**Replication and Extension of Angrist, Joshua D. and Alan B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics* 106, no. 4 (November 1991): 979-1014.**

Alejandro Lapuente Tenorio

Impact Evaluation Methods

August 2023

**Abstract**

This paper replicates the results estimated in 1991 by Joshua D. Angrist and Alan B. Krueger in their work “Does compulsory school attendance affect schooling and earnings?” Additionally, this paper extends their work by implementing metropolitan differentiation and finds that the effect of compulsory schooling laws doesn’t vary significantly across urban or rural areas.

Estimating the return to education is very problematic due to the unobserved factors that participate in the determination of earnings. Labor economists have been pioneers in the use of different econometric techniques to obtain an unbiased and consistent estimate of this causal relationship. Among their works, a very important and famous study was published by Joshua D. Angrist and Alan B. Krueger in 1991 called “Does compulsory school attendance affect schooling and earnings?”[[1]](#footnote-1) In this paper, the authors address this problematic relationship by using quarter of birth as an instrument for years of education in the estimation of the return to education. They’re interested in the causal effect that compulsory schooling laws[[2]](#footnote-2) have in years of education and how these additional years of education translate into higher earnings for the individuals. Their findings are that, in their own words, “compulsory schooling laws are effective in compelling some students to attend school” and these students “earn higher wages as a result of their extra schooling.”[[3]](#footnote-3)

In summary, the causal relationship of interest in Angrist and Krueger’s work is the effect of compulsory schooling laws on earnings. This effect is carried out through the additional education generated by these laws. Which implies that, apart from the usual problems that come with estimating the return to education, in this case, the treatment itself is unobservable. In this context, the authors introduce the quarter of birth both as an instrumental variable for years of education and as a proxy for compulsory schooling laws.

Once this idea has been presented, it’s needed to demonstrate that season of birth is a good instrument for years of education. A first step is fulfilled because of the undeniable randomness associated with the season of birth. However, a good instrument also needs to be relevant and valid. To achieve relevance, there must be a relationship between the season of birth and years of education (this is between the instrument and the independent variable). The authors show this in Figures 1, 2 and 3. As remarked by the black squares, being born early in the year is associated with completing fewer years of education.

A graph showing the growth of a number

Description automatically generated with medium confidence

Figure 1: Years of Education and Season of Birth (1970 Census).

Note: quarter of birth is listed above each observation.

A graph showing the growth of a baby

Description automatically generated

Figure 2: Years of Education and Season of Birth (1980 Census).

Note: quarter of birth is listed above each observation, the first quarter of the year is marked with black.

A graph of a line with numbers and points

Description automatically generated with medium confidence

Figure 3: Years of Education and Season of Birth (1980 Census).

Note: quarter of birth is listed above each observation, the first quarter of the year is marked with black.

Now, for the instrument to be valid, season of birth must affect earnings only through years of education (i.e., there aren’t any other pathways through which the season a person is born increases or decreases their future earnings). The authors explain why this is the case attacking different theories that link non-education aspects of season of birth with earnings, however, the most convincing argument towards validity is that there’s no relationship between season of birth and earnings for college graduates. Meaning that, if the relationship is nonexistent for individuals whose schooling wasn’t prolonged due to compulsory schooling laws, then the observed variation in earnings from season of birth must be through years of education caused by compulsory schooling laws. The variation of earnings within season of birth can be seen in Figure 4.

Therefore, their identification strategy consists in exploiting the variation generated by season of birth on years of education to obtain an unbiased estimate of the return to education. Furthermore, as this variation is assumed to be a consequence of compulsory schooling laws, the return to education estimated is interpreted as the causal effect of compulsory schooling laws on future earnings.

A graph of growth and growth

Description automatically generated with medium confidence

Figure 4: Log of Weekly Earning and Season of Birth (1970 and 1980 Censuses).

Note: quarter of birth is listed above each observation, the first quarter of the year is marked with black.

The assumption that the variation of years of education within season of birth is generated by compulsory schooling laws is founded on the following reasoning. Schools in the United States require children to have turned the age of 6 by the end of the calendar year for beginning their education around the month of August. This requirement leads individuals born early in the year to enter school at an older age than individuals born late in the year. Compulsory schooling laws come into scene when the children grow to the school leaving age given by the law and decide to drop out. Given that individuals born later in the year had started their education earlier in their lives, when dropping out after turning the legal age to do so, they’ve completed more years of education than individuals going through the same process but having been born early in the year. This is the underlying assumption that explains the variation displayed in Figures 1, 2 and 3.

In addition to the empirical research that motivates their research, the authors also compare the unbiased results they obtain with the results coming from a simple OLS regression of earnings on education. This comparison is done with the intention of measuring the bias that unobservable determinants have on the estimation of the return to education. They find that the difference isn’t statistically significant and conclude the following: “This evidence casts doubt on the importance of omitted variables bias in OLS estimates of the return to education, at least for years of schooling around the compulsory schooling level.”[[4]](#footnote-4)

To extend the analysis done in their paper, I propose looking at the difference in the return to education for men living in rural areas compared with the return to education for the general men population. For this analysis, I take rural areas as non-SMSA’s (Standard Metropolitan Statistical Areas). As defined by the United States Census Bureau, SMSA’s were constituted by counties located in urban areas.[[5]](#footnote-5) This concept was used to assure that suburbs of central cities weren’t seen as non-urban areas when recollecting statistics. Therefore, men living in rural areas were classified as non-SMSA’s residents. Using this metric, I find that men from these areas don’t get a return to education statistically different than the average male population.

But, before getting into more detail about this extension, it’s important to dive deeper into Angrist and Krueger’s work. This is a study motivated by the efforts in improving public policy related to education. The importance of estimating the return to compulsory schooling can be examined from two main perspectives: first, it provides policymakers with more accurate information about the consequences of compulsory schooling laws, and, second, it can be seen as a way of estimating the return to education in a more general manner. The latter is of particular interest because the identification strategy proposed in this work results in an estimate clean of unobservable determinants bias.

To conduct this analysis, the authors use data from the 1960, 1970 and 1980 decennial Censuses of the United States. The most important variables they use are date of birth (season/quarter), years of education and weekly earnings. Other information is also used for different sets of controls that will be shown later in this work. Although the authors use, at the beginning of the paper, information from the 1960 Census for demonstrating the relationship between compulsory schooling laws and years of education, the econometric analysis that leads to the results is performed using only data from the 1970 and 1980 Censuses. They look at men born in the decades of 1920 (information from 1970 Census), 1930 and 1940 (information from 1980 Census) when they’re around 40 to 50 years old to assure they’ve completed their studies and have consolidated in the job market.

With this data, Angrist and Krueger perform a Two Stage Least Squares estimation constructed in the following way:

A close-up of a math symbol

Description automatically generated

where denotes the years of completed education of the individual , is a vector of covariates, indicates if the individual was born in quarter , and indicates if the individual was born in year of the decade. Here, the coefficient is the return to education.[[6]](#footnote-6)

This model was estimated for each of the three cohorts: men born between the years 1920-1929, 1930-1939 and 1940-1949. The results for men with years of education around the compulsory schooling level are displayed in Table 1 and can be interpreted as follows. Men born between 1920 and 1929 earn around 10% more per completed year of education when they’re 40 to 49 years old. The estimation for individuals born between 1930 and 1939 is around a 6% increase around the same ages. And, finally, for the 1940 to 1949 cohort, a 7.8% wage increase is estimated at ages 30 to 39. Therefore, it’s concluded that compulsory schooling laws are effective in increasing years of education and future earnings.

In terms of the difference between the TSLS and the OLS estimates displayed in Tables 1 and 2 Angrist and Krueger note that there is only a small difference that suggests a downward bias in the OLS estimate. They attribute it to omitted variables or measurement errors. However, this difference is not statistically significant. These results serve as evidence that suggests that the omitted variables bias is not significant when estimating the return to education. It’s worth noting that the replication of the study leads to the same results for both OLS and TSLS estimates.[[7]](#footnote-7)

A table of numbers and a number of numbers

Description automatically generated with medium confidence

Table 1: TSLS estimates.

Significance codes: 0.01 \*\*\*, 0.05 \*\* and 0.1 \*.

Note: standard errors are in parenthesis, each equation also includes an intercept and further controls.

A table of numbers and a number of numbers

Description automatically generated with medium confidence

Table 2: OLS estimates.

Significance codes: 0.01 \*\*\*, 0.05 \*\* and 0.1 \*.

Note: standard errors are in parenthesis, each equation also includes an intercept and further controls.

To take the analysis even further, Angrist and Krueger explore two more paths. The first is allowing the seasonal pattern in education to vary by state of birth. The motivation behind the deepening is that compulsory schooling laws and school start age are determined at the state level. This generates additional variation that leads to more precise results. Looking exclusively at the 1930-1939 cohort and using data from the 1980 Census, the authors describe their methodology and results as follows: “freeing up the instruments by state of birth and including 50 state-of-birth dummies in the wage equation results in approximately a 40 percent reduction in the standard errors of the TSLS estimates.”[[8]](#footnote-8)

Then, for the second deepening path, they look at racial differentiation in the return to education. From Table 1, it is visible that black men earn lower wages than the average male population, however, the relationship between these lower wages and the individual’s level of education needs to be estimated independently. When conducting this estimation for the cohort of black men born between the years of 1930 to 1939 the results show that the return to education is also lower for black men (around 3.9%[[9]](#footnote-9) ) than for the average male population (around 6%, as noted before). [[10]](#footnote-10)

Inspired by this final analysis made by Angrist and Krueger, and as mentioned before in this paper, I performed an extension study looking into the differences in the return to education within geographical area of residence. The motivation behind it comes from the idea that the effect of compulsory schooling laws and the return to education could vary depending on geographical characteristics. This, with the expectation that the context in which an individual develops should play a role in determining their education and future earnings (e. g. one could expect a smaller return to education for men living in rural areas because of their access to job opportunities that don’t require advanced technical specialization).

Table 3 presents TSLS estimates for men living in rural areas. There are two main things to note. First, compared to the results from Table 1, the direction of the change is inconsistent through cohorts. And second, the differences between Tables 1 and 3 estimates of the return to education aren’t statistically significant for any cohort.

A table of numbers and a number of black and white numbers

Description automatically generated with medium confidence

Table 3: TSLS estimates for men living in rural areas.

Significance codes: 0.01 \*\*\*, 0.05 \*\* and 0.1 \*.

Note: standard errors are in parenthesis, each equation also includes an intercept and further controls.

In summary, the results are as follows. For the cohort of men born between 1920 and 1929 living in rural areas, the estimated return to education is slightly lower than for the average male population. In the 1930 to 1939 cohort, the return to education is around half for men living in rural areas, however, this estimate is not statistically significant as it’s almost the same value as its standard error. For men born in the 1940s, the return to education is slightly higher for the ones living in rural areas. As mentioned before, none of these differences are statistically significant. And, also worth mentioning, the variation across cohorts isn’t statistically different either.

These results imply that there isn´t a clear indication of the return to education being different from men living in rural areas. However, limitations on the availability of data could play an important role. Based on Angrist and Krueger’s findings, I would expect more precise estimates allowing for the seasonal pattern to vary by state of birth. Nonetheless, it cannot be concluded that the added precision could make the difference significant. Despite this caveat, these results point towards equal relevance of compulsory schooling laws in different contexts, at least in terms of the return to education.

To conclude this paper, it’s important to point out the main findings. First, Angrist and Krueger show that compulsory schooling laws are effective in increasing education attainment and that this education has a positive impact on the future earnings of men. To obtain these unbiased results, they exploit the variation caused by years of education within season of birth by instrumenting the former with the latter. Surprisingly, their findings show that the difference between the instrumental variable TSLS and OLS estimates is very small and not statistically significant. This is evidence against the relevance of omitted variable bias in measurements of the return to education, at least for education years around the compulsory schooling level. I was able to replicate these estimates with accuracy because the data used by the authors was publicly available. Then, I ran the same models using the same data.

Motivated by looking into the relevance of geographical area of residence in the return to education, I extended their analysis by getting the estimates of the return to education for men living in rural areas. And, although it’s important to point out the limitations of these results, I found that there aren’t any differences in the return to education for rural workers when compared to the overall male population. Meaning that an extra year of education around the compulsory schooling level generates statistically equal relative increases in future earnings for rural and urban workers. Then, does this mean that policymakers should ignore geographical characteristics when arguing for compulsory schooling laws? Based only on this paper, they should. However, reality is more complex, and additional research on the implications should be done to adequately answer the question.

The possibilities of extension analysis in this paper were limited by the availability of data to conduct it. An idea I found interesting was to look for different effects of compulsory schooling laws throughout the lifespan of men. For example, instead of only estimating the effects on men born between 1920 and 1929 when they were aged 40 to 49, I would’ve liked to estimate their effects earlier and later in their lives. Here, finding when the positive effects of education on wages kick in would be essential to understand the incentivization process that young men face around the compulsory schooling age.

**References**

Angrist, Joshua D. and Alan B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics* 106, no. 4 (November 1991): 979-1014.

“Glossary.” *United States Census Bureau*. Revised June 23, 2023. <https://www.census.gov/programs-surveys/metro-micro/about/glossary.html>.

“Standard Statistical Metropolitan Areas in the United States as Defined October 18, 1963.” United States Census Bureau. December 05, 1963. “Glossary,” *United States Census Bureau*, revised June 23, 2023, <https://www.census.gov/programs-surveys/metro-micro/about/glossary.html>.

1. Joshua D. Angrist and Alan B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics* 106, no. 4 (Nov. 1991): 979-1014. [↑](#footnote-ref-1)
2. The laws that compel students to stay in school until they turn a specific age (16 in must US states but up to 18 in some). [↑](#footnote-ref-2)
3. Angrist and Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, 1010. [↑](#footnote-ref-3)
4. Angrist and Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, 1010. [↑](#footnote-ref-4)
5. “Glossary,” *United States Census Bureau*, revised June 23, 2023, <https://www.census.gov/programs-surveys/metro-micro/about/glossary.html>. [↑](#footnote-ref-5)
6. Angrist and Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, 997. [↑](#footnote-ref-6)
7. The replication study was performed using data provided by the authors, obtaining the same results was the expected outcome. [↑](#footnote-ref-7)
8. Angrist and Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, 1004. [↑](#footnote-ref-8)
9. Angrist and Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, 1006. [↑](#footnote-ref-9)
10. Neither of the two estimations (allowing seasonal pattern to vary by state of birth and estimates for black men) were replicated in this paper because the state of birth data wasn’t made publicly available by the authors. [↑](#footnote-ref-10)