

An Empirical Analysis on the Impact of Higher Education on Income Inequality

Xinhua Hui¹

Abstract

The study on the relationship between the higher education and income inequality is of great importance to exploring ways to reduce income inequality. With the macro-level time-series data of the United States from 1967 to 2015, this paper empirically tested the relationship between higher education and income inequality. The result indicated that there is a significant inverted-U relationship between higher education and income inequality, that is, when the higher education is not widely available, the bonus of higher education is significant, which can aggravate income inequality. When the higher education is widely available, the education expansion will narrow the income gap. At the same time, the model also verified the impact of such variables as financialization, trade union density, trade dependence, the proportion of female labor participation, and business cycle fluctuations on the evolution of income inequality in the United States. Hopefully, the result of this research can offer some helpful references for developing countries to narrow their income gap by educational expansion.

Keywords: Higher education, income inequality, inverted-U relationship, macro-level time-series data of the United States.

¹ School of Social Science, Tsinghua University.

1. Introduction

The development of education is the main driving force and influential factor for sustained economic growth and improvement of productivity. It is also one of the main ways for the middle and lower classes to enter the upper classes. Educational investment is the most direct and major human capital investment, which will affect economic growth and changes in the pattern of income distribution. Therefore, educational expansion is one of the important factors affecting income distribution. The educational attainment of American citizens is affected by many factors, among which gender, race, place of birth (native or nonnative), physical condition (disability or not) and family financial status are more influential ones. At present, the difference in educational attainment between different genders has been almost eliminated in the United States, but the income gap between genders still exists. Race, as a relatively unique phenomenon in the United States, is related to historical issues. Even though after long-term struggles and reforms, the educational attainment of the people of all races in the United States is increasing, the differences between races are still remarkable. The United States is a nation of immigrants and therefore there are a large number of foreign-born citizens, and native-born citizens are significantly better educated than foreign-born citizens. Physical disability can also significantly affect educational attainment. Besides, the family financial status and the educational attainment are negatively correlated. Education has been making great contributions to the rapid development of the United States, by cultivating a large number of talents. However, differences in the educational attainment of citizens are also one of the main factors that widen the gap in income distribution in the United States. Therefore, the study on the impact of educational development in the United States on income inequality has always been a hot topic in both the academic community and the society. This paper intends to use the U.S. macro-level time-series data to verify the dynamic relationship between higher education and income inequality in the United States, thus providing a helpful supplement to relevant research.

2. Literature Review

The results of the existing literature on the impact of educational expansion on the income gap can be roughly divided into four categories.

First, some scholars believe that educational expansion may widen the gap in income distribution. For example, Bhagwati (1973) believed that the educational expansion will increase the income gap, since it will allow the low-income groups with higher educational attainment to get better-paid jobs than those with lower educational attainment, especially in the countries with low economic development levels. Sylwester (2000) pointed out that higher education means higher income in the future, so the cost is higher. Therefore, opportunities for higher education are more likely to be obtained by people with higher income, while the poor can't afford higher education and thus can't get out of the poverty trap. The Matthew Effect can make the income gap wider and wider.

Second, some others think just the opposite: they argue that educational expansion will narrow the gap in income distribution. Ahluwalia (1976) pointed out that, according to the Human Capital Theory, in the case of increased supply of skilled labor and high marginal productivity of labor, it is possible to improve the productivity of low-income population by providing more education opportunities for them and improving their educational level, and, consequently, increase the income of low-income population, bridging their income gap with high-income population. Psacharopoulos (1982) believed that with the continuous improvement of the educational level of the female, women can may get better-paid jobs, which will be able to narrow the income gap caused by gender.

Third, some scholars believe that the impact of educational expansion on income distribution is uncertain. Mincer (1974) pointed out whether increasing the average years of schooling for citizens can narrow the income gap or not depends on changes in the rate of return to education: the result may be positive or negative. Alesina and Perotti (1996) pointed out that whether the educational expansion will reduce the income inequality or not depends on the relationship between the cost of education and the value of per capita income. When the cost of education exceeds the per capita income, the rich can afford higher education, while the poor can't. So the stock of human capital of the rich increases and their future income also increases, thus widening the income gap between the rich and the poor. However, when the education cost is lower than the per capita income, the poor can afford the education expenses like the rich, so that the income gap with the rich can be reduced. Gregorio and Lee(2002) pointed out that the inequality in education will worsen the inequality in income, but the effect of improving education on income distribution is uncertain in the case that the education distribution remains unchanged.

Last, some scholars believe that the relationship between the educational expansion and the income distribution gap is in line with the inverted-U curve: in the early stage of education expansion, the gap in income distribution tend to increase, while after reaching a certain inflection point, the income gap began to narrow with educational expansion. The inverted-U relationship between education and income inequality was first proposed by Londono (1990) and Ram(1990). Basing on cross-sectional data from more than 90 countries, Ram empirically showed that the inflection point of the inverted-U curve was about 7 years of schooling on average. Thomas (2002) et al. extended the data to 140 countries for empirical analysis and verified Ram's conclusion of "7-year inflection".

In addition, other views exist that public and private educations should be researched separately. For example, Eckstein and Zilcha (1991) proposed that the lower limit of fund provided by the government should be set to support compulsory education, which can help to narrow the income gap. Dablanorris et al.(2004) believed that increasing the budget for public education requires the government to be the strong backup force to reduce the income distribution gap. The model analysis of Fernandez and Rogerson (1995) showed that the public education expenditure affects the opportunities for the poor to receive education.

3. Model Specification and Data Presentations

With the macro time series data of the United States from 1967 to 2015, this paper empirically tested the non-linear relationship between the higher education and income inequality in the United States. This paper collected a relatively comprehensive data on control variables affecting the income inequality from multiple databases, which can better separate and verify the impact of factors other than education on income distribution.

3.1 Model Specification

Based on the existing literature, the following regression model is established:

$$ineq = \beta_0 + \beta_1 edu + \beta_2 edu^2 + \beta_3 control + \varepsilon$$

The explained variable “ineq” is the Gini coefficient, which represents income inequality, and the explanatory variable “edu” represents the higher educational attainment in the United States. “Control” represents other control variables that have an impact on income inequality other than educational factors; if β_2 is significant, it confirms the non-linear relationship between the higher education level and income inequality in the United States. In particular when $\beta_1 > 0, \beta_2 < 0$, it shows that there is an inverted-U relationship between the higher educational attainment and income inequality in the United States.

3.2 Variable Selection Description and Data Sources

The explained variable “Gini” represents the degree of income inequality. Gini coefficient was used to measure in this paper and the Data comes from the Current Population Survey published on website of the United States Census Bureau.

The explanatory variable “edu.” refers to the proportion of people aged 25 and over who have a university degree or above. Data Sources: the website of United States Census Bureau.

Control variables refers to other factors that have an impact on income inequality in the US economic development. The following 7 variables are selected in this paper. The natural logarithm of gross domestic product per capita is “LnGDPpercapita”. Gross domestic product per capita was selected as an indicator of economic growth to control the impact of economic growth on income inequality, while eliminating heteroscedasticity by taking natural logarithms. Data sources: the website of the World Bank .

Business Cycle is referred to as “inverseu”. This paper uses the reciprocal of unemployment rate lagging two periods^{1/U_{t-2}}. Higher unemployment and a more severe economic recession might lead to an increase in income inequality, which is the comprehensive result of the direct impact of loss of income due to unemployment and the indirect impact of falling income due to the economic

recession. However, this effect was not caused by labor income alone. The economic downturn would also reduce capital utilization and reduce capital income. Therefore, it is impossible to directly judge the final change in income inequality. In the early stages of the cyclical recovery after the economic recession, income inequality will increase due to the coexistence of rapid recovery of profits and the stagnation of wages. The reciprocal of the unemployment rate is generally used to measure the role of the business cycle, and empirical experience indicates that the unemployment rate in the two periods better showed the deviation of profit and labor income after the economic recession. The data on unemployment rate comes from the website of the U.S. Bureau of Labor Statistics.

Trade Union Density is referred to as “*Union*”: “*Union*” = the number of union members (non-agricultural) / total number of workers. Trade union organizations in the United States play a pivotal role in wage negotiations. The greater the density of trade unions, the stronger the bargaining power of workers and the more favorable to the increase in workers’ income. Therefore, there is a positive correlation between trade union density and workers’ labor income. Beginning in the late 1960s, the density of trade unions in the United States began to decrease severely. This phenomenon particularly affected industries dominated by collective bargaining negotiations (Fichtenbaum’s (2011), resulting in the decrease of the workers’ wages, and thus increasing income inequality. The Data comes from the website of trade union membership and coverage database.

Foreign trade dependence is referred to as “*trade*”. “*Trade*”= total net export/GDP. The Stolper-Samuelson Theorem states that international trade affects the relative price of factors, increases the price of sufficient factors in the country, and lowers the price of scarce factors in the country. The United States has relatively abundant technological and capital factors. International trade will increase the income of elites with more capital and highly skilled workers, while decreasing the income of unskilled workers. Therefore, international trade will increase income inequality. Data comes from the the website of Bureau of Economic Analysis of the United States.

Import share is referred to as “*importshare*”. “*importshare*” = total import / GDP. In the 1970s, the United States is transformed from a net exporter to a net importer. Many export industries with higher wage levels saw a decline in its business, while low-cost imports increase the competition between cheap foreign labor and domestic labor, resulting in lower wages for American workers. Therefore, the import share will increase income inequality. Although the import share and foreign trade dependence affects the income inequality differently, there might be strong collinearity between the two. Therefore, in the regression analysis, these two variables can be used for verification respectively. Data comes from the website of Bureau of Economic Analysis of the United States.

The proportion of female labor participation is referred to as “*femaleLF*”. In the 1960s and 1970s, the proportion of female labor participation increased significantly in the United States, but then the income inequality increased significantly in the 1980s. Therefore, there may be certain relationship between the

proportion of female labor participation and income inequality. One of the main reasons may be that the increase in the female labor participation results mainly from families with higher income. Therefore, the increase in female labor participation will further increase the income of families with high income, which will in turn increase income inequality. The data is from the website of US Bureau of Labor Statistics. Financialization is referred to as as “*fir*”. It mainly reflects the improvement of capital allocation efficiency in capital market by financial development, but unfortunately, the official measurement indicators failed to be found. Therefore, based on the research results of the existing literature, this paper defines two definitions of “*fir*”: I. Financial related ratio, i.e. “*fir1*” = Financial related total assets (finance, insurance, real estate, leasing, etc.)/GDP, and the data comes from the website of US Federal Reserve System; II. Output value ratio of financial related industry, i.e. “*fir2*” = total output value of financial related industry (finance, insurance, real estate, leasing, etc.)/total output value of all industries. The data comes from the website of US Federal Reserve System.

The statistical characteristics of each variable are shown in Table 1.

Table 1: Statistical Characteristics of Major Variables

Names of Variables	Abbreviation	Number of Variables	Mean Value	Standard Deviation (SD)	Median Value	Minimum Value	Maximum Value
Gini Coefficient	GINI	49	43.55	3.09	43.1	38.8	48.2
Financialization 1	FIR1	69	0.16	0.03	0.16	0.1	0.2
Financialization 2	FIR2	69	0.13	0.03	0.11	0.08	0.18
Educational Level	EDU	58	19.33	7.84	19.65	5.4	32.5
Foreign Trade Dependence	TRADE	56	-0.02	0.02	-0.01	-0.05	0.01
Trade Union Density	NONMEM	42	12.49	5.57	10.65	6.6	24.6
The proportion of Female Labor Participation	FEMALE	41	0.53	0.04	0.54	0.42	0.58
Business Cycle	INVERSEU	67	0.19	0.05	0.18	0.1	0.34
Natural Logarithm of Per Capita GDP	LNGDP	56	9.71	0.94	9.94	8.01	10.93
Import	IMPORT	24	0.14	0.02	0.14	0.10	0.17

Based on the above variables, the following multivariate regression model can be established.

$$\begin{aligned} Ineq = & \beta_0 + \beta_1 edu + \beta_2 edu^2 + \beta_3 LnGDPpercapita + \beta_4 (1/U_{t-2}) \\ & + \beta_5 Union + \beta_6 trade + \beta_7 femaleLF + \beta_8 fir + \varepsilon \end{aligned}$$

4. Results of Empirical Analysis

4.1 Multicollinearity Test

The Variance Inflation Factor method (VIF) was used to perform a multicollinearity test on the explanatory variables, and the test showed that there was a significant multicollinearity between “*trade*” and “*LnGDPpercapita*”, a result which was similar to some of the previous literature studies. Because “*trade*” contains the influence of “*LnGDPpercapita*”, *LnGDPpercapita* was removed in the regression.

4.2 Unit Root Test

Macro-level time-series data have obvious time trend, possibly causing false regression results, so the stationarity of the data should be checked before regression. ADF test showed that the explained variables, explanatory variables, and most of the control variables (except for “*union*” and “*inverseu*”) had unit roots, which were non-stationary time series. The ADF test showed that the series after the difference was stationary, indicating that the original series is $I(1)$. The test results are shown in Table 2.

Table 2: Unit Root Test for Related Variables (DF Test)

	Indicator	Variable	DF Statistics	Critical Value Significant level at 5%	Stationarity
Original Series	Inequality	Gini	-0.204	-0.292	Non-stationary
	Financialization	fir1	-1.482	-2.905	Non-stationary
		fir1 ²	-0.823	-2.905	Non-stationary
	Control Variable	Union	-3.607	-2.937	Stationary
		Trade	-1.454	-2.916	Non-stationary
		femaleLF	-2.555	-2.939	Non-stationary
		edu	0.011	-2.920	Non-stationary
		Inverseu	-3.598	-2.907	Stationary
Series after First Order Difference	Inequality	DGini	-6.561	-2.927	Stationary
	Financialization	Dfir1	-9.068	-2.906	Stationary
		Dfir1 ²	-9.386	-2.906	Stationary
	Control Variable	DUnion	-5.511	-2.941	Stationary
		DTrade	-6.423	-2.917	Stationary
		DfemaleLF	-2.935	-2.939	Stationary
		Dedu	-6.324	-2.921	Stationary
		DInverseu	-7.605	-2.907	Stationary

In order to prevent the inaccuracy of the results brought by the single test method, the paper also used the PP test to test the stationarity of the data. The test results are consistent with the DF test, and will not be repeated here.

4.3 Co-integration Test

Since the original series was a non-stationary time series, the co-integration relationship between the variables should be tested. The results of Johansen test are shown in Table 3 below. The results are significant at 5%, the null hypothesis that “co-integration rank is 0” can be rejected, that is, there does exist a co-integration relationship.

Table 3: Co-integration Test Results

Trend: trend Sample:1978-2015		Johansen tests	Number of obs=34 Lags=2	
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5% Critical Value	Prob.**
None *	1.00	498.19	159.53	0.00
At most 1 *	0.88	263.42	125.62	0.00
At most 2 *	0.83	192.38	95.75	0.00
At most 3 *	0.79	132.47	69.82	0.00
At most 4 *	0.64	78.78	47.86	0.00
At most 5 *	0.53	43.70	29.80	0.00
At most 6 *	0.41	18.06	15.49	0.02
At most 7 *	0.00	0.03	3.84	0.86
Hypothesized No. of CE(s)	Eigenvalue	Max-eigen Statistic	5% Critical Value	Prob.**
None *	1.00	234.77	52.36	0.00
At most 1 *	0.88	71.04	46.23	0.00
At most 2 *	0.83	59.92	40.08	0.00
At most 3 *	0.79	53.69	33.88	0.00
At most 4 *	0.64	35.08	27.58	0.00
At most 5 *	0.53	25.65	21.13	0.01
At most 6 *	0.41	18.03	14.26	0.01
At most 7 *	0.00	0.03	3.84	0.86

Because of the co-integration relationship between variables, co-integration regression was used to test the long-term relationship between income inequality and educational level. This paper didn't use the first-order difference series of each variable for regression, because the use of the difference model can ensure the stationarity of the data, the economic significance of the regression model is very different.

4.4 Regression Analysis

The results of the co-integration regression are shown in Table 4, and the results showed that the variables had a long-term equilibrium relationship. In the regression analysis, the model containing only the explanatory variables was firstly regressed, and then the influence of the control variables on the explained variables was tested by adding control variables step by step

Table 4: Co-integration Regression Results

Indicator	Explanatory Variable	Explained Variable Gini			
		Model 1	Model 2	Model 3	Model 4
Educational Level	edu	0.036***	0.014***	0.014***	0.005***
	edu ²	-0.001***	-2.00E-04**	-1.78E-04***	-8.85E-06*
Control Variable	Fir		0.970***	0.576	0.874***
	Union Trade		0.004***	0.003***	0.003***
	Female			0.318***	0.295***
	LF Inverseu			0.136***	0.294***
Sample Size		49	40	38	38
R ²		0.258	0.933	0.940	0.949
Note: *, **, *** indicate significance at significant levels of 10%, 5%, and 1% respectively.					

The regression results of model 1 showed that the primary regression coefficient of the variable “*edu*” was positive, and the quadratic regression coefficient was negative, and both were significant at the 1% significance level, indicating that there was a significant inverted-U relationship between income inequality and educational attainment. That is to say, in the early stage of education expansion, the income distribution gap was widened, while after reaching a certain inflection point, the income gap began to narrow with the education expansion. The empirical results were consistent with the actual economic situation. At the lower level of education, the smaller groups receiving higher education can obtain better-paid jobs in the employment market. Because of the low mobility between different types of work, the bonus of higher education is significant, which can increase income inequality. At a higher level of education, however, a large proportion of people have access to higher education, so the participants in the highly competitive job market were roughly equal in their ability. Therefore, the bonus of higher education was no longer remarkable, and the education expansion narrowed the income gap in this stage.

The control variables “*fir*” and “*Union*” were added to the model 2. The regression results showed that the effect of “*edu*” stay the same after adding the control variables: they only reduced the coefficient to some extent. The impact of financialization and trade union density on income inequality was positive, and consistent with the relevant literature conclusions.

The control variables “*femaleLF*” and “*inverseU*” were added to the model 2. The regression results showed that the effect of “*edu*” stay the same after adding the control variables: they only reduced the coefficient to some extent. Among them, the regression coefficient of trade was significantly positive, indicating that the development of international trade can increase income inequality. International trade has increased the price of the relatively abundant capital and technology in the United States, which increased the income of high-income population who had the advantages in these two factors, thus increasing income inequality. “*Trade*” was

replaced by the variable “*importshare*” for regression, and the sign and significance of the regression coefficients remained basically unchanged. The coefficient of variable “*femaleLF*” was significantly positive. With the development of economy, more American women entered the labor market, which happened more in high-income families, so the proportion of female labor participation will increase income inequality. The regression coefficient of “*inverseU*” was significantly negative, indicating that the income inequality would decline with the business cycle fluctuations, mainly because in the economic depression, both the labor income and the capital income would decline due to the decrease of capital utilization.

The regression coefficient indicated that the control variables did exert an impact on income inequality. In terms of coefficient, “*trade*” and “*femaleLF*” had a greater effect on income inequality than “*edu*” and trade union density.

4.5 Error Correction Model and Its Test

Co-integration regression results showed long-term equilibrium relationship among income inequality, explanatory variables and control variables (inverted-U relationship). In order to test the short-term equilibrium relationship between variables, an error correction model should be used.

The regression results of the error correction model showed an adjustment factor of 0.053, which was significant at the 10% significance level. That is to say, when the short-term effects of the education level, financialization, and trade union density on income inequality deviate from their long-term equilibrium, they will return to the long-term equilibrium state with an adjustment of 0.053.

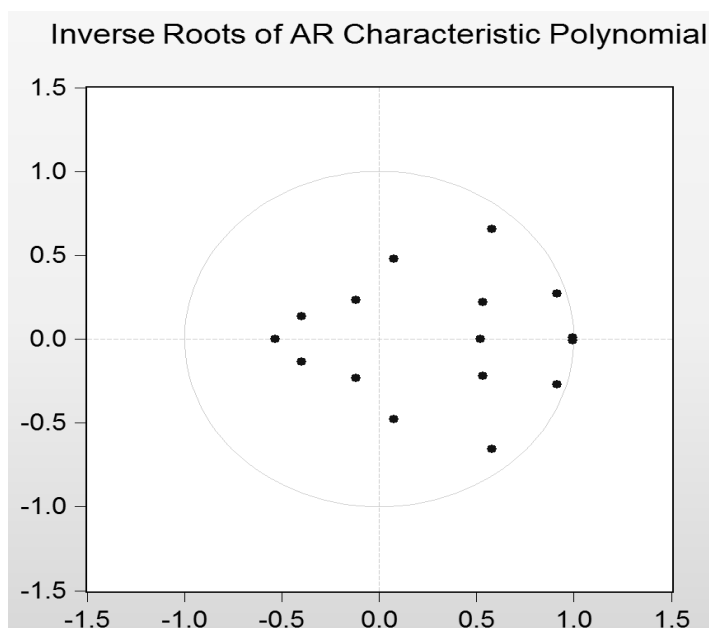


Figure 1: VECM System Stability Determination

The reliability of the regression model was verified by the VECM system stability test. The results are shown in Figure 1. Not only the hypothetical unit root of the model was inside the unit circle, but all the eigenvalues of the adjoint matrix fell within the unit circle, indicating that the system was stable.

5. Conclusion

The development and reform of higher education play an important role in economic growth, income distribution and social stability. Therefore, the analysis of their influence modes and relations has important theoretical and practical significance. With the macro time series data of the United States from 1967 to 2015, this paper tested the relationship between income inequality and the higher education, showing that there is a significant inverted “U” model relationship between the two. That is to say, when the higher education is not widely available, the bonus of higher education is tremendous, which can increase income inequality. When the higher education is widely available, a large proportion of people can have access to higher education, the participants in the highly competitive job market are roughly equal in their ability. Therefore, the education expansion will narrow the income gap in this stage. This conclusion is helpful for developing countries. Higher education reform is a focus issue in the development of developing countries. The income inequality of residents can be narrowed by having more people receive higher education.

Moreover, this paper also verified the positive impact of variables such as financialization, trade union density, trade dependence and the proportion of female labor participation, and negative impact of business cycle fluctuations on the evolution of income inequality in the United States. There is still room for further analysis of the relationship between education inequality and income inequality in the future.

References

- [1] A. Alesina and R. Perotti, Income Distribution, Political Instability, and Investment, *European economic review*, 40(6), (1996), 1203-1228.
- [2] E. Dablanorris and M. Gradstein, *The Distributional Bias of Public Education: Causes and Consequences*, International Monetary Fund, 2004.
- [3] G. Psacharopoulos, Earnings and education in Greece, 1960–1977, *European Economic Review*, 17(3), (1982), 333 - 347.
- [4] J. A. Mincer, *Schooling and Earnings*, NBER Chapters, (1974), 41- 63.
- [5] J. D. Gregorio and J. W. Lee, Education and Income Inequality: New Evidence From Cross-Country Data, *Review of Income and Wealth*, 48(3), (2002), 395 - 416.
- [6] J.N.Bhagwati, Education, Class Structure, and Income Equality, *World Development*, 1(5), (1973), 21 - 36.
- [7] K. Sylwester, Income Inequality, Education Expenditures, and Growth, *Journal of Development Economics*, 63(2), (2000), 379 - 398.

- [8] M. S. Ahluwalia, Inequality, Poverty and Development, *Journal of Development Economics*, 3(4), (1976), 307 - 342.
- [9] R. Fernandez and R. Rogerson, On the Political Economy of Education Subsidies. *Review of Economic Studies*, 62(2), 1995, 249 - 262.
- [10] R. Ram, Educational Expansion and Schooling Inequality: International Evidence and Some Implications, *Review of Economics & Statistics*, 72(2), 1990, 266 - 274.
- [11] V. Thomas, W. Yan and X. Fan, A New Dataset on Inequality in Education: Gini and Theil Indices of Schooling for 140 Countries 1960-2000, World Bank, 2002.
- [12] Z.Eckstein and I. Zilcha, The Effects of Compulsory Schooling On Growth, Income Distribution and Welfare. *Economic Quarterly*, 40(2)1993, 243 - 251.