Economics & Management Series EMS-2024-02

Educational Expansion and Educational Inequality

Takahiro Akita
International University of Japan

March 2024

IUJ Research Institute
International University of Japan

Educational Expansion and Educational Inequality

Takahiro Akita*

Professor Emeritus, International University of Japan, akita@iuj.ac.jp

Abstract

This study explores the relationship between educational expansion and educational inequality through the use of hypothetical examples. It also examines the relationship empirically based on a Barro and Lee dataset on educational attainment for Asian countries and economies. If individuals without formal education are assigned 0 years of education, the education Gini coefficient is likely to decline monotonically with educational expansion. In contrast, if we assume that they receive some sort of informal education equivalent to a small amount of formal education, then the education Gini coefficient is likely to exhibit an inverted U-shaped pattern. Transforming years of education into human capital using an exponential function could lead to the Gini coefficient of human capital exhibiting an inverted U-shaped pattern with respect to human capital expansion. On the other hand, the standard deviation of education is likely to display an inverted U-shaped pattern, whether individuals without formal education are assigned 0 years of education or not. The Barro and Lee dataset reveals that the standard deviation of education follows an inverted U-shaped pattern, even when individuals without formal education are assigned 0 years of education. In contrast, the education Gini coefficient demonstrates a downward-sloping pattern when individuals without formal education are assigned 0 years of education. However, when assigning one year to individuals without formal education, the education Gini coefficient displays an inverted U-shaped pattern. These empirical observations align with the conclusions drawn from hypothetical examples.

Key words: educational expansion, educational inequality, education Gini coefficient, inverted U-shaped pattern, human capital inequality, Asian countries and economies

* The author is grateful to the Japan Society for the Promotion of Science for its financial support (Grant-in-Aid for Scientific Research: 23K01409).

1. Introduction

It is widely believed that the expansion of education serves as an effective means for promoting the dissemination of economic opportunities. Numerous empirical studies have thus been conducted to explore the relationship between education and the distribution of income using cross-section or panel datasets.[1] Though the methods and the datasets differ, most studies have found that an increase in the average level of education seems to reduce income inequality (Ram, 1989; Park, 1996; De Gregorio and Lee, 2002; Park, 2017; Lee and Lee, 2018). Meanwhile, the majority of studies investigating the effect of educational inequality on income distribution have discovered evidence that supports a direct relationship between inequality of education and income inequality, implying that a rise in educational inequality seems to increase income inequality (Park, 1996; De Gregorio and Lee, 2002; Park, 2017; Lee and Lee, 2018).

As the average level of education and educational inequality both appear to influence income inequality, an interesting question arises: How does educational inequality evolve with the expansion of education (Ram, 1990). While the partial effect of the expansion of education seems to reduce income inequality, this effect could be counteracted or strengthened by changes in educational inequality associated with educational expansion (Ram, 1990). Against this backdrop, this study examines the relationship between educational expansion and educational inequality.[2]

This paper is organized as follows. Section 2 introduces measures of educational inequality, while section 3 discusses the evolution of educational inequality associated with the expansion of education. Section 4 empirically analyzes the relationship between educational expansion and educational inequality in Asian countries and economies, utilizing a Barro and

[1] For example, Knight and Sabot (1983), Ram (1989), Lam and Levison (1992), Park (1996), De Gregorio and Lee (2002), Fordvari and van Leeuwen (2011), Park (2017), Lee and Lee (2018) and Castello-Climent and Domenech (2021).

[2] A number of studies have explored the relationship between educational expansion and educational inequality (or human capital expansion and human capita inequality). They include Ram (1990), Lam and Levison (1992), Park (1996), Castello and Domenech (2002), Lin (2007), Lim and Tang (2008), Hojo (2009), Lin and Yang (2009), Agrawal (2014), Meschi and Scervini (2014), Yang, et al. (2014), Coady and Dizioli (2018), Banzragch, et al. (2019), Shukla and Mishra (2019), Castello-Climent and Domenech (2021), Almeida, et al. (2022) and Luo, Zeng and Baležentis (2022).

Lee dataset on educational attainment for the population aged 25 to 64 over the period 1950-2015. The concluding section summarizes the findings of this study.

2. Measures of Educational Inequality

To measure inequality in the distribution of educational attainment (educational inequality), some researchers have used the variance or standard deviation (Ram, 1990; Lam and Levison, 1992; Park, 1996; De Gregorio and Lee, 2002; Meschi and Scervini, 2014; Park 2017).

To define the variance and standard deviation of educational attainment, we consider a country with four levels of education: (0) no formal education; (1) primary education; (2) secondary education; and (3) tertiary education. Let e_i and p_i be the cumulative years of education for *i*th education level and the proportion of people who have completed *i*th level of education as their highest attainment level, respectively (i = 0, 1, 2, and 3). Then, the mean years of education for the country is given by

$$\mu = \sum_{i=0}^{3} p_i e_i, \ \ (\sum_{i=0}^{3} p_i = 1). \qquad (1)$$

Using e_i, p_i and μ, the variance and standard deviation of educational attainment are defined respectively by

$$V = \sum_{i=0}^{3} p_i (e_i - \mu)^2, \qquad (2)$$

$$\sigma = \sqrt{V} = \sqrt{\sum_{i=0}^{3} p_i (e_i - \mu)^2}. \qquad (3)$$

We should note that as a measure of *income* inequality, the variance or standard deviation satisfies three desirable properties: anonymity, population homogeneity (or population independence), and the Pigou-Dalton principle of transfers (Anand, 1983; Fields, 2001; Akita and Kataoka, 2022). Anonymity means that an inequality measure should not depend on who has a higher income or lower income, while population homogeneity denotes that an inequality measure remains unchanged if the number of individuals at each income level is changed by the same proportion. The Pigou-Dalton transfer principle implies that any income transfer from a richer to a poorer individual that does not reverse their relative ranks in income lowers the value of an inequality index.

Since Thomas, et al. (2001) introduced the Gini coefficient of education as an extension of the Gini coefficient of income distribution, many researchers have employed the Gini coefficient to measure educational inequality (Castello and Domenech, 2002; Lin, 2007; Lim and Tang, 2008; Hojo, 2009; Fordvari and van Leeuwen, 2011; Agrawal, 2014; Meschi and Scervini, 2014; Coady and Dizioli, 2018; Banzragch, et al., 2019; Shukla and Mishra, 2019; Castello-Climent and Domenech, 2021; Almeida, et al., 2022; and Luo, et al., 2022). Using e_i, p_i and μ, the Gini coefficient of education is defined by

$$G = \frac{1}{2\mu}\sum_{i=0}^{3}\sum_{j=0}^{3} p_i p_j \left|e_j - e_i\right| = \frac{1}{\mu}\sum_{i=0}^{3}\sum_{j>i}^{3} p_i p_j \left|e_j - e_i\right| . \quad (4)$$

It should be noted that as a measure of *income* inequality, the Gini coefficient satisfies anonymity, population homogeneity, and the Pigou-Dalton principle of transfers. Moreover, it satisfies income homogeneity (or mean independence), which implies that an inequality measure remains unchanged if everyone's income is changed by the same proportion. The Gini coefficient is a *relative* inequality measure, as relative income shares are sufficient to calculate the Gini. On the other hand, the variance and standard deviation do not satisfy income homogeneity; thus, they are *absolute* inequality measures.

To examine the characteristics of the Gini coefficient, we consider two villages, each consisting of 5 individuals with different levels of educational attainment (Table 1). In village 1, there are three individuals with incomplete primary education (3 years of education) and two individuals with primary education (6 years of education). On the other hand, in village 2, there are three individuals with primary education (6 years of education) and two individuals with secondary education (12 years of education). Years of education in village 2 are two times those in village 1. But both villages have the same Gini coefficient at 0.171 as follows, exemplifying that the Gini coefficient satisfies income homogeneity.

Village 1: $G = \frac{1}{4.2}((0.6)(0.4)|6 - 3|) = 0.171.$

Village 2: $G = \frac{1}{8.4}((0.6)(0.4)|12 - 6|) = 0.171.$

Table 1. Distributions of Educational Attainment for Villages 1 and 2

	Village 1		Village 2	
Individual	Years of education	Share of years of education (%)	Years of education	Share of years of education (%)
1	3	14.3	6	14.3
2	3	14.3	6	14.3
3	3	14.3	6	14.3
4	6	28.6	12	28.6
5	6	28.6	12	28.6
Total	21	100.0	42	100.0
Mean	4.20		8.40	
Standard deviation	1.47		2.94	
Gini coefficient	0.171		0.171	
Coefficient of variation (*CV*)	0.350		0.350	

It is important to note that the coefficient of variation, defined by the ratio of the standard deviation to the mean (*CV*), can also be used as a measure of educational inequality.

$$CV = \frac{\sigma}{\mu} \tag{5}$$

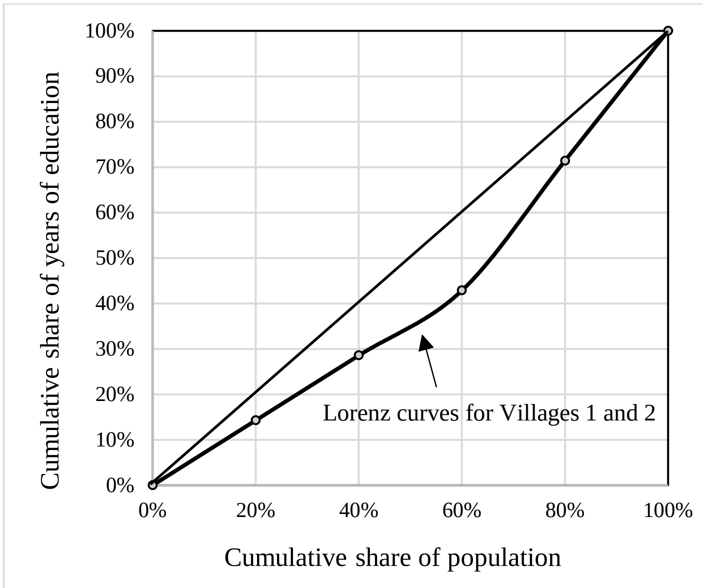
The *CV* satisfies anonymity, population homogeneity, and the Pigou-Dalton principle of transfers. Since two villages have the same *CV*, the *CV* also satisfies income homogeneity; thus, it is a *relative* inequality measure.

Village 1: $CV = \frac{1.47}{4.2} = 0.350.$

Village 2: $CV = \frac{2.94}{8.4} = 0.350.$

As shown in Fig. 1, villages 1 and 2 have the same Lorenz curve for the distribution of educational attainment. The Lorenz curve is often used to visualize an income distribution (Lorenz, 1905). To draw the Lorenz curve, we first order all individuals from the poorest to the richest. On the horizontal axis, we plot the cumulative population shares of individuals from the poorest to the richest, while on the vertical axis, we plot the cumulative income shares. In the case of the education Lorenz curve, we first order individuals from the least educated to the most educated. Then, on the vertical axis, we plot the cumulative shares of years of education. It is important to note that using the Lorenz curve, the Gini coefficient is defined by the ratio of the area between the diagonal line and the Lorenz curve to the area of triangle under the diagonal line (see Appendix 1 for the definition of the Gini coefficient). The Gini coefficient ranges between 0 and 1.

Figure 1. Education Lorenz Curves for Villages 1 and 2



On the other hand, the standard deviation differs in two villages. Village 2 has 2.94, which is two times that in village 1 as follows, implying that the standard deviation does not meet income homogeneity. The lowest value of the standard deviation is 0.

Village 1: $\sigma = \sqrt{(0.6)(3-4.2)^2 + (0.4)(6-4.2)^2} = 1.47.$

Village 2: $\sigma = \sqrt{(0.6)(6-8.4)^2 + (0.4)(12-8.4)^2} = 2.94.$

3. Educational Expansion and Educational Inequality: Some Hypothetical Examples

How does educational inequality evolve with the expansion of education? To explore this question, let us consider a hypothetical country comprising 10 individuals. Over an extended time period, the country has undergone a significant expansion in education. Table 2 illustrates the changes in the education structure associated with this expansion, where the numbers are years of education attained by each individual: 0 no formal education; 6 primary education; 12 secondary education; and 16 tertiary education (Example 1). For instance, at time 5, there are 2 individuals with no formal education, 4 with primary education, 3 with secondary education, and 1 with tertiary education.

Table 2. Changes in the Structure of Education (Example 1)

Individual	Time 0	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
1	0	0	0	0	0	0	0	0	6	6	6
2	0	0	0	0	0	0	6	6	6	6	12
3	0	0	0	0	0	6	6	6	6	12	12
4	0	0	0	0	6	6	6	6	12	12	12
5	0	0	0	0	6	6	6	12	12	12	12
6	0	0	0	6	6	6	12	12	12	12	16
7	0	0	0	6	6	12	12	12	12	16	16
8	0	0	6	6	12	12	12	12	16	16	16
9	0	6	6	6	12	12	16	16	16	16	16
10	6	6	6	12	12	16	16	16	16	16	16
Total	6	12	18	36	60	76	92	98	114	124	134
Mean	0.6	1.2	1.8	3.6	6.0	7.6	9.2	9.8	11.4	12.4	13.4
STD	1.80	2.40	2.75	3.98	4.65	5.04	4.92	4.85	3.90	3.67	3.10
Gini	0.90	0.80	0.70	0.57	0.42	0.36	0.29	0.27	0.18	0.15	0.11
CV	3.00	2.00	1.53	1.11	0.77	0.66	0.53	0.50	0.34	0.30	0.23

(Note) STD and CV stands for the standard deviation and the coefficient of variation, respectively.

Table 2 presents the standard deviation of years of education. To examine the evolution of educational inequality with the expansion of education, the standard deviation is plotted against mean years of education in Fig. 2. Educational inequality as measured by the standard deviation exhibits an inverted U-shaped pattern with respect to educational expansion. The standard deviation starts at 1.8 at time 0 when 90% of individuals have no formal education with the mean years of education of 0.6. As mean years of education increase, the standard deviation rises, but after reaching the peak at around 5.0, it declines. At time 10, it is 3.1 with the mean years of education of 13.4, where there are 1 individual with primary education, 4 with secondary education and 5 with tertiary education. This result is consistent with the results of previous empirical studies based on the standard deviation or variance (Ram, 1990; Lam and Levison, 1992; Park, 1996; De Gregorio and Lee, 2002; Meschi and Scervini, 2014; Park, 2017; and Shukla and Mishra, 2019). It should be noted that under a specific condition, we can demonstrate an inverted U-shaped pattern when educational inequality is measured by the variance or standard deviation (see Appendix 2).

Figure 2. Evolution of Educational Inequality by the Standard Deviation (Example 1)

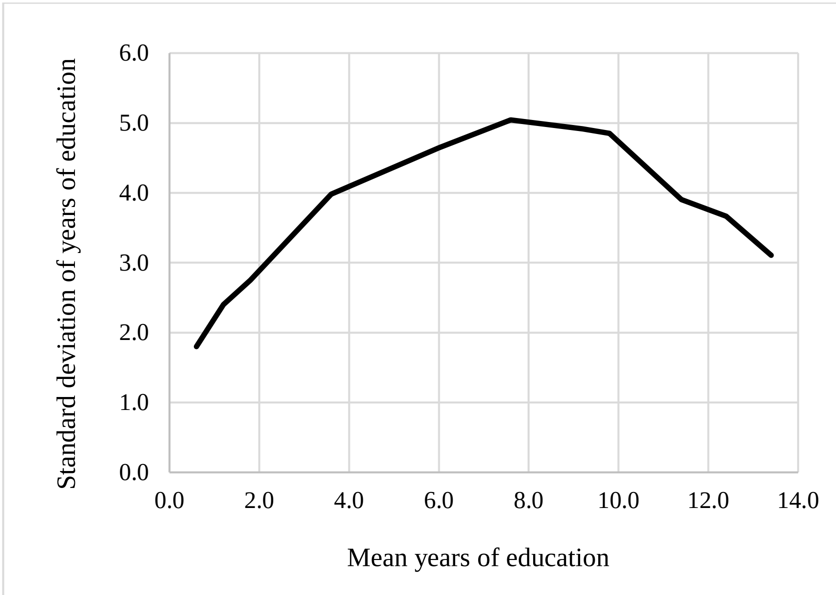


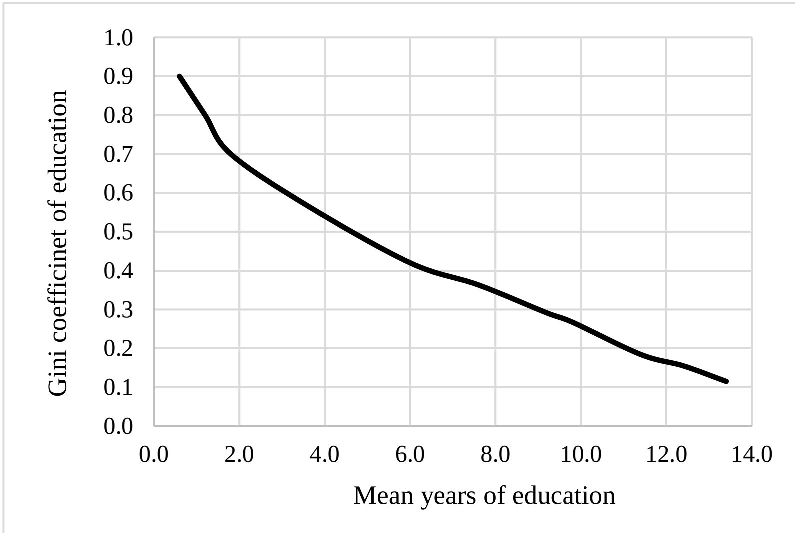
Table 2 also provides the Gini coefficient of education for the country. If educational inequality is measured by the Gini coefficient, it exhibits a downward-sloping pattern rather than an inverted U-shaped pattern, that is, educational inequality declines monotonically with educational expansion (Fig. 3).[3] This result is consistent with the results of previous empirical studies which use the Gini coefficient as a measure of educational inequality (Castello and Domenech, 2002; Lin, 2007; Lim and Tang, 2008; Fordvari and van Leeuwen, 2011; Agrawal, 2014; Meschi and Scervini, 2014; Banzragch, et al., 2019; Shukla and Mishra, 2019; Castello-Climent and Domenech, 2021; and Luo, Zeng and Baležentis, 2022). We should note that under a specific condition, we can show a downward-sloping pattern if educational inequality is measured by the Gini coefficient (see Appendix 3). We can also demonstrate that if there is no inequality among the educated, the Gini coefficient of education is the same as the proportion of individuals without formal education (for example, Gini at time 0, 1, and 2 in Table 2).[4]

---

[3] The coefficient of variation (*CV*) also exhibits a downward-sloping pattern rather than an inverted U-shaped pattern (see Table 2).

[4] In Appendix 3, we will show that when individuals without formal education are assigned 0 years of education ($e_0 = 0$), the overall education Gini coefficient is given by $p_0 + (1 - p_0)G_1$ where G_1 is the education Gini coefficient among the educated. Therefore, if G_1 is 0, then the overall Gini is the same as the proportion of individuals without formal education.

Figure 3. Evolution of Educational Inequality by the Gini Coefficient (Example 1)



In contrast to the distribution of income or expenditure, the distribution of educational attainment possesses two distinctive characteristics. First, an individual without formal education is usually assigned 0 years of education. Therefore, when all individuals do not have any formal education, the mean years of education becomes 0, making the Gini coefficient and the *CV* undefined.[5] This limitation arises from the presence of the mean years of education (μ) in the denominator for both the Gini coefficient and the *CV*. When only a few individuals obtain some formal education with the other individuals having no formal education, the Gini coefficient and the *CV* have a very large value. For example, the Lorenz curve for the distribution of educational attainment at time 0 in Table 2 is presented in Fig. 4; the Gini coefficient and the *CV* are very large at 0.90 and 3.0, respectively. However, the Gini coefficient and the *CV* decrease as education expands. The education Lorenz curve at time 5 in Table 2 is depicted in Fig. 5; the Gini coefficient and the *CV* decrease substantially to 0.36 and 0.66, respectively. In contrast, even when all individuals have no formal education, the standard

[5] When measuring inequality in the distribution of income or expenditure, household incomes or expenditures are usually positive; thus, the mean income or expenditure is positive.

deviation can be computed and equals to 0. Consequently, in the initial phases of educational expansion, the standard deviation rises (refer to Fig. 2).

Figure 4. Education Lorenz Curve at Time 0 in Table 2

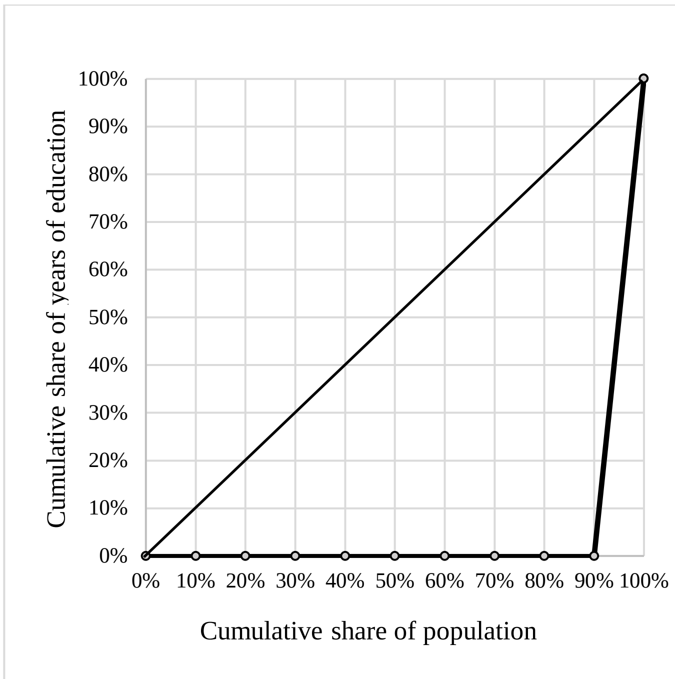
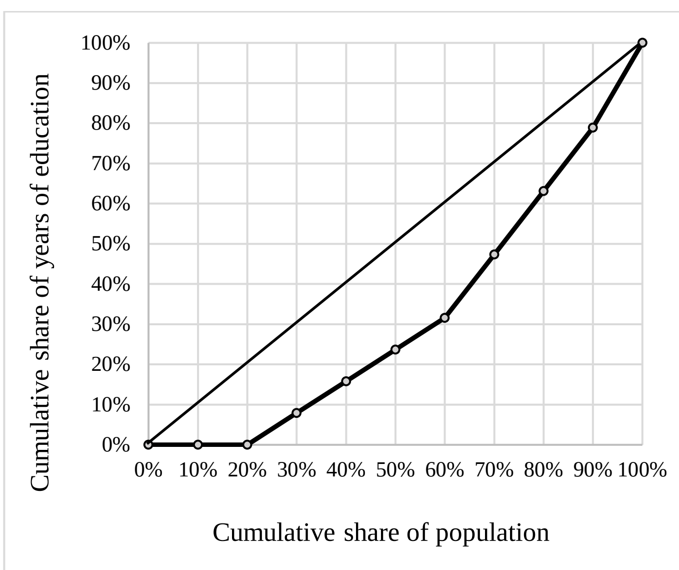


Figure 5. Education Lorenz Curve at Time 5 in Table 2



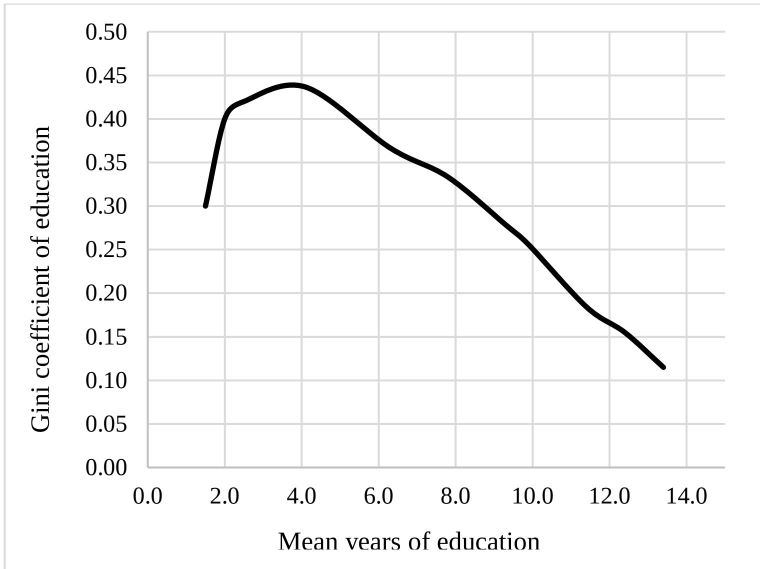
The second distinctive feature of the distribution of educational attainment arises from the presence of an upper limit on the number of years of education. When an individual obtains a doctorate degree, the number of years of education is around 20-22, establishing the maximum limit. Consequently, whether measuring educational inequality by the standard deviation or the Gini coefficient, educational inequality declines during the later stages of educational expansion. If all individuals attain the highest degree, educational inequality is reduced to 0.

We now suppose that individuals without formal education receive some sort of informal education (e.g., family education and/or learning by doing) equivalent to a small amount of formal education. How does educational inequality change with the expansion of education? Table 3 illustrates the changes in the education structure where one year of education is given to individuals without formal education (Example 2). Fig. 6 depicts the Gini coefficient against the mean years of education. In contrast to Example 1 outlined in Table 2, the Gini exhibits an inverted U-shaped pattern in relation to educational expansion. Though not illustrated in Table 3, if all individuals receive no formal education, then the Gini coefficient is 0 under the assumption that they are assigned one year of education. But it rises sharply, and the peak value is attained when mean years of education is around 4. This result is consistent with the result by Coady and Dizioli (2018), in which individuals without formal education are assigned one year to estimate the Gini coefficient.

Table 3. Changes in the Structure of Education (Example 2)

Individual	Time 0	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
1	1	1	1	1	1	1	1	1	6	6	6
2	1	1	1	1	1	1	6	6	6	6	12
3	1	1	1	1	1	6	6	6	6	12	12
4	1	1	1	1	6	6	6	6	12	12	12
5	1	1	1	1	6	6	6	12	12	12	12
6	1	1	1	6	6	6	12	12	12	12	16
7	1	1	1	6	6	12	12	12	12	16	16
8	1	1	6	6	12	12	12	12	16	16	16
9	1	6	6	6	12	12	16	16	16	16	16
10	6	6	6	12	12	16	16	16	16	16	16
Total	15	20	25	41	63	78	93	99	114	124	134
Mean	1.5	2.0	2.5	4.1	6.3	7.8	9.3	9.9	11.4	12.4	13.4
STD	1.50	2.00	2.29	3.53	4.27	4.75	4.73	4.66	3.90	3.67	3.10
Gini	0.30	0.40	0.42	0.44	0.37	0.33	0.28	0.26	0.18	0.15	0.11
CV	1.00	1.00	0.92	0.86	0.68	0.61	0.51	0.47	0.34	0.30	0.23

Figure 6. Evolution of Educational Inequality by the Gini Coefficient (Example 2)



Some researchers use the concept of human capital and measured human capital inequality by the Gini coefficient. In their studies, the number of years of education is converted to human capital, usually, using the following exponential equation (Lim and Tang, 2008; Morrisson and Murtin, 2013).[6]

$$h_i = \exp(re_i), \tag{6}$$

where h_i and r are human capital for education level i and returns to formal education, respectively. Here, for simplicity, returns to formal education is assumed to be constant across education levels. If this equation is used to obtain human capital for each education level, then $h_i > 0$ for $e_i \geq 0$ (i = 0, 1, 2, and 3). In particular, we have $h_0 = 1$ when an individual without formal education is assigned 0 years of education (that is, $e_0 = 0$).

In order to investigate the evolution of human capital inequality, we transform the years of education presented in Table 2 into human capital using Eq. (6) where the returns to formal education is set at 0.1 (Table 4). Fig. 7 presents the evolution of the Gini coefficient of human

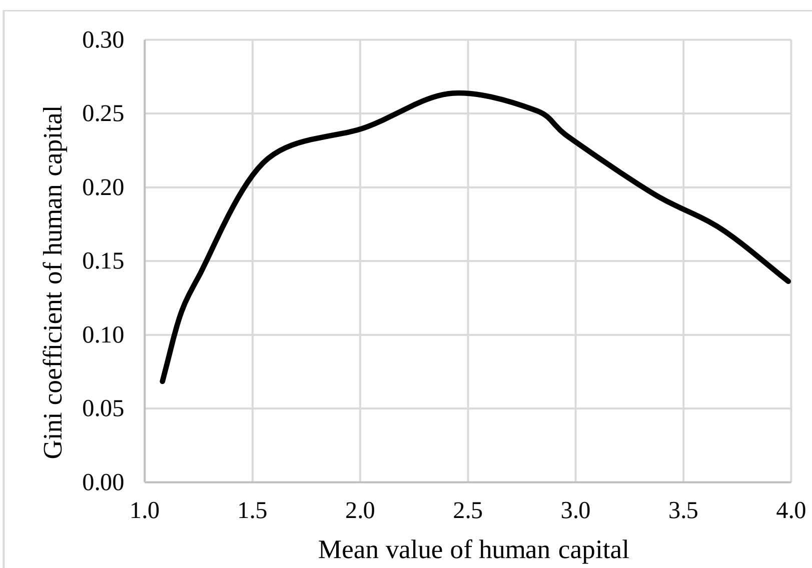
[6] Though Castello and Domenech (2002) and Castello-Climent and Domenech (2021) employed the concept of human capital, they used years of education as a proxy for human capital and assigned 0 to individuals without formal education. Consequently, they observed a downward sloping pattern for the relationship between the expansion of human capital and the Gini coefficient of human capital.

capital with respect to the expansion of human capital. Despite assigning 0 years to individuals without formal education, the Gini coefficient of human capital displays an inverted U-shaped pattern because $h_0 = 1$, which markedly contrasts with the pattern depicted in Fig. 3. This outcome is consistent with the results of Lim and Tang (2008) and Morrisson and Murtin (2013), who found an inverted U-shaped pattern for the relationship between the expansion of human capital and the Gini coefficient of human capital.

Table 4. Changes in the Structure of Human Capital (Constructed Based on Example 1)

Individual	Time 0	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.82	1.82	1.82
2	1.00	1.00	1.00	1.00	1.00	1.00	1.82	1.82	1.82	1.82	3.32
3	1.00	1.00	1.00	1.00	1.00	1.82	1.82	1.82	1.82	3.32	3.32
4	1.00	1.00	1.00	1.00	1.82	1.82	1.82	1.82	3.32	3.32	3.32
5	1.00	1.00	1.00	1.00	1.82	1.82	1.82	3.32	3.32	3.32	3.32
6	1.00	1.00	1.00	1.82	1.82	1.82	3.32	3.32	3.32	3.32	4.95
7	1.00	1.00	1.00	1.82	1.82	3.32	3.32	3.32	3.32	4.95	4.95
8	1.00	1.00	1.82	1.82	3.32	3.32	3.32	3.32	4.95	4.95	4.95
9	1.00	1.82	1.82	1.82	3.32	3.32	4.95	4.95	4.95	4.95	4.95
10	1.82	1.82	1.82	3.32	3.32	4.95	4.95	4.95	4.95	4.95	4.95
Total	10.82	11.64	12.47	15.61	20.25	24.20	28.15	29.65	33.61	36.74	39.87
Mean	1.08	1.16	1.25	1.56	2.02	2.42	2.82	2.97	3.36	3.67	3.99
STD	0.25	0.33	0.38	0.70	0.91	1.20	1.31	1.27	1.21	1.18	1.06
Gini	0.07	0.11	0.14	0.22	0.24	0.26	0.25	0.23	0.20	0.17	0.14
CV	0.23	0.28	0.30	0.45	0.45	0.49	0.47	0.43	0.36	0.32	0.26

Figure 7. Evolution of Human Capital Inequality by the Gini Coefficient (Example 1)



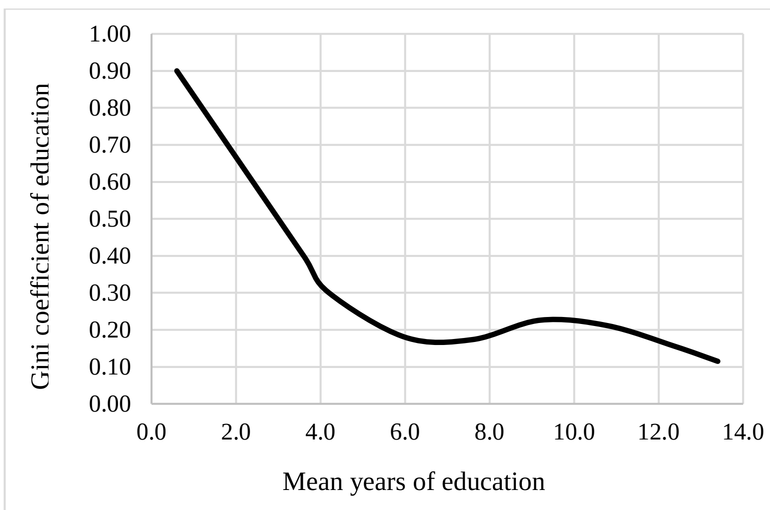
In summary, the results vary significantly depending on how individuals without formal education are treated. If individuals without formal education are assigned 0 years of education, it is likely that the Gini coefficient of education declines monotonically with educational expansion (see Fig. 3). In contrast, if we assume that they receive some sort of informal education equivalent to a small amount of formal education (one year or two years), then the Gini coefficient of education is likely to exhibit an inverted U-shaped pattern with respect to educational expansion (see Fig. 6). If we transform years of education to human capital using an exponential function with positive returns to education, then the Gini coefficient of human capital is likely to exhibit an inverted U-shaped pattern with respect to human capital expansion. This is due to the fact that individuals without formal education possess positive human capital (see Fig. 7).

Let us next consider another example (Example 3), which is presented in Table 5. Fig. 8 shows the evolution of the Gini coefficient of education with educational expansion. From time 0 to 4, there exist only two education levels: no formal education and primary education. Because there is no disparity among individuals with primary education during this period, the Gini coefficient remains identical to the proportion of individuals without formal education (see footnote 4). The Gini declines monotonically during the early stages of educational expansion. But, in contrast to Example 1, it displays a subtle inverted U-shaped pattern after reaching its bottom at time 5 to 6 (mean years of education of 6.0 to 7.6). This implies that if a country, on average, commence with a primary to junior secondary education level, the Gini coefficient of education might display an inverted U-shaped pattern.

Table 5. Changes in the Structure of Education (Example 3)

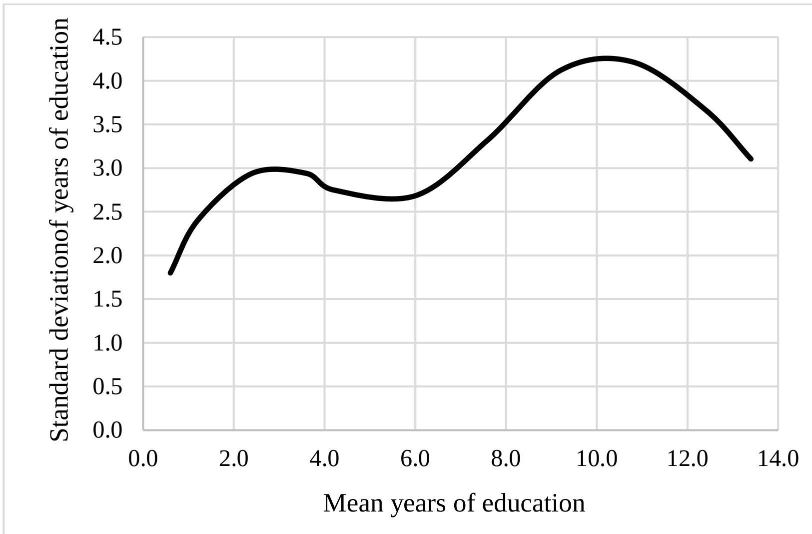
Individual	Time 0	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
1	0	0	0	0	0	0	6	6	6	6	6
2	0	0	0	0	0	6	6	6	6	6	12
3	0	0	0	0	0	6	6	6	6	12	12
4	0	0	0	0	6	6	6	6	6	12	12
5	0	0	0	6	6	6	6	6	12	12	12
6	0	0	0	6	6	6	6	6	12	12	16
7	0	0	6	6	6	6	6	12	12	16	16
8	0	0	6	6	6	6	6	12	16	16	16
9	0	6	6	6	6	6	12	16	16	16	16
10	6	6	6	6	6	12	16	16	16	16	16
Total	6	12	24	36	42	60	76	92	108	124	134
Mean	0.6	1.2	2.4	3.6	4.2	6.0	7.6	9.2	10.8	12.4	13.4
STD	1.80	2.40	2.94	2.94	2.75	2.68	3.32	4.12	4.21	3.67	3.10
Gini	0.90	0.80	0.60	0.40	0.30	0.18	0.17	0.23	0.21	0.15	0.11
CV	3.00	2.00	1.22	0.82	0.65	0.45	0.44	0.45	0.39	0.30	0.23

Figure 8. Evolution of Educational Inequality by the Gini Coefficient (Example 3)



We should note that even when individuals without formal education are given one year of education, the education Gini coefficient shows a similar pattern to the one presented in Fig. 8 after time 5 to 6. On the other hand, Fig. 9 depicts the evolution of the standard deviation of education with the expansion of education. It displays an inverted W-shaped pattern instead of an inverted U-shaped one. But, after time 5 (mean years of education of 6.0), it exhibits an inverted U-shaped pattern. Because the standard deviation does not satisfy income homogeneity, meaning that it is dependent on the mean years of education, the inverted U-shaped pattern is magnified.

Figure 9. Evolution of Educational Inequality by the Standard Deviation (Example 3)



4. Educational Expansion and Educational Inequality: Empirical Evidence

Using a Barro and Lee dataset on educational attainment over the period 1950-2015, this section empirically analyzes the relationship between educational expansion and educational inequality (Barro and Lee, 2013). The dataset provides proportions of the population aged 25-64 with seven levels of education: (0) no formal education, (1) incomplete primary, (2) complete primary, (3) lower secondary, (4) upper secondary, (5) incomplete tertiary, and (6) complete tertiary education. Using Eqs. (3) and (4), the standard deviation and Gini coefficient are estimated by assigning 0, 3, 6, 9, 12, 14 and 16 years for these respective education levels (refer to Appendix 4 for more details).

.

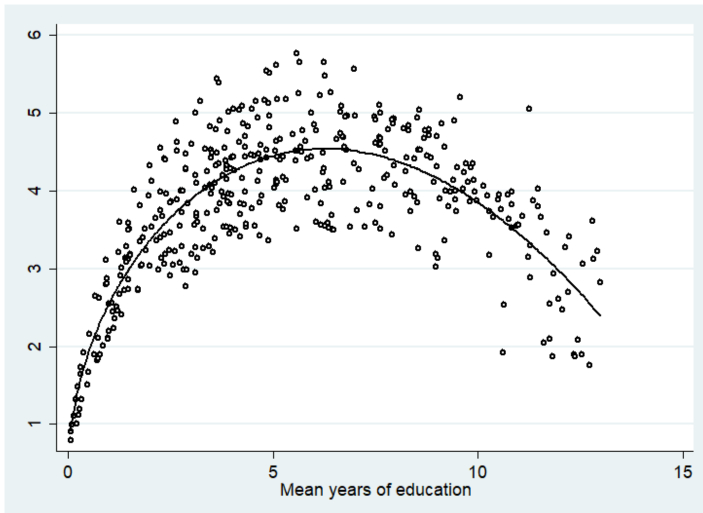
4.1. Empirical Evidence from Asian Countries and Economies

Table 6 presents a list of Asian countries and economies included in the Barro and Lee dataset. Based on the data for these countries and economies, Fig. 10 depicts the evolution of educational inequality by the standard deviation. The standard deviation of education exhibits an inverted U-shaped pattern and attains the peak value when the mean years of education is around 7 years. This inverted U-shaped pattern coincides with the one observed in Example 1 of the previous section (Fig. 2).

Table 6. Asian Countries and Economies in the Barro and Lee Dataset

Region	Countries and economies
South Asia	Afghanistan Bangladesh India Maldives Nepal Pakistan Sri Lanka
Southeast Asia	Brunei Darussalam Cambodia Indonesia Lao People's Democratic Republic Malaysia Myanmar Philippines Singapore Thailand Viet Nam
Central Asia	Kazakhstan Kyrgyzstan Tajikistan
China & Mongolia	China China, Hong Kong Special Administrative region China, Macao Special Administrative Region Mongolia
Pacific	Fiji Papua New Guinea Tonga
Developed Asia	Japan Republic of Korea Taiwan

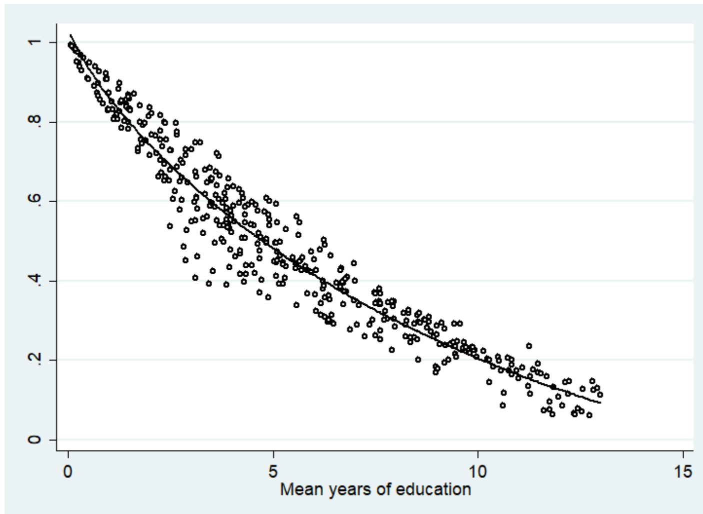
Figure 10. Evolution of Educational Inequality by the Standard Deviation (Asian Countries and Economies)



(Note) Vertical axis: standard deviation of education.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

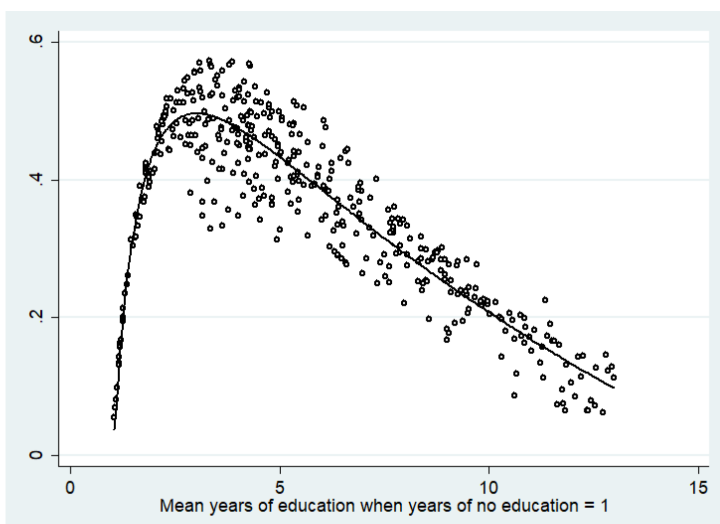
When educational inequality is measured by the Gini coefficient, it displays a downward sloping pattern (Fig. 11). This pattern coincides with the one observed in Example 1 of the previous section (Fig. 3). If individuals without formal education are assigned one year, the Gini coefficient of education shows an inverted U-shaped pattern (Fig. 12), which resembles the pattern observed in Example 2 of the previous section (Fig. 6). However, in contrast to the standard deviation, the peak is reached at a significantly smaller number of years of education (around 3 years).

Figure 11. Evolution of Educational Inequality by the Gini Coefficient (Asian Countries and Economies)



(Note) Vertical axis: education Gini coefficient.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

Figure 12. Evolution of Educational Inequality by the Gini Coefficient (Asian Countries and Economies): Years of No Formal Education = 1

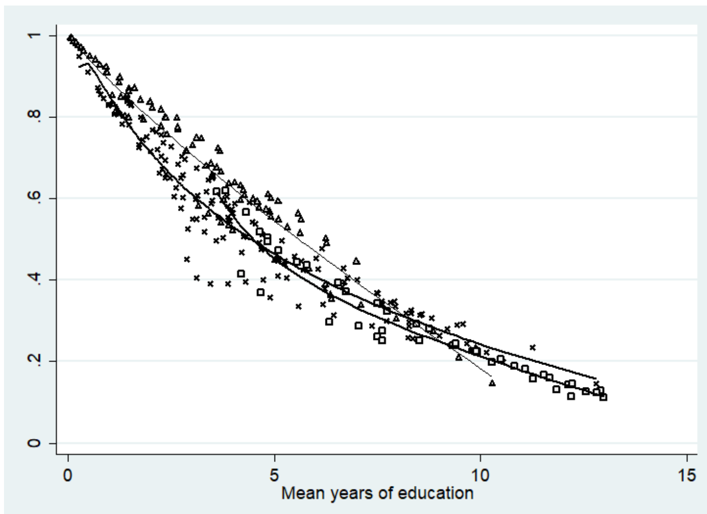


(Note) Vertical axis: education Gini coefficient.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

4.2. A Comparison between Three Asian Sub-regions (South Asia, Southeast Asia and Developed Asia)

It is intriguing to compare between three sub-regions: South Asia, Southeast Asia, and Developed Asia. Fig. 13 shows the Gini coefficient of education for these three sub-regions, which exhibits a downward sloping pattern. But, while South Asia displays a linear trend, Southeast Asia and Developed Asia demonstrate a slight bending pattern.[7]

Figure 13. Evolution of Educational Inequality by the Gini Coefficient: A Comparison between Three Asian Sub-regions



(Note) (1) Δ South Asia, x Southeast Asia, and □ Developed Asia, (2) vertical axis: education Gini coefficient.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

With the exception of Maldives and Sri Lanka, South Asian countries had a very high proportion of individuals without formal education during the early decades (1950s to 1970s); the mean proportion was around 0.5. As a result, the Gini coefficient of education is highly correlated with the proportion of individuals without formal education in South Asia with the correlation coefficient of 0.99.[8] Meanwhile, Developed Asia started with a relatively high level

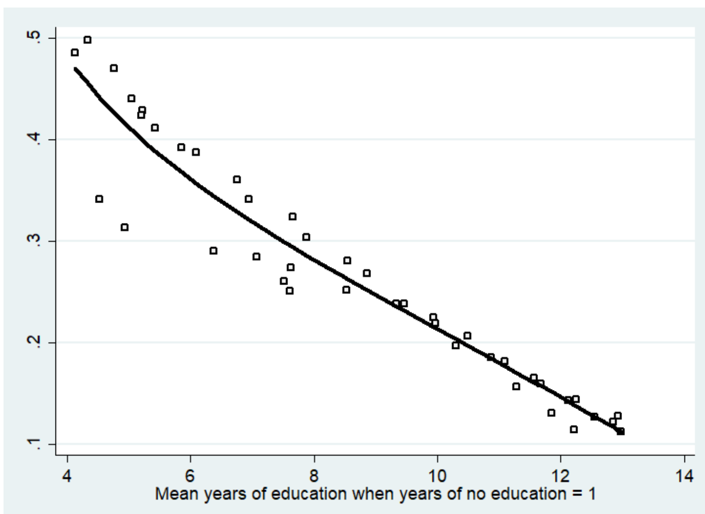
[7] The Gini coefficient for China follows a similar trend pattern to that of South Asia, albeit at a significantly lower magnitude at each level of education.

[8] As we will discuss in Appendix 3, when individuals without formal education are assigned 0 years of education ($e_0 = 0$), the overall education Gini coefficient is given by $p_0 + (1 - p_0)G_1$ where G_1 is the education Gini coefficient among the educated. Because $0.125 \leq G_1 \leq 0.384$ for South Asia, the overall Gini is highly correlated with the proportion of individuals without formal education.

of education in the 1950s with the mean years of education of 4.8. Its Gini coefficient declines gradually with educational expansion, but at a decelerating rate. Southeast Asia shows a similar pattern to Developed Asia though started with a lower level of education.

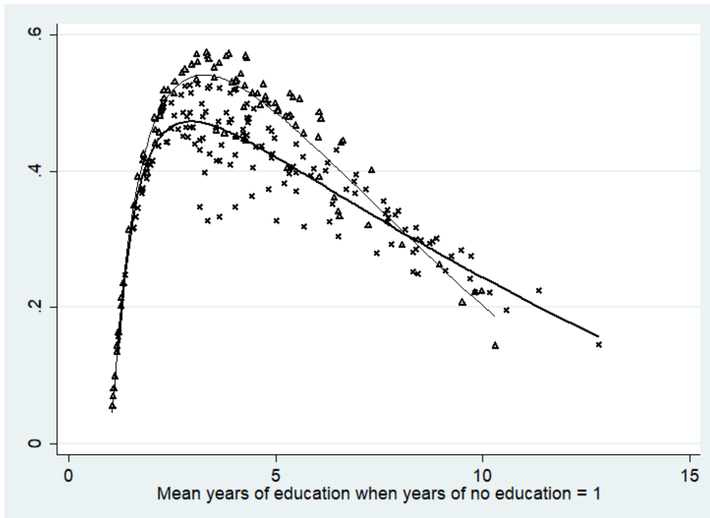
It is important to note that because Developed Asia started with a relatively high level of education, its Gini coefficient shows a declining trend even when a value of one is assigned to individuals without formal education (Fig. 14). This is in sharp contrast to the patterns observed for South Asia and Southeast Asia (Fig. 15).

Figure 14. Evolution of Educational Inequality by the Gini Coefficient (Developed Asia): Years of No Education = 1



(Note) Vertical axis: education Gini coefficient.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

Figure 15. Evolution of Educational Inequality by the Gini Coefficient (South Asia and Southeast Asia): Years of No Education = 1



(Note) (1) Δ South Asia and x Southeast Asia, (2) Vertical axis: education Gini coefficient.
(Source) Barro and Lee dataset (Barro and Lee, 2013)

5. Concluding Remarks

This study explored the relationship between educational expansion and educational inequality through the use of hypothetical examples. It also examined the relationship empirically based on a Barro and Lee dataset on educational attainment for Asian countries and economies.

The major findings are summarized as follows. If individuals without formal education are assigned 0 years of education, the Gini coefficient of education is likely to decline monotonically with educational expansion. In contrast, if we assume that they receive some sort of informal education equivalent to a small amount of formal education, then the education Gini coefficient is likely to exhibit an inverted U-shaped pattern with respect to educational expansion. Transforming years of education into human capital using an exponential function with positive returns to education could lead to the Gini coefficient of human capital exhibiting an inverted U-shaped pattern with respect to human capital expansion. On the other hand, the standard deviation (or variance) of education is likely to display an inverted U-shaped pattern, whether individuals without formal education are assigned 0 years of education or a positive number of years of education.

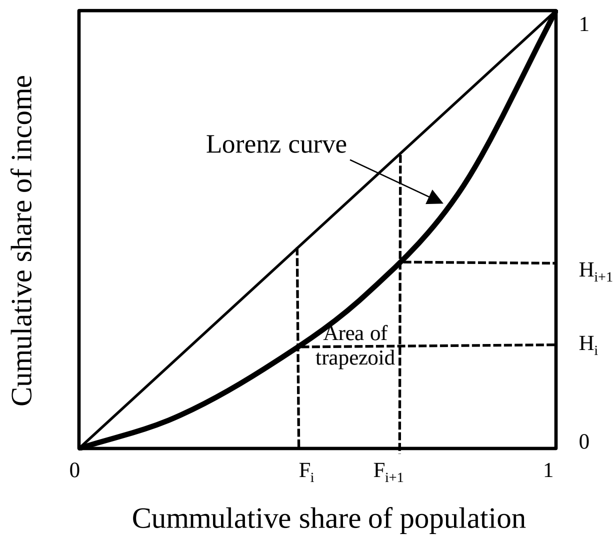
The Barro and Lee dataset for Asian countries and economies reveals that the standard deviation of education follows an inverted U-shaped pattern in relation to educational expansion, even when individuals without formal education are assigned 0 years of education. In contrast, the dataset shows that the Gini coefficient of education demonstrates a downward-sloping pattern when individuals without formal education are assigned 0 years of education. A comparison among South Asia, Southeast Asia, and Developed Asia reveals that while South Asia exhibits a linear downward trend, Southeast Asia and Developed Asia both demonstrate a slight bending pattern. However, when assigning one year to individuals without formal education, the Gini coefficient of education displays an inverted U-shaped pattern, though it attains the peak value at much earlier stages of educational expansion than the standard deviation of education. These empirical observations align with the conclusions drawn from hypothetical examples.

Appendix 1: Definitions of the Gini Coefficient and the Education Gini Coefficient

The Gini coefficient can be defined as the ratio of the area between the diagonal line and the Lorenz curve to the area of the triangle beneath the diagonal line (see Fig. 16). To mathematically establish this definition, consider a country with *n* individuals arranged in non-descending order of their incomes: $0 \le y_1 \le y_2 \le \cdots \le y_n$. Let F_i and H_i be the cumulative shares of population and income up to *i*th individual, respectively, where $F_0 = H_0 = 0$ and $F_n = H_n = 1$. Then, the area of the trapezoid under the Lorenz curve between the *i*th and (i+1)th individuals is given by $\frac{1}{2}(F_{i+1} - F_i)(H_{i+1} + H_i)$. Summing this from $i = 0$ to $i = n - 1$, we obtain the area under the Lorenz curve as follows: $\frac{1}{2}\sum_{i=0}^{n-1}(F_{i+1} - F_i)(H_{i+1} + H_i)$. Therefore, the area between the diagonal line and the Lorenz curve is given by $\frac{1}{2} - \frac{1}{2}\sum_{i=0}^{n-1}(F_{i+1} - F_i)(H_{i+1} + H_i)$. Since the Gini coefficient is the ratio of this area to the area of the triangle under the diagonal line, we obtain

$$G = 1 - \sum_{i=0}^{n-1}(F_{i+1} - F_i)(H_{i+1} + H_i). \qquad (7)$$

Figure 16. Lorenz Curve and Gini Coefficient



By definition, we have the following relationships: $F_i = \frac{i}{n}$ and $H_i = \frac{1}{n\mu}\sum_{j=1}^{i} y_j$, where $\mu = \frac{1}{n}\sum_{i=1}^{n} y_i$. Since we have $F_{i+1} - F_i = \frac{1}{n}$, Eq. (7) can be rewritten as follows.

$$G = 1 - \frac{1}{n}\sum_{i=0}^{n-1}(H_{i+1} + H_i). \tag{8}$$

We will now demonstrate the equivalence of this definition with an alternative definition given in the following proposition.

<u>Proposition 1</u>

The Gini coefficient, as defined by Eq. (7), is equivalent to the following definition.

$$G = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right| = \frac{1}{n^2\mu}\sum_{i=1}^{n}\sum_{j\le i}\left(y_i - y_j\right). \tag{9}$$

Proof

Because we have $\sum_{i=0}^{n-1}(F_{i+1}H_{i+1} - F_iH_i) = F_nH_n - F_0H_0 = 1$, Eq. (7) can be modified as follows,

$$G = \sum_{i=0}^{n-1}(F_iH_{i+1} - F_{i+1}H_i)$$
$$= \sum_{i=1}^{n}(F_{i-1}H_i - F_iH_{i-1})$$

$$= \sum_{i=1}^{n}(F_i(H_i - H_{i-1}) - (F_i - F_{i-1})H_i)$$

$$= \sum_{i=1}^{n}\left(F_i(H_i - H_{i-1}) - \frac{1}{n}H_i\right)$$

$$= \frac{1}{n^2\mu}\sum_{i=1}^{n}(n^2\mu F_i(H_i - H_{i-1}) - n\mu H_i).$$

Since we have $F_i = \frac{i}{n}$ (or $nF_i = i$), $H_i = \frac{1}{n\mu}\sum_{j=1}^{i} y_j$ and $H_i - H_{i-1} = \frac{y_i}{n\mu}$, the equation above is rewritten as follows.

$$G = \frac{1}{n^2\mu}\sum_{i=1}^{n}\left(iy_i - \sum_{j=1}^{i} y_j\right). \tag{10}$$

Since we have $\sum_{i=1}^{n} iy_i = \sum_{i=1}^{n}\sum_{j\le i} y_i$ and $\sum_{i=1}^{n}\sum_{j=1}^{i} y_j = \sum_{i=1}^{n}\sum_{j\le i} y_j$, this equation can be modified to

$$G = \frac{1}{n^2\mu}\sum_{i=1}^{n}\sum_{j\le i}\left(y_i - y_j\right)$$

$$= \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|.$$

Hence, we establish the equivalence between Eq. (7) and Eq. (9).

$$G = 1 - \sum_{i=0}^{n-1}(F_{i+1} - F_i)(H_{i+1} + H_i) = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|.$$

We will next show the equivalence of this definition with another definition presented in the following proposition.

Proposition 2

The Gini coefficient, as defined by Eq. (9), is equivalent to the following definition.

$$\mathrm{G} = \frac{2}{n\mu}cov(i(y_i), y_i) \tag{11}$$

where $i(y_i)$ is the ranking of individuals in terms of their incomes.

Proof

By the definition of covariance, we can modify Eq. (11) as follows.

$$\mathrm{G} = \frac{2}{n\mu}\frac{1}{n}\sum_{i=1}^{n}(i - \bar{\imath})(y_i - \mu)$$

$$= \frac{2}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \mu\sum_{i=1}^{n} i - \bar{\imath}\sum_{i=1}^{n} y_i + \bar{\imath}n\mu\right)$$

$$= \frac{2}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \mu\sum_{i=1}^{n} i\right)$$

$$= \frac{1}{n^2\mu}\left(2\sum_{i=1}^{n} iy_i - \frac{2}{n}\frac{n(n+1)}{2}\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2\mu}\left(2\sum_{i=1}^{n} iy_i - (n+1)\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \sum_{i=1}^{n}(n+1-i)y_i\right)$$

On the other hand Eq. (10) can be modified to

$$G = \frac{1}{n^2\mu}\sum_{i=1}^{n}\left(iy_i - \sum_{j=1}^{i} y_j\right)$$

$$= \frac{1}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \sum_{i=1}^{n}\sum_{j=1}^{i} y_j\right)$$

$$= \frac{1}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \sum_{i=1}^{n}(n+1-i)y_i\right)$$

because we have

$$\sum_{i=1}^{n}\sum_{j=1}^{i} y_j = (y_1) + (y_1 + y_2) + \cdots + (y_1 + y_2 + y_3 + \cdots + y_n) = \sum_{i=1}^{n}(n+1-i)y_i.$$

Therefore, Eq. (9) is equivalent to Eq. (11).

$$G = \frac{1}{n^2\mu}\sum_{i=1}^{n}\sum_{j\le i}\left(y_i - y_j\right) = \frac{2}{n\mu}cov(i(y_i), y_i).$$

From propositions 1 and 2, we conclude that Eq. (7), (9) and (11) are equivalent.

It is important to note that in the case of the distribution of educational attainment, grouped data can be used to calculate the Gini coefficient. Suppose that there are four levels of education: (0) no formal education; (1) primary education; (2) secondary education; and (3) tertiary education. If we let e_i and p_i be the cumulative years of education for *i*th level of education and the proportion of individuals with *i*th level of education, respectively, then the Gini coefficient of education can be defined, based on Eq. (9), by

$$G = \frac{1}{2\mu}\sum_{i=0}^{3}\sum_{j=0}^{3} p_i p_j \left|e_j - e_i\right|,$$

where $\mu = \sum_{i=0}^{3} p_i e_i$ and $\sum_{i=0}^{3} p_i = 1$.

Appendix 2: Educational Expansion and Educational Inequality by the Variance or Standard Deviation

Consider a country with four levels of education: (0) no formal education; (1) primary education; (2) secondary education; and (3) tertiary education. Using e_i and p_i, the variance and standard deviation of educational attainment can be defined, respectively, by

$V = \sum_{i=0}^{3} p_i(e_i - \mu)^2 \geq 0,$

$\sigma = \sqrt{V} = \sqrt{\sum_{i=0}^{3} p_i(e_i - \mu)^2} \geq 0,$

where $\mu = \sum_{i=0}^{3} p_i e_i$ and $\sum_{i=0}^{3} p_i = 1$.

Let V_0 and V_1 be the variance of education when all individuals are included (that is, $V_0 = V$) and the variance of education when individuals without formal education are excluded, respectively. Then, we can obtain the following proposition.

Proposition 3

$$V_0 = p_0(e_0 - \mu)^2 + (1 - p_0)((\mu^* - \mu)^2 + V_1)$$
$$= \frac{p_0}{1-p_0}(e_0 - \mu)^2 + (1 - p_0)V_1 \text{ for } 0 \leq p_0 < 1, \quad (12)$$

$$V_1 = \sum_{i=1}^{3} p_i^*(e_i - \mu^*)^2 = \frac{p_1^*}{1-p_1^*}(e_1 - \mu^*)^2 + (1 - p_1^*)V_2 \text{ for } 0 \leq p_1^* < 1, \quad (13)$$

where

$p_i^* = \frac{p_i}{\sum_{j=1}^{3} p_j} = \frac{p_i}{1-p_0}$ for $0 \leq p_0 < 1$ (i = 1, 2, and 3),

$\mu^* = \sum_{i=1}^{3} p_i^* e_i$, and

$1 = \sum_{i=1}^{3} p_i^*$.

V_2 is the variance when individuals without formal education and individuals with primary education are excluded.

Proof

First, we have

$V_0 = \sum_{i=0}^{3} p_i(e_i - \mu)^2 = p_0(e_0 - \mu)^2 + \sum_{i=1}^{3} p_i(e_i - \mu)^2$

$= p_0(e_0 - \mu)^2 + (1 - p_0)\sum_{i=1}^{3} p_i^*(e_i - \mu)^2$

$= p_0(e_0 - \mu)^2 + (1 - p_0)\sum_{i=1}^{3} p_i^*\big((\mu^* - \mu) + (e_i - \mu^*)\big)^2$

$= p_0(e_0 - \mu)^2 + (1 - p_0)((\mu^* - \mu)^2 + \sum_{i=1}^{3} p_i^*(e_i - \mu^*)^2)$, since we have $\sum_{i=1}^{3} p_i^*(\mu^* - \mu)^2 = (\mu^* - \mu)^2$ and $\sum_{i=1}^{3} p_i^*(\mu^* - \mu)(e_i - \mu^*) = (\mu^* - \mu)\sum_{i=1}^{3} p_i^*(e_i - \mu^*) = 0$.

Next, $(\mu^* - \mu)$ can be modified to

$\mu^* - \mu = \sum_{i=1}^{3} p_i^* e_i - \sum_{i=0}^{3} p_i e_i = -p_0 e_0 + \frac{p_0}{1-p_0}(\mu - p_0 e_0) = \frac{p_0}{1-p_0}(\mu - e_0).$

By substituting this into the equation above, we obtain

$V_0 = p_0(e_0 - \mu)^2 + (1 - p_0)((\mu^* - \mu)^2 + V_1) = \frac{p_0}{1-p_0}(e_0 - \mu)^2 + (1 - p_0)V_1.$

Similarly, we can obtain

$V_1 = \sum_{i=1}^{3} p_i^*(e_i - \mu^*)^2 = \frac{p_1^*}{1-p_1^*}(e_1 - \mu^*)^2 + (1 - p_1^*)V_2.$

When $p_0 = 1$, we have $p_1 = p_2 = p_3 = 0$ and $\mu = e_0$; thus, we have $V_0 = 0$. On the other hand, when $p_0 = 0$, we have $\mu = \mu^*$ and $V_0 = V_1 \geq 0$. Next, we have the following proposition.

Proposition 4

Under the assumption that p_1^*, p_2^* and p_3^* remain constant regardless of the value of p_0, that is, μ^* and V_1 are constant, V_0 rises as μ increases, but after reaching a peak at $\mu = \frac{1}{2}\left((\mu^* + e_0) + \frac{V_1}{\mu^*-e_0}\right)$, V_0 declines as μ increases ($e_0 \leq \mu \leq \mu^*$). When $\mu = e_0$, $V_0 = 0$, while when $\mu = \mu^*$, $V_0 = V_1$. That is, educational inequality, as measured by the variance, exhibits an inverted U-shaped pattern with respect to educational expansion (that is, educational Kuznets curve).

Proof.

Because we have $\frac{p_0}{1-p_0} = \frac{\mu^*-\mu}{\mu-e_0}$ and $1 - p_0 = \frac{\mu-e_0}{\mu^*-e_0}$, Eq. (12) can be rewritten as $V_0 = (\mu^* - \mu)(\mu - e_0) + \left(\frac{\mu-e_0}{\mu^*-e_0}\right)V_1$. Differentiating this equation with respect to μ, we obtain $\frac{\partial V_0}{\partial \mu} = \left((\mu^* - \mu) - (\mu - e_0)\right) + \frac{V_1}{\mu^*-e_0}$ under the assumption that μ^* and V_1 remain constant regardless of the value of p_0. When $\mu = e_0$, $\left.\frac{\partial V_0}{\partial \mu}\right|_{\mu=e_0} = (\mu^* - e_0) + \frac{V_1}{\mu^*-e_0} > 0$. When $= \mu^*$, $\left.\frac{\partial V_0}{\partial \mu}\right|_{\mu=\mu^*} = -(\mu^* - e_0) + \frac{V_1}{\mu^*-e_0} = \frac{-(\mu^*-e_0)^2+V_1}{\mu^*-e_0} < 0$, because we have $-(\mu^* - e_0)^2 + V_1 = \sum_{i=1}^{3} p_i^*((\mu^* - e_i)^2 - (\mu^* - e_0)^2) < 0$ for $0 \leq e_0 < e_1 < e_2 < e_3$. Since $\frac{\partial^2 V_0}{\partial \mu^2} = -2 < 0$, V_0 takes the maximum at $\mu = \frac{1}{2}\left((\mu^* + e_0) + \frac{V_1}{\mu^*-e_0}\right)$.

Since the standard deviation of educational attainment is a monotonic transformation of the variance, it follows the same pattern as the variance.

Appendix 3: Educational Expansion and Educational Inequality by the Gini Coefficient

Consider a country with four levels of education: (0) no formal education; (1) primary education; (2) secondary education; and (3) tertiary education. Using e_i and p_i, the Gini coefficient of education is defined by

$$G = \frac{1}{2\mu}\sum_{i=0}^{3}\sum_{j=0}^{3} p_i p_j \left|e_j - e_i\right| = \frac{1}{\mu}\sum_{i=0}^{3}\sum_{j>i}^{3} p_i p_j \left|e_j - e_i\right| \quad (0 < \mu).$$

where $\mu = \sum_{i=0}^{3} p_i e_i$ and $\sum_{i=0}^{3} p_i = 1$. Since we have $0 \le e_0 < e_1 < e_2 < e_3$, the Gini coefficient can be rewritten as follows.

$$G = \frac{1}{\mu}\sum_{i=0}^{3}\sum_{j>i}^{3} p_i p_j \left(e_j - e_i\right). \tag{14}$$

Let G_0 and G_1 be the Gini coefficient of education when all individuals are included and the Gini coefficient of education when individuals without formal education are excluded, respectively (that is, $G_0 = G$). Then, we can obtain the following proposition.

<u>Proposition 5</u>

$$G_0 = p_0\left(1 - \frac{e_0}{\mu}\right) + (1 - p_0)\left(1 - \frac{p_0 e_0}{\mu}\right)G_1 \quad (0 < \mu), \tag{15}$$

$$\begin{aligned} G_1 &= \frac{1}{\mu^*}\sum_{i=1}^{3}\sum_{j>i}^{3} p_i^* p_j^* \left(e_j - e_i\right) \\ &= p_1^*\left(1 - \frac{e_1}{\mu^*}\right) + (1 - p_1^*)\left(1 - \frac{p_1^* e_1}{\mu^*}\right)G_2 \quad (0 < \mu^*), \end{aligned} \tag{16}$$

where

$p_i^* = \frac{p_i}{\sum_{j=1}^{3} p_j} = \frac{p_i}{1-p_0}$ for $0 \le p_0 < 1$ (i = 1, 2, and 3),

$\mu^* = \sum_{i=1}^{3} p_i^* e_i$, and

$1 = \sum_{i=1}^{3} p_i^*$.

G_2 is the Gini coefficient when individuals without formal education and individuals with primary education are excluded.

Proof

By definition, we have

$\mu G_0 = [p_0 p_1(e_1 - e_0) + p_0 p_2(e_2 - e_0) + p_0 p_3(e_3 - e_0)] + [p_1 p_2(e_2 - e_1) + p_1 p_3(e_3 - e_1) + p_2 p_3(e_3 - e_2)]$

$= [p_0(\mu - p_0 e_0) - p_0 e_0(1 - p_0)] + (1 - p_0)^2[p_1^* p_2^*(e_2 - e_1) + p_1^* p_3^*(e_3 - e_1) + p_2^* p_3^*(e_3 - e_2)]$

$= p_0(\mu - e_0) + (1 - p_0)^2 \mu^* G_1$

$= p_0(\mu - e_0) + (1 - p_0)(\mu - p_0 e_0)G_1,$

because we have $(1 - p_0)\mu^* = \mu - p_0 e_0$. Dividing both sides of this equation by μ, we obtain

$G_0 = p_0\left(1 - \frac{e_0}{\mu}\right) + (1 - p_0)\left(1 - \frac{p_0 e_0}{\mu}\right)G_1.$

Similarly, we can obtain

$G_1 = p_1^*\left(1 - \frac{e_1}{\mu^*}\right) + (1 - p_1^*)\left(1 - \frac{p_1^* e_1}{\mu^*}\right)G_2$

When individuals without formal education are assigned 0 years of education ($e_0 = 0$), Eq. (15) is reduced to

$$G_0 = p_0 + (1 - p_0)G_1. \tag{17}$$

In other words, G_0 is a weighted average of 1 and G_1. If $G_1 = 0$, that is, there is no educational inequality among the educated, then $G_0 = p_0$, meaning that the educational Gini coefficient coincides with the proportion of individuals without formal education.

The following proposition presents the evolution of G_0 with respect to μ when $e_0 = 0$.

Proposition 6

Under the assumption that $e_0 = 0$ and that p_1^*, p_2^* and p_3^* remain constant regardless of the value of p_0, that is, μ^* and G_1 are constant, G_0 declines monotonically as μ increases and reaches G_1 when $\mu = \mu^*$. When $\mu = 0$ ($p_0 = 1$), G_0 is undefined; thus, we have $1 > G_0 \geq G_1 \geq 0$ for $0 < \mu \leq \mu^*$.

Proof

Since $e_0 = 0$, we have $\mu = \sum_{i=1}^{3} p_i e_i = (1 - p_0)\mu^*$, where μ^* is constant by assumption. Therefore, we have $p_0 = 1 - \frac{\mu}{\mu^*}$. Substituting this into Eq. (17), we obtain $G_0 = \left(1 - \frac{\mu}{\mu^*}\right) +$

$\frac{\mu}{\mu^*}G_1$. Differentiating this equation with respect to μ, we obtain $\frac{\partial G_0}{\partial \mu} = \frac{1}{\mu^*}(G_1 - 1) < 0$ if $1 > G_1 \geq 0$, meaning that G_0 declines monotonically as μ increases and reaches G_1 when $\mu = \mu^*$.

This proposition suggests that if the education Gini coefficient among the educated individuals remains constant regardless of the value of the proportion of individuals without formal education, then the education Gini coefficient for all individuals declines monotonically with the expansion of education.

Now, suppose that individuals without formal education receive some informal education equivalent to α years of formal education where $e_1 > e_0 = \alpha > 0$. Then, when $p_0 = 1$, we have $\mu = \alpha$ and $G_0 = 0$. On the other hand, when $p_0 = 0$, we have $\mu = \mu^*$ and $G_0 = G_1 \geq 0$. The following proposition presents the evolution of G_0 with respect to μ when $e_0 = \alpha > 0$.

Proposition 7

Under the assumptions that p_1^*, p_2^* and p_3^* remain constant regardless of the value of p_0, that is, μ^* and G_1 are constant and that $\mu^* > e_0 = \alpha > 0$ and $G_1 < \frac{\mu^* - e_0}{\mu^* + e_0}$, G_0 rises as μ increases, but after reaching a peak, G_0 declines as μ increases ($0 < e_0 \leq \mu \leq \mu^*$). When $\mu = e_0 = \alpha > 0$, $G_0 = 0$, while when $\mu = \mu^*$, $G_0 = G_1 \geq 0$. That is, educational inequality, as measured by the Gini coefficient, exhibits an inverted U-shaped pattern with respect to educational expansion (that is, educational Kuznets curve).

Proof

Because we have $\mu - e_0 = (1 - p_0)(\mu^* - e_0)$ and $\mu - p_0 e_0 = (1 - p_0)\mu^*$, Eq. (15) is rewritten as follows

$$G_0 = \frac{1}{\mu}\left[\mu^* G_1 + \big((\mu^* - e_0) - 2\mu^* G_1\big)p_0 + \big(\mu^* G_1 - (\mu^* - e_0)\big)p_0^2\right],$$

where $p_0 = \frac{\mu^* - \mu}{\mu^* - e_0}$. If we define $f(\mu)$ as follows

$$f(\mu) = \mu^* G_1 + \big((\mu^* - e_0) - 2\mu^* G_1\big)p_0 + \big(\mu^* G_1 - (\mu^* - e_0)\big)p_0^2,$$

then we have $G_0 = \frac{1}{\mu} f(\mu)$. Differentiating this equation with respect to μ, we obtain

$$\frac{dG_0}{d\mu} = \frac{1}{\mu^2}\left(\mu \frac{df(\mu)}{d\mu} - f(\mu)\right)$$

where $\mu \frac{df(\mu)}{d\mu} = \left(\frac{\mu}{\mu^*-e_0}\right)\left[\left(2\mu^* G_1 - (\mu^* - e_0)\right) + 2\left((\mu^* - e_0) - \mu^* G_1\right)\left(\frac{\mu^*-\mu}{\mu^*-e_0}\right)\right]$.

When $\mu = e_0 = \alpha > 0$ (that is, $p_0 = 1$), we have $\left.\frac{\partial G_0}{\partial \mu}\right|_{\mu=e_0} = \frac{1}{e_0} > 0$ because $e_0 = \alpha > 0$. On the other hand, when $\mu = \mu^*$ (that is, $p_0 = 0$), we have $\left.\frac{\partial G_0}{\partial \mu}\right|_{\mu=\mu^*} = \frac{G_1(\mu^*+e_0)-(\mu^*-e_0)}{\mu^*(\mu^*-e_0)}$. Therefore, if $G_1 < \frac{\mu^*-e_0}{\mu^*+e_0}$, then $\left.\frac{\partial G_0}{\partial \mu}\right|_{\mu=\mu^*} < 0$. This implies that if $G_1 < \frac{\mu^*-e_0}{\mu^*+e_0}$ holds true, then G_0 rises as μ increases, but after reaching a peak, G_0 declines as μ increases ($0 < e_0 \leq \mu \leq \mu^*$). When $\mu = e_0 = \alpha > 0$, $G_0 = 0$, while when $\mu = \mu^*$, $G_0 = G_1 \geq 0$.

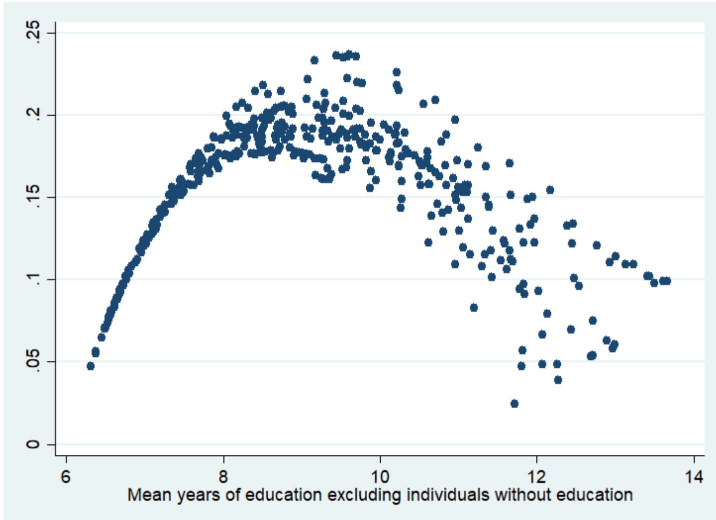
This proposition implies that even if educational inequality is measured by the Gini coefficient, educational inequality could exhibit an inverted U-shaped pattern with respect to educational expansion when individuals without formal education receive informal education equivalent to a small amount of formal education. Because the slope of the G_0 function at $\mu = e_0 = \alpha > 0$ (that is, $p_0 = 1$) is given by $\frac{1}{e_0}$, G_0 experiences a sharp increase during the initial stages of educational expansion when individuals without formal education are assigned a small number of years of education.

We should note that according to the Barro and Lee dataset for Asia, there is an inverted U-shaped relationship between G_1 and μ^* (refer to Fig. 17), where G_1 is estimated by assigning 6, 12, and 16 years for primary, secondary and tertiary education levels, respectively.

$$\mu^* G_1 = p_1^* p_2^*(12 - 6) + p_1^* p_3^*(16 - 6) + p_2^* p_3^*(16 - 12),$$

where $\mu^* = 6p_1^* + 12p_2^* + 16p_3^*$. G_1 varies between 0.02 and 0.24 as μ^* increases from 6.3 to 13.7. Suppose that individuals without formal education are assigned one year as their education (e_0 =1). Then, $\frac{\mu^*-e_0}{\mu^*+e_0}$ increases from 0.73 to 0.86 as μ^* rises from 6.3 to 13.7. Therefore, $G_1 < \frac{\mu^*-e_0}{\mu^*+e_0}$ is always satisfied in Asia.

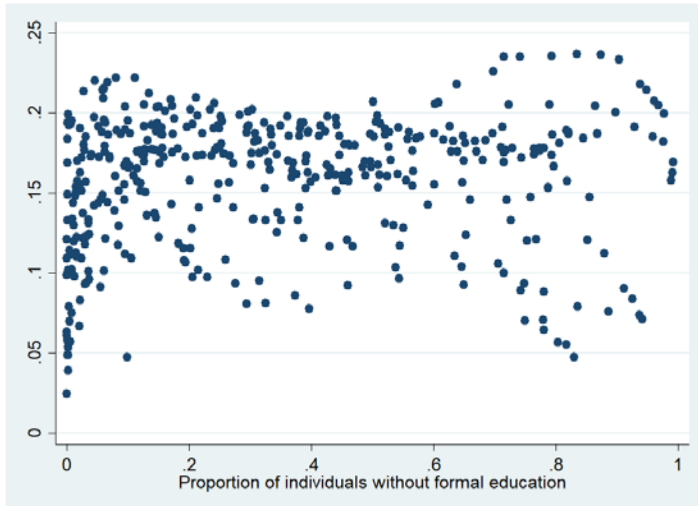
Figure 17. Evolution of G_1 with respect to μ^*



(Note) Vertical axis: education Gini when individuals without formal education excluded (G_1)
(Source) Barro and Lee dataset (Barro and Lee, 2013)

The question arises how G_1 changes as p_0 increases from 0 to 1. Based on the Barro and Lee data set for Asia, Fig. 18 presents a scatter plot for the relationship between G_1 and p_0. G_1 varies between 0.02 and 0.24 as p_0 increases from 0 to 1. Correlation coefficient between these two variables is 0.13. While G_1 is positively correlated with p_0 to some extent, its variation is not large as compared to the variation of p_0, thereby supporting the validity of Propositions 6 and 7.

Figure 18. Relationship between G_1 and p_0



(Note) Vertical axis: education Gini when individuals without formal education excluded (G_1)
(Source) Barro and Lee dataset (Barro and Lee, 2013)

Appendix 4: Estimating the Variance and Gini Coefficient of Education using the Barro and Lee Dataset

Consider a country with seven levels of education: (0) no formal education, (1) incomplete primary, (2) complete primary, (3) lower secondary, (4) upper secondary, (5) incomplete tertiary, and (6) complete tertiary education. Then, using e_i and p_i, the variance is defined by

$$V = \sum_{i=0}^{6} p_i (e_i - \mu)^2,$$

where $\mu = \sum_{i=0}^{6} p_i e_i$ and $\sum_{i=0}^{6} p_i = 1$. By using the Barro and Lee data on the proportion of individuals with *i*th level of education (p_i), the variance of education can be calculated for each country or economy as follows, where 0, 3, 6, 9, 12, 14, and 16 years are assigned for the seven education levels described above.

$$V = p_0(0-\mu)^2 + p_1(3-\mu)^2 + p_2(6-\mu)^2 + p_3(9-\mu)^2 + p_4(12-\mu)^2$$
$$+ p_5(14-\mu)^2 + p_6(16-\mu)^2$$

where $\mu = 0p_0 + 3p_1 + 6p_2 + 9p_3 + 12p_4 + 14p_5 + 16p_6$.

On the other hand, the Gini coefficient of education is defined, based on Eq. (9), by

$$G = \frac{1}{2\mu} \sum_{i=0}^{6} \sum_{j=0}^{6} p_i p_j \left| e_j - e_i \right|,$$

Since we have $0 \leq e_0 < e_1 < e_2 < \cdots < e_6$, the education Gini coefficient can be rewritten as follows.

$$G = \frac{1}{\mu} \sum_{i=0}^{6} \sum_{j>i}^{6} p_i p_j \left(e_j - e_i\right).$$

By using the Barro and Lee data on p_i, the Gini coefficient of education can be calculated as follows, where 0, 3, 6, 9, 12, 14, and 16 years are assigned for the seven education levels described above.

$$\begin{aligned}\mu G = {} & p_0 p_1 (3) + p_0 p_2 (6) + p_0 p_3 (9) + p_0 p_4 (12) + p_0 p_5 (14) + p_0 p_6 (16) \\ & + p_1 p_2 (6-3) + p_1 p_3 (9-3) + p_1 p_4 (12-3) + p_1 p_5 (14-3) + p_1 p_6 (16-3) \\ & + p_2 p_3 (9-6) + p_2 p_4 (12-6) + p_2 p_5 (14-6) + p_2 p_6 (16-6) \\ & + p_3 p_4 (12-9) + p_3 p_5 (14-9) + p_3 p_6 (16-9) \\ & + p_4 p_5 (14-12) + p_4 p_6 (16-12) + p_5 p_6 (16-14),\end{aligned}$$

References

Agrawal, T. (2014). Educational inequality in rural and urban India. *International Journal of Educational Development*. 34, 11-19.

Akita, T., & Kataoka, M. (2022). *Regional Inequality and Development: Measurement and Applications in Indonesia*, Springer.

Almeida, D. R. C., Andrade, J. A. S., Duarte, A., & Simões, M. (2022). Human capital disparities and earnings inequality in the Portuguese private labour market. Social Indicators Research, 159, 145–167.

Anand, S. (1983). *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.

Banzragch, O., Mizunoya, S., & Bayarjargal, M. (2019). Education inequality in Mongolia: measurement and causes. *International Journal of Educational Development*, 68, 68-79.

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.

Castello, A., & Domenech, R. (2002). Human capital inequality and economic growth: Some new evidence. *The Economic Journal*, 112, C187-C200.

Castello-Climent, A., & Domenech, R. (2021). Human capital and income inequality revisited. *Education Economics*, 29(2), 194-212.

Coady, D., & Dizioli, A. (2018). Income inequality and education revisited: Persistence, endogeneity and heterogeneity. *Applied Economics*, 50(25), 2747–2761.

De Gregorio, J. D., & Lee, J. W. (2002). Education and income inequality: New evidence from cross-country dData. *Review of Income and Wealth* 48(3), 395-416.

Fields, G. S. (2001). *Distribution and Development: A New Look at the Developing World*, The MIT Press: Cambridge, MA.

Fordvari, P., & van Leeuwen, B. (2011). Should less inequality in education lead to a more equal income distribution? *Education Economics*, 19(5), 537–554.

Hojo, M. (2009). Inequality in Japanese education: Estimation using the Gini education coefficient. *Japanese Economy*. 36(3), 3-27.

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

Lam, D., & Levison, D. (1992). Declining inequality in schooling in Brazil and its effects on inequality in earnings. *Journal of Development Economics*, 37, 199-225.

Lee, J.W. & Lee, H. (2018). Human capital and income inequality. *Journal of the Asia Pacific Economy*, 23(4), 554-583.

Lim, A. S, K., & Tang, K. K. (2008). Human capital inequality and the Kuznets curve. *The Developing Economies*. 46(1), 26-51.

Lin, C. H. A. (2007). Education expansion, educational inequality, and income inequality: Evidence from Taiwan, 1976–2003. *Social Indicators Research*, 80, 601–615.

Lin, C. H. A., & Yang, C. H. (2009). An analysis of educational inequality in Taiwan after the higher education expansion. *Social Indicators Research*, 90, 295–305.

Luo, G., Zeng, S., & Baležentis, T. (2022). Multidimensional measurement and comparison of China's educational inequality. *Social Indicators Research*, 163, 857–874.

Lorenz, M. O. (1905). Methods of measuring concentration of wealth. *Journal of the American Statistical Association*, 9, 209-219.

Meschi, E., & Scervini, F. (2014). Expansion of schooling and educational inequality in Europe: the educational Kuznets curve revisited. *Oxford Economic Papers*, 66(3), 660–680.

Morrisson, C., & Murtin, F. (2013). The Kuznets curve of human capital inequality: 1870–2010. *The Journal of Economic Inequality,* 11, 283–301.

Park, K. H. (1996). Educational expansion and educational inequality on income distribution. *Economics of Education Review,* 15 (1), 51–58.

Park, K. H. (2017). Education, globalization, and income inequality in Asia. ADBI Working Paper, 732.

Ram, R. (1989). Can educational expansion reduce income inequality in less-developed countries? *Economics of Education Review*, 8: 185–9.

Ram, R. (1990). Educational expansion and schooling inequality: International evidence and some implications. *The Review of Economics and Statistics*, 72(2), 266-274.

Shukla, V., & Mishra, U. S. (2019). Educational expansion and schooling Inequality: Testing educational Kuznets Curve for India. *Social Indicators Research*, 141, 1265–1283.

Thomas, V., Wang, Y., & Fan, X. (2001). Measuring educational inequality: Gini coefficient of education. Policy Research Working Paper, 2525, The World Bank.

Yang, J., Huang, X., & Liu, X. (2014). An analysis of education inequality in China. *International Journal of Educational Development*, 37, 2010.