Undergraduate Economic Review

Volume 16 | Issue 1 Article 4

2019

The Relationship Between College Expansion and Income Inequality

Aidan J. Wang
University of California, Berkeley, ajwang@berkeley.edu

The Relationship Between College Expansion and Income Inequality

Abstract

This paper examines the relationship between college expansion and income inequality within a country. Researchers have identified a “composition effect,” “compression effect,” and “dispersion effect.” However, the shape and magnitude of the net relationship remains unclear. I construct a country panel using inequality data from the World Inequality Database and college share data from Barro and Lee. From 0% to 27% college share, the bottom 50% and middle 40% income shares decrease linearly while the top 10% income share increases linearly. The trend shape holds for a sample of only OECD countries, but the magnitude changes, suggesting country-specific factors matter.

Keywords

college, education, income inequality, income distribution

Cover Page Footnote

I am grateful to Emmanuel Saez, Barry Eichengreen, Matthew Tauzer, Junyi Hou, and Troup Howard for helpful guidance and comments.

Income inequality has increased in almost all countries in the past few decades (Alvaredo et al. 2018). Yet, the level of income inequality varies widely between countries and has been changing at different speeds within countries, suggesting that differences in economies have an effect on income inequality. One factor that has long been suggested to matter for income inequality is education, especially college education, which has even been dubbed "the great equalizer." College attainment has also been expanding within countries in the past few decades, yet also varies greatly between countries (Barro and Lee 2013). Various scholars have studied the mechanisms for how college expansion affects income inequality by looking at individual country series. Knight and Sabot (1983), using data from Tanzania and Kenya, identify a "composition effect" resulting from the shift in population from low-skilled to high-skilled labor and a "compression effect" of educated workers' wages being depressed by an increase in relative labor supply. Lemieux (2006) identifies a "dispersion effect" owing to differences in the income distributions of college graduates and non-graduates in the United States.

The question remains: what is the net relationship between college expansion and income inequality within an economy? Income inequality may monotonically increase or decrease with growing college education, or it may increase then decrease, or decrease then increase, as college expands. Having a better understanding of this pattern would help elucidate the potential effects of current trends of expanding college education on income inequality within countries.

To analyze the relationship between college attainment and income inequality among different countries, I construct a panel of countries using bottom 50%, middle 40%, and top 10% income share data from the World Inequality Database (WID) (see Alvaredo et al. 2018) and college share data from Barro and Lee (2013). I run OLS regressions with a quadratic college share term to capture potential nonlinear trends and include country and year fixed effects to analyze variation within each country and control for technological change affecting income inequality. I find that for the range of college shares in the sample, from about 0% to 27%, the bottom 50% and middle 40% income shares decrease linearly while the top 10% income share increases linearly. The shape of this trend is robust when observing the top 10% income share for an enlarged sample of only OECD countries, but the magnitude of the effect changes, suggesting that the strength of the trend is country- or region-specific. Nevertheless, I estimate that a percentage-point increase in college share corresponds to a 0.482 percentage-point decrease in income share for the bottom 50%, a 0.382 percentage-point decrease for the middle 40%, and a 0.864 percentage-point increase for the top 10%, which may be interpreted as average magnitudes of the trend for the countries in the sample. However, these coefficients may be overestimated owing to potentially uncontrolled changes in the college wage premium within countries.

The paper proceeds as follows. Section 1 describes the literature on income inequality and college expansion and explains the mechanisms behind the relationship between them. Section 2 explains the OLS regression methodology used and describes the panel data. In section 3 I discuss my analysis of the overall trend between income inequality and college expansion observed in the data. Section 4 presents conclusions.

I. Literature Review

College education can be roughly thought of as dividing the labor force into two sectors: high-skilled labor (those with college degrees) and low-skilled labor (those without college degrees). Kuznets' (1955) hypothesis of income inequality in a two-sector economy claims that the overall income distribution depends on three factors: the proportions of the populations of the two sectors, their relative average incomes, and the income distributions of each sector. Kuznets famously concludes that under certain circumstances of these parameters, and assuming that relative average incomes and income distributions of each sector are constant, there is an inverted "U" curve of income inequality first increasing then decreasing as the economy's population shifts from one sector to the other. While Kuznets originally thought about agricultural and nonagricultural sectors, Knight and Sabot (1983) recognize that Kuznets' hypothesis can be applied to educational expansion and that these three factors determine how the income distribution changes. I will call the effects of these three factors the "composition effect," the "compression effect" (following Knight and Sabot 1983), and the "dispersion effect."

The "composition effect" is a purely structural one. As more people obtain a college degree and move from the low-skilled labor sector to the high-skilled labor sector, which has a relatively higher average income, more people will earn higher incomes. When the relative proportion of high-skilled labor is small and most people do not have college degrees, the expansion of college education increases inequality as a small but growing proportion of the population earns higher incomes than the rest. On the other hand, when the relative proportion of high-skilled labor is large and most people have college degrees, the increase in people with college degrees decreases inequality as the relatively few people without college degrees catch up to the majority of the population. Somewhere in the middle, a turning point occurs where an increasing college share switches from increasing to decreasing income inequality. This is the intuition behind Kuznets' (1955) original hypothesis of an inverted "U" curve of income inequality, while he assumed that the "compression effect" and "dispersion effect" were nonexistent. Extending this intuition to income shares, the measure I use to examine the income distribution, as the proportion of people with college degrees increases, the top 10% of incomes would increase until this income group "fills up" with college graduates and additional college graduates become part of the middle 40% income group, at which point the middle 40% income share would start to grow. As the college share increases further, college graduates would spill over to the bottom 50% income group, at which point that income share would start to grow. Thus, in a world where college graduate incomes are strictly greater than non-graduate incomes, one would expect a turning point for the top 10% and middle 40% income shares at a 10% college share and for the bottom 50% income share at a 50% college share. Knight and Sabot (1983) estimate the "composition effect" of educational expansion by estimating an earnings function for wage employees in Tanzania in 1971 and 1980 and in Kenya in 1980. Holding the coefficients on education constant, they estimate that a simulated expansion of education increases inequality in all but one of six pairwise comparisons. Machado and Mata (2005) decompose changes in the wage distribution in Portugal from 1986 to 1995 by estimating marginal wage distributions obtained through quantile regressions and find that changes in educational attainment, counterfactually holding returns to education constant, contributed to an increase in wage inequality.

The "compression effect" has garnered special attention in the literature. The basic premise of the "compression effect" is that the relative average income of college graduates versus that of non-graduates may not be constant, as Kuznets originally assumed of his two sectors. A decrease in the relative average income of college graduates versus that of non-graduates clearly decreases overall income inequality even if the college share stays constant. Knight and Sabot (1983) find that changes to coefficients on education of their estimated earnings functions are responsible for consistently reducing wage inequality in their Tanzania and Kenya data, approximately cancelling out the "composition effect" of changes in educational attainment. Meanwhile, Machado and Mata (2005) estimate that in their Portuguese data, changes in returns to educational attributes contributed to an increase in wage inequality of about the same magnitude as did changes in educational attributes. Lemieux (2006), using quantile regressions and a human capital model, shows that most of the increase in wage inequality in the United States from 1973 to 2005 is due to an increase in the return to college education. Some scholars have focused on discovering the determinants of the college wage premium, a measure of the relative average income of college graduates versus non-graduates. Goldin and Katz (2007) find that a supply, demand, and institutions framework very accurately explains the evolution of the college wage premium in the United States in the past century. First, there is a relative supply effect: when college graduates are relatively rare, this scarcity creates value to their high-skilled labor, pushing up their average incomes (Goldin and Katz 2007; Katz and Murphy 1992; Topel 1997). As college education expands, high-skilled labor becomes less scarce, bringing down its value and thus relative average income. Second, growth in the relative demand for high-skilled labor driven by technological change, as economies become more able to utilize the skills of college graduates, increases the college wage premium (Goldin and Katz 2007; Katz and Murphy 1992). Finally, institutional factors such as World War II wage policies played a role in the 1940s and 1950s (Goldin and Katz 2007). Thus, any analysis of the effect of college expansion on the income distribution must also account for technology-driven demand shifts and institutional factors contributing to the "compression effect." However, since in Goldin and Katz's (2007) analysis institutional factors were only present as a 1949 dummy variable, the technological change factor is more important to account for.

There is finally a "dispersion effect" accounting for differences and shifts in the individual income distributions of high-skilled and low-skilled labor. If the income distribution of college graduates is wider than that of non-graduates, then the transition of the population from the latter to the former sector will increase overall income inequality even if the average incomes of the two sectors are the same because a larger fraction of the population will be in a more dispersed sector. Furthermore, the individual income distributions of both sectors may not be static over time. Lemieux (2006) shows that from 1975 to 2003 in the United States, the within-group income dispersion of college graduates was greater than the within-group income dispersion of non-graduates, and furthermore that the within-group income dispersion of college graduates was growing over the period. Xie et al. (2016) find that between 1960 and 2008 in the United States, between-occupation and within-occupation inequality increased for college graduates, both contributing to increasing income dispersion for the college graduate sector.

It remains unclear what the overall relationship between college expansion and income inequality is. Kuznets' hypothesis, considering only the "composition effect," implies an inverted "U" curve, where there is a turning point at which inequality begins to decrease. Adding the "compression effect" implies that as the relative supply of college graduates grows, the college wage premium is also pushed down, flattening the inverted "U" shaped curve. However, other factors, primarily technological change, have contributed to increasing college wage premiums in recent years in some countries, emphasizing the inverted "U" shaped curve. While the "compression effect" changes the magnitude of the relationship between college expansion and income inequality, it should not change the general shape of the trend or the location of a potential turning point. Lastly, the "dispersion effect" may push the potential turning point in the inverted "U" curve to the right, and if it is strong enough, may mitigate the inverted "U" shape altogether in favor of a monotonically increasing trend. There is a need to examine multiple countries over time to uncover what the net trend of income inequality is as college expands within an economy. Gregario and Lee (2002) make an attempt at such an analysis by measuring the Gini indices with respect to years of schooling for various countries using a previous version of the Barro and Lee (2013) dataset I employ.

However, their regressions utilize cross-country variation, which does not accurately uncover the trend *within* countries.

I contribute to the literature by carrying out a multi-country analysis of the relationship between the income distribution and college expansion, accounting for country and time variation. In particular, I look to see whether an inverted "U" trend is observed among different countries. I follow Alvarado et al. (2018) and others in using income shares as a measure of income inequality, since they provide more information and are easier to interpret than indices such as the Gini index commonly used in past studies of income inequality. I use the most updated data series on income shares from the World Inequality Database (WID) (see Alvarado et al. 2018) and on educational attainment from Barro and Lee (2013) to construct longer series and incorporate more countries than previous studies.

II. Methodology and Data

I am primarily interested in describing the observed relationship between college expansion and income inequality, not showing how college attainment causes income inequality. However, my results may be interpreted as causal under certain assumptions. If the relative supply of college graduates in the population is assumed to be predetermined and inelastic in the short run, as is a key assumption in Goldin and Katz's (2007) supply, demand, and institutions framework for the college wage premium, then changing college shares can be said to cause the resulting shifts in the income distribution. More likely, however, there is simultaneous causation between college expansion and income inequality (as argued by Birdsall et al. 1995): the former affects the latter through the mechanisms described above while the latter affects the former by changing the incentives for attending college and the ability of people to do so.

I estimate the following equation separately for each income group—bottom 50%, middle 40%, and top 10%:

$$(1) \qquad S_{it}^{g} = \alpha + \beta_1 C_{it} + \beta_2 {C_{it}}^2 + \beta_3 \log(Y_{it}) + \delta_i^g + \gamma_t^g + \varepsilon_{itg}$$

where *i* indexes the country, *t* indexes the year, and *g* indexes the income group (bottom 50%, middle 40%, or top 10%). *S* is the share of national income, *C* is the proportion of the population with college completed (college share), and *Y* is the per adult national income. I include country and year fixed effects estimators, δ and γ.

I run ordinary least squares regressions to estimate the above equations. I include a quadratic college share term in order to estimate a potentially quadratic relationship between income shares and the college share. The inclusion of country fixed effects eliminates variation due to constant differences in countries' economies, political environments, histories, and social institutions, as well as data differences and other omitted variables. The addition of time fixed effects absorbs worldwide changes in the income distribution and college wage premium within countries due to factors such as technological advancement, as utilized in Goldin and Katz's (2007) analysis of the college wage premium in the United States. This specification depends on the assumption that technological changes affect every country in the same way in any given year, which I discuss further in the following section. The log average income serves as an additional control, since income level has been shown to be correlated with income inequality. Thus, the equations I estimate isolate the relationship between the income distribution and college expansion within each country.

The data I use are a panel of 21 countries observed at five-year intervals from 1950 to 2010. The measures of income inequality are the shares of national income captured by the bottom 50%, middle 40%, and top 10% of income earners, obtained from the World Inequality Database (WID) (see Alvarado et al. 2018). Researchers with WID have constructed comparable income share series for more than seventy countries by combining national accounts, fiscal and wealth data, and surveys in a consistent manner. I follow Alvarado et al. (2018) and others in using income shares as opposed to the Gini index as measures of income inequality because they provide more information and are easier to interpret. The income share data I use are constructed using a pre-tax national income concept (following Piketty et al. 2018) and an age category of adults including the elderly (age 20+). The population categories used to compute the data vary by country between equal-split adults, individuals, and tax units, which does alter measures of the income distribution. However, because the population category is consistent within each country series, these data differences are absorbed by a country fixed effects estimator. Per adult national income in constant 2016 dollars is also obtained from WID and constructed using the same income concept, age category, and corresponding population category as the income share data for each country.

The measure of college attainment is the proportion of the population aged 15-64 with college completed from Barro and Lee (2013). Barro and Lee construct a dataset of measures of educational attainment for 146 countries at five-year intervals from 1950 to 2010 using consistent census data and a backward- and forward-estimation procedure to fill missing observations. Although previous versions of the Barro and Lee dataset received criticism for showing implausible series of educational attainment for some countries, the authors claim to have resolved these problems in their updated 2013 dataset. I merge data from WID with the Barro and Lee dataset, keeping only country-year observations with income share data for all three income groups for better comparability between the three regressions, though WID has more countries with only the top 10% income share data.

Table 1: Summary of Data

Variable	N	Mean	Min	Median	Max
Bottom 50% income share	124	0.165	0.0471	0.152	0.314
Middle 40% income share	124	0.381	0.223	0.369	0.491
Top 10% income share	124	0.453	0.210	0.472	0.730
College share	124	0.0687	0.00321	0.0546	0.268
Average income	124	7.240	1.055	7.055	13.18

Sources and Notes: The income share and average income data are from WID (see Alvaredo et al. 2018) and are constructed using national accounts, fiscal and wealth data, and surveys for different countries. The college share data are from Barro and Lee (2013) and are constructed via census data and backward- and forward-estimation. Only country-year observations with data for all variables are included in the panel. The income share for each income group is the share of national income captured by that section of income earners. The college share is the proportion of the population aged 15-64 with college completed. The average income is per adult national income in constant 2016 dollars.

Table 2: Frequency Table of World Regions Represented in Data

World Region	Frequency	Percent
OECD Economies	32	25.81
East Asia and the Pacific	7	5.65
Europe and Central Asia	16	12.90
Latin America and the Caribbean	2	1.61
Middle East and North Africa	55	44.35
South Asia	12	9.68
Total	124	100.00

Sources and Notes: Countries are categorized by world region in the Barro and Lee (2013) dataset. The table lists the frequency and percentage of country-year observations in each world region.

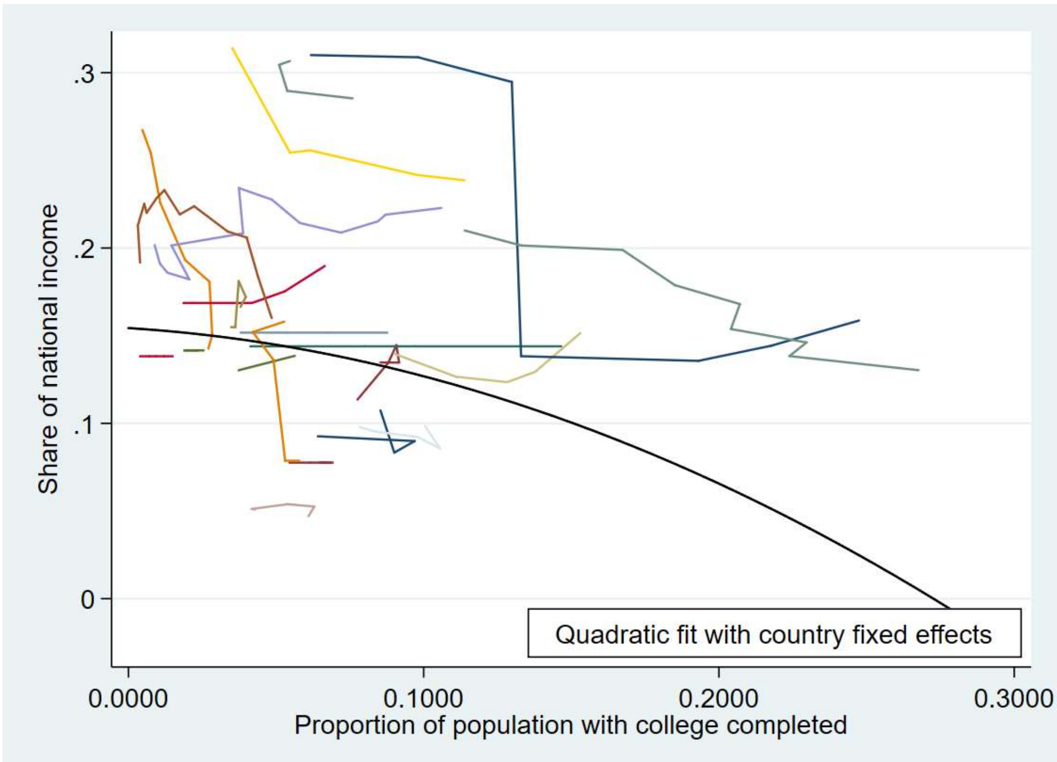
The data are summarized in Table 1. On average, for the countries and time period in my sample, the bottom 50% of the income distribution captures 16.5% of its country's national income, the middle 40% captures 38.1%, and the top 10% captures 45.3%, though income shares vary widely. The proportion of the population with college completed is on average 6.87% and ranges from 0.320% to 26.8% in the sample. The panel is unbalanced, however, with only one country with series extending back to 1950, 18 countries with observations in 1990, and all 21 countries with observations in 2010. Table 2 shows the different world regions represented in the sample. The sample is biased toward the two most represented regions, the Middle East and North Africa and OECD Economies, because WID contains more complete income share data for countries in these regions.

III. Results and Discussion

To analyze the relationship between college attainment and the income distribution, I run separate OLS regressions of the income shares of three income groups (bottom 50%, middle 40%, and top 10%) on the college share with an added quadratic college share term to capture potential nonlinear trends. I include country and year fixed effects and the log average income as controls.

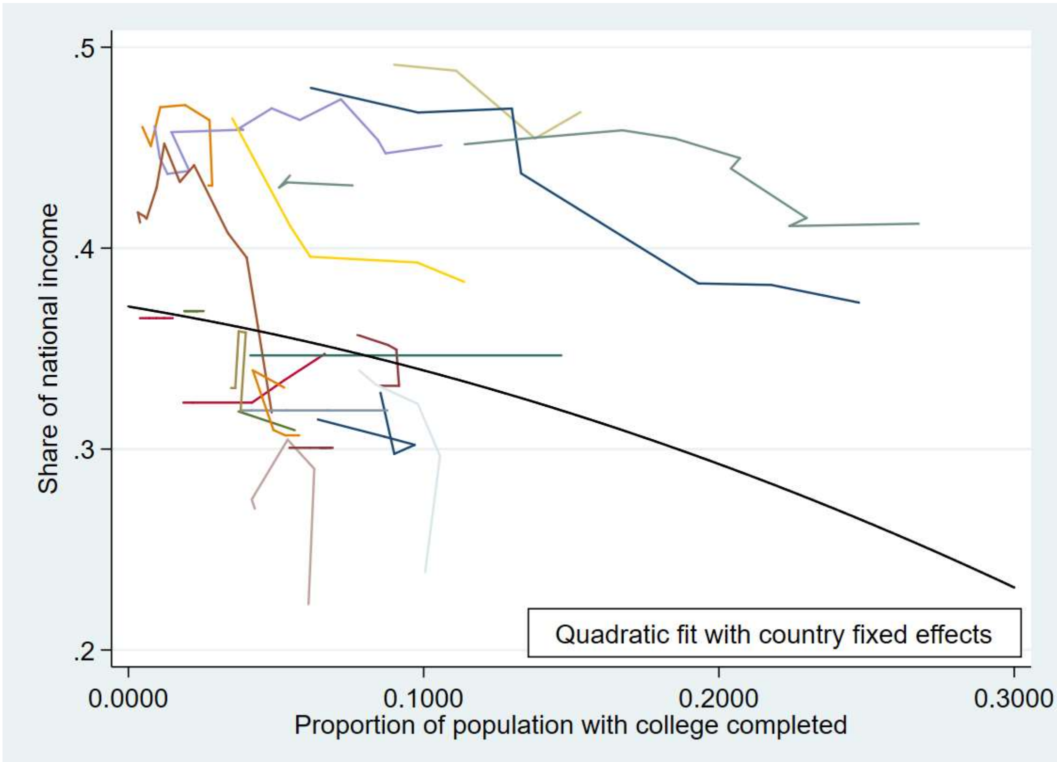
Income share series for the bottom 50%, middle 40%, and top 10% with respect to the college share for different countries and with a quadratic fit line with country fixed effects are shown in Figure 1, Figure 2, and Figure 3, respectively. Since the country series do not have means subtracted, the quadratic fit line with country fixed effects does not pass through country series but rather shows the estimated average shape of all series.

Figure 1: Bottom 50% Income Share Series



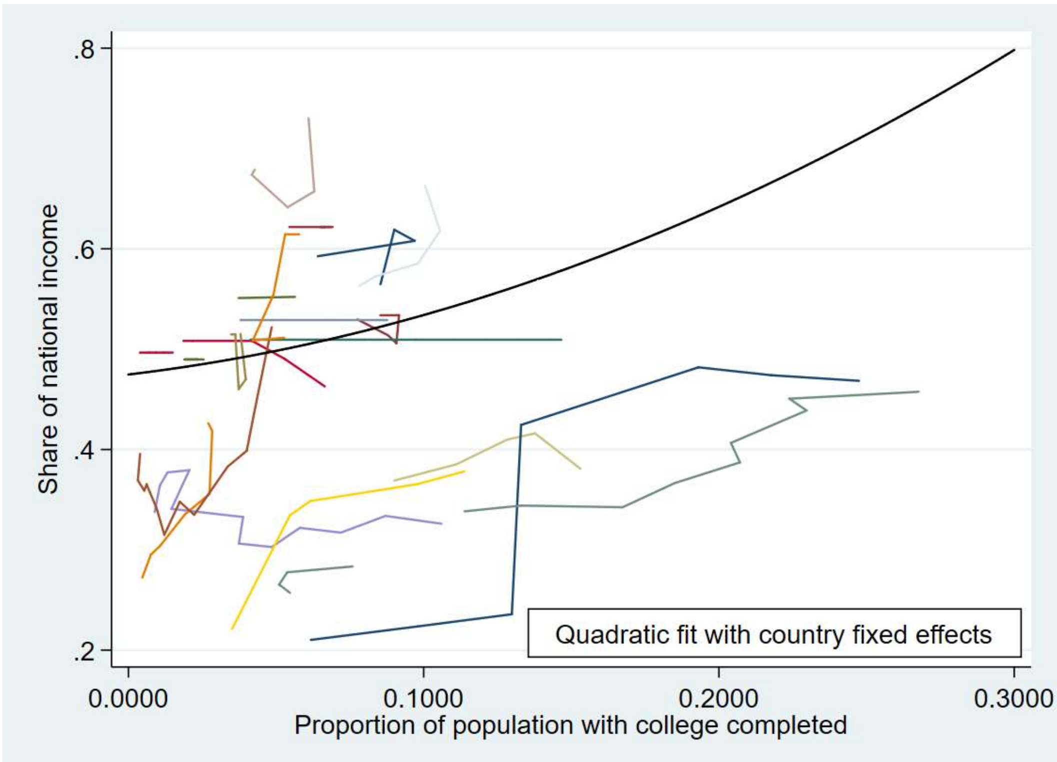
Sources and Notes: Income share data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). Individual country series are plotted in different colors and connected in ascending order of year. A quadratic fit with country fixed effects is plotted in black. Since the country series do not have means subtracted, the quadratic fit line with country fixed effects does not pass through country series but rather shows the estimated average shape of all series.

Figure 2: Middle 40% Income Share Series



Sources and Notes: Income share data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). Individual country series are plotted in different colors and connected in ascending order of year. A quadratic fit with country fixed effects is plotted in black. Since the country series do not have means subtracted, the quadratic fit line with country fixed effects does not pass through country series but rather shows the estimated average shape of all series.

Figure 3: Top 10% Income Share Series



Sources and Notes: Income share data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). Individual country series are plotted in different colors and connected in ascending order of year. A quadratic fit with country fixed effects is plotted in black. Since the country series do not have means subtracted, the quadratic fit line with country fixed effects does not pass through country series but rather shows the estimated average shape of all series.

All three plots reveal large disparities between different country series of income shares with respect to the college share. Figure 1 and Figure 2 show decreasing, possibly quadratic trends for the bottom 50% and middle 40% income shares with respect to the college share. Figure 3 shows an increasing, possibly quadratic trend for the top 10% income share with respect to the college share.

OLS regressions for bottom 50%, middle 40%, and top 10% income shares are shown in Table 3, Table 4, and Table 5, respectively. Specifications (1), (2), and (3) regress income share with respect to only a linear college share term, while columns (4), (5,) and (6) also include a quadratic college share term. To check for robustness to different time specifications, regressions (1) and (4) include no time variable; regressions (2) and (5) contain a linear year variable indexed to 0 at 1950, the first year in the sample; and regressions (3) and (6) include year fixed effects. Column (6) is the equation described in the methodology.

Table 3: Bottom 50% Income Share Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
College share	-0.590***	-0.682***	-0.482***	-0.157	-0.240	-0.057
	(0.147)	(0.168)	(0.165)	(0.357)	(0.345)	(0.287)
College share^2				-1.560	-1.419	-1.380
				(1.101)	(1.034)	(0.915)
Time		0.0004			0.0002	
		(0.0005)			(0.0004)	
Log average income	0.0122	0.0044	0.0011	0.0036	0.0007	-0.0020
	(0.0169)	(0.0206)	(0.0182)	(0.0194)	(0.0219)	(0.0191)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Constant	0.069	0.129	0.140	0.124	0.147	0.156
	(0.152)	(0.180)	(0.149)	(0.166)	(0.186)	(0.154)
Observations	124	124	124	124	124	124
R-squared	0.864	0.865	0.904	0.869	0.869	0.907

Sources and Notes: Income share and average income data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). The time variable is a linear year variable indexed to 0 at 1950, the first year in the sample. Each column is an OLS regression of the bottom 50% income share on the indicated variables. Standard errors are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Middle 40% Income Share Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
College share	-0.414***	-0.471***	-0.382***	-0.127	-0.175	-0.181
	(0.079)	(0.106)	(0.123)	(0.198)	(0.255)	(0.258)
College share^2				-1.033	-0.951	-0.651
				(0.645)	(0.703)	(0.694)
Time		0.0003			0.0001	
		(0.0003)			(0.0003)	
Log average income	-0.0026	-0.0074	-0.0118	-0.0083	-0.0099	-0.0132
	(0.0102)	(0.0085)	(0.0102)	(0.0099)	(0.0083)	(0.0102)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Constant	0.404***	0.442***	0.463***	0.441***	0.454***	0.470***
	(0.094)	(0.079)	(0.084)	(0.090)	(0.077)	(0.084)
Observations	124	124	124	124	124	124
R-squared	0.922	0.923	0.933	0.924	0.924	0.934

Sources and Notes: Income share and average income data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). The time variable is a linear year variable indexed to 0 at 1950, the first year in the sample. Each column is an OLS regression of the middle 40% income share on the indicated variables. Standard errors are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Top 10% Income Share Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
College share	1.003***	1.153***	0.864***	0.284	0.415	0.239
	(0.215)	(0.247)	(0.253)	(0.512)	(0.550)	(0.490)
College				2.592	2.370	2.031
share^2				(1.586)	(1.583)	(1.452)
Time		-0.0007			-0.0003	
		(0.0006)			(0.0006)	
Log average	-0.0096	0.0030	0.0107	0.0047	0.0092	0.0152
income	(0.0243)	(0.0259)	(0.0235)	(0.0260)	(0.0272)	(0.0242)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Constant	0.527**	0.430*	0.397**	0.435*	0.398*	0.373*
	(0.219)	(0.230)	(0.192)	(0.226)	(0.234)	(0.195)
Observations	124	124	124	124	124	124
R-squared	0.906	0.907	0.931	0.910	0.910	0.932

Sources and Notes: Income share and average income data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). The time variable is a linear year variable indexed to 0 at 1950, the first year in the sample. Each column is an OLS regression of the top 10% income share on the indicated variables. Standard errors are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

For all three income share groups, the regression specifications which include a quadratic college-share term yield coefficients insignificant at the 0.1 level. Meanwhile, the regressions on only a linear college-share term yield coefficients on the college share term significant at the 0.01 level. This provides evidence for a negative, linear relationship between bottom 50% and middle 40% income shares with college share and for a positive, linear relationship between top 10% income share and college share for the range of college shares in the sample. For all three income groups, this finding is robust to different time specifications; the three regressions with different time variables all yield coefficients significant at the 0.01 level and which have overlapping 99% confidence intervals. My preferred specification for all three income groups is column (3), which estimates that a percentage-point increase in college share corresponds to a 0.482 percentage-point decrease in income share for the bottom 50%, a 0.382 percentage-point decrease in income share for the bottom 40%, and a 0.864 percentage-point increase in income share for the top 10%. Thus, the bottom 50% and middle 40% both lose linearly as college education expands within an economy for the range of college shares in the sample, but the bottom 50% loses more. Those losses are accrued by the top 10%.

These results provide evidence for a linearly increasing trend of income inequality with respect to the college share for the range of college shares between 0.320% and 26.8%. There is no evidence of an inverted “U” turning point for any income group in this range of college shares, indicating that the “composition effect” is sufficiently offset by a “dispersion effect” to suppress an inverted “U” trend within this range. As more people obtain college degrees, they join a labor sector with a so much more dispersed income distribution that inequality in the overall income distribution increases, at least for this range of college shares. Whether there exists a turning point beyond a 26.8% college share and where such a turning point occurs remains to be discovered. In fact, no country in the Barro and Lee (2013) dataset has achieved a college share above 30% in 2010. It is possible that there is a turning point at perhaps 50%, 60%, or 70% college share, but no country has ever reached such a high college share. It is also possible that a turning point does not exist at all and that income inequality increases monotonically with increasing college share. My results only provide evidence that no such turning point exists below a 26.8% college share.

A natural question is whether the trends observed are truly independent of all country- or region-specific factors. If they are, one should see similar trends when restricting the sample to specific world regions. While my original sample is not large enough to conduct the same analysis within different world regions, WID does have data for more countries for the top 10% income share, which I use to construct an enlarged sample. In particular, this enlarged sample contains 124 observations for OECD countries, which I use to conduct the same analysis as above. The results are shown in Table A1. All six regressions yield coefficients significant at the 0.01 level on both linear and quadratic college share terms. The quadratic trends show a slight “U” shaped—not inverted “U”—curve, though the top 10% income share is only decreasing from a 0.66% to 6.40% college share and is increasing from a 6.40% to 26.8% college share for regression (6). The linear trends show the top 10% income share increasing with college share for the range of college shares in the sample, consistent with my original sample but with a smaller magnitude, a 0.511 percentage-point increase with each percentage-point increase in college share in regression (3). Neither the quadratic nor linear trend show evidence for an inverted “U” curve turning point within the range of college shares in the sample. Thus, these findings agree with the generally increasing trend observed in my original sample of top 10% income shares. However, the magnitude of the effect and to a certain extent the shape of the trend differ substantially from the original sample. Looking at the linear trend specifications, the magnitude of the effect for the OECD countries sample was much smaller than the effect observed in the original sample. This may be explained by country differences in the college wage premium, which cannot be absorbed by a country fixed effects estimator since they affect the slope of the trend line. This implies that the general shape of the observed trend is common to different countries but that the magnitude of the trend varies by country.

Another potential weakness of my regression specification is accurately controlling for changes in the college wage premium within each country. I include time fixed effects to capture technological factors which affect the college wage premium in all countries, as do Goldin and Katz (2007) in their analysis of the determinants of the college wage premium in the United States. This provides accurate estimation under the assumption that technological changes affect each country's college wage premium in the same way in any given year. However, different countries may incorporate technologies such as computers at different times. If this is the case, my regressions do not fully control for shifts in the college wage premium within countries, which may also be correlated with the college share: as a country becomes better educated, it may also incorporate more technologies into the workplace. Thus, the magnitudes of my college share coefficient estimates may be overestimated if they absorb some of the effect of an increasing college wage premium.

IV. Conclusion

Both income inequality and college education have been increasing within countries in the past few decades (Alvaredo et al. 2018; Barro and Lee 2013). Scholars have identified three mechanisms for the effect of college expansion on income inequality: a "composition effect," "compression effect," and "dispersion effect" (Knight and Sabot 1983; Lemieux 2006). In this paper, I analyze the net relationship between college expansion and income inequality using a panel of countries with income share data from the World Inequality Database (see Alvaredo et al. 2018) and college share data from Barro and Lee (2013). I run OLS regressions with a quadratic college share term and country and year fixed effects. I find that bottom 50% and middle 40% income shares decrease linearly with increasing college share while the top 10% income share increases linearly, for the range of college shares in my sample. While the general shape of these trends is consistent when looking at an enlarged sample of only OECD countries, the magnitude of the trend is different, indicating that the relationship depends on country-specific factors such as the college wage premium. My coefficient estimates may be overestimated because of potentially uncontrolled changes in the college wage premium within countries.

These results provide evidence that income inequality within countries increases linearly for the range of college shares in my sample, about 0% to 27%. This finding may be particularly helpful for analyzing a developing country with a currently low proportion of its population with a college degree to understand how trends in college education expansion may increase inequality in its income distribution. However, the magnitude of this effect estimated here should be interpreted with caution, as it varies between different countries and world regions and may be overestimated. While this research describes the relationship between the college share and the income distribution, it does not prove causality of the effect of college expansion on income inequality. Further research exploring the causation of this relationship, such as by instrumenting for college education using government education policies, or through event study analysis using exogenous shocks to college expansion like the post-WWII GI Bill in the United States, will help elucidate the effect of college expansion on income inequality and help policymakers evaluate the impacts of college education policies on income inequality.

References

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2018). *World Inequality Report 2018*. Cambridge, MA: The Belknap Press of Harvard University Press.

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, *104*, 184–198. https://doi.org/10.1016/j.jdeveco.2012.10.001

Birdsall, N., Ross, D., & Sabot, R. (1995). Inequality and Growth Reconsidered: Lessons from East Asia. *The World Bank Economic Review*, *9*(3), 477–508.

Goldin, C., & Katz, L. (2007). *The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005* (No. w12984). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w12984

Gregorio, J. D., & Lee, J.-W. (2002). Education and Income Inequality: New Evidence From Cross-Country Data. *Review of Income and Wealth*, *48*(3), 395–416. https://doi.org/10.1111/1475-4991.00060

Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, *107*(1), 35–78. https://doi.org/10.2307/2118323

Knight, J. B., & Sabot, R. H. (1983). Educational Expansion and the Kuznets Effect. *The American Economic Review*, *73*(5), 1132–1136.

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review*, *45*(1), 1–28.

Lemieux, T. (2006). *Post-Secondary Education and Increasing Wage Inequality* (Working Paper No. 12077). National Bureau of Economic Research. https://doi.org/10.3386/w12077

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*(4), 445–465. https://doi.org/10.1002/jae.788

Piketty, T., Saez, E., & Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States*. *The Quarterly Journal of Economics*, *133*(2), 553–609. https://doi.org/10.1093/qje/qjx043

Topel, R. H. (1997). Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality. *Journal of Economic Perspectives*, *11*(2), 55–74. https://doi.org/10.1257/jep.11.2.55

Xie, Y., Killewald, A., & Near, C. (2016). Between- and Within-Occupation Inequality: The Case of High-Status Professions. *Annals of the American Academy of Political and Social Science*, *663*, 53–79.

Appendix

Table A1: Top 10% Income Share Regressions for Enlarged Sample of OECD Countries

	(1)	(2)	(3)	(4)	(5)	(6)
College share	0.294***	0.522***	0.511***	-0.901***	-0.678***	-0.563***
	(0.108)	(0.146)	(0.135)	(0.207)	(0.229)	(0.198)
College share^2				4.772***	4.884***	4.398***
				(0.775)	(0.793)	(0.628)
Time		-0.001**			-0.001***	
		(0.001)			(0.001)	
Log average income	-0.020	0.008	0.047***	0.013	0.045***	0.071***
	(0.013)	(0.016)	(0.016)	(0.012)	(0.014)	(0.015)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Constant	0.454***	0.159	-0.213	0.171	-0.161	-0.408***
	(0.135)	(0.165)	(0.160)	(0.119)	(0.144)	(0.144)
Observations	124	124	124	124	124	124
R-squared	0.831	0.838	0.888	0.867	0.876	0.915

Sources and Notes: Income share and average income data are from WID (see Alvaredo et al. 2018) and college share data are from Barro and Lee (2013). The enlarged sample is restricted to OECD countries. The time variable is a linear year variable indexed to 0 at 1950, the first year in the sample. Each column is an OLS regression of the top 10% income share on the indicated variables. Standard errors are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.