COMPULSORY SCHOOLING LEGISLATION
AND SCHOOL ATTENDANCE IN
TURN-OF-THE-CENTURY AMERICA: A
"NATURAL EXPERIMENT" APPROACH

Robert A. Margo
T. Aldrich Finegan

COMPULSORY SCHOOLING LEGISLATION
AND SCHOOL ATTENDANCE IN
TURN-OF-THE-CENTURY AMERICA: A
"NATURAL EXPERIMENT" APPROACH

ABSTRACT

Recent research by Joshua Angrist and Alan Krueger has used information on exact dates of birth in the 1960 to 1980 federal censuses to study the impact of compulsory schooling laws on school attendance. This paper modifies their methodology to analyze similar data in the 1900 federal census to measure the impact of turn-of-the-century compulsory schooling laws. Using data on 14-year olds from the 1900 census public use microdata sample we compare attendance rates of children born after January 1, 1990 with those born before, across states with and without compulsory schooling laws. In states that combined school-leaving with child labor laws, we find that compulsion significantly raised attendance rates.

Robert A. Margo
Department of Economics
Vanderbilt University
Nashville, TN 37235
and NBER

T. Aldrich Finegan
Department of Economics
Vanderbilt University
Nashville, TN 37235

1. Introduction

Over the past century there has been a marked increase in school attendance among younger teenagers. In 1900, approximately 65 percent of children aged 14 attended school; the corresponding figure for 1990 was 96 percent.[1] Recent work by Angrist and Krueger (1991) suggests that compulsory schooling legislation (CSL) played an economically significant role in raising teen attendance rates since 1960. The conventional wisdom among economic historians, however, is that CSL had little impact on teen attendance early in the century because the laws were imperfectly enforced (Landes and Solmon, 1972; Eisenberg, 1988, 1994). Consequently, economic historians have downplayed the role of CSL and closely related legislation (for example, child labor laws) in accounting for the long-term trend in schooling, preferring to emphasize other factors (for example, rising parental incomes and secular increases in the returns to high school).

This paper exploits a previously unnoticed feature of the 1900 federal census that permits use of a "natural experiment" to estimate of the impact of CSL, similar to the one used by Angrist and Krueger. We find that CSL in the aggregate did not have a statistically significant impact on school attendance, although the impact was positive. We do find, however, significant positive effects in a sub-group of states that combined CSL with child labor legislation.

1

2. Methodology and Results

Angrist and Krueger (1991; hereafter A-K) examined the impact of CSL by exploiting a so-called "natural experiment" in the census. The 1960-80 censuses collected information on quarter of birth in addition to school attendance as of April 1. In states that permitted school leaving at age 16, A-K reasoned that teenagers would be less likely to attend school if their 16th birthdays occurred before April 1 than after, compared with states that required teenagers to remain in school until age 17 or age 18, under the hypothesis that CSL was effective. A-K (Table 2, p. 990) found positive and statistically significant effects of CSL in 1960 and 1970, but not in 1980, presumably because school attendance was nearly universal by that year.

Prior to 1960 the only census to ask for exact date of birth (as opposed to age in years) was the 1900 census.[2] In 1900 the census recorded both the month and year of birth for the entire population. These data were never published, but they were retained in the public use microdata sample (PUMS) of the 1900 census (Center for Studies in Demography and Ecology, 1980). The 1900 census also collected two pieces of information on school attendance: total months of school attendance during the census year (June 1, 1899 to May 31, 1900) and whether or not a teenager's "gainful occupation" during the census year was "at school" (as opposed to some other use of time).

Because the 1900 census did not collect attendance as of specific date (such as April 1), it is not possible to use Angrist and Krueger's specific estimator. However, it is possible to modify the estimator to accommodate the different information collected in 1900. We focus on children whose fourteenth birthdays occurred during the census year, because most CSL in force at the turn of the century specified age 14 as the school leaving age.

As a point of departure, consider two groups of children, those whose 14th birthday occurred between June 1, 1899 and August 31, 1899 (whom we call "older" age 14s), and those whose birthdays occurred between January 1, 1900 and May 31, 1900 (whom we call "younger" age 14s). Older age 14s experienced their birthdays before the start of the 1900 school year, while younger age 14s experienced their birthdays well after the school year began. Our central hypothesis is that, if CSL were effective at all, rates of school attendance among younger age 14s should have been significantly higher than among older age 14s in states that specified age 14 as the school leaving age, but this should not have been true in states without age-14 CSL.

We specify January 1, 1900 as the a priori cut-off date for younger age 14s for two reasons. First, January 1 is a natural choice since most turn of the century CSL required children to attend a minimum of three to four months -- approximately the length of time between the start of the 1899 school year in a typical state (for example, September or October) and our chosen

cut-off date. Second, as noted earlier, enforcement of CSL was far from perfect in 1900 (Eisenberg 1988, 1994). Had enforcement been complete, children whose 14th birthdays occurred after the start of the school year should have attended some school before possibly dropping out. However, because enforcement was imperfect, a child whose 14th birthday occurred relatively soon after that start of the 1899 school year (for example, November) may have been willing to drop out early rather than attend school for a brief period of time. Children whose 14th birthdays occurred well after the school year began, however, were "at risk" for a longer period and, therefore, if CSL were effective at all -- despite imperfect enforcement -- should have been more likely to attend school than older age 14s. This argument implies that setting an earlier cut-off date should reduce the estimated impact of CSL, a point that we demonstrate empirically later in the paper.

More formally, let dS_t be the difference in school attendance rates between older and younger age-14 children in states with CSL, and dS_c be the difference in school attendance rates between older and younger age-14 children in states without CSL. Following Angrist and Krueger, the "treatment" effect of CSL is estimated by the so-called "differences in differences" (DID) estimator:

$$DID = dS_t - dS_c$$

4

It is important to note that the DID estimator is _not_ the same as the difference in attendance rates between states with and without CSL. The DID estimator first removes common factors associated with higher or lower school attendance by differencing between older and younger age-14s within groups. By design, what is left over is the "treatment" effect of the legislation.

Table 1 shows the results of the basic DID estimator, as applied to the 1900 PUMS.[3] States with age-14 CSL were identified from the descriptions of laws contained in U.S. Office of Education (1901, pp. 2598-2602).[4] We use three measures of attendance: the proportion attending at least one month of school, the proportion attending at least four months, and the proportion reporting their gainful occupation as "at school". The least restrictive definition of school attendance is the first (one month or more). The most restrictive, and perhaps also the most subjective, is the gainful occupation definition. Census enumerators were instructed to report a child's "occupation" as "at school" if the time spent in school exceeded that in the most important alternative use (at home, or in the labor force).[5]

We find a positive impact of CSL on school attendance using the first two measures of school attendance, but virtually no impact using the "at school" occupation indicator. The largest effect (9.9 percentage points) occurs with the least restrictive measure, the proportion attending at least one month. Because the sample sizes from the 1900 PUMS are small, however, we cannot

reject the null hypothesis that the DID estimates are insignificantly different from zero.

A problem with the analysis in Table 1 is the crudeness of the CSL-no CSL dichotomy. As Eisenberg (1988, 1994) points out, turn-of-the-century CSL legislation differed considerably across states in provisions, exemptions, degree of enforcement, and in associated legislation that may have aided (or hindered) enforcement. Ideally, we would compute dS_t for each state with CSL, but the sample sizes in the 1900 PUMS are too small to do so reliably. To gain some insight into differences across states, we selected from states with CSL a sub-group of six states -- Illinois, Connecticut, Indiana, Massachusetts, Michigan, and New York -- that combined CSL with an age-14 comprehensive child labor law.[6] By "comprehensive", we mean child labor laws that restricted or prohibited the paid employment of children under age 14 in a relatively wide array of occupations or industries (for example, manufacturing).

There are two reasons why an age-compatible child labor law might have enhanced the effectiveness of CSL. First, prosecutions of child labor laws were fairly common. In Illinois, for example, there were 1,386 such prosecutions in 1900 (American Academy of Political and Social Science 1907, p. 35). A recent analysis of 1880 census data suggests that Massachusetts' network of factory inspection agents was effective in reducing hours of paid work among children (Atack and Bateman, 1991). By reducing the incentive to seek paid employment, child labor laws might

6

have increased school attendance. Second, child labor laws

adopted by the six-state subgroup required the children satisfy

certain educational requirements (for example, ability to read

and write English) before commencing work. In particular, four of

the six states (the exceptions are Michigan and Indiana)

required documentary proof of age and of completion of the

educational requirements, both of which could be obtained from

school officials (American Academy of Political and Social

Science, 1907, pp. 14-23).[7]

Rows 1,4 and 5 of Column 1 of Table 2 shows the results of

applying our basic DID estimator to the six state sub-group,

compared with states without age-14 CSL. In the sub-group, there

were statistically significant effects ~~of~~ on school attendance

using the first two measures of school attendance, ranging from

18 to 21 percentage points. The impact using the at-school

measure is also positive, but not significant at conventional

levels.

We conduct three sensitivity analyses that are reported in

Table 2. Using the least restrictive measure of attendance (one

month or more), we vary the cutoff date for younger age-14s (rows

2 and 3 in column 1). As hypothesized earlier, setting an

earlier cutoff date results in a substantially smaller estimated

impact of CSL suggesting that, even in the six-state subgroup,

there was considerable room for improved enforcement of

legislation. Column 2 reports the DID estimator for children

aged 15. If our analysis is correct, the DID estimates should be

insignificantly different from zero for these children, because neither CSL nor child labor laws applied to them. In fact, the DID estimates are negative (the wrong sign) and statistically insignificant.

Finally, column 2 reports the results of a multivariate DID estimator. By "multivariate" we mean as derived from regressions of school attendance controlling for factors other than birthday.[8] The logic of our natural experiment is that date of birth is truly exogenous, and therefore is uncorrelated with factors other than CSL that affect the probability of school attendance (Bound, Jaeger, and Baker, 1995). This assumption implies that the multivariate DID estimates should be close in magnitude to the univariate estimates reported in Column 1 of Table 2. The multivariate estimates are generally significant at the 5 percent level, and very similar in magnitude to the univariate estimates, supporting our assumption that date of birth is exogenous.

3. Discussion

We have applied the "natural experiment" methodology of contemporary labor economics to measure the impact of CSL on school attendance using the 1900 PUMS. We found a positive, but statistically insignificant, impact of CSL in the aggregate. For a smaller group of states that combined CSL with an age-compatible comprehensive child labor law, it appears that school

8

attendance -- particularly among children who might not otherwise have attended at all -- was significantly higher than it would have been in the absence of state intervention. Because the sample sizes in the 1900 PUMS are small, however, further research on the impact of CSL (in particular, differences across states) will require the collection of additional observations from the census manuscripts.[9]

1. The 1900 figure was computed from the 1900 Public Use Microdata Sample (Center for the Study of Demography and Ecology, 1980) because the published volumes of the 1900 census do not report school attendance by single years of age. The 1990 figure is from U.S. Bureau of the Census (1994, Table 5, p. 597).

2. Angrist and Krueger (1991, p. 982) asserted that the earliest census to collect information on date of birth was the 1960 census, but evidently they were unaware of the question on month of birth in the 1900 census.

3. The sample is restricted to persons who were (1) age 14 at the time of the census (2) reported their dates of birth to be between June 1885 and May 1886 (3) unmarried at the time of the census (4) living at home in a two-parent household.

4. The "no age-14 CSL" group also excludes states (for example, Minnesota) that required attendance until a later age (for example, age 16). In addition, because we examine the impact of CSL combined with child labor legislation later in the paper, we also excluded from the no age-14 CSL group three states (Missouri, New Jersey, and Louisiana) that had age-14 child labor legislation but no CSL. Unfortunately, the sample sizes in the 1900 PUMS from these three states are too small to test the effectiveness of child labor legislation sans CSL. The substantive results of the paper would not be affected if the excluded states were added to the no age-14 CSL group.

5. See paragraphs 162, 218, and 219 of the instructions to census enumerators in 1900 (Center for Studies in Demography and Ecology, 1980).

6. Child labor states were identified from the same source giving information on CSL, U.S. Office of Education (1901).

7. A-K (1991, p. 993) also hypothesize that child labor laws were a key reason why CSL raised school attendance in 1960 and 1970. A-K (p. 993-94) point out that modern child labor laws require that students obtain age certification for employment that is administered through schools.

8. For each group of states (the six-state subgroup and states with no age-14 CSL) we estimated the following linear probability regression

$$y = \alpha_i + X\delta_i + \beta_i D + \epsilon$$

where i = 1 (six-state subgroup) or 2 (no age-14 CSL), y is school attendance, the X's are covariates (occupational status, homeownership, and literacy of the head of the household; the literacy of the spouse, if present; sex, race, and nativity status of the child; and the location of the household, as proxied by dummy variables indicating the extent of urbanization, whether the household lived on the a farm, and the state of residence); and D is a dummy variable equal to one for older 14-year olds, zero otherwise. The DID estimates in Table 2 are

11

$$DID = ß_2 - ß_1$$

that is, the difference between the coefficients of the dummy variable D in the regression on the six-state subgroup and the regression on states with no age-14 CSL. The substantive findings would not be affected if logit analysis were used instead of a linear probability regression.

9. For a start in this direction, see Finegan and Margo (1995, in progress).

References

American Academy of Poltical and Social Science. 1907.

Supplement, Child Labor Legislation. Philadelphia: American

Academy of Political and Social Science.

Angrist, Joshua and Alan Krueger. 1991. "Does Compulsory School

Attendance Affect Schooling and Earnings?" Quarterly Journal

of Econommics, 106: 979-1014.

Atack, Jeremy and Fred Bateman. 1991. "Whom Did Protective

Legislation Protect? Evidence from 1880." Working Paper

Series on Historical Factors in Long Run Growth, No. 33,

December 1991. Cambridge, MA: National Bureau of Economic

Research.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995.

"Problems with Instrumental Variables Estimation When the

Correlation between the Instruments and the Endogenous

Explanatory Variable is Weak." Unpublished paper, Department

of Economics, University of Michigan.

Center for Studies in Demography and Ecology. 1980. 1900 Public

Use Sample, User's Handbook. Seattle, WA: University of

Washington.

Eisenberg, Martin. 1988. Compulsory Attendance Legislation in

America, 1870-1915. Unpublished Ph.D. dissertation,

Department of Economics, University of Pennsylvania.

Eisenberg, Martin. 1994. "Did Compelling School Attendance Work?

Evidence from the 1900 Census." Unpublished paper,

Department of Economics, Knox College.

Finegan, T. Aldrich and Robert A. Margo. 1995. "The Impact of Turn-of-the-Century Compulsory Schooling Legislation: A Proposal for a New Sample from the 1900 Census Manuscripts." Unpublished paper, Department of Economics, Vanderbilt University, in progress.

Landes, William and Lewis Solmon. 1972. "Compulsory Schooling Legislation: An Economic Analysis of Law and Social Change in the Nineteenth Century," _Journal of Economic History_ 32: 45-91.

U.S. Bureau of the Census. 1994. _1990 Census of Population: Education in the United States_. Washington: U.S. Government Printing Office.

U.S. Office of Education. 1901. _Report of the Commissioner of Education_. Washington: U.S. Government Printing Office.

Table 1

School Attendance and Compulsory Schooling Laws in 1900

Type of Law

	No Age-14 CSL	Age 14 CSL	Col (1) - (2)
Percent Attending >= 1 month			
Date of Birth			
1. 6/99-8/99	59.2	66.3	-7.1
	(4.1)	(3.4)	(5.3)
2. 1/1900-5/1900	60.7	77.7	-17.0
	(2.6)	(2.3)	(3.5)
3. Within-state diff.	-1.5	-11.4	9.9
(row 1 - row 2)	(4.8)	(4.1)	(6.4)
Percent Attending >=4 months			
4. 6/99-8/99	42.9	61.5	-18.6
	(4.1)	(3.6)	(5.4)
5. 1/1900-5/1900	47.5	73.3	-25.8
	(2.7)	(2.4)	(3.6)
6. Within-state diff.	-4.6	-11.8	7.2
(row 1 - row 2)	(4.9)	(4.3)	(6.5)
Percent "At School"			
7. 6/99-8/99	36.1	57.8	-21.7
	(4.0)	(3.6)	(5.4)
8. 1/1900-5/1900	41.9	66.5	-24.6
	(2.7)	(2.6)	(3.7)
9. Within-state diff.	-5.8	-8.7	2.9
(row 1 - row 2)	(4.8)	(4.4)	(6.5)

Source: 1900 Public Use Microdata Sample
Notes: standard errors in parentheses. The DID estimates are in the lower right hand corners of each panel. Sample sizes are 334 for rows 1, 4, and 7; and 678 for rows 2, 5, and 8.

Table 2

School Attendance and Combined Child Labor-Compulsory
Schooling Laws, 1900: DID estimates

Model

Variable	Univariate Age 14	Univariate Age 15	Multivariate Age 14
Percent attended>= 1 month Dates of birth			
1. 1-5/1900 - 6-8/1899	21.1 (7.8)	-6.0 (8.0)	19.3 (7.4)
2. 9-10/1899 - 6-8/1899	4.5 (10.6)		
3. 11-12/1899 - 6-8/1899	9.4 (10.4)		
Percent attended>= 4 months			
4. 1-5/1900 - 6-8/1899	17.6 (7.9)	-5.2 (8.0)	16.1 (7.4)
% At school			
5. 1-5/1900 - 6-8/1899	11.3 (7.9)	-3.1 (8.0)	10.1 (7.6)

Note: see text for definition of DID estimates. Univariate: no controls. Multivariate: derived from linear probability regressions; see footnote 7. Sample sizes are: 756 in rows 1,4 and 5, columns 1 and 3; 405 in row 2, column 1; 391 in row 3, column 1; and 722 in column 2. Standard errors in parentheses.