Microeconometrics: Replication Paper

Win Supanwanid

April 2020

I replicate all 8 Tables in Angrist and Krueger (1991) using STATA and provide a summary and critical review of the aforementioned paper – Does Compulsory School Attendance Affect Schooling and Earning? – in the first and second section. The STATA codes and Log files could be found at .

1 Summary

1.1 Objectives

The objective of Angrist and Krueger (1991) is to state that seasonal pattern of educational attainment is related to the compulsory school attendance law of each state and then study the effect of compulsory schooling laws on earning by using quarter-of-birth as an instrument for years of education.

1.2 Data

Angrist and Krueger (1991) analyzed the data from the 1960, 1970, 1980 census of males born in the United States. I, however, found only the 1970 and 1980 census. The data have 27 variables explaining the personal characteristics of 30-40 and 40-50 years old men born in 1920-1929, 1930-1939, and 1940-1949. The example of those variables are age, years of education, region of residence, year-of-birth, quaarter-of-birth, race, marital status, living in the center of city (SMSA), weekly earnings and so forth.

The data of men born 1920-1929 are selected from the 1 percent sample of the 1970 census with the sample size of 247,199. The data of men born 1930-1939 , however, are selected from the 5 percent sample of the 1980 census with the sample size of 329,509. The data of men born 1940-1949 are similarly selected from the 5 percent sample of the 1980 census with the sample size of 486,926.

1.3 Model

Several equations are used to produce Table I - III. The crux of this paper, however, are Table IV - VIII which adopted the following model (same notation as in the paper).

$$E_i \;\; = \;\; X_i\pi + \sum_c Y_{ic}\delta_c + \sum_c \sum_j Y_{ic}Q_{ij}\theta_{jc} + \epsilon_i \tag{1}$$

$$lnW_i \;\; = \;\; X_i\beta + \sum_c Y_{ic}\xi_c + \rho E_i + \mu_i \tag{2}$$

- E_i : Education of the ith individual

- X_i : Vector of independent variabies

- Y_{ic} : Dummy variables of year-of-birth ($c = 1, 2, ..., 10$)

- Q_{ij} : Dummy variables of quarter-of-birth ($j = 1, 2, 3$)

- W_i : Weekly wage

■ ρ : Return to education

■ $\pi, \delta_c, \theta_{jc}, \beta, \xi_c$: Coeficients

Table IV - VI estimate the coefficient using the interaction of quarter-of-birth and year-of-birth as the intrument variables. Table VII adds state-of-birth as a covariate and intrument variable to the model (1) and (2); Table VIII, however, considers the model only for the black men.

1.4   Results

First, they examine the seasonal pattern in educational attainment. From Table I, the F-statistics are highly statistically significant indicating the strong relationship between quarter-of-birth and educational outcome variables.

Second, they show that seasonal pattern of educational attainment is related to the compulsory school attendance law of each state through difference-in-differences analysis. From Table II, it is shown from the 1944 cohort that enrollment rate is increased by 4 percent for states with compulsory schooling law that require a minimum age of 17 or 18 years old. The effect, however, decreases for the 1954 and 1964 cohort to 2 and 0.5 percent.

Third, they estimate the return to education preliminarily by using Wald estimate. For the 1970 census, Wald estimate shows that one year increases in education leads to an increase of 7.15 percent in weekly wage; the OLS estimate – not statistically different from Wald's – shows 8.01 percent of return to education. For the 1980 census, Wald estimate shows that one year increases in education leads to an increase of 10.20 percent in weekly wage; the OLS estimate – also not statistically different from Wald's – shows 7.09 percent of return to education.

Fourth, they estimate the return to education by using quarter-of-birth as an intrument variable. The estimates for 4 specifications of TSLS and OLS are statistically indistinguishable with return to education for the 1920-1929, 1930-1939, and 1940-1949 cohort of 7.01 percent to 13.10 percent, 6.00 percent to 8.91 percent and 3.93 percent to 9.48 percent.

Fifth, by introducing state-of-birth, they estimate a more efficient TSLS reducing the standard error by approximately 40 percent. The return to education, however, increases for the TSLS

estimate, while decreases for the OLS's leading to a significant gap between the two estimators. They ascribe this result to the high school completion effect.

Lastly, they estimate the return to education for black men to be around 3.91 percent to 6.72 percent which is significantly lower than for white man. Part of the reason is because of the inferior quality education that black men born 1930-1939 were provided.

1.5 Policy Recommendations

Their studies suggest that compulsory school attendance laws have positive effect in driving students to acquire more education. This, however, doesn't mean that compulsory school attendance laws always provide favorable prospects. They recommend that more study is needed on the externalities of these laws.

2 Critical Review

2.1 Contributions

This research studies the effect of compulsory schooling laws on education and earnings. Their unusual experiment is conducted though the seasonal pattern arose in educational attainment. From this discovery, they contribute to the literature by estimating the impact of compulsory schooling laws on earning by using quarter-of-birth as an instrument variable.

Few researches are done concerning the relationship between compulsory schooling laws and other socioeconomic variables. The authors, however, use an unusually natural fact to examine the effect of compulsory schooling laws on education and earnings. Children born in different quarters enter school at different ages; the compulsory school attendance laws, however, require children to study until they reach 16 to 18 years old. This law obliges children born in last quarter to study far longer than their friends born in first quarter. Furthermore, quarter-of-birth is exogeneous to other personal characteristics other than schooling age and thus bring about an apt instrument variable for years of education.

They proceed to estimate the return to education by accounting the effect of compulsory schooling law and conclude that there is a negligible bias between their TSLS estimate and the traditional estimate of OLS.

They further explore the interactions between state-of-birth and quarter-of-birth leading to a more efficient approximation of return to education and end their paper with the estimate of return to education for black workers.

2.2 Limitations

This paper from 1991, appoximately 30 years ago, may not be relevant to the current society. The social perception of education for men born almost centuries ago, actually in 1920-1949, will clearly be different from our generation. Most of the students in the 21st century may value education high enough so that it renders the linkage between season of birth and education unfounded.

Policy Implications will also be limited to countries that adopt compulsory schooling laws by age. In countries such as Thailand, the compulsory schooling law requires every student to study

until Grade 9. This is clearly different from the United States'; thus, we can't determine the effect of Compulsory Schooling law on wage through quarter-of-birth in Thailand since child born in the first quarter can't dropout until his completion of junior high school study.

From Figure III – for men born in 1950-1959 – apart from the downward trend of years of completed education, the seasonal pattern of education shows a different story compare to Figure I and II for men born in 1930-1949. Children born in the last quarter have fewer years of education compare to those born in the first quarter; This finding clearly contradicts the results they provided and the paper doesn't perform further experiment for men born in 1950-1959. They only provided a one sentence explanation that the controversy is due to the Vietnam war and the younger men incomplete schooling.

The paper's limitation is further stated by the author that their results doesn't show that compulsory schooling law is entirely favorable. Although they statistically show that compulsory schooling laws drive some students to study, additional researches on externality of these laws are needed. On the positive side, we may have a reduce in crime rate from schooling attendance; Refractory students – in the absent of compulsory schooling laws will dropout – may inhibit the study of other students.

2.3 Data

This paper used various data sets from the Public Use Census Data; Together they bring about a large enough sample size which makes the statistical results valid under the Central Limit Theorem.

The paper's minor drawback is the vague description of some variables. SMSA – with the only explanation as "center city" – may cause misunderstandings for non-U.S. readers. It is abbreviated from standard metropolitan statistical area which is coined by the U.S. Office of Management and Budget to be a region having at least one urbanized area with population exceeding 50,000.

"Black men" is also an ambiguous term. If, by genetic mutation, I born black while my parents are white, will I be a black men? If my entire family are black but I luckily(?) born white, can I call myself a white man? This paper doesn't clearly state it's classification method of black people. Different criteria regarding who is black are applied for different societies and they even changes across time. It is important to understand this context since the research's interpretation for skin-based classification is entirely different compared to socially based systems of racial classification.

2.4 Model

2.4.1 Selection of Independent Variables

Not stated explicitly in the paper, the Mincer earnings function is adopted as a guideline for selecting the independent variables; this is obvious due to the Age-squared component. Apart from schooling and experience (proxied by Age), Angrist and Krueger (1991) also includes personal characteristics such as race, SMSA, marital status, region of residence which are entirely appropriate as regressors for log of weekly earning.

The interaction of quarter-of-birth and year-of-birth are ingeniously added as instrument variables. The interactions are well-suited since they satisfy the two main requirements for using IVs: the instruments must have a strong first stage, and the instruments must satisfy exclusion restriction.

Quarter-of-birth is strongly correlated to years of education as shown in the first section of Angrist and Krueger (1991), so the interactions have a strong first stage. For the instruments to satisfy

exclusion restriction, the interaction of quarter-of-birth and year-of-birth must be uncorrelated to the residual μ_i in 1.3 Model. In "III. Other Possible Effects of Season of Birth", they verify this requirement by providing indirect evidences from the literature and estimate the income of college graduates by season of birth. No effects of quarter-of-birth are found in the dependent variable; hence, the instruments satisfy the exclusion restriction requirement.

2.4.2 Possible Econometric Problems

As mentioned in the previous subsection, the instrument variables are appropriate since they satisfy the two main requirements for using IVs.

One of the possible problems is that the 1,500 exclusion restrictions of the interactions of state-of-birth-by-quarter-of-birth-by-year-of-birth should be added as the intrument variables for years of education to account for the cross-state seasonal by year variation; however, as mention in footnote 18, estimation that incorporates such intruments is computationally demanding for the 90s.

Another possible problem is the use of linear probability model in Table I for the Outcome Variable "High school graduate". In footnote 7, the paper states that linear probability model is suitable due to the mutually exclusiveness of quarter-of-birth dummies. The problem is that linear probability model has two main drawbacks. First, predicted probabilities are not bounded between zero and one. Second, the error term is not homoskedastic. Although the first drawback is irrelevant in this case, having heteroskedasticity is problematic for inference tasks.

For models in Table IV-VII, we may have heteroskedasticity in years of education since students with more years of education are expected to have more variation in wages (the upper bound increases rapidly). Spatial and temporal autocorrelation are not expected to occur for obvious reasons; Model misspecification is also not expected since their specification is suitable and corresponds to the literature.

2.4.3 Remedial Measures

The 1,500 exclusion restrictions of the interactions of state-of-birth-by-quarter-of-birth-by-year-of-birth could be added as the intrument variables in this era since Moore's Law apparently work out well. In Table I, the standard errors could be provided in the robust form to cope with heteroskedasticity from the linear probability model. Since heteroskedasticity is expected for models in Table IV-VII, robust s.e. should also be provided; this problem may turn out to not happen due to the shrinkage from logarithm of weekly wage.

References

Joshua D. Angrist and Alan B. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

Appendices

A   Tables

A.1   Table I

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect			F-test [P-value]
			I	II	III	
Total years of education[1]	1930-1939	12.79222	-.1242856 (.0166581)	-.0859973 (.0167512)	-.0148872 (.0159604)	25.12 [0.0000]
	1940-1949	13.56001	-.0854568 (.0125193)	-.0352745 (.0125985)	-.0188388 (.012602)	17.36 [0.0000]
High school graduate[2]	1930-1939	.774068	-.0191356 (.0021296)	-.0198344 (.0021415)	-.0038982 (.0020404)	46.60 [0.0000]
	1940-1949	.8636907	-.0145416 (.0014337)	-.0121225 (.0014428)	-.0019522 (.0014432)	51.26 [0.0000]
Years of educ. for high school graduates[3]	1930-1939	14.00601	-.0296008 (.0142548)	.0050956 (.0143257)	.0165048 (.0136098)	3.79 [0.0099]
	1940-1949	14.28134	-.0093476 (.0110349)	.0200636 (.0110922)	.0079357 (.0110816)	2.58 [0.0515]
College graduates[4]	1930-1939	.2356244	-.005028 (.0021646)	.0027638 (.0021767)	.0018581 (.0020739)	5.00 [0.0018]
	1940-1949	.2995881	-.0027701 (.0019175)	.0044954 (.0019297)	-.0000155 (.0019302)	5.01 [0.0018]
Completed master's degree[5]	1930-1939	.0898285	-.0010254 (.0014583)	.0019429 (.0014665)	-.0009199 (.0013972)	1.72 [0.1615]
	1940-1949	.1101511	.0000612 (.0013121)	.0038261 (.0013204)	.0010261 (.0013207)	3.76 [0.0103]
Completed doctoral degree[6]	1930-1939	.0349964	.0015652 (.0009373)	.0024837 (.0009426)	.0004057 (.0008981)	2.88 (0.0343)
	1940-1949	.0360273	-.0017901 (.0007809)	.0009889 (.0007858)	-.0005075 (.0007861)	4.48 [0.0038]

Table 1: The Effect of Quarter of Birth on Various Educational Outcome Variables
[1] From Table I.do line 117, 118, 120, 122
[2] From Table I.do line 132, 133, 200, 202
[3] From Table I.do line 135, 136, 139, 141
[4] From Table I.do line 213, 214, 274, 276
[5] From Table I.do line 290, 291, 350, 352
[6] From Table I.do line 367, 368, 427, 429

A.2 Table II

Date of Birth	Type of state law		Column (1) - (2)
	School-leaving age: 16 (1)	School-leaving age: 17 or 18 (2)	
	Percent enrolled April 1, 1960		
1. Jan 1-Mar 31, 1994	84.89115[1]	85.7213[2]	-0.8301
(age 16)	(0.38272)	(0.77054)	(0.86035)
2. Apr 1-Dec 31, 1994	85.7225[3]	85.99152[4]	-0.2690
(age 15)	(0.21789)	(0.44417)	(0.49473)
3. Within-state diff.	-0.83135[1]	-0.27022[2]	-0.5611
(row 1 - row 2)	(0.44044)	(0.88931)	(0.9924)

Table 2: Percentage of Age Group Enrolled in School By Birthday and Legal Dropout Age
[1] From Table II.do line 61
[2] From Table II.do line 69
[3] From Table II.do line 60
[4] From Table II.do line 68

The result of this table will not be the same as in the paper since the author doesn't provide data for this session. I only used the available data to approximate the result of this table. Results in the last column were manually calculated e.g. $-0.5611 = (-0.8301) - (-0.2690)$

A.3 Table III

PANEL A: WALD ESTIMATES FOR 1970 CENSUS - MEN BORN 1920-1929			
	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (wkly. wage)[1]	5.148471	5.15745	-.0089789 (.0030117)
Education[2]	11.3996	11.52515	-.1255553 (.0155391)
Wald est. of return to education[3]			.0715133 (.0218682)
OLS return to education[4]			.0801112 (.0003549)
PANEL B: WALD ESTIMATES FOR 1980 CENSUS - MEN BORN 1930-1939			
	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (wkly. wage)[5]	5.891596	5.902695	-.0110989 (.0027388)
Education[6]	12.68807	12.79688	-.1088179 (.0132376)
Wald est. of return to education[7]			.101995 (.0239489)
OLS return to education[8]			.070851 (.0003386)

Table 3: Wald Estimates
[1] From Table III.do line 52, 53, 57
[2] From Table III.do line 54, 55, 58
[3] From Table III.do line 61
[4] From Table III.do line 63
[5] From Table III.do line 66, 67, 71
[6] From Table III.do line 68, 69, 72
[7] From Table III.do line 75
[8] From Table III.do line 77

A.4 Table IV

Independent Variables	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802***	0.0769***	0.0802***	0.131***	0.0701***	0.0669***	0.0701***	0.101**
	(0.000355)	(0.0150)	(0.000355)	(0.0334)	(0.000355)	(0.0151)	(0.000355)	(0.0334)
Race(1 = black)					-0.298***	-0.306***	-0.298***	-0.227**
					(0.00434)	(0.0353)	(0.00434)	(0.0776)
SMSA (1 = center city)					-0.134***	-0.136***	-0.134***	-0.116***
					(0.00256)	(0.00924)	(0.00256)	(0.0198)
Married (1 = married)					0.293***	0.294***	0.293***	0.280***
					(0.00374)	(0.00719)	(0.00374)	(0.0141)
Age			0.145*	0.141*			0.116	0.117
			(0.0676)	(0.0704)			(0.0652)	(0.0661)
Age-squared			-0.00154*	-0.00136			-0.00125	-0.00118
			(0.000748)	(0.000787)			(0.000721)	(0.000736)
Observations	247199	247199	247199	247199	247199	247199	247199	247199

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: OLS and TSLS Estimates of the Return to Education for Men Born 1920-1929: 1970 Census
Yes-No Dummies are not kept in the shown figure.

9

A.5 Table V

Independent Variables	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711***	0.0891***	0.0711***	0.0760**	0.0632***	0.0806***	0.0632***	0.0600*
	(0.000339)	(0.0161)	(0.000339)	(0.0290)	(0.000339)	(0.0164)	(0.000339)	(0.0290)
Race(1 = black)					-0.257***	-0.230***	-0.257***	-0.263***
					(0.00404)	(0.0261)	(0.00404)	(0.0458)
SMSA (1 = center city)					-0.176***	-0.158***	-0.176***	-0.180***
					(0.00287)	(0.0174)	(0.00287)	(0.0305)
Married (1 = married)					0.248***	0.244***	0.248***	0.249***
					(0.00317)	(0.00487)	(0.00317)	(0.00726)
Age			-0.0772	-0.0801			-0.0760	-0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared			0.000787	0.000831			0.000770	0.000743
			(0.000688)	(0.000734)			(0.000669)	(0.000712)
Observations	329509	329509	329509	329509	329509	329509	329509	329509

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS and TSLS Estimates of the Return to Education for Men Born 1930-1939: 1980 Census
Yes-No Dummies are not kept in the shown figure.

10

A.6 Table VI

Independent Variables	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0573***	0.0553***	0.0573***	0.0948***	0.0520***	0.0393**	0.0521***	0.0779**
	(0.000298)	(0.0138)	(0.000298)	(0.0223)	(0.000297)	(0.0145)	(0.000297)	(0.0239)
Race(1 = black)					-0.211***	-0.227***	-0.211***	-0.179***
					(0.00322)	(0.0183)	(0.00322)	(0.0299)
SMSA (1 = center city)					-0.142***	-0.154***	-0.142***	-0.118***
					(0.00229)	(0.0135)	(0.00229)	(0.0220)
Married (1 = married)					0.245***	0.244***	0.244***	0.245***
					(0.00220)	(0.00223)	(0.00220)	(0.00229)
Age			0.180***	0.133**			0.152***	0.121*
			(0.0389)	(0.0486)			(0.0379)	(0.0474)
Age-squared			-0.00234***	-0.00158*			-0.00195***	-0.00146*
			(0.000559)	(0.000725)			(0.000545)	(0.000709)
Observations	486926	486926	486926	486926	486926	486926	486926	486926

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: OLS and TSLS Estimates of the Return to Education for Men Born 1940-1949: 1980 Census
Yes-No Dummies are not kept in the shown figure.

Independent Variables	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673***	0.0928***	0.0673***	0.0907***	0.0628***	0.0831***	0.0628***	0.0811***
	(0.000346)	(0.00930)	(0.000346)	(0.0107)	(0.000344)	(0.00949)	(0.000344)	(0.0109)
Race(1 = black)					-0.255***	-0.233***	-0.255***	-0.235***
					(0.00435)	(0.0109)	(0.00435)	(0.0122)
SMSA (1 = center city)					-0.171***	-0.151***	-0.170***	-0.153***
					(0.00289)	(0.00948)	(0.00289)	(0.0107)
Married (1 = married)					0.249***	0.244***	0.249***	0.244***
					(0.00316)	(0.00399)	(0.00316)	(0.00420)
Age			-0.0757	-0.0880			-0.0778	-0.0876
			(0.0617)	(0.0624)			(0.0603)	(0.0609)
Age-squared			0.000752	0.000942			0.000789	0.000938
			(0.000684)	(0.000694)			(0.000669)	(0.000677)
Observations	329509	329509	329509	329509	329509	329509	329509	329509

Standard errors in parentheses

$* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table 7: OLS and TSLS Estimates of the Return to Education for Men Born 1930-1939: 1980 Census
Yes-No Dummies are not kept in the shown figure.

A.8 Table VIII

Independent Variables	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0672***	0.0635***	0.0671***	0.0567**	0.0576***	0.0461*	0.0576***	0.0393*
	(0.00134)	(0.0185)	(0.00134)	(0.0199)	(0.00135)	(0.0187)	(0.00135)	(0.0199)
SMSA (1 = center city)					-0.189***	-0.205***	-0.188***	-0.215***
					(0.0142)	(0.0307)	(0.0142)	(0.0324)
Married (1 = married)					0.222***	0.227***	0.222***	0.231***
					(0.0100)	(0.0136)	(0.0100)	(0.0140)
Age			-0.310	-0.326			-0.298	-0.323
			(0.254)	(0.256)			(0.247)	(0.249)
Age-squared			0.00333	0.00347			0.00323	0.00346
			(0.00282)	(0.00283)			(0.00275)	(0.00276)
Observations	26913	26913	26913	26913	26913	26913	26913	26913

Standard errors in parentheses

$* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Table 8: OLS and TSLS Estimates of the Return to Education for Black Men Born 1930-1939: 1980 Census
Yes-No Dummies are not kept in the shown figure.