' coefficient equals zero. Again, we get p-values equal to 0.0000 for the interaction between "educ_rev" and race, marital status, and age, meaning that there exists a statistically significantly differing effect of years of education on weekly working hours across race, marital status, and age. The p-value for the interaction between gender and weekly working hours is slightly higher at 0.0449. However, this is still below the 5% significant level. Hence, we can conclude that all four interaction terms are useful.

VI. Conclusions

The effect of education on labor income and economic growth has been widely studied in the field of labor economics. Past work focuses on detailed measurement of education attainment and explicit control for students with different backgrounds, such as cognitive ability and family income. Other related work tends to examine how this linkage between education and labor market outcome changes given different socioeconomic backgrounds, races, genders, and health conditions.

In this paper, we studied the relationship between educational level and weekly working hours after controlling for major potential endogenous variables such as sex, income, and health status. We obtained data from IPUMS CPS and used linear regression models to analyze the research question. We used squared years of education as our independent variable by comparing six model specifications. We conducted the hypothesis tests for our control variables and interaction terms and found that all controls are useful in understanding the effect of an additional year of education on weekly working hours.

This is in contrast to our expectation that a higher education level reduces the weekly working hours after one enters the labor market, which should unambiguously lead to better living standards after controlling for income level. We found that, in general, a higher educational level is associated with more hours worked per week. One great limitation of our study is the data. One reasonable consideration is that employed people have to choose between paycheck and leisure time. And the pay from a job is highly dispersed across industries, which requires different levels of educational background. It would be great to include job type or industry as an additional control. However, given the large number of industry codes used in the IPMUS CPS data, we cannot include that in our current regression model. Furthermore, past studies have focused on the fact that one's years of education do not directly speak for one's ability to work and thus suggested including controls in cognitive ability and other clinical or test measurements. Unfortunately, these measurements are not available in our dataset. These are all interesting works left for future studies.

VII. References

Dillard, A. (1989). *The Writing Life*. Harper & Row.

Pryce-Jones, J. (2010). Happiness at Work: Maximizing Your Psychological Capital for Success. Wiley-Blackwell.

Thomas, J. (n.d.). You'll Spend 90,000 Hours of Your Life at Work - Make Them Count. The National News. Retrieved from https://www.thenationalnews.com/opinion/comment/you-ll-spend-90-000-hours-of-your-life-at-work-make-them-count-1.1054726

Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. Journal of Economic Literature, 46(3), 607-668. https://doi.org/10.1257/jel.46.3.607

Zajacova, A., Hummer, R. A., & Rogers, R. G. (2012). Education and Health among U.S. Working-Age Adults: A Detailed Portrait across the Full Educational Attainment Spectrum. Biodemography and Social Biology, 58(1), 40-61. https://doi.org/10.1080/19485565.2012.666122

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler and Michael Westberry. IPUMS CPS: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D030.V11.0

Wirtz, A., Lombardi, D. A., Willetts, J. L., Folkard, S., & Christiani, D. C. (2012). Gender differences in the effect of weekly working hours on occupational injury risk in the United States working population. *Scandinavian Journal of Work, Environment & Health*, *38*(4), 349–357. http://www.jstor.org/stable/41508902

VIII. Appendix

Table 1: Description of Variables

year	Year of the observation
age	Age of the observation
nchlt5	Number of children in the household under the age of 5
uhrsworkt	Hours usually worked per week at all jobs
health	An individual's self perception of one's health status on a five-point scale from 1 to 5, with 1 being excellent, 2 being very good, 3 being good, 4 being fair, and 5 being poor.
female	An indicator of gender of the respondent with 1 being female and 0 being male.
colored	An indicator of race of the individual with 1 means all other races included other than purely white (which includes individuals with a biracial status that contains white as one of it) and 0 means the individual is white.
married	An indicator of marital status of the individual with 1 means the respondent is married and 0 means not currently married, which includes all the cases where the respondent is separated, divorced, widowed, or simply never married (single).
educ_rev	The number of years of education an individual has completed by the time of the survey. For example, "educ_rev" = 12 means a person has successfully graduated from high school. "Educ_rev" = 14 could means that an individual drops out after sophomore in college and could also means that he completed a 2 year of associate degree after high school. We converted all the educational attainment levels into years for ease of interpretation of the regression coefficients.
disabled	An indicator of the disability status of the individual with 1 means the respondent is disabled and 0 means no disable that affects work.
logPersonalInc	The natural log of the total personal income of the individual.
logHouseholdInc	The natural log of the total household income of the individual.
educ_rev_sq	Squared of years of education "educ_rev"
y_2003	Indicator variable that the year of the data is in 2003.
y_2006	Indicator variable that the year of the data is in 2006.
y_2009	Indicator variable that the year of the data is in 2009.
y_2012	Indicator variable that the year of the data is in 2012.
y_2015	Indicator variable that the year of the data is in 2015.
y_2018	Indicator variable that the year of the data is in 2018.
educ_gender	Interaction term between "educ_rev" and the gender of the respondent.
educ_age	Interaction term between "educ_rev" and the age of the respondent.
educ_race	Interaction term between "educ_rev" and the race of the respondent.
educ_married	Interaction term between "educ_rev" and the marital status of the respondent.

Table2: Descriptive Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
year	543,482	2,009	5.760	2,000	2,018
age	543,482	41.15	13.13	15	80
nchlt5	543,482	0.196	0.498	0	5
uhrsworkt	543,482	39.63	11.78	0	100
health	543,482	2.036	0.913	1	5
female	543,482	0.481	0.500	0	1
colored	543,482	0.188	0.391	0	1
married	543,482	0.603	0.489	0	1
educ_rev	543,482	13.70	2.862	0	21
disabled	543,482	0.0202	0.141	0	1
logPersonalInc	543,482	10.33	1.089	0	13.81
logHouseholdInc	543,482	11.14	0.796	0	13.81

Table 3: Descriptive Statistics based on Race

	(1)	(2)	(3)	(4)	(5)
	colored 0		colored 1		
VARIABLES	N	mean	N	mean	Difference
year	441,064	2,009	102,418	2,010	-1
age	441,064	41.19	102,418	40.96	0.23
nchlt5	441,064	0.202	102,418	0.172	0.03
uhrsworkt	441,064	39.68	102,418	39.42	0.26
health	441,064	2.007	102,418	2.161	-0.154
female	441,064	0.471	102,418	0.526	-0.055
married	441,064	0.628	102,418	0.495	0.133
educ_rev	441,064	13.69	102,418	13.78	-0.09
disabled	441,064	0.0201	102,418	0.0203	-0.0002
logPersonalInc	441,064	10.34	102,418	10.27	0.07
logHouseholdInc	441,064	11.17	102,418	11.02	0.15

Table 4: Descriptive Statistics based on Gender

	(1)	(2)	(3)	(4)	(5)
	female 0		female 1		
VARIABLES	N	mean	N	mean	Difference
year	281,964	2,009	261,518	2,009	0
age	281,964	41.26	261,518	41.03	0.23
nchlt5	281,964	0.216	261,518	0.175	0.041
uhrsworkt	281,964	42.13	261,518	36.94	5.19
health	281,964	2.017	261,518	2.057	-0.04
colored	281,964	0.172	261,518	0.206	-0.034
married	281,964	0.641	261,518	0.561	0.08
educ_rev	281,964	13.60	261,518	13.82	-0.22
disabled	281,964	0.0190	261,518	0.0214	-0.0024
logPersonalInc	281,964	10.53	261,518	10.11	0.42
logHouseholdInc	281,964	11.18	261,518	11.10	0.08

Table 5: OLS Simple Regression Results for Different Specifications

VARIABLES	(1) Lin-Lin Model	(2) Quadratic Model	(3) Cubic Model	(4) Log-lin Model	(5) Log-log Model	(6) Lin-log Model
educ_rev	0.619***			0.0187***		
	(0.00573)			(0.000200)		
educ_rev_sq		0.0230***				
		(0.000207)				
educ_rev_cu			0.00100***			
			(9.24e-06)			
leduc_rev					0.223***	7.278***
					(0.00264)	(0.0745)
Constant	31.15***	35.12***	36.71***	3.362***	3.039***	20.74***
	(0.0806)	(0.0439)	(0.0313)	(0.00288)	(0.00698)	(0.195)
Observations	543,482	543,482	543,482	543,284	542,340	542,538
R-squared	0.0226	0.0243	0.0239	0.0171	0.0147	0.0189
				(0.0232)	(0.0202)	

Table 6: OLS Multiple Regression Model Results

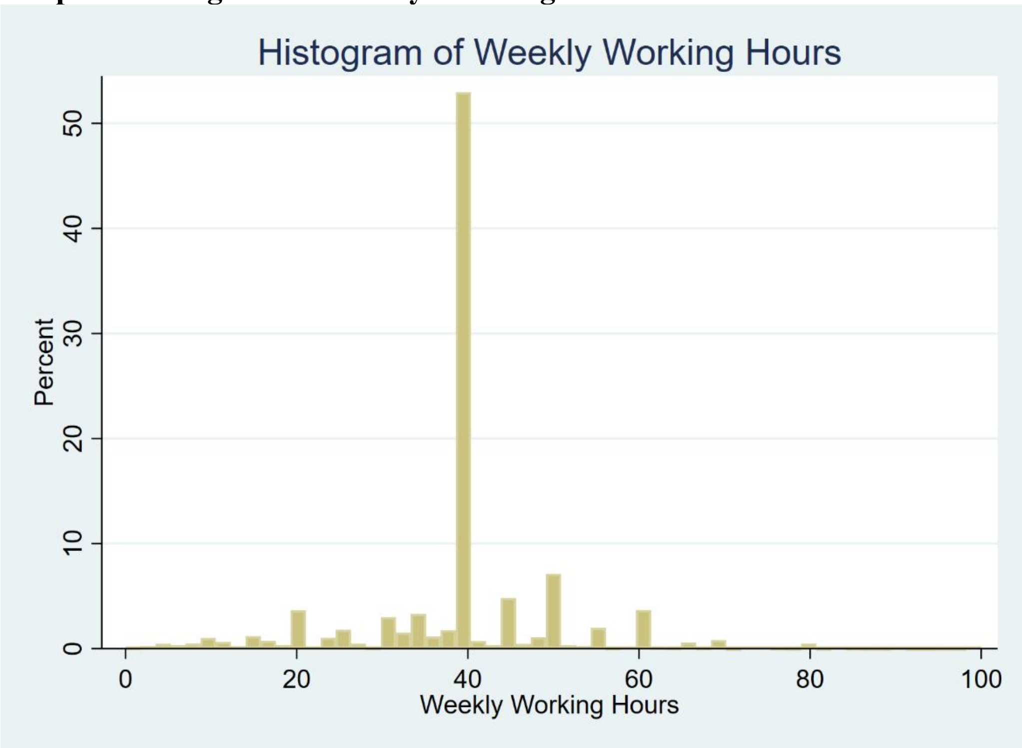
VARIABLES	(1) Multiple Regression without Interactions	(2) Multiple Regression with Interactions
age	-0.0185***	0.0874***
	(0.00160)	(0.00380)
nchlt5	-0.204***	-0.212***
	(0.0309)	(0.0310)
disabled	-5.064***	-5.087***
	(0.135)	(0.135)
health	0.0670***	0.0668***
	(0.0169)	(0.0169)
female	-3.194***	-3.200***
	(0.0312)	(0.0805)
colored	0.213***	0.809***
	(0.0355)	(0.101)
educ_rev_sq	0.00502***	0.0396***
	(0.000240)	(0.000852)
married	1.507***	3.741***
	(0.0332)	(0.0875)
logPersonalInc	4.837***	4.683***
	(0.0348)	(0.0348)
logHouseholdInc	-2.046***	-1.891***
	(0.0321)	(0.0322)
y_2003	-0.963***	-0.949***
	(0.0573)	(0.0571)
y_2006	-0.956***	-0.942***
	(0.0581)	(0.0580)
y_2009	-1.913***	-1.914***
	(0.0582)	(0.0581)
y_2012	-2.005***	-2.023***
	(0.0586)	(0.0584)
y_2015	-1.817***	-1.832***
	(0.0590)	(0.0589)
y_2018	-2.105***	-2.133***
	(0.0605)	(0.0603)
educ_gender		-0.000769**
		(0.000383)
educ_race		-0.00328***
		(0.000484)
educ_married		-0.0122***
		(0.000431)
educ_age		-0.000571***
		(1.80e-05)
Constant	14.28***	7.950***

	(0.332)	(0.352)
Observations	543,482	543,482
R-squared	0.215	0.220

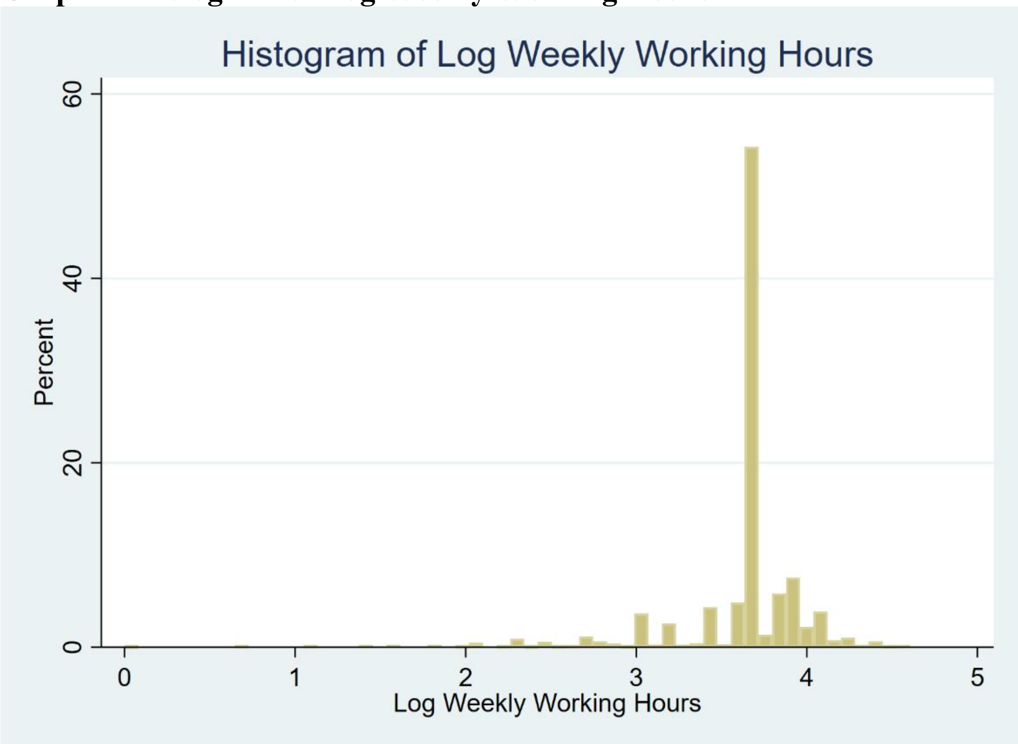
Table 7: Hypothesis Testing

Null Hypothesis	F-stat	Multiple Linear Regression Model with Interaction P-value
$H_0: \beta_2 = 0$	528.82	0.0000
$H_0: \beta_3 = 0$	46.88	0.0000
$H_0: \beta_4 = 0$	1426.44	0.0000
$H_0: \beta_5 = 0$	15.68	0.0000
$H_0: \beta_6 = 0$	1581.46	0.0000
$H_0: \beta_7 = 0$	64.35	0.0000
$H_0: \beta_9 = 0$	1829.43	0.0000
$H_0: \beta_{10} = 0$	18160.57	0.0000
$H_0: \beta_{11} = 0$	3438.80	0.0000
$H_0: \beta_{18} = 0$	4.02	0.0449
$H_0: \beta_{19} = 0$	45.91	0.0000
$H_0: \beta_{20} = 0$	806.38	0.0000
$H_0: \beta_{21} = 0$	1009.00	0.0000
$H_0: \beta_4 = \beta_5$	1416.98	0.0000
$H_0: \beta_{10} = \beta_{11}$	12908.42	0.0000

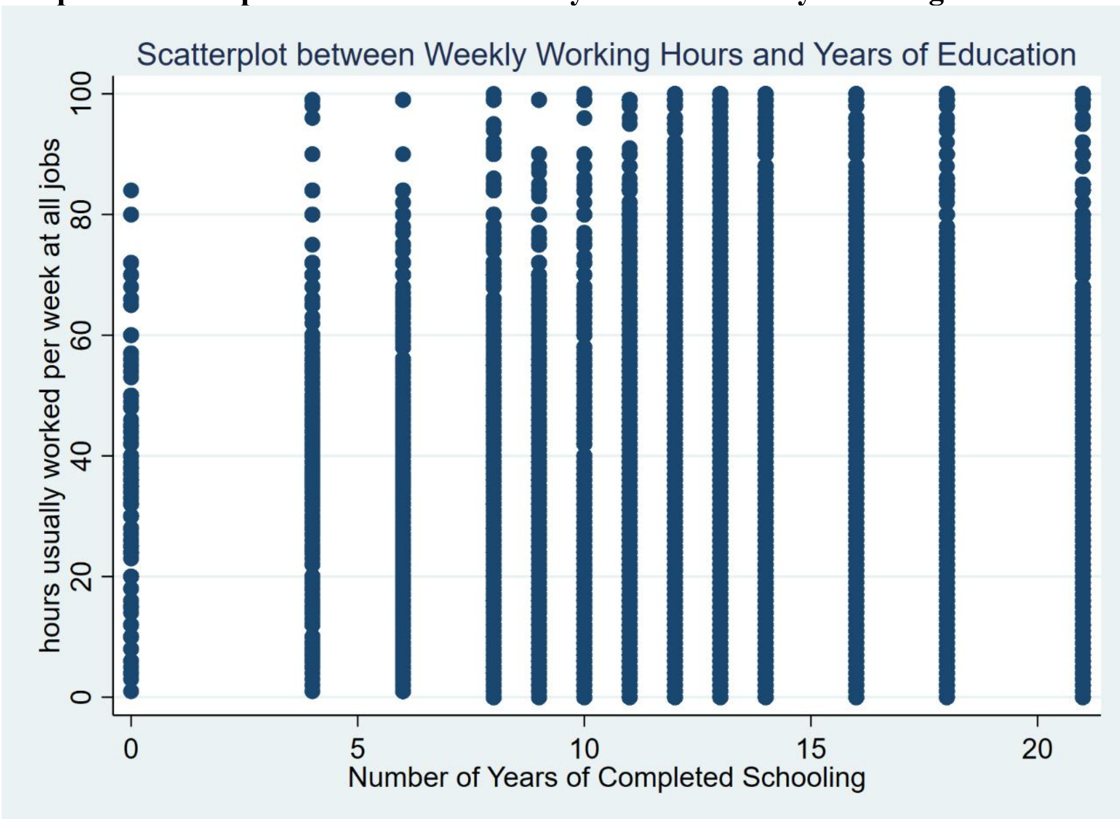
Graph 1: Histogram of Weekly Working Hours



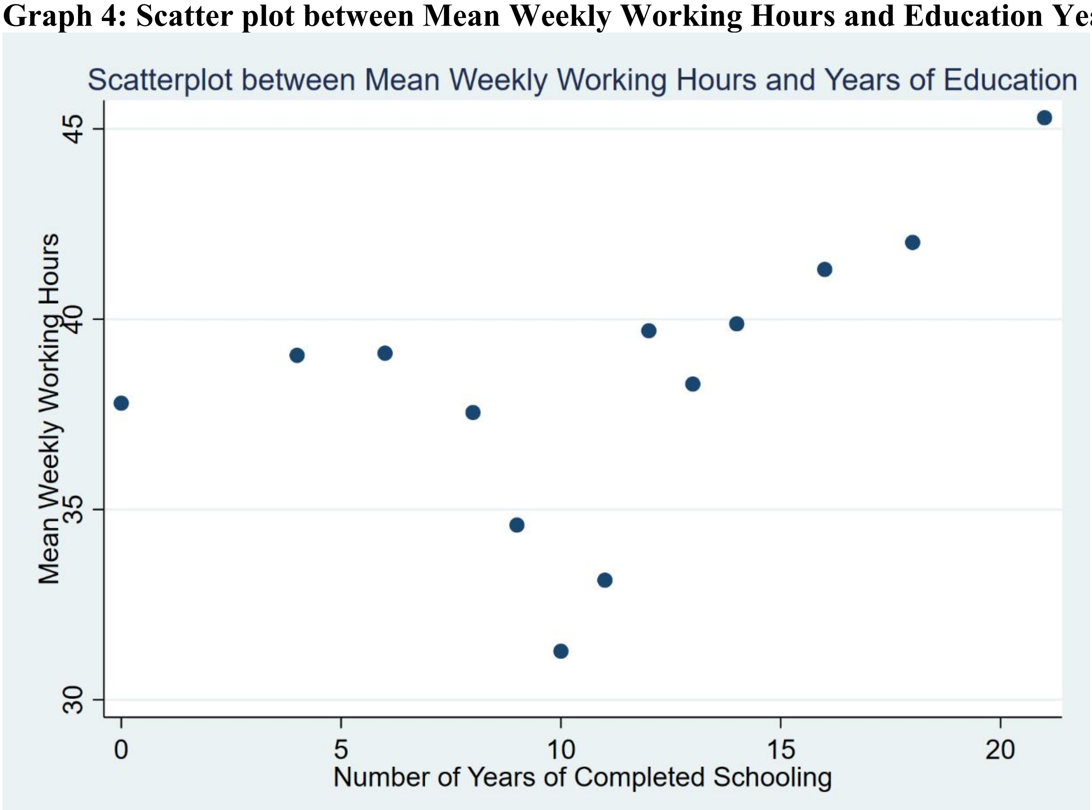
Graph 2: Histogram of Log Weekly Working Hours



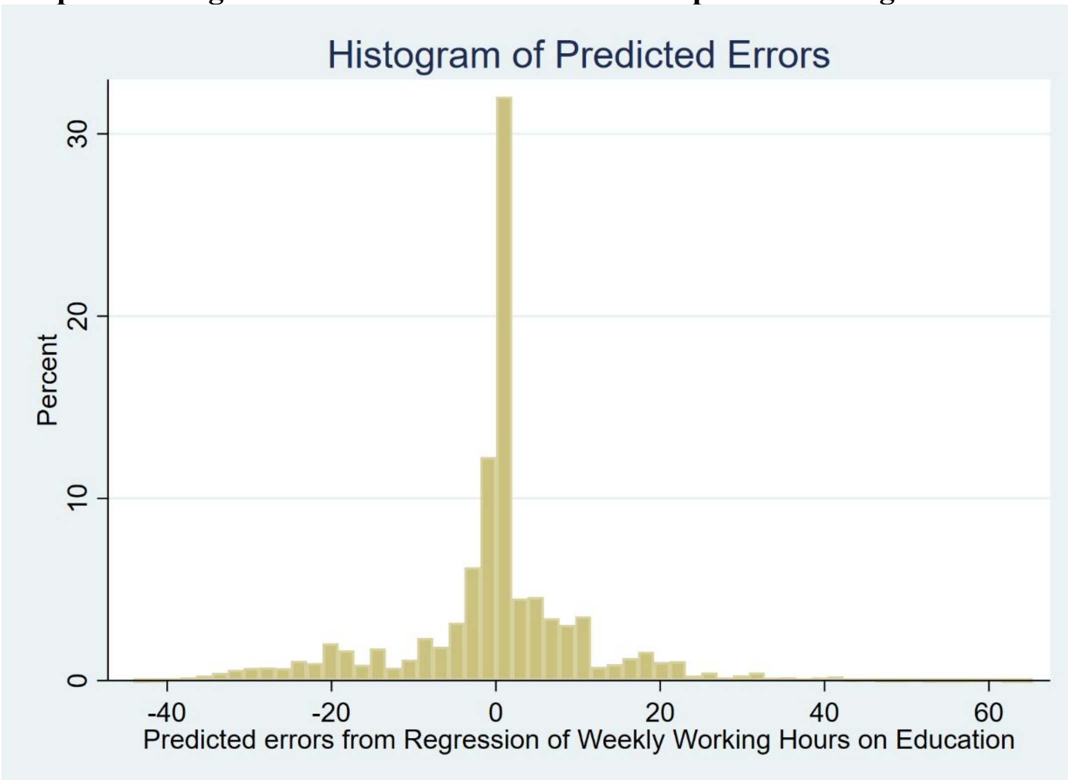
Graph 3: Scatter plot between Education years and Weekly Working Hours



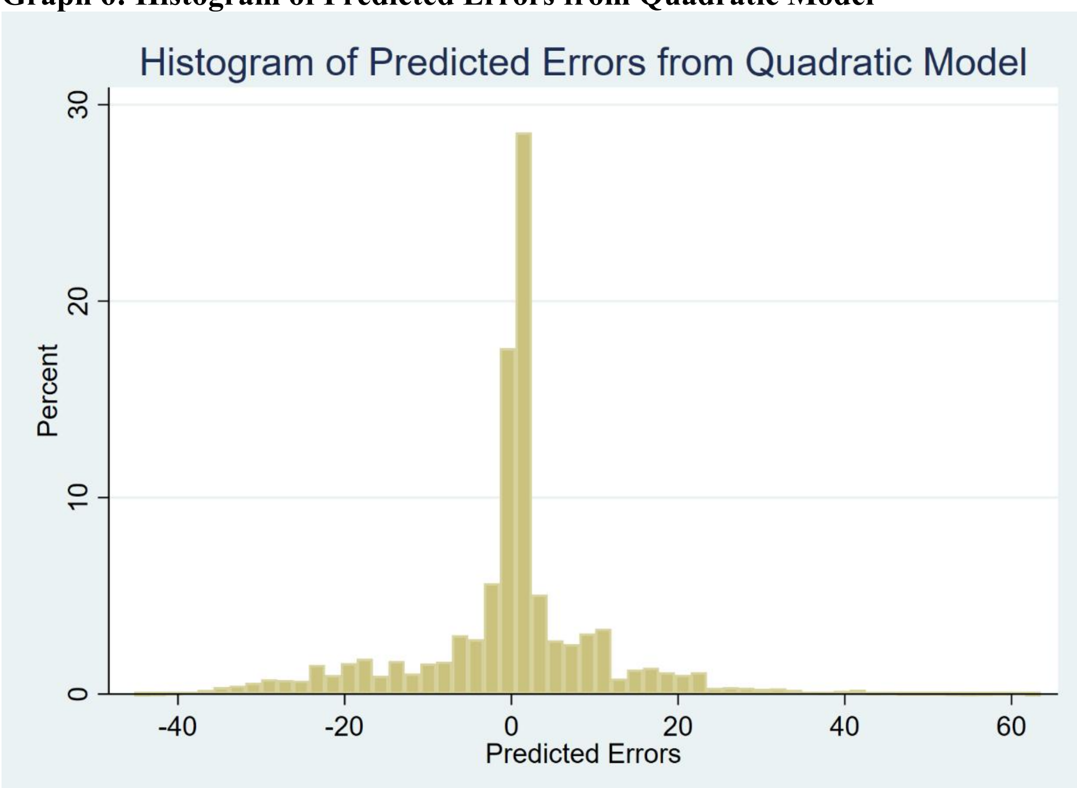
Graph 4: Scatter plot between Mean Weekly Working Hours and Education Years



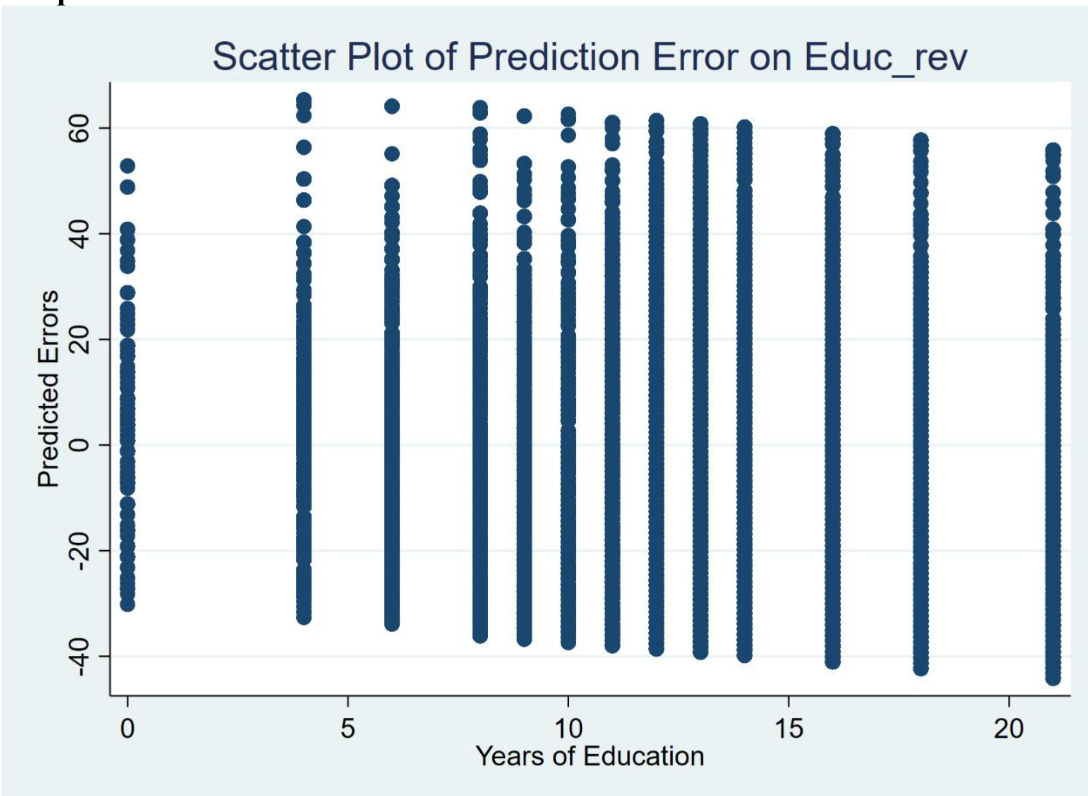
Graph 5: Histogram of Predicted Errors from Simple Linear Regression Model



Graph 6: Histogram of Predicted Errors from Quadratic Model



Graph 7: Scatter Plot of Prediction Error on Education Years



Do files:

```
clear all

log using "/Users/freddie/Desktop/EC 204/final_paper.log", replace

//========================================================================
========//
// Data Cleaning
//========================================================================
========//

// import data downloaded from IPMUS CPS
use "/Users/freddie/Desktop/EC 204/cps_raw.dta", replace

// check variable lists to decide which variables to drop
label list

// drop unrelated variables
drop hflag asecwth asecflag cpsid serial pernum cpsidv cpsidp asecwt educ99 uhrswork1
yrimmig occ month

// Time selection: Pick every 3 years from 2000 to 2018
keep if year == 2000 | year == 2003 | year == 2006 | year == 2009 | year == 2012 | year == 2015
| year == 2018

// drop all missing values
drop if uhrsworkt == 997 | uhrsworkt == 999
drop if educ == 999 | educ == 0 | educ == 1
drop if disabwrk == 0
drop if empstat == 0
drop if marst == 9
drop if race == 999
drop if sex == 9

// Only focus on those that are at worked
keep if empstat == 10

// generate dummies for each controls categories
// gender: let male be the base group
cap gen female = 1 if sex == 2
replace female = 0 if sex ==1

// race: let white be the base group
cap gen colored = 1 if race != 100
replace colored = 0 if race == 100
```

```
// Marital status: let single be base group
cap gen married = 1 if marst == 1 | marst == 2
replace married = 0 if marst != 1 & marst != 2

// Disable status: let no disable be base group
cap gen disabled = 1 if disabwrk == 2
replace disabled = 0 if disabwrk == 1

// Transform eduction into educationg years as done in Do-file#1_Cleaning_data_F23.do
provided in blackboard.
cap gen educ_rev = educ
replace educ_rev = 0 if educ == 2
replace educ_rev = 4 if educ == 10
replace educ_rev = 6 if educ == 20
replace educ_rev = 8 if educ == 30
replace educ_rev = 9 if educ == 40
replace educ_rev = 10 if educ == 50
replace educ_rev = 11 if educ == 60 | educ == 71
replace educ_rev = 12 if educ == 73
replace educ_rev = 13 if educ == 81
replace educ_rev = 14 if educ == 91 | educ == 92
replace educ_rev = 16 if educ == 111
replace educ_rev = 18 if educ == 123
replace educ_rev = 21 if educ == 124 | educ == 125
replace educ_rev = . if educ == .


// drop outliers: millionaires and negative income
drop if inctot >= 1000000
drop if inctot <=0
// do a log transformation
cap gen logPersonalInc = log(inctot)

// Do the same transformation for household income
drop if hhincome >= 1000000
drop if hhincome <=0
cap gen logHouseholdInc = log(hhincome)

// Drop outliers in the weekly working hours
drop if uhrsworkt > 100

// Drop outliers in age
drop if age > 80
```

```
//================================================================================//
//Descriptive Analysis
//================================================================================//

// Summary Statistics
summarize year age nchlt5 uhrsworkt disabled health female colored educ_rev married logPersonalInc logHouseholdInc
outreg2 using SummaryStat_all.doc, replace sum(log) keep(year age nchlt5 uhrsworkt disabled health female colored educ_rev married logPersonalInc logHouseholdInc)

// Summary Statistics by gender
bys female: sum year age nchlt5 uhrsworkt disabled health colored educ_rev married logPersonalInc logHouseholdInc
bys female: outreg2 using SummaryStat_bySex.doc, replace sum(log) eqkeep(N mean) keep(year age nchlt5 uhrsworkt disabled health colored educ_rev married logPersonalInc logHouseholdInc)

bys colored: sum year age nchlt5 uhrsworkt disabled health female educ_rev married logPersonalInc logHouseholdInc
bys colored: outreg2 using SummaryStat_byRace.doc, replace sum(log) eqkeep(N mean)keep(year age nchlt5 uhrsworkt disabled health female educ_rev married logPersonalInc logHouseholdInc)


// Plots of major variables
// twoway (scatter uhrsworkt educ_rev)

hist uhrsworkt, percent ytitle(Percent) xtitle(Weekly Working Hours) title(Histogram of Weekly Working Hours)
graph export "uhrsworkt_hist.png", as(png) replace

// hist inctot, percent
// hist logPersonalInc, percent
// hist logHouseholdInc, percent
// hist age, percent


// Correlation table
corr



// OLS assumption check
```

```
// check the OLS assumptions using these 3 commands
twoway (scatter uhrsworkt educ_rev),  xtitle(Number of Years of Completed Schooling)
title(Scatterplot between Weekly Working Hours and Years of Education, size(medium))
graph export "scatter_model1.png", as(png) replace

// scatterplot that plots mean of weekly earnings at each education level
preserve
collapse (mean) uhrsworkt, by(educ_rev)
twoway (scatter uhrsworkt educ_rev), ytitle(Mean Weekly Working Hours) xtitle(Number of
Years of Completed Schooling) title(Scatterplot between Mean Weekly Working Hours and
Years of Education, size(medium))
restore
graph export "scatter_model2.png", as(png) replace

// Test for heteroskedasticity
reg uhrsworkt educ_rev
hettest

// chi2(1) = 285.17
// Prob > chi2 = 0.0000

* check the mean, variance, skewness and kurtosis
cap predict ehat1, resid
sum ehat1, detail

*  Graph of predicted error on education years
twoway (scatter ehat1 educ_rev), ytitle(Predicted Errors) xtitle(Years of Education) title(Scatter
Plot of Prediction Error on Educ_rev)
graph export "Error on Educ.png", as(png) replace

* check the histogram of residuals to check skewness
hist ehat1, percent ytitle(Percent) xtitle(Predicted errors from Regression of Weekly Working
Hours on Education) title(Histogram of Predicted Errors)
graph export "ehat1.png", as(png) replace


//======================================================================
========//
// Model 1: Regression of uhrsworkt on education (both are in linear terms)
//======================================================================
========//
reg uhrsworkt educ_rev, robust
outreg2 using SingleRegression.doc, replace ctitle(Lin-Lin Model) rdec(4)

* generating the predicted values from the regression
cap predict yhat1, xb
```

```
* generating residuals from the regression
cap predict ehat1, resid


//===========================================================================
========//
// Model 2: Regression of uhrsworkt on education^2 (quadratic in education)
//===========================================================================
========//
* generating a quadratic in education
cap gen educ_rev_sq = educ_rev*educ_rev

* regressing earnings on education square and predicting y
reg uhrsworkt educ_rev_sq, robust
outreg2 using SingleRegression.doc, append ctitle(Quadratic Model) rdec(4)
cap predict yhat2, xb

* (2) Scatterplot of estimated errors and x
quietly reg uhrsworkt educ_rev_sq
cap predict ehat2, resid

hist ehat2, percent ytitle(Percent) xtitle(Predicted Errors) title(Histogram of Predicted Errors
from Quadratic Model)
graph export "Quadratic_error_hist.png", as(png) replace

//===========================================================================
========//
// Model 3: Regression of uhrsworkt on education^3, also called cubic model
//===========================================================================
========//
* generating a cubic in education
cap gen educ_rev_cu = educ_rev*educ_rev*educ_rev

* regressing earnings on education cube
reg uhrsworkt educ_rev_cu, robust
outreg2 using SingleRegression.doc, append ctitle(Cubic Model) rdec(4)

//===========================================================================
========//
// Model 4: Regression of log(uhrsworkt) on education, also called log-lin
//===========================================================================
========//
* creating log earnings
cap gen luhrsworkt = log(uhrsworkt)
```

*comparing the distribution of uhrsworkt and log(uhrsworkt). uhrsworkt has been done before and saved as uhrsworkt_hist.png

hist luhrsworkt, percent ytitle(Percent) xtitle(Log Weekly Working Hours) title(Histogram of Log Weekly Working Hours)
graph export “luhrsworkt.png”, as(png) replace

*regressing log earnings on education
reg luhrsworkt educ_rev, robust
outreg2 using SingleRegression.doc, append ctitle(Log-lin Model) rdec(4)

* Adjust R squared
cap predict lyhat_s, xb
cap gen yhat_s = exp(lyhat_s)
corr uhrsworkt yhat_s
scalar R_sq_loglin_s = r(rho)^2
di R_sq_loglin_s
// R_sq_loglin_s.02317537

//===
========//
// Model 5: Regression of log(uhrsworkt) on log(education), also called log-log
//===
========//
*creating log education
cap gen leduc_rev = log(educ_rev)

*regressing log earnings on log education
reg luhrsworkt leduc_rev, robust
outreg2 using SingleRegression.doc, append ctitle(log-log Model) rdec(4)

* Adjust R squared
cap predict lyhat_m2, xb
cap gen yhat_m2 = exp(lyhat_m2)
corr uhrsworkt yhat_m2
scalar R_sq_loglog = r(rho)^2
di R_sq_loglog
// R_sq_loglog = .02017149

//===
========//
// Model 6: Regression of uhrsworkt on log(education), also called lin-log
//===
========//
reg uhrsworkt leduc_rev, robust

```
outreg2 using SingleRegression.doc, append ctitle(lin-log Model) rdec(4)


//==========================================================================
========//
// Multiple Regression Model with control variables
//==========================================================================
========//

// Generate Year dummy with base year set to 2000.
cap gen y_2003 = 1 if year == 2003
replace y_2003 = 0 if year != 2003

cap gen y_2006 = 1 if year == 2006
replace y_2006 = 0 if year != 2006

cap gen y_2009 = 1 if year == 2009
replace y_2009 = 0 if year != 2009

cap gen y_2012 = 1 if year == 2012
replace y_2012 = 0 if year != 2012

cap gen y_2015 = 1 if year == 2015
replace y_2015 = 0 if year != 2015

cap gen y_2018 = 1 if year == 2018
replace y_2018 = 0 if year != 2018


// First multiple regression model that includes everything (only includes squared educ years)
reg uhrswork t age nchlt5 disabled health female colored educ_rev_sq married logPersonalInc
logHouseholdInc y_2003 y_2006 y_2009 y_2012 y_2015 y_2018, robust
// Output multiple regression results into a table
outreg2 using MultipleRegression.doc, replace ctitle(Multiple Regression without Interactions)


// Interaction terms between educ_rev and gender
cap gen educ_gender = educ_rev_sq*female

// Interaction terms between educ_rev and race
cap gen educ_race = educ_rev_sq*colored

// Interaction terms between educ_rev and marital status
cap gen educ_married = educ_rev_sq*married
```

```
// Interaction terms between educ_rev and age
cap gen educ_age = educ_rev_sq*age

// Multiple regression model with all the interaction terms
reg uhrsworkt age nchlt5 disabled health female colored educ_rev_sq married logPersonalInc
logHouseholdInc y_2003 y_2006 y_2009 y_2012 y_2015 y_2018 educ_gender educ_race
educ_married educ_age, robust
// Output multiple regression with interaction results into a table
outreg2 using MultipleRegression.doc, append ctitle(Multiple Regression with Interactions)



//=====================================================================
========//
// Hypothesis Testing
//=====================================================================
========//
* Test if each control variable has significant impact over weekly working hours
test age
test nchlt5
test disabled
test health
test female
test colored
test married
test logPersonalInc
test logHouseholdInc


* Test if interaction terms are significant
test educ_gender
test educ_race
test educ_married
test educ_age

* Test if health has the same effect as diable
test health = disabled

* Test if personal income has the same effect as household income
test logHouseholdInc = logPersonalInc

save "/Users/freddie/Desktop/EC 204/cps_raw_new.dta", replace

log close
```