Long-Run Impacts of Agricultural Shocks on Educational Attainment: Evidence from the Boll Weevil

RICHARD B. BAKER, JOHN BLANCHETTE, AND KATHERINE ERIKSSON

The boll weevil spread across the South from 1892 to 1922 with devastating effect on cotton cultivation. The resulting shift away from this child labor–intensive crop lowered the opportunity cost of school attendance. We investigate the insect's long-run effect on educational attainment using a sample of adults from the 1940 census linked back to their childhood census records. Both white and black children who were young (ages 4 to 9) when the weevil arrived saw increased educational attainment by 0.24 to 0.36 years. Our results demonstrate the potential for conflict between child labor in agriculture and educational attainment.

A substantial body of research on developing countries documents the trade-off parents face in choosing between sending their children to work or to school by showing that child labor reduces various measures of educational attainment and achievement, including attendance, test scores, and years of schooling (see, e.g., Beegle, Dehejia, and Gatti 2009; Boozer and Suri 2001; Gunnarsson, Orazem, and Sánchez 2006; Emerson, Ponczek, and Souza 2017). However, this literature often ignores unique features, such as the informal employment of children on family farms and the seasonality of labor demands, of child labor in agricultural regions, which make up a majority of the developing world. Those studies focusing on rural areas suggest that negative agricultural shocks have negative (Jensen 2000; Bandara, Dehejia, and

The Journal of Economic History, Vol. 80, No. 1 (March 2020). doi: 10.1017/S0022050719000779

Richard B. Baker is Assistant Professor of Economics, The College of New Jersey, School of Business, PO Box 7718, 2000 Pennington Road, Ewing, NJ 08628–0718. E-mail: bakerr@tcnj.edu. John Blanchette is Ph.D. candidate in Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616-8617. E-mail: jblanchette@ucdavis.edu. Katherine Eriksson is Assistant Professor of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616–8617; Research Associate at the National Bureau of Economic Research; and Research Affiliate at Laboratory for the Economics of Africa's Past, University of Stellenbosch. E-mail: kaeriksson@ucdavis.edu.

We are grateful to James Feigenbaum, Paul Rhode, and Marianne H. Wanamaker for providing feedback on earlier versions of this paper. We thank workshop participants at the University of California Davis, The College of New Jersey, the 2018 World Economic History Congress, and North American Summer Meeting of the Econometric Society for their helpful comments and suggestions. We owe special thanks to Celeste K. Carruthers and Marianne H. Wanamaker for generously sharing with us their county-level administrative data on schools.

Lavie-Rouse 2015) or neutral (Dammert 2008) effects on educational outcomes. However, educational responses to agricultural shocks are likely to be highly dependent on crop mix and the labor requirements of affected crops. A negative agricultural shock affecting a widely grown child labor–intensive crop will both reduce household income and the productivity of child labor in agriculture. If the income effect is limited by the ability to switch to other, similarly profitable, crops, then positive effects on educational outcomes may be observed.

In addition, there is a sizable literature examining both the short- and long-run effects, on educational outcomes, of various "treatments" expected to improve schooling, in developing and developed countries. Studies within this literature consistently find positive short-run effects on student outcomes, but results are mixed with respect to long-run outcomes. Baird et al. (2016), Heckman et al. (2010), and Chetty et al. (2011) find both short- and long-run effects. However, other papers find short-run effects that fade out over the medium to long run. For example, Bleakley (2007) finds immediate positive effects of hookworm eradication on school enrollment, but does not find evidence that educational attainment increased in the long run for affected cohorts. Similarly, Evans and Ngatia (2018) show that a program to provide uniforms to school children in Kenya had positive effects on contemporaneous enrollment, but analysis of a follow-up survey, taken eight years after the intervention, finds no effect on primary school completion or years of education. There are at least three reasons why short-run effects may not translate into long-term gains: First, as Evans and Ngatia (2018) posit, a lack of long-run impact would be consistent with dynamic complementarities whereby short-run investments are not reinforced through further investment in later years in a situation with low school quality. Second, a treatment may affect enrollment and attendance in the short run without affecting educational attainment if the treatment does not change the child's targeted level of years of schooling. In this case, enrollment effects merely reflect intertemporal substitution to take advantage of the treatment, say a complimentary school uniform. Third, increased enrollment and attendance do not necessarily lead to grade completion, the latter of which has greater value as a signal of human capital in the labor market.

In contribution to these literatures, we exploit a shift in agricultural production, away from a child labor–intensive crop (cotton), that occurred in the early twentieth-century American South to demonstrate that a negative agricultural shock can have both a short- and long-run positive impact on educational outcomes. This shock to cotton production was caused by the boll weevil infestation which spread from the southern tip of Texas in 1892 until it affected nearly the whole Cotton Belt in 1922 (Hunter and Coad 1923). Due to reduced returns to cotton cultivation under boll weevil conditions, and at the encouragement of the state and federal agricultural agencies, farmers substituted away from cotton to alternative crops (e.g., corn, peanuts, and sweet potatoes). The tasks involved in the cultivation of these alternatives were less suitable for children, as compared to chopping, hoeing, and picking cotton. The resulting, and plausibly exogenous, fall in the marginal product of child labor in rural areas represents a reduction in the opportunity cost of schooling. This would suggest that the boll weevil increased educational attainment for children who were young when the infestation began. However, the boll weevil also represents a negative income shock, which clouds expectations regarding its overall impact on educational attainment.[1]

To examine the boll weevil's impact on educational attainment, we match men in the 1940 census who were born in Southern states comprising the Cotton Belt to their childhood census records in 1900, 1910, and 1920.[2] As the 1940 census was the first to inquire about educational attainment, the result is a matched sample linking childhood location and family background information with years of schooling as reported in adulthood. Identifying childhood county of residence allows us to calculate the age of each individual when the boll weevil arrived, or the age at which they were first exposed to the weevil. We use this information to compare the educational attainment of those exposed to the boll weevil at young ages with that of older cohorts, whose education should be unaffected by the insect, in a differences-in-differences framework.

The results show that white children who were 7–9 years of age when the boll weevil arrived attained approximately 0.236 more years of schooling on average, relative to those that were 19–30 when the boll weevil infestation began. We find similar gains of approximately 0.243 years of schooling for comparatively aged black children. Overall, there

[1] Other channels through which the boll weevil may have affected educational attainment, such as improved health, increased probability of migration, and higher returns to education, are discussed later.

[2] States with territory in the Cotton Belt include: Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Mississippi, Missouri, New Mexico, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia. We, however, exclude Missouri, Kansas, and New Mexico because they are not in the Southern United States and only small portions of these states were considered to be in the Cotton Belt. Moreover, the latter two were not shown to be infested by the boll weevil as of 1922 (Hunter and Coad 1923).

is a lack of evidence to suggest that black and white children benefited differently from the boll weevil infestation with respect to educational attainment. The gains in educational attainment, however, decrease for older children of both races, with those aged 16–18 when the weevil arrived seeing a comparatively modest gain of approximately one-twentieth of a year of schooling. Our results are robust to changing the comparison age cohorts as well as controlling for other potentially confounding factors, such as the introduction of Rosenwald schools and compulsory schooling laws.

Our findings demonstrate that the seasonal demands for child labor in agriculture can have substantial negative impacts on educational attainment. Child labor–intensive crops, such as coffee, cotton, tea, and tobacco, are primary agricultural products of many regions of the developing world.[3] Thus, our results are suggestive of the potential educational benefits of programs that encourage the production of alternative (less child labor–intensive) crops and the adoption of technologies that reduce demand for child labor in agriculture.

Additionally, this work adds to a growing literature on the broader impacts of the boll weevil. Early efforts to understand the effect of the boll weevil on the Southern economy were largely limited to examinations of state-level variation (Higgs 1976; Osband 1985). Lange, Olmstead, and Rhode (2009) revived interest in the boll weevil with their analysis of the insect's effect on crop production in the South using county-level data. They show that cotton production dramatically decreased in the years following the arrival of the boll weevil with farmers shifting resources to the production of other crops. Ager, Brueckner, and Herz (2017) reveal the boll weevil reduced labor force participation (particularly among females), farm wages, and the number of fixed-rent tenant farms, with counties more reliant on cotton experiencing greater declines, illustrating the broader impact of the boll weevil on local economies. Bloome, Feigenbaum, and Muller (2017) show that the boll weevil infestation reduced the proportion of farms worked by tenants, which in turn altered incentives to marry and reduced the proportion of African Americans who were wed at a young age, revealing the weevil's effect on life-altering decisions.

[3] The U.S. Department of Labor (1995) report *By the Sweat and Toil of Children* finds these crops to be particularly suited to the use of child labor and details the role of children in farming them in different regions of the developing world. A recent report by the U.S. Department of Labor (2018), known as the *List*, identifies 15 developing countries in which cotton is produced by means of child labor. An additional 23 developing countries are also on the *List* for exploiting the labor of children in the production of coffee, tea, or tobacco.

Finally, and most closely related to this article, Baker (2015) finds school enrollment rates for African-American children in Georgia increased following the boll weevil's arrival, resulting from the insect's negative effect on production of child labor–intensive cotton.[4] Estimated effects for white children, in comparison, are close to zero and not statistically significant. In this article, we first confirm that the enrollment effects hold when adding five more states, as well as broader range of years, to the analysis. In doing so, we find positive enrollment effects for both black and white children, but black children still have a stronger enrollment response to the boll weevil. Then, we move on to our main analysis estimating the impact of childhood exposure to the boll weevil at various ages on educational attainment as measured in adulthood, between 18 and 40 years after exposure. Establishing the long-run gains in educational attainment in response to the boll weevil infestation, a negative agricultural shock, is the primary contribution of this work.

BACKGROUND

Cotton, Children, and Schooling

At the beginning of the twentieth century, the South was still an agrarian economy, as agriculture employed 57 percent of the labor force in 1910. And cotton was its staple. Cotton was the single most valuable crop in 10 of the 16 Southern states at the dawn of the twentieth century. In Alabama, Arkansas, Georgia, Mississippi, South Carolina, and Texas—the states that formed the heart of the Cotton Belt—cotton comprised more than half of the value of all crops produced (U.S. Bureau of the Census 1913).

The widespread cultivation of cotton and its long harvest season, which commonly ran from September through December, generated a seasonal increase in demand for labor, as harvesting cotton was not mechanized until the mid-twentieth century. Since harvesting cotton involves repetitively picking lightweight fibers from their bolls and transferring them to a sack, it is a tedious task that is performed reasonably well by anyone over the age of five. In the cotton harvest, therefore, there was a high degree

[4] Using census data from 1920 and 1930, Lombardi (2019) finds that short-run weather shocks negatively affecting the cotton crop resulted in reduced school enrollment for blacks in the South. Meanwhile, Baker (2015), using summer rainfall instead of the boll weevil as an instrument for cotton production, demonstrates that black school enrollment rates declined in response to increased cotton production. We focus, instead, on long-run responses to a persistent negative agricultural shock.

of substitutability between adult and child labor (Bradley and Williamson 1918; Matthews and Dart 1924).[5] It is not surprising, then, that agriculture was by far the largest employer of children. In 1910, 34.4 percent of 10 to 15 year olds living in the South worked, of which 86.7 percent were employed in farming. Moreover, these youth made up 17 percent of the agricultural labor force in the South (U.S. Bureau of the Census 1924).[6] The majority of these child laborers undoubtedly worked the cotton fields.

While some adjustments were made to the school calendar in an attempt to accommodate farming cotton (Collins and Margo 2006), the lengthy harvest period could not be avoided altogether. Baker (2015) provides anecdotal evidence of the conflict between the demand for child labor in farming cotton and schooling in Georgia, but this conflict is not unique to that state. The superintendent of West Baton Rouge Parish, Louisiana, noted, "a falling off in attendance at several schools during the harvesting season ... due to the scarcity of labor and the need of the children in the cotton fields" (Louisiana Department of Education 1908, p. 48). His colleague in Calcasieu Parish expressed similar sentiments, observing, "times to pick cotton, the harvesting of rice and other crops take many out of school for a good part of the term" (Louisiana Department of Education 1902, pp. 60–61). To accommodate the surge in attendance following the conclusion of the cotton picking season, the superintendent of Tunica County, Mississippi, planned to hire an additional teacher in each African-American district (Mississippi Department of Public Education 1907, p. 95). Even Texas Superintendent Arthur Lefevre noted, "the average daily attendance being unusually low during 1902–3 on account of the high price and long picking season of the cotton crop of that year" (Texas Department of Education 1905, p. 7).

The Boll Weevil

The adult cotton boll weevil, *Anthonomus grandis*, is a small beetle, about six millimeters in length, and grayish in color, with a long snout

[5] Adult and child labor were also largely substitutable for one another in the tasks of chopping and hoeing. Chopping involves thinning cotton seedlings to achieve the desired interval between plants, while hoeing involves the removal of grass and weeds from around the seedlings. Together these are the most time consuming tasks, next to harvesting, accounting for nearly half of the pre-harvest labor in cotton production (Holley and Arnold 1938).

[6] In 1910, census enumerators were given specific instructions to inquire about the occupations of women and children, as well as men. Moreover, enumerators were instructed to record children working for their parents on a farm as farm laborers. Therefore, in comparison to earlier censuses, the 1910 census is an unusually good source of data on the labor force participation of children (Moehling 1999, p. 82, 2004, p. 79). Still, these figures might understate the extent of child labor in agriculture during the fall harvest since the 1910 census recorded employment on 15 April, at the beginning of the agricultural season when the demand for child labor was comparatively low.

and wings.[7] The westward expansion of cotton production in the United States during the nineteenth century eventually linked the Cotton Belt with the native habitat of the boll weevil, Mexico and Central America. The boll weevil first appeared in cotton fields near Brownsville, Texas, in 1892. It then spread north and east at a steady rate. By 1922, the boll weevil could be found in virtually all cotton counties in the United States, from Texas to North Carolina (Hunter and Coad 1923).

The life cycle of the boll weevil is closely intertwined with the cotton plant. Indeed, the insect lives inside the squares and bolls of the cotton plant for three of the four stages of its life cycle (egg, larvae, and pupae), and as an adult it feeds almost exclusively on the cotton plant.[8] The reason for the boll weevil's narrow appetite is that cotton is one of only a few plants (the others being wild flora with geographically small habitats) that provide the weevil with the nutrients required to produce the pheromones necessary for its reproduction. This dependence on cotton causes the insect to spend its entire life in or near cotton fields (Giesen 2011).

The boll weevil's spread had a disastrous effect on cotton production in the South. Damaged squares and bolls usually dropped from the plant after being fed upon and following the deposition of eggs by the boll weevil. As a result, heavily infested areas saw significant reductions in cotton output. In a county-level analysis, Lange, Olmstead, and Rhode (2009) demonstrate that within five years of the boll weevil's arrival cotton production fell by approximately 50 percent, due primarily to reduced yield but also reduced cotton acreage.

The boll weevil's negative effect on cotton yields decreased the returns to farming cotton relative to other crops. This caused farmers to substitute away from cotton in favor of more profitable alternatives, including corn, peanuts, potatoes, and sweet potatoes. These alternative crops generated less demand for child labor than cotton (Bradley and Williamson 1918; Matthews and Dart 1924).[9] Farmers also responded to the boll weevil by growing earlier maturing varieties of cotton and moving up the planting date, as the insect did the most damage late in the season

[7] Lange, Olmstead, and Rhode (2009) provide a description and history of the boll weevil.

[8] A cotton square refers to a young flower bud of the cotton plant, and a cotton boll is the fiber-producing fruit.

[9] A survey of Louisiana hill farms reports the labor requirements of various tasks in making different crops. The hours of labor required per acre in harvest, excluding plowing and hauling as tasks clearly unsuitable for children, was 46.7 hours for cotton, 5.7 for corn, 35.0 for peanuts, and 42.4 for sweet potatoes (Oates and Reynoldson 1921). Thus, even if cultivated acreage remained unchanged and the harvesting tasks of alternative crops were as equally suited to child labor as picking cotton, a shift away from cotton production would reduce the demand for child labor during the fall harvest season.

(Lange, Olmstead, and Rhode 2009). This reduced the conflict between the timing of the cotton harvest and fall schooling. Therefore, the arrival of the boll weevil, and the subsequent reduction in cotton production, had significant implications for the whole household.

DATA

To estimate the boll weevil's effect on educational attainment, as reported in the 1940 census, we need to know the year of arrival of the boll weevil in each Cotton Belt county and where men in 1940 were living as children when the infestation began. Therefore, we convert a USDA map tracking the boll weevil's progress into a machine-readable format and construct a linked sample of men in the 1940 census matched to themselves as children in either the 1900, 1910, or 1920 censuses.[10]

Mapping the Progress of the Boll Weevil

Information on the presence of the boll weevil was collected from the USDA-produced map tracking the insect's spread provided by Hunter and Coad (1923). For each year from 1892 through 1922 the counties of the Cotton Belt were assessed as being either boll weevil free, partially infested, or fully infested. A county was considered to be partially infested in a given year if and only if the line denoting the frontier of the boll weevil infestation in that year passed through the county leaving parts of the county visible on both sides of the line. Using this panel data, we then define three measures for each county that capture the timing of the arrival of the boll weevil: We define the *first arrival year* as the first year in which the county is coded as either partially or fully infested. The *complete infestation year* is the first year in which the county is coded as fully infested and not subsequently found to be boll weevil free or partially infested. The *year of infestation* is the ceiling of the average of the *first arrival year* and *complete infestation year*. We use *year of infestation* for each county to construct indicators for the presence of and exposure to the boll weevil throughout our analysis, which is consistent with the methodology of Lange, Olmstead, and Rhode (2009).[11]

In our examination of the long-run effects of the boll weevil on education, we restrict our analysis to certain counties based on characteristics

[10] Code and data for replication are provided by Baker, Blanchette, and Eriksson (2019).

[11] Online Appendix Table A.1 demonstrates that our main results are robust to alternatively constructing exposure to the boll weevil based on *first arrival year* and *complete infestation year*.

of the timing of the boll weevil's arrival, to increase the precision with which we date treatment. First, we begin with all Southern counties that Hunter and Coad (1923) identify as having territory within the Cotton Belt: 993 counties spread over 13 states.[12] Second, we exclude 102 counties that were boll weevil free and 47 counties that were only partially infested in 1922, approximately half of which are in the Western High Plains and Southwestern Tablelands regions of Texas and Oklahoma.[13] This restriction limits our sample to 844 counties, across 12 states, of the Cotton Belt that were fully infested by the boll weevil in 1922. Third, we remove counties that experienced a full retreat of the boll weevil (assessed as being boll weevil free in a year subsequent to being partially or fully infested), which drops 120 counties.[14] Fourth, we exclude 38 additional counties where the absolute difference between the *first arrival year* and the *complete infestation year* is greater than four years; 30 of these are in Texas, 6 in Arkansas, and 2 in Oklahoma.[15] Finally, we require that childhood location of individuals in our linked census data be observed prior to the boll weevil's arrival in that location, to ensure that our results are not biased by weevil induced migration.[16] This requirement further excludes from our sample 33 counties, which are all in Texas, where the boll weevil infestation began prior to 1900. This leaves us with a sample of 653 Southern counties across 12 states.[17]

Linked Census Data

We start by extracting all men born in the 13 Southern states of the Cotton Belt from the 1940 full count census data, provided by IPUMS and accessed through the NBER server. Then, we link these men to themselves as children in the 1900, 1910, or 1920 censuses. We restrict our search to individuals who were between the ages of 3 and 18 in the earlier census years; this means that they were between 23 and 58 years old in

[12] For the purpose of exposition, we discuss counties in this paragraph according to their 1920 borders.

[13] Climate—the dry summers and cold winters in particular—prevented the boll weevil from maintaining a presence in these regions (Hunter 1916, p. 23).

[14] Each of these 120 counties was reinfested by the boll weevil before 1922 and hence not excluded by the prior restriction. They are removed, however, because the proper treatment date is ambiguous. The median of these counties remained boll weevil free for two years before reinfestation.

[15] We show in Online Appendix Tables A.2 and A.3 that our main results are robust to dropping these restrictions, as well as imposing additional sample restrictions.

[16] Specifically, we impose the restriction that childhood census year be less than or equal to the *year of infestation* based on childhood county of residence.

[17] A map of the sample counties and those excluded due to these various restrictions is provided in Online Appendix Figure A.1.

1940.[18] In the case where men are found as children in two census years, we keep the earlier observation.[19]

We link individuals on the basis of first name, last name, birth year, race, and birth state using the standard procedure developed by Abramitzky, Boustan, and Eriksson (2012). The algorithm is described in Online Appendix Section A.1. For our main results, we aim to reduce false positive matches by requiring individuals to be unique, in terms of standardized first and last name, race, and birth state, in both censuses within a three-year age band (plus or minus one year). In particular, given that the census was not taken on the same date each year, we worry that, even with accurately reported age, calculated year of birth will be off by one year for individuals born in the months surrounding census dates.[20] Requiring uniqueness within three-year age bands eliminates the possibility of false positive matches due to this anomaly.

Our linking method results in a match rate of 27.37 percent for whites and 18.58 percent for blacks.[21] These numbers are somewhat higher than the literature because some individuals have a chance to match twice. For example, someone who is 43 years old in 1940 would have been 13 in 1910 and 3 in 1900. As mentioned earlier, we keep the individual at the youngest age we find him, but having two chances to find a match mechanically increases match rates.

Robustness samples carry out a range of alternative matching procedures: First, we relax the restriction that individuals be unique within three-year age bands; this maximizes the sample size at the risk of increasing false positive matches. Second, we require uniqueness by name, race, and birth state within five-year age bands (plus or minus two years of age). Third, we exclude matches whose calculated birth years differ by more than one year across the two censuses. Fourth, we impose the five-year age band uniqueness requirement and require that birth years differ by no

[18] When matching from 1940 backwards we do allow for the possibility that, for example, 23 year olds in 1940 could best match 1 or 2 year olds in 1920. However, we only retain in our linked sample individuals between the ages of 3 and 18, inclusive, in childhood census years.

[19] While this rule may seem arbitrary, it helps maximize the size of our final matched sample because, as mentioned earlier, we subsequently restrict the sample to individuals whose childhood location we observe prior to the boll weevil's arrival. However, our estimates are not diminished in magnitude or significance by keeping the later, and dropping the earlier, childhood observation.

[20] The 1900 census was conducted as of 1 June; the 1910 census as of 15 April; the 1920 census as of 1 January; and the 1940 census as of 1 April.

[21] Match rates are calculated as the proportion of individuals in the relevant group in 1940 for whom we find a match in at least one of the childhood census years, 1900, 1910, and 1920. For example, there were 2,390,606 black men aged 23 to 58 born in one of the 13 Southern states in the Cotton Belt recorded in the 1940 census. Of these, we find matches for 444,236 in one or more childhood census year. This yields a match rate of , or 18.58 percent.

more than one year. Finally, we return to just our initial three-year age band uniqueness restriction and match based on exact reported names instead of standardizing them. Balancing the trade-off between higher accuracy (limiting false matches) and match rates, as documented by Bailey et al. (2017), leads to our preferred choice of matching procedure, but our results are robust to a wide variety of methods.

We show in Table 1 that men in the matched sample come from higher socioeconomic backgrounds on average than the population at risk to match, which is common in the matching literature (see, e.g., Abramitzky, Boustan, and Eriksson 2012; Eriksson forthcoming). Specifically, childhood household heads of matched individuals are more likely to be literate, own their homes, and have higher occupational income scores, relative to those of the population.[22] Additionally, matched individuals themselves completed more years of schooling on average than the population. Because of this, we later weight linked individuals using inverse probability weights constructed so that the preferred matched sample is representative of the population on the variables shown in Table 1. Our results are not driven by differential selection into the matched sample.

ESTIMATION AND RESULTS

The Boll Weevil, Enrollment, and Attendance

Baker (2015) showed that the boll weevil increased the enrollment rate of black children by 4 percent, and had a positive but not statistically significant effect on the white enrollment rate, in Georgia. In this section, we confirm that the boll weevil had a positive effect on enrollment[23] using county-level data from 1900 to 1934 for six Southern states of the Cotton Belt: Alabama, Arkansas, Georgia, North Carolina, South Carolina, and

[22] We use the IPUMS-provided, 1950-based occupational income score, which is assigned to individuals based on reported occupation and gives the median total annual income for those employed in that occupation in 1950 according to published census data. We refer the interested reader to IPUMS documentation on the variable OCCSCORE for more details.

[23] County-level totals for enrollment are used as reported by the state departments of education, and cover those of all ages enrolled in public elementary and secondary schools. Intuitively, the enrollment of a school is a count of the number of pupils listed on the rolls of teachers employed at that school. For example, South Carolina law required teachers to "keep and furnish annually to the Trustees of the school district a list of all pupils that have attended the school during the preceding scholastic year" (S.C. Code § 1715 [1912]) for the purpose of computing enrollment. Since enrollment was calculated after the conclusion of the school term, it can then be interpreted as a count of the number of pupils attending school for at least one day during the term.

TABLE 1
COMPARING THE MATCHED SAMPLE TO THE FULL POPULATION

	White				Black			
	(1) Matched Sample		(2) Difference from Pop.		(3) Matched Sample		(4) Difference from Pop.	
Panel (A): Childhood Characteristics								
Age	9.4013	[4.4445]	–0.1121***	(0.0068)	9.4639	[4.4162]	–0.0710***	(0.0104)
Farm status	0.6398	[0.4801]	0.0048***	(0.0007)	0.6685	[0.4708]	0.0358***	(0.0011)
Urban status	0.1433	[0.3504]	0.0071***	(0.0005)	0.1055	[0.3072]	–0.0043***	(0.0007)
Year of BW arrival	1915.87	[5.6518]	0.3137***	(0.0087)	1915.42	[5.3168]	0.3897***	(0.0125)
Household head:								
Literate	0.8804	[0.3245]	0.0202***	(0.0005)	0.5391	[0.4985]	0.0500***	(0.0012)
Homeowner	0.5699	[0.4951]	0.0315***	(0.0008)	0.2665	[0.4421]	0.0368***	(0.0010)
Occupation score	17.2352	[11.0573]	0.3949***	(0.0168)	13.9372	[5.7784]	0.1867***	(0.0136)
Census year:								
1900	0.3738	[0.4838]	–0.0499***	(0.0007)	0.3895	[0.4876]	–0.0672***	(0.0011)
1910	0.4466	[0.4971]	0.0259***	(0.0008)	0.4707	[0.4991]	0.0230***	(0.0012)
1920	0.1795	[0.3838]	0.0240***	(0.0006)	0.1398	[0.3468]	0.0442***	(0.0008)
State of residence:								
Alabama	0.1552	[0.3621]	–0.0055***	(0.0006)	0.1515	[0.3585]	–0.0075***	(0.0008)
Arkansas	0.0413	[0.1989]	–0.0008*	(0.0003)	0.0354	[0.1849]	0.0014***	(0.0004)
Florida	0.0379	[0.1909]	0.0044***	(0.0003)	0.0409	[0.1980]	0.0039***	(0.0005)
Georgia	0.1532	[0.3601]	–0.0079***	(0.0005)	0.1501	[0.3572]	–0.0231***	(0.0008)
Kentucky	0.0016	[0.0402]	–0.0001*	(0.0001)	0.0007	[0.0267]	0.0001	(0.0001)
Louisiana	0.0558	[0.2296]	0.0010**	(0.0004)	0.0544	[0.2269]	–0.0083***	(0.0005)
Mississippi	0.0866	[0.2812]	–0.0002	(0.0004)	0.1361	[0.3429]	–0.0127***	(0.0008)
North Carolina	0.1934	[0.3949]	0.0101***	(0.0006)	0.1941	[0.3955]	0.0459***	(0.0009)
Oklahoma	0.0036	[0.0600]	0.0007***	(0.0001)	0.0007	[0.0255]	0.0004***	(0.0001)
South Carolina	0.1114	[0.3147]	0.0077***	(0.0005)	0.1423	[0.3494]	–0.0138***	(0.0008)
Tennessee	0.0911	[0.2878]	0.0022***	(0.0004)	0.0515	[0.2210]	0.0108***	(0.0005)
Texas	0.0689	[0.2532]	–0.0116***	(0.0004)	0.0423	[0.2013]	0.0029***	(0.0005)
Observations	513,378		3,114,232		197,679		2,355,430	
Panel (B): Adulthood Characteristics								
Age	37.3596	[9.2200]	–0.2292***	(0.0080)	37.8861	[9.1852]	0.2393***	(0.0152)
Years of schooling	8.5412	[3.5866]	0.2481***	(0.0031)	5.5008	[3.3625]	0.2558***	(0.0055)
Occupation score	22.4712	[11.5376]	0.5602***	(0.0098)	16.7011	[8.0959]	0.2816***	(0.0133)
Observations	1,761,616		8,196,775		444,236		2,834,842	

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.
Notes: The matched sample includes male children between the ages of 3 and 18 in 1900, 1910, and 1920, matched to adult educational outcomes in 1940. Columns (1) and (3) report means for the matched sample with standard deviations in brackets. Coefficients in Columns (2) and (4) are from regressions of individual characteristics of interest on an indicator for being in the matched sample. Thus, these columns show the difference between the matched sample and population with standard errors in parentheses.
Source: See the text.

Tennessee.[24] This yields a sample of 327 geographically consistent counties which were invaded by the boll weevil between 1906 and 1922.[25]

To begin, we estimate the effect of the boll weevil on enrollment using the following specification:

$$ln(enrollment)_{ct} = \beta boll\ weevil_{ct} + \gamma ln(teachers)_{ct} + \theta_c + \theta_t + \phi_c t + \varepsilon_{ct}, \quad (1)$$

where *boll weevil*$_{ct}$ is an indicator for the presence of the boll weevil which takes the value one if year t is greater than or equal to *year of infestation* in county c, and zero otherwise.[26] Additionally, we include the natural log of the number of teachers, as a county-level control for school quality; county fixed effects θ_c; year fixed effects θ_t; and county-specific linear time trends $\phi_c t$. Equation (1), therefore, represents a difference-in-difference approach with identification of β exploiting the differential timing of the *year of infestation* across counties. For example, the change in log enrollment, from before and after the boll weevil's arrival, in newly infested counties is compared with the change in log enrollment over the same time period in as yet boll weevil–free counties, while controlling for unobserved county-level variables that remain fixed over time, to capture the level shift in log enrollment associated with the boll weevil infestation.

Table 2 presents the results. Columns (1) through (4) provide estimates for whites, while Columns (5) through (8) show estimates for blacks. The coefficients on the presence of the boll weevil indicator suggest the infestation had a positive and statistically significant effect on enrollment of both races. The coefficient in Column (4) of 0.027 suggests the boll weevil infestation led to a 2.7 percent ($e^{0.027} - 1$) increase in school enrollment of white children. That is an increase in enrollment of 107 white children in the average county. The coefficient in Column (8) of

[24] Data for the period 1910–1934 were generously provided by Celeste K. Carruthers and Marianne H. Wanamaker (see, for a description of the data, Carruthers and Wanamaker 2017). Data for 1900–1909 were collected from annual or biennial reports of the state departments of education for those years.

[25] Of the 545 counties that existed at some point during the 1900–1934 period in the six-state sample, the boll weevil affected 466. Four additional restrictions are imposed on this sample of 466 boll weevil affected counties: First, 58 counties are dropped due to border changes. Second, 50 additional counties where the boll weevil entered and then fully retreated before reentering are excluded, because the treatment date is ambiguous. Third, a further 19 counties that were only partially infested in 1922 are also excluded. Fourth, 12 counties with little to no black population are excluded for comparability. A map of the sample counties and those excluded due to these various restrictions is provided in Online Appendix Figure A.2.

[26] We use log enrollment, rather than enrollment rate, as the dependent variable because the school-age population is not available for all states in all years of our sample.

TABLE 2
ESTIMATES OF THE BOLL WEEVIL'S EFFECT ON LN(ENROLLMENT)

	White				Black			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Boll weevil	0.049***	0.022*	0.034***	0.027***	0.062***	0.024*	0.065***	0.050***
	(0.011)	(0.009)	(0.008)	(0.007)	(0.017)	(0.012)	(0.014)	(0.012)
ln(teachers)?	No	Yes	No	Yes	No	Yes	No	Yes
Time trends?	No	No	Yes	Yes	No	No	Yes	Yes
Observations	7,808	7,658	7,808	7,658	7,769	7,597	7,769	7,597
Number of counties	327	327	327	327	327	327	327	327
R^2	0.962	0.975	0.984	0.987	0.958	0.976	0.976	0.982

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.
Notes: The dependent variable is the natural log of enrollment. Standard errors adjusted for clustering by county in parentheses. All regressions include year and county fixed effects. Columns (2), (4), (6), and (8) include the natural log of the number of teachers employed as a control for the supply of education. Columns (3), (4), (7), and (8) also include county-specific linear time trends.
Source: See the text.

0.050 suggests the boll weevil infestation led to a 5.1 percent increase in school enrollment of black children, an increase of 123 black children in the average county. The estimated coefficients for blacks are consistently higher than those for whites, and in specifications including county-specific linear time trends they are statistically different from one another at the 90 percent confidence level.

To better understand the timing of the weevil's effect on enrollment, we replace the boll weevil indicator with 6 leads and 12 lags for the year of the insect's arrival. The specification becomes

$$ln(enrollment)_{ct} = \sum_{k\geq -6, k\neq 0}^{k\leq 12} \beta_k * \mathbf{1}\{t - BW_c = k\} + \gamma ln(teachers)_{ct} + \theta_c + \theta_t + \phi_c t + \varepsilon_{ct}, \quad (2)$$

where BW_c represents the *year of infestation* of county *c*. Equation (2) also includes the natural log of the number of teachers (to control for the supply of education), county fixed effects, year fixed effects, and county-specific linear time trends. The indicator for the initial year of arrival is omitted; therefore, all effects are relative to the arrival date of the boll weevil. The coefficient β_{12} represents the average effect 12 or more years after being infested (relative to the year of arrival), while β_{-6} gives the effect six or more years before infestation.

Figure 1 presents transformed coefficients of the leads and lags. Solid lines provide the main effects, while dashed lines represent 95 percent

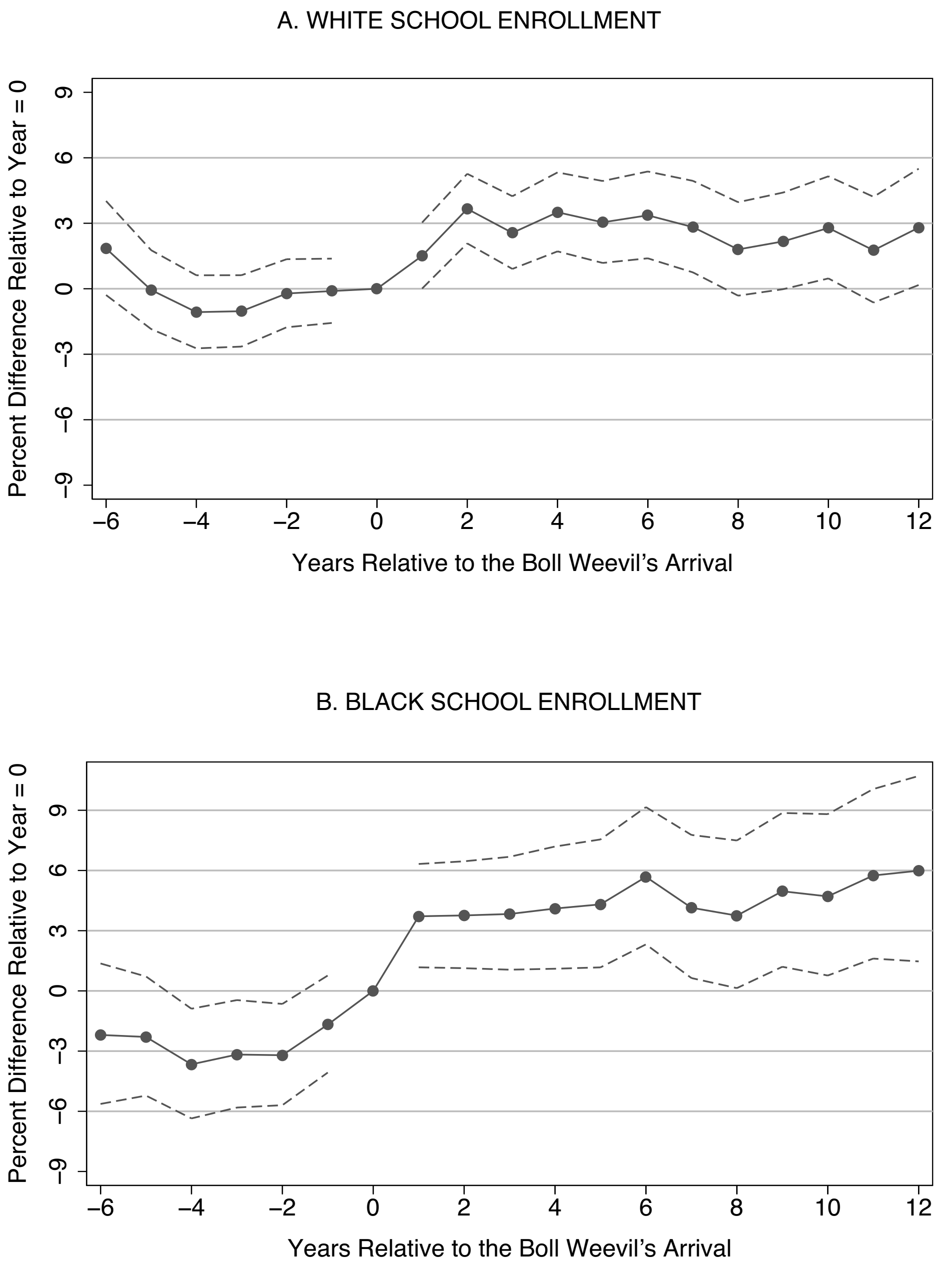

FIGURE 1
SCHOOL ENROLLMENT RELATIVE TO THE BOLL WEEVIL'S ARRIVAL

Notes: The y-axis shows the percent change in enrollment relative to the boll weevil's year of arrival. The dashed lines indicate 95 percent confidence intervals.
Source: See the text.

confidence intervals. Panel (A) shows school enrollment of whites remained fairly flat in the years leading up to the boll weevil's arrival, with some decline in enrollment occurring six or more years prior to contact. Relative to the *year of infestation*, white enrollment had increased by approximately 3 percent two years later, with the estimated effect remaining near 3 percent up to 12 years later. Panel (B) shows the school enrollment of blacks was also relatively steady prior to the boll weevil, with an increase in enrollment occurring in the *year of infestation*. Three years after the arrival date, black enrollment had increased by approximately 4 percent compared to the *year of infestation*. These enrollment gains climb further to 6 percent and persist up to 12 years after the initial boll weevil infestation.

While the boll weevil had a positive effect on enrollment, it is unclear if this effect translated to higher attendance. In Table 3, we resolve this uncertainty by showing results of specifications replacing enrollment in Equation (1) with average daily attendance. The coefficient in Column (4) of 0.015 suggests the boll weevil infestation led to a 1.5 percent ($e^{0.015} - 1$) increase in school attendance of white children, yet this coefficient is not statistically significant at the 90 percent confidence level. Meanwhile, the coefficient of 0.033 in Column (8) suggests the boll weevil's arrival led to a statistically significant 3.4 percent increase in school attendance of black children. Comparable coefficients across race are not statistically different from one another, even though the point estimates are higher for blacks.

Unlike Baker (2015), who only finds a statistically significant response for blacks, we find statistically significant effects of the boll weevil on enrollment and attendance for both races when considering a broader set of states. The differences in our estimates are perhaps accounted for by heterogeneous responses to the boll weevil across states or greater precision due to the larger sample size.[27] Nevertheless, many white children also worked in the cotton fields, so finding that the boll weevil had a positive effect on white enrollment is not particularly surprising. However, it should also be noted that our results are not directly comparable to those of Baker, as his dependent variable is enrollment rate

[27] In Online Appendix Table A.4 we provide estimates of the boll weevil's effect on enrollment after restricting the included counties to match Baker's (2015) sample for Georgia. The results show school enrollment of blacks increased by a statistically significant 5.2 percent following the boll weevil infestation. White enrollment, on the other hand, increased by 0.6 percent according to the point estimate, but this result is not statistically different from zero at the 90 percent confidence level. These estimates are consistent with Baker's findings for Georgia.

TABLE 3
ESTIMATES OF THE BOLL WEEVIL'S EFFECT ON LN(ATTENDANCE)

	White				Black			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Boll weevil	0.042**	0.011	0.024*	0.015	0.053*	0.010	0.048**	0.033*
	(0.014)	(0.013)	(0.011)	(0.011)	(0.022)	(0.017)	(0.017)	(0.015)
ln(teachers)?	No	Yes	No	Yes	No	Yes	No	Yes
Time trends?	No	No	Yes	Yes	No	No	Yes	Yes
Observations	7,699	7,621	7,699	7,621	7,665	7,560	7,665	7,560
Number of counties	327	327	327	327	327	327	327	327
R^2	0.944	0.958	0.970	0.973	0.937	0.959	0.960	0.969

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.
Notes: The dependent variable is the natural log of average daily attendance. Standard errors adjusted for clustering by county in parentheses. All regressions include year and county fixed effects. Columns (2), (4), (6), and (8) include the natural log of the number of teachers employed as a control for the supply of education. Columns (3), (4), (7), and (8) also include county-specific linear time trends.
Source: See the text.

while we consider enrollment without controlling for the school-age population.[28]

In fact, the results presented in this section must be interpreted with some degree of caution with respect to mechanisms, because we do not account for the school-age population. Boll weevil–driven in-migration prior to infestation and out-migration after infestation would both bias our estimates of the insect's impact on enrollment toward zero, assuming that some migrants were enrolled in school. While the opposite, out-migration prior to infestation and in-migration after infestation, would bias our estimates upward. While we cannot rule out the possibility that our estimates are biased by boll weevil–induced migration, Lange, Olmstead, and Rhode (2009) show that population increased in advance of the boll weevil infestation, and declined afterward. This suggests that our estimates of the weevil's impact on enrollment are biased toward zero, or conservative.[29] Thus, they provide suggestive evidence that children,

[28] The difference between Baker's (2015) results and those presented here could be explained, for example, by increasing school-age population coincident with the onset of the boll weevil infestation, with the increase in white population being relatively greater than that of the black population.

[29] If a boll weevil induced increase in fertility explained the uptick in enrollment, we would expect enrollment to increase with a lag of five or more years. Furthermore, Bloome, Feigenbaum, and Muller (2017) find a decline in the share of African Americans who married at young ages in response to the boll weevil, which would likely have a negative effect on fertility rates.

previously not in school, enrolled after the arrival of the boll weevil, or already enrolled children stayed in school longer than they would have in the absence of the boll weevil infestation.

While increased enrollment following the boll weevil's arrival is suggestive of improved student outcomes, it does not directly follow that the weevil caused increased educational attainment, which is of much greater economic importance than school enrollment. Even if we were able to control for the school-age population, it would be necessary to assume that newly enrolled students were attending classes and completing grades to use the enrollment results to derive implications for educational attainment. As enrollment merely implies attendance for at least one day during the school term, these are not trivial assumptions. Indeed, the estimates showing average daily attendance was less responsive than enrollment to the boll weevil suggest that new enrollees had worse than average attendance records. It is certainly possible that the boll weevil increased enrollment while having little to no effect on grade completion, and thereby educational attainment. Therefore, we use our linked-census sample to directly investigate the weevil's impact on years of schooling, in the next section.

Boll Weevil and Long-Run Schooling Outcomes

In our linked census data we follow individuals over time, but observe their years of schooling only in the 1940 census manuscripts, when they are adults. Therefore, we examine the boll weevil's impact on educational attainment by comparing adjacent birth cohorts. Our main specification is a differences-in-differences approach where we compare those who were children (ages 4 through 18), and still eligible for public schooling, when the boll weevil arrived to older cohorts (ages 19 through 30), who likely completed their schooling prior to the infestation.[30] This approach is represented by the following regression equation:

$$y_{ica} = \sum_{k=0}^{4} \beta_k * \mathbf{1}\{3(5-k)+1 \leq BW_c - a \leq 3(6-k)\} + \gamma X_i + \theta_c + \theta_a + \varepsilon_{ica}, \qquad (3)$$

[30] Because we only include in the sample individuals observed between the ages of 3 and 18 in 1900, 1910, or 1920, age at exposure is necessarily based on childhood county of residence. Therefore, we are comparing those exposed to the boll weevil as children to individuals who grew up in the same county but were not exposed to the boll weevil until adulthood. Limiting the sample to those aged 3–18 in early census years has two primary benefits: First, we are able to include controls for important family background characteristics. Second, this avoids concerns that the comparison group is affected by migration, which would bias our estimates.

where y_{ica} is an educational outcome of interest for individual i, who was born in year a and observed in county c in childhood (between the ages of 3 and 18, inclusive).[31] Boll weevil exposure occurs at age $BW_c - a$, where BW_c represents the *year of infestation* of county c.[32] Therefore, the indicator function ($\mathbf{1}\{3(5-k)+1 \leq BW_c - a \leq 3(6-k)\}$) returns a one if the individual was aged $3(5-k)+1$ through $3(6-k)$ in the *year of infestation*, and zero otherwise. The θ_c and θ_a are county and birth cohort fixed effects, respectively, while X_i is a vector of individual-level family background controls as measured in the childhood census year. Family background controls include childhood household head's occupational income score, homeownership status, and literacy, as well as indicators for urban location and farm residence. The β_k are the coefficients of interest, representing the average treatment effect of the boll weevil on individuals in the relevant exposure cohort. Because the boll weevil arrived at different times in different counties we are able to separately identify the full set of birth cohort and county fixed effects as well as the treatment effects for those exposed at age 18 or under.

The first two columns of Table 4 show the results of regressions following Equation (3) with years of schooling as the dependent variable. Children who were first exposed to the boll weevil between the ages of 4 and 6, inclusive, saw the greatest gains in years of schooling as a result. White children exposed at ages 4–6 attained 0.2669 more years of schooling on average, as compared with those exposed at ages 19–30 (shown in Column [1]). White children exposed at 7–9 years of age experienced similar relative gains of 0.2364 years of schooling. Both results are statistically significant at the 99.9 percent confidence level. Whites exposed to the boll weevil at older ages of 10–12 and 13–15 saw smaller increases in educational attainment of 0.1481 and 0.1019

[31] The restriction that individuals be observed at ages 3–18 in early census years is necessary to determine childhood location and household characteristics. However, it also imposes a binding upper bound on the oldest age-at-exposure cohort included in our sample for counties with a *year of infestation* between 1900 and 1911, which represent 38 percent of sample counties. Online Appendix Figure A.3 indicates for each county the oldest age-at-exposure cohort included in our sample. While including childhood county of residence fixed effects should yield unbiased estimates of the β_k, in the appendix we also provide estimates excluding early infested counties in order to generate a more balanced sample on this dimension. Online Appendix Table A.5 demonstrates the robustness of our main results to limiting the sample to counties where the boll weevil arrived after 1906, 1909, and 1912. Relatedly, the restriction that we observe childhood residence prior to the boll weevil's arrival in that location also generates cross-county heterogeneity in age-at-exposure cohorts included in the sample, by imposing a lower bound. Online Appendix Figure A.4 indicates for each county the youngest age-at-exposure cohort included in our sample.

[32] Throughout the remainder of the text, $BW_c - a$ we refer to as age at exposure. Additionally, we define exposure cohorts as groups of individuals that were first exposed to the boll weevil at the same age.

TABLE 4
ESTIMATES OF THE BOLL WEEVIL'S EFFECT ON LONG-RUN EDUCATIONAL OUTCOMES

	Years of Schooling		Completed 8th Grade	
	(1) White	(2) Black	(3) White	(4) Black
Age exposed:				
4–6	0.2669***	0.3579***	0.0177*	0.0327***
	(0.0608)	(0.0724)	(0.0086)	(0.0093)
7–9	0.2364***	0.2427***	0.0263***	0.0249***
	(0.0500)	(0.0601)	(0.0069)	(0.0074)
10–12	0.1481***	0.1514**	0.0175***	0.0128*
	(0.0378)	(0.0467)	(0.0052)	(0.0058)
13–15	0.1019***	0.1409***	0.0140***	0.0116*
	(0.0287)	(0.0390)	(0.0040)	(0.0047)
16–18	0.0452*	0.0609*	0.0058+	0.0035
	(0.0226)	(0.0304)	(0.0031)	(0.0036)
Observations	429,757	170,839	429,757	170,839
R^2	0.1627	0.0908	0.1169	0.0566
Dependent variable:				
Mean	8.3557	4.9768	0.5735	0.1980
Standard deviation	3.6740	3.2239	0.4946	0.3985

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.
Notes: The dependent variables are given in the column headings. Standard errors adjusted for clustering by childhood county of residence are in parentheses. All specifications include year of birth fixed effects, childhood county of residence fixed effects, and controls for family background. Family background controls include childhood household head's occupational score, homeownership status, and literacy, as well as indicators for urban location and farm residence.
Source: See the text.

years of schooling, respectively. If the opportunity cost of schooling is increasing in years of schooling and work experience, and the timing of the transition from schooling to working maximizes expected present value of lifetime earnings, then a modest negative shock to the opportunity cost of schooling will induce those in school to stay in school longer but will affect few who have already left school for work.[33] Thus, not surprisingly, whites exposed to the boll weevil between the ages of 16 and 18 saw an increase in educational attainment of only 0.0452 years on average, relative to those exposed at ages 19–30, and this result is statistically different from zero only at the 95 percent confidence level.

[33] Rather than the opportunity cost of schooling increasing in work experience, the same predictions can be generated by the presence of a cost associated with returning to school after having left school for work (e.g., having to repeat the last grade completed).

Column (2) shows remarkably similar gains in educational attainment for blacks resulting from the boll weevil infestation. The only coefficient that differs notably, but not statistically, from those of whites is that for the youngest exposure cohort; black children exposed at ages 4–6 attained 0.3579 more years of schooling on average, compared to those exposed at ages 19–30.

Columns (3) and (4) of Table 4 instead consider the boll weevil's impact on eighth grade completion.[34] Again, children exposed to the boll weevil at younger ages were more responsive in their likelihood of completing at least eight years of schooling. As shown in Column (3), white children exposed at ages 7–9 were more likely to complete the eighth grade by 2.63 percentage points, compared with those exposed at ages 19–30. Black children saw a similar increase, of 2.49 percentage points, in the likelihood of graduating grammar school due to boll weevil exposure at ages 7–9, as indicated in Column (4). Differences in point estimates are not statistically significant across race and are generally quite small.

In interpreting the β_k as average treatment effects, we are implicitly making the assumption that individuals between the ages of 19 and 30 when the boll weevil arrives in their childhood county of residence do not alter their years of schooling or grade completion rates in response. Indeed, there are structural reasons to expect 19 to 30 year olds to be little affected by the weevil; children generally attend school, while adults generally work. When the outcome of interest is eighth grade completion, this seems like a quite reasonable assumption as the typical age for finishing eighth grade is 14. While age-for-grade statistics were not commonly published by Southern states during the early twentieth century, a report on the 1923 school year in Tennessee reveals that approximately 1.3 percent of whites and 3.3 percent of blacks ages 19–21 were enrolled in grades one through eight (Tennessee Department of Education 1924). Thus, the boll weevil is unlikely to have had a discernible impact on eighth grade completion for exposure cohorts over the age of 19. However, with

[34] Eighth grade completion is an indicator that takes the value one if years of schooling is greater than or equal to eight, and zero otherwise. We focus on eighth grade completion because it represents graduation from grammar school in most states in our sample (exceptions include Georgia and Louisiana, where grammar school extends through seventh grade, and Alabama, where the cutoff is sixth grade). We believe this to be a more relevant milestone for the period than high school graduation, since access to high schools was limited, especially for black children. Online Appendix Tables A.6 and A.7 provide, for white and black children respectively, comparable estimates with indicators for completion of grades 1–12 as dependent variables. White children are less responsive than black children at lower grade levels, which is likely due to already high grade completion rates for whites prior to boll weevil exposure. Meanwhile, black children are less responsive than white children at higher grade levels, possibly due to limited high school access, as well as differences in the distribution of years of schooling across race.

years of schooling as the outcome, the validity of this assumption is not as clear. The 1920 census shows that 4.3 (2.2) percent of white (black) Southern-born men ages 19–30 were still attending school.[35] If the boll weevil had a positive (negative) effect on the educational attainment of individuals in older exposure cohorts, then the coefficients shown in Table 4 may understate (overstate) the true effects. Therefore, we next consider the possibility of pre-trends (or the validity of 19–30 year old exposure cohorts as a comparison group).[36]

It is common for studies taking an empirical approach of the form specified in Equation (3) to also show results of a regression equation with both leads and lags on treatment, to allay concerns regarding the possibility of preexisting trends, like the following:

$$y_{ica} = \sum_{k\geq -4, k\neq -1}^{k\leq 4} \beta_k * \mathbf{1}\{3(5-k)+1 \leq BW_c - a \leq 3(6-k)\} + \gamma X_i + \theta_c + \theta_a + \varepsilon_{ica}. \quad (4)$$

Such results are often shown graphically. However, Borusyak and Jaravel (2017) point out the results of such a regression are underidentified due to the absence of a truly untreated control group. In our case, it is not possible to fully disentangle the birth cohort effects θ_a from exposure cohort effects β_k in the presence of county fixed effects θ_c. The simplest solution to this underidentification problem would be to drop the county fixed effects. However, county fixed effects are necessary for identification since the timing of the boll weevil's arrival is not independent of time invariant county characteristics (e.g., longitude, latitude, and climate).[37]

Instead, we follow the recommendation of Borusyak and Jaravel (2017), imposing the restrictions that $\hat{\beta}_{-4} = \hat{\beta}_{-1} = 0$. In practice, this means dropping two indicators for older exposure cohorts: those exposed at 19–21 and 28–30. Then, an F-test on the joint significance of the remaining pre-trends provides a test for the existence of non-linear pre-trends. Given the nonlinearity in attendance rates between 19–21 year olds and older age groups, we would expect to see evidence of non-linear pre-trends if those

[35] Calculated using the USA Full Count sample for 1920 (Ruggles et al. 2018). We exclude those born in Kentucky, Oklahoma, Virginia, West Virginia, Maryland, Delaware, and the District of Columbia, to better compare with our sample of boll weevil–affected counties.

[36] Using a wide age-at-exposure range for the comparison group is crucial to our ability to test for pre-trends. However, we show in Online Appendix Table A.8 that the results presented in Table 4 are robust to restricting the comparison group age-at-exposure range to 19–27 or 19–24.

[37] In the case of Equation (2), with results shown in Figure 1, we address the underidentification problem by having β_{12} represent the average effect 12 *or more years* after, and β_{-6} 6 *or more years* before, being infested. This approach is not possible here because we cannot observe the location of individuals prior to their birth.

TABLE 5
THE BOLL WEEVIL'S EFFECT ON LONG-RUN EDUCATIONAL OUTCOMES WITH PRE-TRENDS

	Years of Schooling		Completed 8th Grade	
	(1) White	(2) Black	(3) White	(4) Black
Age exposed:				
4–6	0.2695***	0.3648***	0.0167+	0.0370***
	(0.0635)	(0.0742)	(0.0089)	(0.0097)
7–9	0.2374***	0.2494***	0.0253***	0.0285***
	(0.0516)	(0.0623)	(0.0071)	(0.0078)
10–12	0.1467***	0.1587**	0.0163**	0.0156**
	(0.0391)	(0.0483)	(0.0053)	(0.0060)
13–15	0.0980**	0.1486***	0.0129**	0.0135**
	(0.0300)	(0.0404)	(0.0042)	(0.0050)
16–18	0.0389	0.0688*	0.0046	0.0047
	(0.0238)	(0.0314)	(0.0032)	(0.0038)
22–24	–0.0227	0.0233	–0.0037	0.0020
	(0.0194)	(0.0274)	(0.0027)	(0.0035)
25–27	–0.0281	0.0104	–0.0019	–0.0057
	(0.0233)	(0.0308)	(0.0033)	(0.0037)
Observations	429,757	170,839	429,757	170,839
R^2	0.1627	0.0908	0.1169	0.0567
F-stat on pre-trend	1.1172	0.3632	0.9300	1.8519

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.
Notes: The dependent variables are given in the column headings. Standard errors adjusted for clustering by childhood county of residence are in parentheses. All specifications include year of birth fixed effects, childhood county of residence fixed effects, and controls for family background. Family background controls include childhood household head's occupational score, homeownership status, and literacy, as well as indicators for urban location and farm residence.
Source: See the text.

exposed at 19–21 were modifying their schooling behavior in response to the boll weevil. The results of these restricted regressions are shown in Table 5.[38] The fact that the coefficients on the 4–6 to 16–18 exposure cohort indicators are little affected, compared to those presented in Table 4, is encouraging. The coefficients on the indicators for exposure to the

[38] Additionally, we show in Online Appendix Table A.9 that our main results are robust to excluding the 19–21 exposure cohort from the omitted category and top coding years of schooling at 12. This provides additional support for interpreting the coefficients presented in Table 4 as average treatment effects.

boll weevil at ages 22–24 and 25–27 are small in magnitude and not statistically different from zero. Moreover, the F-statistics on their joint significance are below 2.31, the critical value of F for the $\alpha = 0.1$ significance level, so we cannot reject the null hypothesis that the coefficients on the remaining pre-trends jointly equal zero. This provides reasonable reassurance against the presence of non-linear pre-trends, as would be likely if older exposure cohorts modified their schooling behavior due to, or in anticipation of, the boll weevil's arrival.[39]

We also investigate whether the boll weevil's effect on years of schooling differed according to the intensity of cotton production in one's childhood county of residence. We use the ratio of acres in cotton to improved farm acreage in 1889 as a measure of cotton intensity prior to the boll weevil's arrival.[40] Figure 2 reveals, for our linked sample, the county-level distribution of the share of improved acreage planted in cotton in 1889. The median county in our sample grew cotton on 30.4 percent of improved acreage, while the 95th percentile was 50.6 percent.

We modify Equation (3) by interacting each exposure cohort indicator with a 4th-order polynomial in cotton intensity.[41] Figure 3 reveals the predicted impact of exposure to the boll weevil at ages 4–6 (Panel [A]), 7–9 (Panel [B]), 10–12 (Panel [C]), and 13–15 (Panel [D]) on years of schooling as it varies by cotton intensity in childhood county of residence.[42] Figure 3 shows substantial positive effects on educational attainment, even in counties with little cotton production in 1889. The boll weevil's effect on years of schooling increases further as intensity of cotton production rises to approximately 15 percent. The youngest exposure cohort (4–6 year olds in Panel [A]) experienced a 0.15 year increase

[39] We can examine these educational outcomes because years of schooling and completion of grammar school are largely determined before the boll weevil arrives for those exposed at 19 years of age and older, making them an arguably valid comparison group for the estimation of average treatment effects. However, labor market outcomes of those in the 19–30 year old exposure cohorts were likely affected by the boll weevil, as were those of younger exposure cohorts. Therefore, we lack a valid comparison group for quantifying the boll weevil's effect on occupational income score, wages, employment status, etc., using this data.

[40] This approach is similar to Lange, Olmstead, and Rhode (2009), but they use total farm acres as the denominator. The U.S. Census Office (1890) clarified for enumerators of the farm schedules that "land once plowed is *improved* unless afterward abandoned for cultivation" and "rocky, hill, and mountain pastures are *not improved*" (p. 36). However, such pasture land, and otherwise uncultivated land, was to be included in total farm acreage. Therefore, using improved acreage to calculate cotton intensity allows us to better capture the share of cultivated land devoted to cotton and improves cross-county comparability.

[41] We use a 4th-order polynomial because it maximizes *Adj.* R^2. Online Appendix Figures A.5 and A.6 show results using instead a 3rd-order and 2nd-order polynomial in cotton intensity, respectively. The 3rd-order and 4th-order polynomials in cotton intensity produce strikingly similar estimates.

[42] Comparable graphs of the boll weevil's impact on eighth grade by cotton intensity completion are shown in Online Appendix Figure A.7.

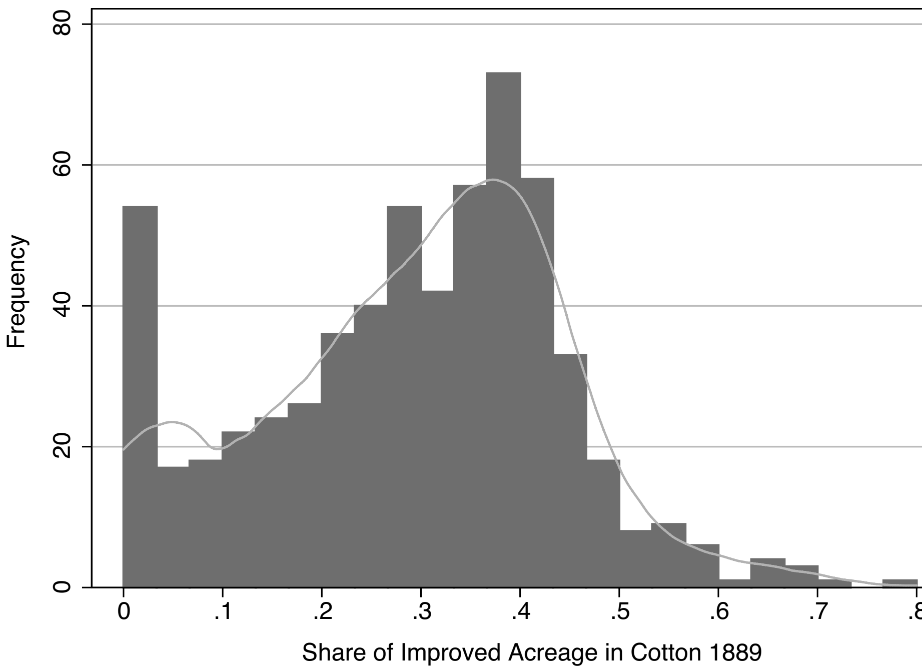

FIGURE 2
DISTRIBUTION OF SAMPLE COUNTIES BY INTENSITY OF COTTON PRODUCTION

Notes: Cotton intensity is defined as the ratio of acres planted in cotton to improved farm acreage as measured in 1889. The solid line provides the kernel density estimate of the distribution.
Source: Haines and ICPSR (2010).

in schooling as cotton acreage went from 0 to 15 percent, that is a 50 percent increase in the estimated effect relative to those in counties with no cotton acreage. Then, the impact of the weevil declines from its peak as cotton intensity increases from 15 to 50 percent of farm acreage. Panel (A) shows that the decline for the youngest exposure cohort mirrors the rise, falling by approximately 0.15 years of schooling. However, the 7–9 (Panel [B]) and 10–12 (Panel [C]) exposure cohorts saw larger declines relative to the preceding rise. For example, point estimates for the 7–9 exposure cohort suggest gains in educational attainment of 0.32 years in counties with little cotton acreage, 0.40 years in counties with 15 percent of acreage in cotton, and 0.18 in counties with 50 percent of improved land in cotton.

The predicted effects by cotton intensity suggest that children in counties where cotton production was marginally optimal benefited the most from the weevil with respect to educational attainment, while children in counties that were heavily dependent on cotton cultivation benefited to a lesser extent. At first glance, this might seem counterintuitive.

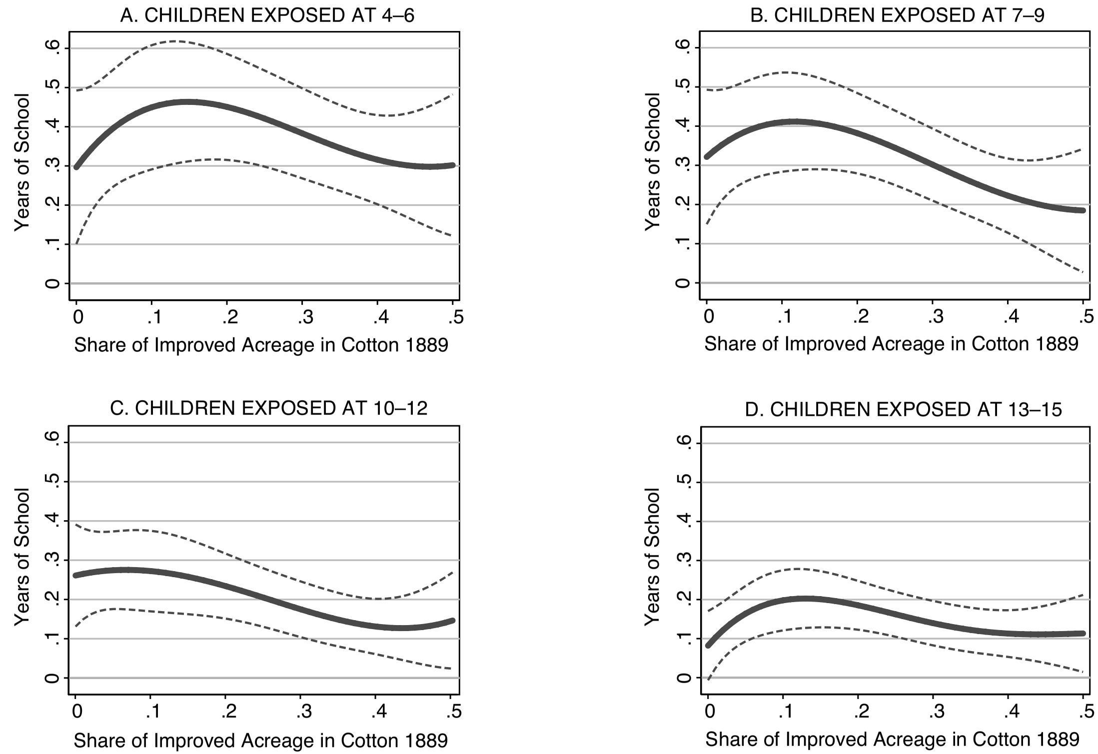

FIGURE 3

IMPACT OF THE BOLL WEEVIL ON YEARS OF SCHOOLING BY INTENSITY OF COTTON FARMING IN 1889

Notes: The y-axis shows the difference in years of schooling relative to individuals exposed to the boll weevil between the ages of 19 and 30, inclusive. The dashed lines indicate 95 percent confidence intervals. Adjusted $R^2 = 0.2038$.
Source: See the text.

However, bear in mind that these estimates reflect the overall impact of the boll weevil on educational attainment; the expected gains in attainment resulting from the decline in the opportunity cost of schooling, due to substitution away from cotton, are partially offset by an increased desire for children to generate earnings as a result of the weevil's negative effect on household income. Therefore, the estimated effects shown in Figure 3 would be expected if the propensity to switch crops after infestation decreased, or the magnitude of the income shock due to the weevil increased, with cotton intensity. Indeed, empirical and anecdotal evidence suggests that the boll weevil did greater damage where cotton was more intensely grown and substitution away from cotton, in terms of acreage, was greatest where the crop was marginally optimal (Giesen 2011; Lange, Olmstead, and Rhode 2009). The estimates presented in Figure 3, then, are consistent with expectations that education gains were most pronounced in locations where substitution away from cotton was greatest and income effects were limited; while counties where the boll weevil had large negative effects on income, and yet substitution away from cotton was relatively limited, saw more modest gains in educational attainment.

There is also a possibility that migration plays a role in generating the patterns observed in Figure 3. Lange, Olmstead, and Rhode (2009) show that counties with high levels of cotton intensity, relative to those with moderate levels, experienced greater declines in population, by approximately a factor of two, following the boll weevil's arrival. They suggest these population declines could be explained by migration from infested counties to cotton counties not yet affected by the boll weevil. If this accurately describes the pattern of migration, then a larger share of children whom we observe in high-cotton-intensity counties, relative to children observed in moderate-cotton-intensity counties, would have moved just after the weevil's arrival to boll weevil–free cotton-producing counties. These children would not have experienced a reduction in the opportunity cost of schooling generated by substitution away from cotton to alternatives less suited to child labor, as would the non-migrant population, and thus we would not expect to see gains in educational attainment due to boll weevil exposure for these children. Therefore, the estimated effect of the boll weevil on the educational attainment of children located in high-cotton-intensity counties before infestation could be lower than in moderate-cotton-intensity counties, even if the weevil's effect among the non-migrant population was constant in cotton intensity. Alternatively, it is possible that these children migrated, instead, to urban areas, or the North, which would have provided improved educational opportunities.

Regardless, boll-weevil induced migration should not be interpreted as biasing our results since our goal is to quantify the overall effect of the weevil on educational attainment, inclusive of effects running through the migration channel.

ROBUSTNESS OF MAIN RESULTS

Concurrent Shocks to Education

A possible threat to validity regards shocks to educational resources and requirements contemporaneous with the weevil infestation. Events that overlapped in timing with the boll weevil's spread and significantly affected educational attainment include the passage of compulsory school attendance laws (Lleras-Muney 2002) and the construction of Rosenwald schools (Aaronson and Mazumder 2011). If the passage of compulsory schooling laws or Rosenwald school construction were correlated with the onset of the boll weevil infestation, this could result in spurious correlation between boll weevil exposure and student outcomes.

Therefore, we assign to individuals in our sample a measure of Rosenwald school exposure, based on childhood county of residence and year of birth, and include this and its interaction with an indicator for rural status as controls. Additionally, we calculate the number of years of schooling required, by state and birth cohort, from the ages for which compulsory schooling laws required attendance and the effective dates of those laws, adding this measure as a control.[43] Panel (B) of Table 6 shows the results of our main specifications modified to add controls for compulsory schooling and Rosenwald school exposure. The coefficients presented are comparable in magnitude, if not slightly larger, and statistical significance to the baseline results, which are repeated in Panel (A) for convenience, ameliorating concern that omitted variable bias is driving our results.

Robustness to Matching

Table 7 presents results of specifications, following Equation (3) with years of schooling as the dependent variable, utilizing samples generated by several different matching procedures. Columns (1) through (6) show results for whites, while Columns (7) through (12) display

[43] For children under the school entrance age when compulsory schooling laws took effect, this is simply the minimum age at which exit was legally allowed less the entrance age.

TABLE 6
ROBUSTNESS OF THE BOLL WEEVIL'S EFFECT ON EDUCATIONAL OUTCOMES TO CONCURRENT SHOCKS

	Years of Schooling		Completed 8th Grade	
	(1) White	(2) Black	(3) White	(4) Black
Panel (A): Baseline				
Age exposed:				
4–6	0.2669***	0.3579***	0.0177*	0.0327***
	(0.0608)	(0.0724)	(0.0086)	(0.0093)
7–9	0.2364***	0.2427***	0.0263***	0.0249***
	(0.0500)	(0.0601)	(0.0069)	(0.0074)
10–12	0.1481***	0.1514**	0.0175***	0.0128*
	(0.0378)	(0.0467)	(0.0052)	(0.0058)
13–15	0.1019***	0.1409***	0.0140***	0.0116*
	(0.0287)	(0.0390)	(0.0040)	(0.0047)
16–18	0.0452*	0.0609*	0.0058+	0.0035
	(0.0226)	(0.0304)	(0.0031)	(0.0036)
Observations	429,757	170,839	429,757	170,839
R^2	0.1627	0.0908	0.1169	0.0566
Panel (B): Adding Controls for Concurrent Shocks				
Age exposed:				
4–6	0.3042***	0.3871***	0.0218*	0.0364***
	(0.0623)	(0.0729)	(0.0088)	(0.0094)
7–9	0.2627***	0.2596***	0.0291***	0.0270***
	(0.0502)	(0.0602)	(0.0070)	(0.0074)
10–12	0.1720***	0.1677***	0.0199***	0.0148*
	(0.0384)	(0.0477)	(0.0053)	(0.0060)
13–15	0.1256***	0.1548***	0.0166***	0.0133**
	(0.0293)	(0.0400)	(0.0041)	(0.0049)
16–18	0.0578*	0.0676*	0.0071*	0.0044
	(0.0232)	(0.0305)	(0.0032)	(0.0036)
Rosenwald exposure?	Yes	Yes	Yes	Yes
Compulsory schooling laws?	Yes	Yes	Yes	Yes
Observations	429,673	170,827	429,673	170,827
R^2	0.1627	0.0910	0.1170	0.0569

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.

Notes: The dependent variables are given in the column headings. Standard errors adjusted for clustering by childhood county of residence are in parentheses. All specifications include year of birth fixed effects, childhood county of residence fixed effects, and controls for family background. Family background controls include childhood household head's occupational score, homeownership status, and literacy, as well as indicators for urban location and farm residence. Additionally, specifications in Panel (B) include a measure of Rosenwald school exposure, its interaction with an indicator for rural status, and a measure of exposure to compulsory school attendance laws as controls for contemporaneous shocks to education.
Source: See the text.

TABLE 7
ROBUSTNESS OF THE BOLL WEEVIL'S EFFECT ON YEARS OF SCHOOLING TO MATCHING

	White					
	(1)	(2)	(3)	(4)	(5)	(6)
Age exposed:						
4–6	0.2669***	0.2102***	0.2451***	0.2746***	0.2475***	0.2147***
	(0.0608)	(0.0541)	(0.0668)	(0.0643)	(0.0706)	(0.0648)
7–9	0.2364***	0.2003***	0.2273***	0.2354***	0.2264***	0.1665**
	(0.0500)	(0.0449)	(0.0529)	(0.0528)	(0.0558)	(0.0528)
10–12	0.1481***	0.1165***	0.1308**	0.1400***	0.1228**	0.1200**
	(0.0378)	(0.0342)	(0.0413)	(0.0405)	(0.0445)	(0.0418)
13–15	0.1019***	0.0922***	0.0892**	0.0926**	0.0802*	0.0667*
	(0.0287)	(0.0253)	(0.0314)	(0.0307)	(0.0336)	(0.0314)
16–18	0.0452*	0.0308	0.0413+	0.0413+	0.0356	0.0307
	(0.0226)	(0.0208)	(0.0246)	(0.0240)	(0.0265)	(0.0255)
Match requirements:						
Unique within 3 years?	Yes	No	Yes	Yes	Yes	Yes
Unique within 5 years?	No	No	Yes	No	Yes	No
Difference in birth year ≤ 1?	No	No	No	Yes	Yes	No
Match on exact names?	No	No	No	No	No	Yes
Match rate	27.3749	35.7069	23.5834	25.0809	21.4880	24.6369
Observations	429,757	567,158	367,879	382,444	325,555	358,976
R^2	0.1627	0.1372	0.1714	0.1697	0.1788	0.1654
	Black					
	(7)	(8)	(9)	(10)	(11)	(12)
Age exposed:						
4–6	0.3579***	0.2543***	0.4061***	0.3410***	0.3697***	0.3644***
	(0.0724)	(0.0627)	(0.0798)	(0.0821)	(0.0936)	(0.0829)
7–9	0.2427***	0.1923***	0.2421***	0.2900***	0.2711***	0.2340***
	(0.0601)	(0.0512)	(0.0660)	(0.0697)	(0.0801)	(0.0706)
10–12	0.1514**	0.1268**	0.1699***	0.1812***	0.1807**	0.1634**
	(0.0467)	(0.0398)	(0.0510)	(0.0545)	(0.0612)	(0.0544)
13–15	0.1409***	0.0994**	0.1482***	0.1370**	0.1270*	0.1550***
	(0.0390)	(0.0338)	(0.0431)	(0.0434)	(0.0498)	(0.0439)
16–18	0.0609*	0.0269	0.0738*	0.0814*	0.0935*	0.0664+
	(0.0304)	(0.0258)	(0.0344)	(0.0336)	(0.0393)	(0.0365)
Match requirements:						
Unique within 3 years?	Yes	No	Yes	Yes	Yes	Yes
Unique within 5 years?	No	No	Yes	No	Yes	No
Difference in birth year ≤ 1?	No	No	No	Yes	Yes	No
Match on exact names?	No	No	No	No	No	Yes
Match rate	18.5826	26.5591	15.1410	14.8900	11.8161	15.3941
Observations	170,839	254,015	136,112	129,810	100,120	133,857
R^2	0.0908	0.0705	0.1054	0.0974	0.1147	0.0952

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.

Notes: The dependent variable is years of schooling. Standard errors adjusted for clustering by childhood county of residence are in parentheses. All specifications include year of birth fixed effects, childhood county of residence fixed effects, and controls for family background. Family background controls include childhood household head's occupational score, homeownership status, and literacy, as well as indicators for urban location and farm residence. Columns (1) and (7) repeat the baseline results shown in Table 4 for ease of comparison. Match rates are calculated relative to the 1940 population of men between the ages of 23 and 58, inclusive, with recorded level of education.
Source: See the text.

results for blacks. The first columns for each race, (1) and (7), reproduce results utilizing our baseline sample, those shown in Table 4, for ease of comparison. Columns (2) and (8) relax match restrictions slightly by allowing matches between individuals that are not unique within three-year age bands. The methodology in this sample is, therefore, identical to Abramitzky, Boustan, and Eriksson (2012). This method results in a match rate of 35.71 (26.56) percent for the white (black) sample, significantly higher than the 27.37 (18.58) percent match rate of the baseline. Relaxing restrictions yields point estimates that are smaller in magnitude by 15 to 29 percent relative to the baseline for the youngest three exposure cohorts, but still statistically significant. This is likely due to attenuating noise introduced by a higher level of false positive matches.

Matching restrictions are increased relative to the baseline, to further reduce the likelihood of false positive matches, in Columns (3) and (9), which require individuals to be unique within five-year age bands. The match rate falls to 23.58 (15.14) percent for the white (black) sample, yet the coefficients are not meaningfully changed compared with the baseline. Columns (4) and (10) show results from a matched sample requiring that individuals be unique within three-year age bands and have absolute differences in year of birth of less than or equal to one year. The latter restriction is also aimed at reducing the frequency of false positive matches, as those with two-year discrepancies in year of birth are more likely to be incorrect. Again, the coefficients are remarkably similar to the baseline. Columns (5) and (11) additionally restrict matches to individuals that are unique within five-year age bands, and present results that are nearly identical. Columns (6) and (12) return to the three-year age band uniqueness restriction, but require matches on exact names rather than standardized versions of names. Results are little changed for blacks here, but point estimates are reduced in magnitude for whites by 19 to 30 percent relative to the baseline for the youngest three exposure cohorts.[44]

Robustness to Weighting

Recall that the baseline sample is not representative of the population, as shown in Table 1. In Table 8, we examine whether the results are robust to weighting the baseline sample to be representative of the population along the characteristics presented in Table 1. Panel (A)

[44] Online Appendix Table A.10 shows the baseline results with eighth grade completion as the dependent variable to be similarly robust. Our estimates are also robust to using only observations in the intersection of the matched samples generated by linking on standardized names and exact names (results not shown).

TABLE 8
ROBUSTNESS OF THE BOLL WEEVIL'S EFFECT ON LONG-RUN EDUCATIONAL OUTCOMES TO WEIGHTING

	Years of Schooling		Completed 8th Grade	
	(1) White	(2) Black	(3) White	(4) Black
Panel (A): Weighting on Childhood Characteristics				
Age exposed:				
4–6	0.2753***	0.3586***	0.0205*	0.0324***
	(0.0626)	(0.0727)	(0.0090)	(0.0095)
7–9	0.2414***	0.2445***	0.0271***	0.0248***
	(0.0514)	(0.0600)	(0.0071)	(0.0074)
10–12	0.1516***	0.1533***	0.0179***	0.0125*
	(0.0390)	(0.0462)	(0.0053)	(0.0058)
13–15	0.1064***	0.1464***	0.0142***	0.0123*
	(0.0296)	(0.0390)	(0.0042)	(0.0048)
16–18	0.0468*	0.0600+	0.0056+	0.0035
	(0.0232)	(0.0307)	(0.0032)	(0.0036)
Observations	429,629	170,835	429,629	170,835
R^2	0.1608	0.0837	0.1166	0.0515
Panel (B): Weighting on Adulthood Characteristics				
Age exposed:				
4–6	0.2359***	0.3453***	0.0173*	0.0311***
	(0.0630)	(0.0734)	(0.0087)	(0.0088)
7–9	0.2281***	0.2311***	0.0269***	0.0229**
	(0.0522)	(0.0602)	(0.0070)	(0.0071)
10–12	0.1415***	0.1477**	0.0168**	0.0119*
	(0.0402)	(0.0474)	(0.0052)	(0.0055)
13–15	0.1030***	0.1407***	0.0137***	0.0114**
	(0.0305)	(0.0387)	(0.0041)	(0.0043)
16–18	0.0385	0.0596+	0.0048	0.0036
	(0.0241)	(0.0308)	(0.0032)	(0.0035)
Observations	427,500	169,335	427,500	169,335
R^2	0.1663	0.0918	0.1198	0.0548

+ = Significant at the 10 percent level.
* = Significant at the 5 percent level.
** = Significant at the 1 percent level.
*** = Significant at the 0.1 percent level.

Notes: The dependent variables are given in the column headings. Standard errors adjusted for clustering by childhood county of residence are in parentheses. All specifications include year of birth fixed effects, childhood county of residence fixed effects, and controls for family background. Family background controls include childhood household head's occupational score, homeownership status, and literacy, as well as indicators for urban location and farm residence. The specifications in Panel (A) use an inverse proportional weighting method to weight our matched sample to be reflective of the population with respect to observable childhood characteristics: census year; state of residence; race; age; household head's occupational score, homeownership, and literacy; household farm and urban status; and year of infestation of county of residence. The specifications in Panel (B) use an inverse proportional weighting method to weight our matched sample to be reflective of the population with respect to observable adulthood characteristics: state of residence; race; age; years of schooling; and occupational income score.
Source: See the text.

presents specifications analogous to the baseline shown in Table 4, but uses inverse probability weights calculated based on family background characteristics provided by childhood census data. Instead, Panel (B) uses inverse probability weights estimated from 1940 census characteristics. The results presented do not meaningfully differ from the baseline results. Therefore, our results cannot be explained by observable differences between the population and the matched sample.

CONCLUSION

We add to an expanding literature exploring the boll weevil's impact beyond its direct effect on cotton production. The spread of the infestation through the South, whose agrarian economy was heavily dependent on cotton, provided an exogenous shock to agricultural productivity, particularly for women and children. Our findings reveal gains in educational attainment for those at young ages when the boll weevil arrived. Whites exposed at ages 4–6 gained 0.2669 years of schooling on average, while comparatively aged blacks gained 0.3579 years. White and black children ages 7–9 also saw significant gains of 0.2364 and 0.2427 years of schooling, respectively. Slightly older exposure cohorts experienced gains as well, but these decline as age at exposure increased to 18.

The magnitude of these estimates can be compared to the findings of several studies of contemporaneous shocks to schooling. Our finding that exposure to the boll weevil at age 7–9 increased educational attainment by nearly a quarter of a year, is comparable to the imposition of compulsory schooling and child labor laws requiring school entrance at age 7 and allowing work permits beginning at age 12.[45] It is important to note, however, that passage of such a law does not imply compliance with the law. Compulsory attendance laws were not well enforced in many locales and often provided myriad exceptions. Another comparison, for black children in particular, is instructive: Aaronson and Mazumder (2011) estimate that going from no (Rosenwald exposure of 0) to full (Rosenwald exposure of 1) coverage of black 7 through 17 year olds by Rosenwald teachers led to a gain in educational attainment of 1.186 years on average for blacks. Therefore, our result is roughly equivalent to having enough Rosenwald teachers to teach 20 percent of black children in one's childhood county during ages 7 through 17, or Rosenwald exposure of 0.20 where the mean Rosenwald exposure in 1930 was 0.27.

[45] Lleras-Muney (2002) shows that each additional year between the school entrance age and work permit age increased years of schooling by 0.05 years. This makes our result roughly equivalent to a five-year gap between the school entrance age and permit age.

Finally, Baker (2019) shows that a 1 percent increase in school resources for the first three years of schooling increased educational attainment by 0.0378 years for white children in early twentieth-century Georgia. Thus, white children would experience approximately the same gains in years of schooling from a 6.25 percent increase in school financial resources. Therefore, our estimates represent economically meaningful gains on the order of a significant funding boost, which seems quite reasonable for an event that so dramatically changed agricultural production in the region.

As the boll weevil itself only directly affects cotton production, any impact of the boll weevil on student outcomes must run through its devastation of the cotton crop. Contemporary observations and empirical evidence have demonstrated that Southern farmers shifted away from cotton production after infestation (Lange, Olmstead, and Rhode 2009; Giesen 2011). This shift from a child labor–intensive crop to alternatives that generated less demand for child labor provides a likely mechanism: a fall in the value of the marginal product of child labor in agriculture, or the opportunity cost of schooling, led to increased enrollment and attendance. For younger exposure cohorts, this accumulated into higher levels of average educational attainment. A potential second mechanism is that the boll weevil made farming a less attractive occupation, causing children at the margin to shift their occupational aspirations and their preparations accordingly. Where fieldwork might have provided suitable training for the farm profession, schooling made available a wider set of occupations.

Both of these mechanisms likely played a role in increasing educational attainment following the weevil's arrival. If the first was at play, then our results are suggestive of the benefits of programs encouraging farmers to switch cultivation to less child labor–intensive crops and to adopt child labor–reducing technologies, which would decrease the opportunity cost of schooling in rural areas. However, if the second mechanism was also at play, then our results might overstate the potential gains from such programs, as such programs generally have neutral to positive effects on the returns to farming. Getting farmers to switch crops would be achieved by compensating them for losses incurred due to switching, and mechanization likely has a positive effect on the productivity of adult farm labor. Still, it should be noted that our estimates are net of the income effect of the boll weevil, suggesting that our results could in fact understate the benefits of these programs.

A third potential mechanism is suggested by the work of Clay, Schmick, and Troesken (2017), who claim that the diversification of crops following the boll weevil infestation reduced the incidence of pellagra, a disease caused by having insufficient niacin, by increasing the availability of locally-grown vegetables that had higher niacin content than imported foods. If those susceptible to the disease were randomly selected from the population or the disease was not commonly fatal, then we would expect increased health might have some positive effect on educational attainment on average for those surviving to adulthood. However, pellagra disproportionately affected the poor, because cheap shelf-stable foods had lower niacin content prior to enrichment. The Louisiana Board of Health noted in their 1928/1929 report, well after the boll weevil infestation began: "As this disease [pellagra] rises and falls with the economic situation, there seems little we can do to prevent its prevalence in localities where crops fail or employment is not remunerative" (1930, p. 10). That same report shows the case fatality rate for pellagra over the 1926–1929 period to be 47 percent in Louisiana. Given the positive relationship between family income and educational attainment (see, e.g., Blanden and Gregg 2004; Taubman 1989), it is unclear whether this channel has a strengthening or attenuating effect on our estimates.[46]

While we find that the boll weevil had a positive effect on educational outcomes overall, through its devastating impact on cotton production in the South, it is important to note that we find positive effects on two distinct outcomes: school enrollment in the short run and educational attainment over the long run. For positive enrollment effects to be interpreted as generating educational attainment gains it is necessary to assume that enrolled students are attending school and completing grades, but this need not be the case. Moreover, it is possible to see educational attainment gains without observing changes in enrollment by decreasing retention rates. Therefore, it is necessary to consider these as separate outcomes, but in many contexts it is not possible, time consuming, or costly to measure educational attainment (which is especially true for studies of modern interventions). Whether observed gains in educational attainment due to boll weevil exposure are generated by changes in enrollment, attendance, retention rates, or all three is unclear, and not discernible from available data. Given the importance of educational attainment in the labor market as a measure of human capital, however, directly estimating the boll weevil's effect on years of schooling and demonstrating sizable gains for both whites and blacks is a significant contribution to the literature.

[46] Indeed, studies of the income neutral health interventions, which disproportionately affected the poor, of iron fortification (Niemesh 2015) and hookworm eradication (Bleakley 2007) fail to find statistically significant effects on years of schooling when accounting for mean reversion. Rather, they find these interventions had substantial educational benefits at the intensive margin.

There is still much work to be done to understand the broader effects of the boll weevil on the Southern economy. Whether the spread of the infestation just prior to an unprecedented wave of migration out of the region represents a causal or coincidental relationship remains unexplored. Additionally, tracing the insects' impact on occupational choice would be instructive, given its likely negative impact on the attractiveness of farming as a profession.

REFERENCES

Aaronson, Daniel, and Bhashkar Mazumder. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* 119, no. 5 (2011): 821–88.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, no. 5 (2012): 1832–56.

Ager, Philipp, Markus Brueckner, and Benedikt Herz. "The Boll Weevil Plague and its Effect on the Southern Agricultural Sector, 1899–1929." *Explorations in Economic History* 65 (2017): 94–105.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform in Historical Data? Evidence from New U.S. Ground Truth." NBER Working Paper No. 24019, Cambridge, MA, November 2017.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. "Worms at Work: Long-Run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131, no. 4 (2016): 1637–80.

Baker, Richard B. "From the Field to the Classroom: The Boll Weevil's Impact on Education in Rural Georgia." *Journal of Economic History* 75, no. 4 (2015): 1128–60.

———. "School Resources and Labor Market Outcomes: Evidence from Early Twentieth-Century Georgia." *Economics of Education Review* 70 (2019): 35–47.

Baker, Richard B., John Blanchette, and Katherine Eriksson. "Replication: Long-Run Impacts of Agricultural Shocks on Educational Attainment: Evidence from the Boll Weevil." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-12-02. https://doi.org/10.3886/E115906V1.

Bandara, Amarakroon, Rajeev Dehejia, and Shaheen Lavie-Rouse. "The Impact of Income and Non-Income Shocks on Child Labor: Evidence from a Panel Survey of Tanzania." *World Development* 67 (2015): 218–37.

Beegle, Kathleen, Rajeev Dehejia, and Roberta Gatti. "Why Should We Care about Child Labor? The Education, Labor Market, and Health Consequences of Child Labor." *Journal of Human Resources* 44, no. 4 (2009): 871–89.

Blanden, Jo, and Paul Gregg. "Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain." *Oxford Review of Economic Policy* 20, no. 2 (2004): 245–63.

Bleakley, Hoyt. "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics* 122, no. 1 (2007): 73–117.

Bloome, Deirdre, James Feigenbaum, and Christopher Muller. "Tenancy, Marriage, and the Boll Weevil Infestation, 1892–1930." *Demography* 54, no. 3 (2017): 1029–49.

Boozer, Michael A., and Tavneet K. Suri. "Child Labor and Schooling Decisions in Ghana." Unpublished manuscript, Yale University, New Haven, CT, September 2001.

Borusyak, Kirill, and Xavier Jaravel. "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume." Working paper, 2017. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2826228.

Bradley, Frances Sage, and Margaretta A. Williamson. *Rural Children in Selected Counties of North Carolina*. U.S. Children's Bureau, Rural Child Welfare Series No. 2. Washington, DC: GPO, 1918.

Carruthers, Celeste K., and Marianne H. Wanamaker. "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap." *Journal of Labor Economics* 35, no. 3 (2017): 655–96.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *Quarterly Journal of Economics* 126, no. 4 (2011): 1593–660.

Clay, Karen, Ethan Schmick, and Werner Troesken. "The Rise and Fall of Pellagra in the American South." NBER Working Paper No. 23730, Cambridge, MA, August 2017.

Collins, William J., and Robert A. Margo. "Historical Perspectives on Racial Differences in Schooling in the United States." In *Handbook of the Economics of Education*, vol. 1, edited by Eric A. Hanushek and Finis Welch, 107–54. Amsterdam, Netherlands: Elsevier, 2006.

Dammert, Ana C. "Child Labor and Schooling Response to Changes in Coca Production in Rural Peru." *Journal of Development Economics* 86, no. 1 (2008): 164–80.

Emerson, Patrick M., Vladimir Ponczek, and André Portela Souza. "Child Labor and Learning." *Economic Development and Cultural Change* 65, no. 2 (2017): 265–96.

Eriksson, Katherine. "Education and Incarceration in the Jim Crow South: Evidence from Rosenwald Schools." *Journal of Human Resources*, forthcoming.

Evans, David K., and Muthoni Ngatia. "School Costs, Short-Run Participation, and Long-Run Outcomes: Evidence from Kenya." World Bank Policy Research Working Paper No. 8421, April 2018.

Giesen, James C. *Boll Weevil Blues: Cotton, Myth, and Power in the American South*. Chicago: University of Chicago Press, 2011.

Gunnarsson, Victoria, Peter F. Orazem, and Mario A. Sánchez. "Child Labor and School Achievement in Latin America." *World Bank Economic Review* 20, no. 1 (2006): 31–54.

Haines, Michael R., and Inter-university Consortium for Political and Social Research (ICPSR). Historical, Demographic, Economic, and Social Data: The United States, 1790–2002. ICPSR02896-v3. Ann Arbor, MI: ICPSR, 2010.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94, no. 1–2 (2010): 114–28.

Higgs, Robert. "The Boll Weevil, the Cotton Economy, and Black Migration 1910–1930." *Agricultural History* 50, no. 2 (1976): 334–50.

Holley, William C., and Lloyd E. Arnold. *Changes in Technology and Labor Requirements in Crop Production: Cotton*. Works Progress Administration, National Research Project Report No. A-7. Philadelphia, PA, 1938.

Hunter, W. D. "Statement of Mr. W. D. Hunter, in Charge of Southern Field-Crop Insect Investigations, Bureau of Entomology." In *Agriculture Appropriation Bill: Hearings Before the Committee on Agriculture, House of Representatives*. Washington, DC: GPO, 1916.

Hunter, W. D., and B. R. Coad. *The Boll Weevil Problem*. Washington, DC: GPO, 1923.

Jensen, Robert. "Agricultural Volatility and Investments in Children." *American Economic Review* 90, no. 2 (2000): 399–404.

Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode. "The Impact of the Boll Weevil, 1892–1932." *Journal of Economic History* 69, no. 3 (2009): 685–718.

Lleras-Muney, Adriana. "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939." *Journal of Law and Economics* 45, no. 2 (2002): 401–35.

Lombardi, Paul. "Examining the Effect of Economic Shocks on the Schooling Choices of Southern Farmers." *European Review of Economic History* 23, no. 2 (2019): 214–40.

Louisiana Department of Education. *Biennial Report of the State Superintendent of Education of Public Education to the General Assembly*. Baton Rouge, LA: The Advocate, 1902.

———. *Proceedings of the Conference of Parish Superintendents of Public Education*. Baton Rouge, LA: Daily State Publishing, 1908.

Louisiana Board of Health. *Biennial Report of the Louisiana State Board of Health to the Legislature of the State of Louisiana, 1928–1929*. New Orleans, LA: State Printer, 1930.

Matthews, Ellen Nathalie, and Helen M. Dart. *The Welfare of Children in Cotton-Growing Areas of Texas*. U.S. Children's Bureau, Publication No. 134. Washington, DC: GPO, 1924.

Mississippi Department of Public Education. *Biennial Report and Recommendations of the State Superintendent of Public Education to the Legislature of Mississippi*. Nashville, TN: Brandon Printing, 1907.

Moehling, Carolyn M. "State Child Labor Laws and the Decline of Child Labor." *Explorations in Economic History* 36, no. 1 (1999): 72–106.

———. "Family Structure, School Attendance, and Child Labor in the American South in 1900 and 1910." *Explorations in Economic History* 41, no. 1 (2004): 73–100.

Niemesh, Gregory T. "Ironing Out Deficiencies: Evidence from the United States on the Economic Effects of Iron Deficiency." *Journal of Human Resources* 50, no. 4 (2015): 910–58.

Oates, M. Bruce, and L. A. Reynoldson. *Standards of Labor on the Hill Farms of Louisiana*. U.S. Department of Agriculture, Bulletin No. 961. Washington, DC: GPO, 1921.

Osband, Kent. "The Boll Weevil Versus 'King Cotton'." *Journal of Economic History* 45, no. 3 (1985): 627–43.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. IPUMS USA: Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2018.

Taubman, Paul. "Role of Parental Income in Educational Attainment." *American Economic Review* 79, no. 2 (1989): 57–61.

Tennessee Department of Education. *Annual Report of the Department of Education*. Nashville, TN: Ambrose-Nashville, 1924.

Texas Department of Education. *Biennial Report of the State Superintendent of Public Instruction*. Austin, TX: Gammel-Statesman Publishing, 1905.

U.S. Bureau of the Census. *Thirteenth Census of the United States: 1910*. 11 vols. Washington, DC: GPO, 1913.

———. *Children in Gainful Occupations at the Fourteenth Census of the United States*. Washington, DC: GPO, 1924.

U.S. Census Office. *Eleventh Census of the United States: Instructions to Enumerators*. Washington, DC: GPO, 1890.

U.S. Department of Labor. *By the Sweat and Toil of Children Volume II: The Use of Child Labor in U.S. Agricultural Imports and Forced and Bonded Child Labor*. Washington, DC, 1995.

———. *List of Goods Produced by Child Labor or Forced Labor*. Washington, DC, 2018.