Effect of education on household disposable income

Sheng (Jeremy) Jiang, Qiluoao(Dylan) Guo, Yile Tan
University of Washington

Abstract

Using time series data from several sources, this research paper expects to find the correlation between education, percentage of low pay wage level, percentage of foreigners in total population, and consumer price index, and etc. with household disposable income. This paper will indicate the significance of those data indicators affecting overall household disposable income. In the whole background, a deeper analysis of those data will assist household, organization and government more effectively to increase household disposable income through improved performances.

Keywords: household disposable income, education

Introduction

It's believed in the society that there is a strongly connection between knowledge and financial success. There is an assumption most people telling and being told that one of keys to accomplish financial success is to have better education. Is there a direct connection between knowledge and financial status? For the term of knowledge, we have focused on education level as a better measurable indicator. For the financial status, we will use the household disposable income as the indicator to represent the status. So our mission of the research is to find the effectiveness of education on household disposable income. Does a higher level of education can lead to a higher household disposable income or are there additional indicators to the connection of education and household disposable income?

We put the United States as the example country in the research. We sought to draw connection between the adult education level and household disposable income in the United States, and draw new indicators which may link with education and income. Then we sought to find if indicators of the connection fits the linear relationship, or doesn't qualify for linear correlation.

Any correlation from the research can be used as clues to increase household disposable income, as well as improve quality of line in the United States. In this research, we sought to find most important indicator which affects income. And we expected collect more information from those correlations to provide solutions, suggestions on improving household disposable income.

Literature review

Income, as one of the most real-life related economic factor, has been investigated and analyzed for a long time. Many factors such as characteristics of worker, situation of the workplace, and economic status of the country would have great effect on individual income.

Not only in the recent years, economists have done a large quantity of research on what would affect personal income and how such factors influence it. Among those works, "COLLEGE FOR ALL?", an article from Anthony P. Carnevale discusses that corresponding to a deep cultural preconception "college for all", undergraduate degree promotes individual's career.

Among all the factor in career field, higher degree not only gives more chance for finding a job, but also implies higher personal income. In the growth industries, jobs requiring postsecondary education are significantly increasing, leading to a rising demand for people with higher degree. By all appearance, the most important signal of the rising demand is the income gap between employees with different education background. By the supply-demand model, the demand for higher educated worker increases, the salary that firms pay to those workers would also increases. Carnevale show a time series statistics of disposable income by level of education since 1976 displays that people of high level of education earn more than people with relative lower education background. Besides, he discusses reasons of this phenomenon. Postsecondary education is becoming core workforce- development system and postsecondary-educated workers are likely to be more effective in high-performance work organization (Carnevale 2008). As a result, competition for high-educated worker accelerate the income rising of those workers, eventually amplifying the income gap between them, which is also supported by work of Purcell, Iams and Shoffner.

In their research paper, "EDUCATION, EARNINGS INEQUALITY, AND FUTURE SOCIAL SECURITY BENEFITS: A MICROSIMULATION ANALYSIS" (2015), Purcell, Iams, and Shoffner demonstrate the correlation between whether receiving college education and the income growth rate and its effect on annual income. A line chart directly shows the gap between the income. With the detailed data about workers' earning from Social Security record,

they found out that different educational attainment would give rise to the gap between rate of income growth and considerably enlarge the gap in annual income between college graduates and high school graduates.

Moreover, Social Security benefits are used for presents the influence of income rising by college education attainment. That is, they also predict how the educational background would affect future earning. They conduct their research by using an SSA micro-simulation model called Modeling Income in the Near Term to estimate the effect of income-growth differentials by educational attainment on the future earning of individual born during 1965-1979(Purcell, Iams, and Shoffner 2015). Result shows that, in general, people with college education are predicted to have an income rise nearly as twice as that of those who do not attend college.

Above that, separating data and analyzing by gender, they surprisingly find out that generally people with college education would be expected to receive higher income, but there is a distinct difference in income rise across gender. Their finding of sexual difference between income is expanded and deepened by Emily Murphy and Daniel Oesch in their work titled "THE FEMINIZATION OF OCCUPATIONS AND CHANGE IN WAGES: A PANEL ANALYSIS OF BRITAIN, GERMANY, AND SWITZERLAND."

Murphy and Oesch analyze the gender income gap over 1990s and 2000s in Britain, Germany and Switzerland and finally come to a result that occupational feminization is accompanied by a decline in workers' income (Murphy and Oesch 2016). They first objectively talked about the difference cross gender then test their hypothesis about the impact of gender and those cross-sexual differences in people's career. Wage disparities between male and female occupations are, to some extent, rooted in unequal gender norms that accord higher value to male work effort, and thus imply labor-market discrimination against women.

Their analysis is based on longitudinal panel data from the British Household Panel (BHPS), the German Socio-Economic Panel (SOEP), and the Swiss Household Panel (SHP) to estimate individual fixed effects for male and female and regress income on variable derive from hypothesis, percentage of feminization in level of job-specific skills, time investment in housework and human capital, separately. It is represented in the result that if moving to a more female occupational workplace, workers would face a loss in their earning. Considering other factor that are formerly considered effective factors of income, level of job-specific skills, time investment in housework and human capital are not statistically significant to explain the income gap between female and male.

Besides, another interesting article, "DOES IMMIGRATION HARM THE POOR?", by Steven Camarota discuss the effect of immigration of natives in the United States. He mentions a study using government survey data to analyze the effect on natives' working opportunity and disposable income. An increase in unskilled worker, from immigrants, improve the competition of low-wage jobs and then due to the demand-supply model, salary of low-wage jobs gradually decreases. In contrast, workers who do not involve in the competition with immigrants would come across a pay rise.

Summarizing from their works, many factors, educational background, gender and foreign immigration, contribute to the change in household income. Under this circumstance, we could establish our research on testing their conclusion with data we found, and also pick other factor that may affect disposable income.

Econometric models

In total, four models are used in this project to test the factors that can affect wage: simple linear regression model; multiple regression model; restricted and unrestricted model; and time series regression model.

In the simple linear regression model, we try to explain household disposable income in terms of variable US's adult education level. It shows how household disposable income varies with changes in US's adult education level. Apparently, in the real world, household disposable income cannot only be determined by US's adult education level. However, the result can give us an orientation about the general relation between US's adult education level and household disposable income and what other variables can affect the household disposable income. We use Ordinary Least Squares (OLS) estimates.

In multiple regression model, we incorporate more explanatory factors into the model: M1, M2, M3 and M4. We believe that Standard assumptions hold: 1. All parameters are linear; 2. All samples are randomly selected; 3. There is no perfect collinearity; 4. The conditional mean is zero; 5. Homoskedasticity: the value of the explanatory variables must contain no information about the variance of the unobserved factors. 6. It is assumed that the dependent variable is normally distributed. We use T-test to test our hypothesis that whether these variable a statistically significant.

In the restricted and unrestricted models, we want to test whether the variables: percentage of worker with low pay, percentage of foreign worker, and percentage of female in population are jointly significant and whether there is multicollinearity between them. We assume that the six Standard assumptions hold as well. We use F-test to test our hypothesis that whether these variable are jointly significant.

We also have time series regression models since we want to know how time trend affects the household disposable income as well. In this model, we still have six assumptions: 1.All parameters are linear; 2.There is no perfect collinearity; 3. The conditional mean is zero; 4.The volatility of the errors must not be related to the explanatory variables in any of the period; 5.Conditional on the explanatory variables, the unobserved factors must not be correlated over time. 6. It is assumed that the dependent variable is normally distributed.

Description of data

Independent and Dependent Variables

The research focused on finding a correlation between US's adult education level and household disposable income. For adult education level, we used the data on the perspective of tertiary level in the percentage of 25 to 64-year-old group. The hypothesis of the research was that they have a positive correlation, when the adult education level increases, the household disposable income rises. Through the cause-and-effect relationship, adult education level is the independent variable we used. Since adult education level indicates technology, infrastructure, social development of a country rather depending on those developments, adult education level was independent variable in the case. Household disposable income, on the other hand, affecting by those development, was the dependent variable.

We also found there were other variables besides education could be explanatory variables in the relationship between education and income. Some other explanatory variable that affected income were low paid wage as a percent of total wage, percent of foreign population in total population, female percentage in population, and annual growth rate, etc. There were also insignificant indicators that didn't satisfy the assumptions as they had perfect correlation with household disposable income.

Source of Data

All of research data comes from the OECD, Organization for Economic Co-Operation and Development, from year of 1980 to 2015. Since we only focused on United States, there are 36 observations. In additional to dependent variable, we also considered average wage, crude oil import price, crop production, hours employee worked, and tax revenue. However, all of those indicators does not qualify for linear correlation.

Empirical result

Simple Linear Regression

Model S1:

Household Disposable Income= −24670.221+1527.8905 education +u (other factor)

```
constant:   beta0 =      -24670.221    standard error =       1607.2096                t-stat =      -15.349722
education:        beta1 =       1527.8905    standard error =       46.490975                t-stat =       32.864238


Sample size =       36.000000
R2 =              0.96948095
```

Model S2: Log (Household Disposable Income) = 7.9993062+ 0.062382502education +u (other factor)

```
constant:   beta0 =       7.9993062    standard error =      0.042540178                t-stat =       188.04120
education:        beta1 =      0.062382502    standard error =     0.0012305392                t-stat =       50.695258


Sample size =       36.000000
R2 =              0.98694321
```

Based on the result from regression, model S1 found β1 was 1527.8905. It meant that as the percentage of tertiary education level of 24 to 65-year-old group increased by 1%, the household disposable income increased by 1527.8905 USD per capita. In the model S2, we used log function of y. β1 in model S2 was 0.062382502, meaning as the percentage of tertiary education level of 24 to 65-year-old group increased by 1%, the household disposable income increased by 0.0062382502. The R-squared value in model S1 was 96.9488095%, while in model S2 was 98.694321%. These two high R-squared values in two different models indicated that around 95-98% of data could be explained. And model S2 seemed more reasonable as model 1's β1 was overstated. Because 1% change of education level realistically didn't have such huge impact on household disposable income. And model S2 also had positive B2 =7.9993, smaller standard errors as well as smaller t-statistics. So we preferred model S2 as the simple regression model. However, the model's t stats on B0 and B1 were huge numbers, which were odd in the research.

From the t-stats in model S2, we learned that t-stat of b0 is 188.04120, t-stat of b1 is 50.69. Both indicators were at significance of 1%. Therefore, education level is significant at 1% of significance level in the correlation with household disposable income.

Multiple Linear Regression

Model M1

Log (Household Disposable Income)= 7.9764237+ 0.062239374 education + 0.0011597210 low pay W +u (other factor)

constant:	beta 0 =	7.9764237	standard error =	0.23787932	t-stat =	33.531388
educ:	beta1 =	0.062239374	standard error =	0.0019237026	t-stat =	32.353948
low pay:	beta2 =	0.0011597210	standard error =	0.011855880	t-stat =	0.097818210

Sample size = 36.000000
R2 = 0.98694700
F = 4.4639837
p-value = 0.019238321
LM = 7.6656906
p-value = 0.021647933

We sought to find explanatory variables to better fit of the regression equation as well as increase R-square. Then multiple regression model is formed with the second explanatory variable, percentage of low pay wage in total wage level. B2 =0.0011597210, meaning as percentage of low pay wage increased by 1 %, then household disposable income increased by 0.0011597210 USD/capita. And t stat for B2 is 0.097818210, which is very small. For the F test, P value is 0.01923 which is smaller than 0.05 common significance level.

Model M2:

Log (Household Disposable Income) = c+ B1 education +B2 low pay W+ B3 Foreign +u(other factor)

constant: beta 0 = 8.0181569 standard error = 0.24367912 t-stat = 32.904571
educ: beta1 = 0.062724588 standard error = 0.0020117316 t-stat = 31.179401
low pay: beta2 = -0.0013315732 standard error = 0.012248816 t-stat = -0.10871036
foreign: beta3 = 6.8836693e-06 standard error = 7.9910968e-06 t-stat = 0.86141733

Sample size = 36.000000
R2 = 0.98724282
F = 3.0622510
p-value = 0.042090031
LM = 8.0298418
p-value = 0.045398976

In Model M2, we included another explanatory variable, foreign population percentage in total population. We found R-squared value increased from 0.98694700 in Model S2 to 0.98724282. And in the model M2, B2 coefficient turned to negative while B1 and B3 coefficient are positive. All the beta coefficients make sense in the equation. As education level,

foreign population increased, and percentage of low pay wage decreased, then household disposable income rose with those changes. In a more specific word, as adult tertiary education level increased by 1%, household disposable income increased by 0.062724588 USD/capita; as percentage of low pay wage decreased by 1%, household disposable income rose by 0.0013315732 USD/capita; as percent of foreign population increased by 1%, household disposable income increased by 6.8836693e-06 USD/capita.

Model M3:

Log (Household Disposable Income)= c+ B1 education +B2 low pay W+ B3 foreign +B4 female + u

constant:	beta 0 =	8.0313031	standard error =	0.66961075	t-stat =	11.993988
educ:	beta1 =	0.063043593	standard error =	0.015235101	t-stat =	4.1380488
low pay:	beta2 =	-0.0014350316	standard error =	0.013373288	t-stat =	-0.10730581
foreign:	beta3 =	6.8796696e-06	standard error =	8.1211102e-06	t-stat =	0.84713413
female:	beta4 =	-0.00015370580	standard error =	0.0072743296	t-stat =	-0.021129893

Sample size = 36.000000
R2 = 0.98724300
F = 2.8042198
p-value = 0.042677284
LM = 9.5650757
p-value = 0.048427125

In Model 3, we added another explanatory variable in succession. We added percentage of female in total population as we believed there were connection between household disposable income and female population. Then we found B4 is -0.00015370580 which meant female was negatively correlated with household disposable income. As percentage of female population decreased by 1%, then household disposable income increased by 0.00015370580 USD/capita. And we also tried the F test comparing with result of S1, we found P value is 0.042677284 which is at 5% of common significance level. Therefore, we can safely say that these variables are jointly significant.

Model M5:

Log (Household Disposable Income) = c+ B1 education +B2 low pay W+ B3 foreign +B4 female+B5 CPI+u

constant: beta 0 = 7.8760588 standard error = 0.62766152 t-stat = 12.548258
educ: beta1 = 0.052127892 standard error = 0.014925176 t-stat = 3.4926150
low pay: beta2 = -0.0029691695 standard error = 0.012484291 t-stat = -0.23783244
foreign: beta3 = 6.3479567e-06 standard error = 7.5744412e-06 t-stat = 0.83807591
female: beta4 = 0.0041391936 standard error = 0.0070173690 t-stat = 0.58984979
CPI: beta5 = -0.011331715 standard error = 0.0047600459 t-stat = -2.3805895

Sample size = 36.000000
R2 = 0.98926998
F = 1.0623692
p-value = 0.40054119
LM = 5.4153628
p-value = 0.36731593

In the Model M5, we added the CPI as another additional explanatory variable. We learned the B5 is -0.011331715. It meant that as CPI decreased by 1%, the household disposable income increased 0.011331715 USD/capita. And t-stats we found for B5 is -2.3805895. It was the largest t stat of b2,b3,b4, and b5. It probably meant CPI was more statistically significant than other indicators, but less significant than education on household disposable income. However, the F test, p value is 0.40054119 which is larger than 5% common significance level.

Time trend effect and detrending

Regression model: LOG (Household Disposable Income) = c + B1 education + B2 time+ u (other factor)

In this model, we discuss the influence of time trend to household disposable income. We add a time trend as a variable in our regression and define year 1980 as the base year (starting

from 1). Below is the result of the regression.

```
Without Time Trends

Constant:           7.9993062     (S.E.):       0.042540178
t-statistics:         188.04120
Education:        0.062382502     (S.E.):      0.0012305392
t-statistics:         50.695258
Sample Size:        36.000000
R2:                 0.98694321
Adjusted R2:        0.98655919

With Time Trends

Constant:           8.3905457     (S.E.):        0.25112052
t-statistics:         33.412425
Education:        0.043752028     (S.E.):       0.011853978
t-statistics:         3.6909151
Time Trends:      0.012910829     (S.E.):      0.0081722437
t-statistics:         1.5798390
Sample Size:        36.000000
R2:                 0.98786130
Adjusted R2:        0.98712562
```

Analyzing our first model regressing only on education we can find out that the elasticity of household disposable income with respect to education level is very large and statistically significant. With a t-statistics about 60, we are confident enough to reject the null hypothesis "H_0: $B_1 = 0$" with a degree of freedom of 36.

And looking through data there is an upward tendency for both household disposable income and education level. It is apparently that with the development of the society, more and more people would have the chance to receive education. In order to account for the trending behavior of education, we add a time trend in our regression. Comparing two regressions, it is surprising that an apparently numerical change in the coefficient of education does not exist (B_1 = 0.06238 and B_1`= 0.04375). However, when we compute the t-statistics for the new coefficient of education, it decreases significantly. Although we can still safely reject the null that "H_0: B_1 = 0" with a degree of freedom of 36 with a t-statistics about 4 with any significance level, we

should not ignore the decline in t-statistics when doing regression. Considering the r square and adjusted r square for both regression model, there is not significant change between them. Though there is not much change for the coefficient of education, when we compute the t-statistics for time trends, the t-statistics for it is about 1.5798, so we can say that at a significance level of 20%, the null hypothesis that "B_2= 0" at a degree of freedom of 36. In such case, the effect of time trend cannot be neglected. Therefore, we continue our research on detrending the regression model to remove the trend from our time series. Result is presented as below.

```
Detrending Time Series

Constant:       -7.9865393e-16     (S.E.):       0.0084891362
Education:      -2.9873483e-06     (S.E.):      7.9475042e-07
Sample Size:         36.000000
R2:                  0.29356463
Adjusted R2:         0.27278712
```

Now, a 1% increase in percentage of workers who receive postsecondary education is estimated to decrease household disposable income in a numerically small percentage,
and the estimate is numerically and statistically different from one. The adjusted R‑squared shows that the growth in percentage of higher-educated worker explains about 27.28% of the growth in household disposable income.

Conclusion

To conclude, according to research, and as stated in the literature review, US's adult education level affects household disposable income up until a certain threshold. After adding other variables, the relation among the variables changes. Our analysis backs the claim. As we can see from our different regression model, it is very hard to conclude how exactly these variables affect the household disposable income. It is reasonable since even though we have already had many information from the data, it is possible that we still have something unknown,

like some other variables that can affect household disposable income, or some intangible factors.

However, our regression is still instructive. We know that household disposable income has positive relation with education. Also, as the XXXX gets higher, the XXXX decreases. Our model supports the claim that education level is a significant factor to affect the household disposable income.

Restriction and further study

The data we used in our regression is an annual statistic from 1980 to 2015 of United States, a pretty small sample size in analysis, which means there would be a relatively large variance among our regression. And also, results derived from regression using the data of United States may not be applied to other countries, which indicates a direction for future study. To conclude a more general relation between education and household income, a dataset collected worldwide could be collected for deepened research.

Also, in our model with time series, detrending the data results a numerically significant change in the coefficient of education. Before detrending our regression, population education level is positively related with household disposable income, but later, after detrending, an increase in population education level would lead to a decline in household disposable income, which contrast to previous study and our original hypothesis. Thus, more data can be collected to test the correlation between education and disposable income, and also, why detrend flip the effect of education on disposable income.

Reference

Carnevale, A. P.(2008). College For all?. *Change*, 40(1), 22-31.

Purcell, P. J., Iams, H. M., & Shoffner, D. (2015). EDUCATION, EARNINGS INEQUALITY, AND FUTURE SOCIAL SECURITY BENEFITS: A MICROSIMULATION ANALYSIS. *Social Security Bulletin*, 75(3), 15-33.

Murphy, E., & Oesch, D. (2016). The Feminization of Occupations and Change in Wages: A panel Analysis of Britain, Germany and Switzerland. *Social Forces*, 94(3), 1221-1255.

CAMAROTA, S. A. (1998). Does immigration harm the poor?. *Public Interest*, (133), 23-32