

NBER WORKING PAPER SERIES

THE BOLL WEEVIL'S IMPACT ON RACIAL INCOME GAPS IN THE EARLY  
TWENTIETH CENTURY

Karen Clay  
Ethan J. Schmick  
Werner Troesken

Working Paper 27101  
<http://www.nber.org/papers/w27101>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2020, Revised February 2021

We would like to thank Andy Ferrara, Daniel Jones, Jessica LaVoice, Allison Shertzer, and Randy Walsh for helpful comments and suggestions. This paper also benefited from comments by seminar participants at Marquette University, the American Economic Association Annual Conference, the Liberal Arts Colleges Development Economics Conference, and the North American Regional Science Association Conference. Karen Clay acknowledges financial support from Heinz College, Carnegie Mellon University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Karen Clay, Ethan J. Schmick, and Werner Troesken. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Boll Weevil's Impact on Racial Income Gaps in the Early Twentieth Century  
Karen Clay, Ethan J. Schmick, and Werner Troesken  
NBER Working Paper No. 27101  
May 2020, Revised February 2021  
JEL No. I24,J10,J62,N32

### **ABSTRACT**

This paper investigates the effect of a large negative agricultural shock, the boll weevil, on racial income gaps in the first half of the twentieth century. We draw on complete count census data to generate a new large linked sample of Black and white fathers and their sons. Fathers are observed before and after the arrival of the boll weevil in their county, and their sons are observed in their father's household as children and again in 1940 as adults. In race specific difference-in-differences specifications and in triple differences specifications, we find that the boll weevil differentially affected wages of Black sons born after its arrival. Relative to white sons born after the boll weevil, Black sons born after the boll weevil saw a 6% increase in their wages. The magnitude of the effect is similar when the sample is constrained to sons whose father stayed in the South and to sons who stayed in the South. Evidence on changes in fertility and heights suggests that the relative gains were driven by improvements in early life conditions for Black sons born after the weevil's arrival. The magnitude of the effect of the boll weevil on the Black-white wage gap can be better understood by comparing it with changes between 1940 and 1950, a period of rapid convergence. The boll weevil caused the Black-white wage gap to fall by roughly half of the decline between 1940 and 1950 in the South or a third of the decline nationally.

Karen Clay  
Heinz College  
Carnegie Mellon University  
5000 Forbes Avenue  
Pittsburgh, PA 15213  
and NBER  
kclay@andrew.cmu.edu

Werner Troesken  
N/A user is deceased

Ethan J. Schmick  
Marquette University  
David A. Straz, Jr., Hall 1225  
W. Wisconsin Ave  
Milwaukee, WI 53202  
ethan.schmick@marquette.edu

# 1 Introduction

This paper investigates the effect of a large negative agricultural shock, the boll weevil, on racial income gaps in the first half of the twentieth century. Shocks are of interest, because they can lead to rapid changes in an economy, shifting modes and types of production, occupations, and the location of residence. Further, these shocks can have implications for income inequality. Specifically, shocks can widen racial income gaps if they disproportionately harm less advantaged groups or narrow gaps if they disproportionately help less advantaged groups. These effects may be seen immediately in the current generation of adults or in the next generation of children who are born around the time of the shock.

The boll weevil was a cotton pest that destroyed cotton crops and spread slowly throughout the American South between 1892 and 1922. By 1922 all cotton growing regions of the United States had been infested by the weevil. Within 5 years of the arrival of the boll weevil in a county, total cotton production fell 39-50%.<sup>1</sup> The spread of the boll weevil in the South changed the mix of crops, agricultural tenancy arrangements, occupations, and induced substantial migration (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017; Feigenbaum, Mazumder and Smith, 2019). While there were many other large shocks in the first half of the twentieth century including the Mississippi Floods of 1927 and the Dust Bowl (Hornbeck and Naidu, 2014; Hornbeck, 2012; Arthi, 2018), the boll weevil was a very large shock, affecting the approximately 22% of the U.S. population and 75% of Black men.

Drawing on a large newly linked census data set, this paper examines the effects of the boll weevil on the wages of Black and white sons born before and after its arrival. Any impact of the boll weevil on sons could come through at least two channels – changes in early life conditions and the migration of families. Changes in early life conditions would primarily affect cohorts born after the boll weevil and may have differed by race. Further, migration of families, which we proxy through father migration, may affect both early life conditions of sons born after the boll weevil and schooling and employment outcomes

---

<sup>1</sup>See Lange, Olmstead and Rhode (2009) and Ager, Brueckner and Herz (2017).

of all sons. We begin by observing Black and white fathers in the 1900 or 1910 U.S. Censuses in the years before the boll weevil arrived in the county they resided in. These fathers are then linked to the next decadal census (1910 or 1920) after the arrival of the boll weevil in their original county. This allows us to observe fathers' characteristics and whether the father migrated. Sons are observed in their father's household in 1900, 1910, or 1920 and are linked to the 1940 Census, which allows us to observe their wage income, occupation, years of schooling, and whether they have migrated out of the South or out of the state or county where their father was initially observed. Fewer than 20 percent of Black and white sons are observed outside the South in 1940.

Although one might have expected the boll weevil to have had adverse effects on children born after its arrival, we find that Black sons born after its arrival experienced significant relative wage gains. In race-specific difference-in-differences specifications that control for the county in which the father was originally observed, we find that Black sons born immediately after the arrival of the boll weevil had wages that were 8% higher than Black sons born before its arrival. In contrast, white sons born after the weevil's arrival had wages that were similar to white sons born before its arrival. In triple difference specifications, Black sons born after the arrival of the weevil saw a 6% increase in their wages relative to white sons born after its arrival. The magnitude of the effect is similar when the sample is restricted to sons whose fathers stayed in the South or to sons who themselves stayed in the South, suggesting little role for migration. To get a sense of the magnitude of the decline in the Black-white wage gap, it is useful to compare the effect to the decline between 1940 and 1950, a period of rapid convergence. The Black-white wage gap fell by 11% in the South, and 19% in the US as a whole (Boustan, 2009). Thus, the boll weevil caused the wage gap to fall by roughly half of the decline between 1940 and 1950 in the South or a third of the decline nationally.

Relative improvements in early life conditions appear to have occurred through three channels: fertility, female labor force participation, and nutrition. We find that Black fathers had slightly fewer children than white fathers after the boll weevil, which may have allowed greater investment in child quality. Ager, Brueckner and Herz (2017) find

that Black female labor force participation was lower after the arrival of the boll weevil. This may have reduced the stress on pregnant and nursing Black women. Clay, Schmick and Troesken (2019) document reductions in pellagra, a nutritional disease, and increases in the production of high nutritive value foods following the arrival of the boll weevil. Increased production of nutritious food may have differentially benefited low economic status households, which were disproportionately Black. In line with general improvements in nutrition, we find differential increases in the heights of Black World War II enlistees born after the boll weevil. Taken together, this evidence is consistent with relative improvements in early life conditions for Black men born after the boll weevil's arrival.

Our paper contributes to the literature in two ways. First, it highlights the role that a large negative shock, the boll weevil, played in helping close the Black-white wage gap for sons born after its arrival. The literature on Black-white inequality has focused on a number of factors, most prominently migration (Boustan, 2009; Collins and Wanamaker, 2014; Hornbeck and Naidu, 2014; Derenoncourt, 2019) and the Great Compression (Goldin and Margo, 1992; Margo, 1995). Our estimates suggest that the boll weevil caused the Black-white wage gap to fall substantially when compared to declines between 1940 and 1950 in the South, where more than 80% of the sons in our sample lived in 1940, and nationally. Thus, the boll weevil played an important role in narrowing the Black-white wage gap.

Second, it provides new detailed evidence on the effect of a large negative shock that occurred in the late nineteenth and early twentieth centuries on early life conditions for Black and white Southerners. Although there is a large literature on the effect of early life shocks on long run outcomes (Almond and Currie, 2011; Almond, Currie and Duque, 2018), a much smaller subliteration focuses on Black-white inequality (Almond, Currie and Herrmann, 2012; Bhalotra and Venkataramani, 2015; Almond, Currie and Duque, 2018). Because of data constraints, we know relatively little about the early life conditions of individuals who were children in the late nineteenth and early twentieth centuries and adults in the mid-twentieth century. The mid twentieth century is important, because it

is a period of considerable change in Black-white inequality. Our findings suggest that changes in early life conditions that occurred because of the boll weevil may have helped set the stage for later wage gains.

## 2 Background and Conceptual Framework

The arrival of the boll weevil in the cotton belt during the early 1900s acted as an exogenous shock that disrupted cotton production and broadly impacted the Southern economy. The boll weevil, native to Mexico, first migrated to Texas in 1892. From there, it progressed North and East through the cotton belt. By 1922, the boll weevil had spread throughout the entire cotton growing region of the United States.<sup>2</sup> The U.S. Department of Agriculture (1951), Ransom and Sutch (2001), Lange, Olmstead and Rhode (2009), and Ager, Brueckner and Herz (2017) all find that the arrival of the boll weevil had large negative effects on cotton yields and production. Within 5 years of the arrival of the boll weevil in a county, total cotton production fell 39-50%. Ager, Brueckner and Herz (2017), Bloome, Feigenbaum and Muller (2017), and Ager, Brueckner and Herz (2020) show that the arrival of the boll weevil had an adverse effect on tenancy and local labor markets. These adverse effects on tenancy and local labor markets induced substantial migration (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017; Feigenbaum, Mazumder and Smith, 2019).

While previous literature has tended to look at effects at the population level, we examine a specific group – fathers who had a son under the age of 10 in the household prior to the arrival of the boll weevil in their county. Specifically, we study fathers who were initially observed in 1900 or 1910 and observed again in the next census, 1910 or 1920. We then follow their sons who were born within 10 years of the arrival of the boll weevil in the father’s original county forward to the 1940 Census.

As noted previously, any impact of the boll weevil on sons could come through at

---

<sup>2</sup>See Lange, Olmstead and Rhode (2009) or Hunter and Coad (1923) for a year-by-year map of the boll weevils’ progression through the cotton belt.

least two channels – changes in early life conditions and migration of fathers. We will first discuss early life conditions and then turn to migration.

## 2.1 Early Life Conditions

How might the boll weevil have caused improvements in early life conditions for Black men? There are at least three possible channels: parental fertility, female labor force participation, and changes in nutrition. First, analysis of fertility, presented in section 5.6, indicates that Black fathers had slightly fewer sons after the arrival of the boll weevil than white fathers in the same county. This may have allowed some households to focus on child quality rather than quantity.

Second, the boll weevil appears to have reduced the labor force participation of Black women, which may have improved the health of pregnant Black women and new mothers. Ager, Brueckner and Herz (2017) find that the arrival of the boll weevil reduced the labor force participation of women. Comparisons of the labor force participation of Black and white women show that white women experienced small statistically insignificant declines, while Black women experienced large statistically significant declines. Because their data comes from repeated cross sections, some of this decline may reflect the differential migration of employed Black women. Regardless, improvements in the health of pregnant Black women and new mothers may have had benefits for their sons.

Third, evidence suggests that the boll weevil affected nutrition and diets. Prior to the boll weevil's arrival, the diet for poor Black and white Southerners consisted primarily of salt pork, molasses, and cornmeal. Ager, Brueckner and Herz (2017), Clay, Schmick and Troesken (2019), and Lange, Olmstead and Rhode (2009) provide evidence that the arrival of the boll weevil led to increases in the local production of food crops, since cotton was not as productive after the weevil's arrival. Specifically, there were increases in corn, peanut, and sweet potato acres per capita. Clay, Schmick and Troesken (2019) study the decline of pellagra, a nutritional disease associated with niacin deficiency, in the American South in the first half of the twentieth century. Building on this, in section

5.6, we show that the declines in pellagra were bigger in counties with larger shares of the population that were Black.

Height data from World War II enlistees suggests nutrition may have improved, specifically for Black sons born after the boll weevil. Nutrition and other early life conditions have been linked to adult height (Currie and Almond, 2011). Our analysis of height in section 5.6 has three findings. First, Black enlistees born immediately after the arrival of the boll weevil were taller than Black enlistees born before its arrival. Second, white enlistees born and after its arrival had similar heights to white enlistees born before its arrival. Third, Black enlistees born after the boll weevil experienced gains in height relative to white enlistees born born after the boll weevil.

The boll weevil may have also affected schooling either directly or through improvements in early life conditions. Using detailed administrative data from Georgia, Baker (2015) found that the boll weevil affected school enrollment with differential affects for Black children. Reductions in cotton production reduced the demand for child labor, allowing more Black children to attend school. More recent work by Baker, Blanchette and Eriksson (2020) and Ager, Brueckner and Herz (2020) has found increases in schooling for some age groups, but not differential effects by race. We examine the effects of the boll weevil on schooling further in section 5.4.

In summary, the available evidence suggests that early life conditions could have been affected by the arrival of the boll weevil and these effects may have differed by race.

## **2.2 Fathers' Migration**

The boll weevil set off a wave of migration. Using county level panel data on population and a difference-in-differences strategy, Lange, Olmstead and Rhode (2009) estimate sizeable net migration associated with the boll weevil. Counties with the highest cotton shares saw the greatest declines in population, while counties with lower cotton shares saw smaller declines. Ager, Brueckner and Herz (2017) and Feigenbaum, Mazumder and Smith (2019) also find that high cotton counties experienced significant out migration.



The latter study links this migration to a decrease in the violence and repression experienced by Black Southerners in cotton producing counties.

Panel A of Table 1 reports migration by fathers in our sample between the first and second census. Just 4-5% of Black fathers and 9-10% of white fathers, moved out of the South; 24% of Black fathers and 30% of white fathers moved out of state; and 61-63% of Black fathers and 59% of white fathers moved out of their original county. While more than half of fathers moved out of their original county, a large number stayed within the same state.

Fathers' migration may have affected both early life conditions and access to schooling and opportunities in the local labor market. If the effects of the boll weevil are only through early life conditions, then we should observe differences in outcomes for sons born after the arrival of the boll weevil. These effects might differ by fathers' migration status and race. On the other hand, if the effects are only through changes in access to schooling and opportunities in the labor market, then the effects may differ across sons by fathers' migration status and race but not by whether the sons were born before or after the boll weevil.

## 3 Data

This section describes the data on the boll weevil, the construction of the linked sample, measures of income, and summary statistics for the sample.

### 3.1 Boll Weevil

Data on the year the boll weevil first arrived in a county are taken from Lange, Olmstead and Rhode (2009), which originally came from USDA boll weevil maps. Counties invaded by the boll weevil between 1892 and 1922 are shown in Figure 1. We require that the fathers that we initially observe in 1900 or 1910 be living in a county that is invaded by the boll weevil in the next ten years. Counties shaded in green were invaded by the

boll weevil between 1901 and 1920 and, therefore constitute the set of counties that we initially observe fathers residing in. Counties shaded in red were invaded by the boll weevil, but not during 1901-1920 time period and, therefore, fathers initially residing in these counties are not in our sample.

## 3.2 Linking

To study the impact of the boll weevil on Black-white inequality, we generate a linked sample of fathers and their sons. Appendix Figure B.1 provides details on the linking procedure that we use to generate our sample of fathers and sons. We begin by looking for fathers in the 1900 or 1910 Censuses, provided by IPUMS (Ruggles et al. (2020)), that had a son aged 10 or younger and who were living in a county that would be invaded by the boll weevil in the next ten years (i.e. fathers in the 1900 Census must have been living in county that would be invaded by the boll weevil between 1901 and 1910; step 1). We then link these fathers to the next decadal census (step 2). By linking these fathers we are able to observe whether they migrated out of the South or their initial state of residence. We then take the set of sons of successfully linked fathers (step 3) and link them from the 1900, 1910, or 1920 Censuses to the 1940 Census to obtain adult outcomes (step 4). We assign sons to being born either before or after the boll weevil's arrival based on the year the weevil first arrived in the county their father initially resided in (i.e. the county their father lived in during 1900, for 1900-1910 linked fathers, or 1910, for 1910-1920 linked fathers). Our final data set contains linked sons from the 1891 to 1920 birth cohorts. To perform all of the linking we use the ABE linking algorithm (Abramitzky, Boustan and Eriksson, 2012, 2014, 2019), which uses first name (phonetically cleaned), surname (phonetically cleaned), birthplace, birth year, and race within a 5-year age-band. Linking algorithms and match rates are discussed further in Appendix A.

Appendix Table B.1 examines the representativeness of our sample and shows that differences between the linked sample and the sample that we attempted to link are small in magnitude. We are significantly more likely to find sons who lived in owner-occupied

housing in the census they were initially observed in. We are also significantly more likely to find sons who were born prior to 1905 (birth cohorts 1891 to 1904). There are significant differences between the linked sample and the sample that we attempted to link along numerous other dimensions, but most of these differences are small in magnitude. For example, 11.8% of Black sons that we linked lived in an urban area in the initial census, while 11.1% of Black sons in the entire sample we attempted to link lived in an urban area.

### 3.3 Measures of Income

The 1940 census contains information on annual wage income and weeks worked for sons, which we use to calculate weekly wages. This was the first census to ask about income, although it only asked about wage or salary income earned as an employee. Thus, it does not report self-employed income from farming or owning a business. Weekly wages are defined as an individual's yearly income in 1939 divided by the number of weeks they reported working in 1939.<sup>3</sup> We impose a number of restrictions on who is included in our baseline sample for our main empirical analysis when weekly wages is the dependent variable. The restrictions are discussed further in Appendix A, but they include dropping individuals that were unemployed, not in the labor force, on work relief, worked fewer than 30 weeks, or were in the top 3% of the income distribution.<sup>4</sup> Thus, our empirical specifications with weekly wage as the dependent variable contain 12,902 Black sons and 46,199 white sons.<sup>5</sup>

We use two alternative measures of income that assign income to farm owner-operators, tenants/sharecroppers, and other self-employed individuals. The first measure

---

<sup>3</sup>The 1940 Census had individuals report income and weeks worked for the previous year.

<sup>4</sup>We explore the sensitivity of our results to a number of other restrictions in section 5.5.

<sup>5</sup>Appendix Figure B.2 plots the number of observations by birth cohort. The number of observations is, generally, increasing up to the 1910 birth cohort for two reasons. First, is the result of the way the sample was constructed. For example, the 1891 birth cohort only contains sons who were 10 years old in 1900 and whose father was living in a county invaded in 1901. The 1892 birth cohort contains sons that were either 10 years old in 1900 and whose father lived in a county invaded in 1902 or who were nine years old in 1900 and whose father lived in a county invaded in 1901. Second, counties invaded later were often more populous.

is the commonly used occupational income score (OCC score). OCC scores are defined by IPUMS (Ruggles et al. (2020)) as the median income received by persons employed in a particular occupation in the 1950 Census. For the second measure we follow the method described in Collins and Wanamaker (2017) to construct income scores within a region, race, occupation, sex, and worker class cell.<sup>6</sup> In particular, this method allows us to assign income scores for both tenant farmers/sharecroppers and farm owner-operators. We have data on both OCC score and constructed income score for 23,505 Black sons and 87,385 white sons.

### 3.4 Summary Statistics

Table 1 provides summary statistics for fathers and sons. Columns 1 and 2 examine Black fathers and sons, while columns 3 and 4 are for white fathers and sons. Columns 1 and 3 use the sample of sons (and their fathers in Panel A) that are included in our wage regressions. Most of our analysis focuses on these sons. Columns 2 and 4 use the entire sample of sons (and their fathers) that we were able to construct an income score for. As shown in Panel A, fathers were largely employed in agriculture. In the first census 72-75% of Black fathers and 61-65% of white fathers worked in agriculture. The share declines slightly in the second census to 66-69% of Black fathers and 56-61% of white fathers. Very small shares of fathers work in manufacturing. As we noted in our previous discussion, while more than half of Black and white fathers moved out of their original county, many stayed within the same state, and relatively few moved out of the South.

Panel B of Table 1 provides summary statistics for sons. The first three measures are wages, OCC score, and constructed income. Both weekly wages and constructed income scores are in 1939 dollars. The constructed income score is much larger than weekly wages because it reflects annual income as opposed to weekly wage. OCC score was originally provided in hundreds of 1950 dollars, but for the summary statistics table we convert this to hundreds of 1939 dollars using the CPI from Officer and Williamson (2021). Across

---

<sup>6</sup>Worker class can either be “wage-worker” or “self-employed”.

all three measures, Black sons earn considerably less than white sons.<sup>7</sup> Black sons report fewer years of schooling. The migration patterns of Black and white sons are very similar. Most remain in the state that their father was originally observed in, although very few remain in the original county. Fewer than one fifth of the sons moved out of the South.

## 4 Empirical Strategy

Our analysis takes three approaches to estimating the effects of the boll weevil on weekly wage outcomes. First, we estimate a race-specific difference-in-differences model that includes father’s-initial-county time trends. Second, we estimate a difference-in-differences model with both Black and white sons. This approach reveals whether the effects we find in the race-specific difference-in-differences vary significantly by race. Finally, we estimate a triple differences model that takes advantage of variation across counties, birth years, and racial groups. We describe these approaches in the next three subsections and discuss the assumptions required to identify a causal effect.

### 4.1 Race-specific Difference-in-Differences

Our race specific difference-in-differences specification takes the following form:

$$\begin{aligned} \text{Log}(\textit{weekly wage})_{ict} = & \beta[\textit{Born post boll weevil}_{ct} = 1] + \theta_c + \theta_t + \theta_c * \textit{time} \\ & + \theta_b + \theta_e + \epsilon_{ict} \end{aligned} \tag{1}$$

In the above equation,  $i$  indexes a son,  $c$  indexes the county that son’s father was initially living in (in 1900 or 1910), and  $t$  indexes birth year. Thus,  $\text{Log}(\textit{weekly wage})_{ict}$  is the weekly wage, as observed in the 1940 census, of son  $i$ , whose father initially lived in county  $c$ , and who was born in year  $t$ . The requirements for a son to be included in

---

<sup>7</sup>The smallest Black-white difference in percentage terms is for OCC score because race is not taken into account when computing OCC scores; it is when computing constructed income.

our baseline sample were discussed in Appendix A.

$[Born\ post\ boll\ weevil_{ct} = 1]$  is a dummy variable that takes a value of one if individual  $i$  was born after the arrival of the boll weevil in their father’s initial county. We measure treatment based on the son’s birth year relative to when the boll weevil arrived in their father’s initial county.<sup>8</sup> As previously mentioned, we restrict our analysis to sons who were born within 10 years, plus or minus, of the weevil’s arrival in their fathers’ initial county.

The remaining controls are: county, birth year, census enumeration year, and birth order fixed effects along with county specific time-trends.  $\theta_c$  are fixed effects for father’s initial county and  $\theta_t$  are birth year fixed effects.  $\theta_c * time$  are county specific linear time-trends.  $\theta_b$  is a dummy variable for individual  $i$ ’s location in the birth order of his family. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account siblings who either moved out of the house or died before the censuses were taken. It also does not take into sisters.  $\theta_e$  are initial census enumeration year fixed effects. We estimate Equation 1 separately for Black and white sons and cluster standard errors at the father’s initial county-of-residence level.

Several assumptions are required for the estimates of  $\beta$  in equation 1 to be interpreted as the causal impact of being born after the boll weevil’s arrival. First, the arrival of the boll weevil in a father’s initial county-of-residence must be exogenous. That is, there must not be time-varying factors that correlate with both the arrival of the boll weevil and sons’ weekly wages. Prior literature has always considered the arrival of the boll weevil in a county to be exogenous; while farmers tried to take steps to prevent the invasion of the boll weevil, in practice there was little that could be done to stop its spread.

Second, Goodman-Bacon (2018) shows that the difference-in-differences estimator with time varying treatment (our case) is a weighted average of all 2x2 difference-in-differences estimators. In our context, counties invaded early by the boll weevil will serve

---

<sup>8</sup>For sons whose fathers move, we do not observe the timing of the birth and the move relative to the arrival of the boll weevil. This is particularly true for fathers who move within their original state, because we only observe a son’s state of birth, not the timing of their father’s move. Sons born in the year the boll weevil first arrived are coded as not being treated (i.e.  $[Born\ post\ boll\ weevil_{ct} = 0]$ ).

as controls for counties invaded later and vice versa. Thus, we must assume pairwise common trends in outcomes (within racial groups) between all counties in our sample, conditional on controls. To examine this common trends assumption, we present event study estimates of equation 1. Finally, as also shown by Goodman-Bacon (2018), we need to assume that treatment effects do not vary over time (i.e. treatment effects do not vary between counties invaded earlier versus later). To examine the extent to which treatment effects do vary over time, we present results using different years of boll weevil arrival.<sup>9</sup>

## 4.2 Complete Interaction Difference-in-Differences

To examine whether there is a significant difference in  $\beta$ s for Black sons relative to white sons we estimate the following complete interaction difference-in-differences specification:

$$\begin{aligned} \text{Log}(\textit{weekly wage})_{ict} = & \lambda[\textit{Born post boll weevil}_{ct} = 1] * [\textit{Black}_i = 1] + \theta_c * [\textit{Black}_i = 1] \\ & + \theta_t * [\textit{Black}_i = 1] + \theta_c * \textit{time} * [\textit{Black}_i = 1] + \theta_b * [\textit{Black}_i = 1] \\ & + \theta_e * [\textit{Black}_i = 1] + \epsilon_{ict} \end{aligned} \quad (2)$$

This specification is identical to Equation 1, but it interacts all variables with  $[\textit{Black}_i]$ , which is a dummy variable that takes a value of one if individual  $i$ 's reported race was Black. The reported  $\lambda$ s are simply the difference between the  $\beta$  for Black sons and the  $\beta$  for white sons estimated using the race specific difference-in-differences from Equation 1. Standard errors are, again, clustered at the father's initial county-of-residence level.

---

<sup>9</sup>The Goodman-Bacon decomposition is intended for use with balanced panel data. We do not use the decomposition in our context since our data is individual-level and not a balanced panel.

### 4.3 Triple Differences: Comparing Black Sons and White Sons

Our final empirical approach involves triple differences and uses the following specification:

$$\begin{aligned} \text{Log}(\textit{weekly wage})_{ict} = & \gamma[\textit{Born post boll weevil}_{ct} = 1] * [\textit{Black}_i = 1] \\ & + \theta_c + \theta_c * [\textit{Black}_i = 1] + \theta_t + \theta_t * [\textit{Black}_i] + \theta_c * \theta_t + \theta_b + \theta_e + \epsilon_{ict} \end{aligned} \tag{3}$$

The triple difference specification allows us to leverage all three dimensions of our data: (1) birth years, (2) father’s initial county-of-residence, and (3) race. The specification is similar to equation 1, but includes interactions of birth year and father’s initial county-of-residence fixed effects with the Black dummy variable ( $\theta_t * [\textit{Black}_i]$  and  $\theta_c * [\textit{Black}_i]$ ). It, also, includes birth year and father’s initial county-of-residence fixed effects interacted with each other ( $\theta_c * \theta_t$ ). Standard errors are, again, clustered at the father’s initial county-of-residence level. It is worth noting that equation 3 differs from equation 1 in the inclusion of the county-by-year fixed effects instead of county-specific time-trends. Thus,  $\gamma$  will not simply be the difference between the estimated  $\beta$ s for Black and white sons in equation 1.

$\gamma$  in equation 3 estimates the effect of being born after the boll weevil for a Black son relative to a white son whose fathers initially resided in the same county. Despite this Black-white comparison, equation 3 does not require Black and white sons to have common trends in outcomes prior the arrival of the boll weevil. Other interactions, namely birth year interacted with Black, father’s initial county interacted with Black, and father’s initial county interacted with birth year, allow for different pre-trends between Black and white sons. It does, however, require that no factor, other than the boll weevil, generated a contemporaneous change in the trend in outcomes between Black and white sons whose fathers were initially observed in the same county. The fact that the boll weevil invaded



counties in our sample at different times over a 20-year period reduces concerns that a single contemporaneous change is driving our results.

## 5 Black-White Inequality and the Boll Weevil

We begin in section 5.1 by presenting our main results on the effect of the boll weevil on the wages of Black and white sons. In section 5.2, we show that our results are robust to excluding sons of fathers who migrate out of the South and sons who migrate out of the South. In section 5.3, we consider other measures of income and selection. In section 5.4, we investigate the effect of the boll weevil on schooling. In section 5.5, we demonstrate the robustness of our results to different samples, time periods, and linking procedures. Having established the effects of the boll weevil on wages, we then turn in section 5.6 to the mechanisms through which the boll weevil affected wages.

### 5.1 Main Results

Table 2 presents the difference-in-differences results for Black sons in Panel A and white sons in Panel B. The difference-in-differences estimates using all sons and the interaction of all variables with the Black dummy is displayed in Panel C. Panel D shows the triple difference estimates comparing Black and white sons. We begin with the simplest specification and then add additional controls. Column 1 includes only county and birth year fixed effects. Column 2 adds birth order and census year fixed effects. Column 3 adds county time trends for Panels A through C and county-by-race, year-by-race, and county-by-year fixed effects for Panel D. We consider column 3 to be our preferred empirical specification and reference it through the remainder to the paper. One question of obvious interest is whether the effects in columns 1-3 hold once we control for years of schooling. Thus, column 4 includes years of schooling fixed effects, keeping in mind that schooling itself might be impacted by the arrival of the boll weevil. We explore schooling further in Section 5.4.

Black sons born after the arrival of the boll weevil in their father's original county had significantly higher wages than Black sons born before its arrival. In Panel A columns 1 and 2 of Table 2, Black sons born after the boll weevil had wages that are 4.6-4.9% higher than Black sons born before the boll weevil. Inclusion of county time trends in column 3 increases the magnitude to 7.7%.<sup>10</sup> Controlling for years of schooling fixed effects in column 4 decreases the magnitude slightly to 7.5%. The event study in Panel A of Appendix Figure B.3 shows positive wage effects for Black sons born after the boll weevil. We estimate this event study on the same sample used in Panel A of Table 2. However, we replace the post boll weevil dummy variable with a series of dummy variables that indicate when a son was born relative to the arrival of the boll weevil in their father's original county. The -5 dummy variable includes individuals born 5 to 10 years prior to the arrival of the boll weevil, while the 5 dummy variable includes individuals born 5 to 10 years after the arrival of the boll weevil.

In contrast to the results for Black sons, the arrival of the boll weevil appears to have had no effect on wage outcomes for white sons. The effects in Panel B of Table 2 are uniformly small and statistically insignificant. The event study in Panel B of Appendix Figure B.3 shows no wage effects for white sons born after the the boll weevil. Panel C of Table 2 shows that the difference in effects of the boll weevil for Black sons and white sons in Panels A and B is statistically significant in all cases.

The triple difference results in Panel D show that, conditional on their fathers being initially observed in the same county, Black sons born after the arrival of the boll weevil experienced wage gains relative to white sons born after its arrival. In columns 1 and 2, Black sons born after the boll weevil had wages that were 3.7% higher than white sons born after the boll weevil. Inclusion of the triple difference fixed effect interactions in column 3 increases the magnitude to 6.3%. The inclusion of years of schooling fixed effects in column 4 has almost no effect on the coefficient, which is 6%. The Black-white

---

<sup>10</sup>In Appendix Table B.4, we show that the effects in Panel A are larger for Black sons born after the boll weevil in higher cotton intensity counties (column 1). These result in high cotton intensity counties hold if the sample is restricted to sons whose fathers remained in the South or in their original state (columns 2 and 4). The results also hold if the sample is restricted to sons who remained in the South or in their father's initial state (columns 3 and 5).

weekly wage gap in our sample is about 0.66 log points.<sup>11</sup> Thus, the Black-white wage gap fell by 0.06 log points or about 9% (0.06/0.66).<sup>12</sup>

The arrival of the boll weevil had a substantial impact on the Black-white wage gap of sons in our sample, when compared with gains from 1940 to 1950. Boustan (2009) shows that between 1940 and 1950 the Black white wage gap fell by 0.11 log points in the South and by 0.19 log points nationally. Thus, the boll weevil caused the Black-white wage gap to fall by roughly half of the decline between 1940 and 1950 in the South or a third of the decline nationally. The decline is particularly notable given that less than one fifth of the sons in our sample migrated out of the South. The discussion in Margo (1995) highlights a number of factors that led to the gains between 1940 and 1950, including the Great Compression, the opening of jobs previously unavailable to Black individuals, migration out of the South, and relative increases in schooling for Black individuals. The gains that we observe in 1940 as a result of the boll weevil are all the more striking, because so few of these factors were at play.

## 5.2 Migration

Table 3 demonstrates that the wage gains experienced by Black sons born after the boll weevil are not driven by migration. For reference, column 1 reproduces column 3 from Table 2. Column 2 restricts attention to sons whose father remained in the South, while column 3 focuses on sons who remained in the South. We consider a son to have moved out of the South if they were not observed in the South census region in the 1940 census. A father moved out of the South if they were not in the South census region in the second census we observe them in (i.e. in either 1910 or 1920.) The results in column 2 and 3 are

---

<sup>11</sup>To get this number, we ran a regression of our sample from Table 2 using county and birth year fixed effects and a dummy variable indicating if an individual was Black. The coefficient estimate on the Black dummy variable is -0.66. For the remainder of this section we are going to refer to log points and use these to calculate percentages. However, throughout the rest of the paper we refer to log points and percentages interchangeably.

<sup>12</sup>Appendix Tables B.5 and B.6 decompose the weekly wage effect to determine if they are driven by increases in annual income or changes in weeks worked. The results in these tables suggest that the weekly wage effect shown in Table 2 are driven by increases in income rather than changes in weeks worked. The magnitudes for increases in income in Appendix Table B.5 are similar to the magnitudes in Table 2. Appendix Table B.6 shows no change weeks worked for Black or white sons.

very similar to the results in column 1. For example, the relative wage gains of Black sons compared to white sons in Panel D column 1 (Full sample), column 2 (Father remained in South) and column 3 (Son remained in South) are 6.3%, 6.0% and 7.2%. The results in columns 4 and 5, where we restrict attention to sons whose father remained in the same state or to sons who remained in the same state their father was initially living in, are similar to our main estimates, although reductions in sample size reduces the statistical power.

To further explore the effects of the boll weevil and migration on weekly wages, we first show that sons' migration is not statistically significantly related to the boll weevil. Column 4 of Appendix Table B.7 investigates the relationship between the boll weevil and a son being observed outside the South in 1940. In all of the specifications, the effects of being born after the arrival of the boll weevil are small and not statistically significant.

Appendix Table B.8 decomposes the effects of the boll weevil on weekly wages by migration status. Column 1 reproduces column 3 of Table 2. Column 2 examines if sons that moved out of the South were differentially impacted by the boll weevil and column 3 examines if sons whose father moved out of the South were differentially impacted. The effect for Black sons born after the boll weevil in all columns of panels A, C, and D is very similar in magnitude to Tables 2 and 3. Thus, Black sons who did not migrate, or whose father did not migrate, saw significant increases in their weekly wages. This is not to say that there were not returns to migrating. In fact, the returns to migration appear to have been large. For example in panel D column 2, sons who moved out of the South had wages that were 0.23 log points higher, and Black sons who moved out of the South had wages that were differentially higher by 0.44 log points. The return to migration for Black sons in column 2 of panel A (0.68 log points) are similar to estimates of real wage differentials between migrants and non-migrants in the Collins and Wanamaker (2014) sample of 5,000 Southern born Black men who were linked from 1910 to 1930. Thus, Appendix Table B.8 shows that while Black sons received a large return to migrating out of the South, the wage gains experienced by Black sons born after the boll weevil were not driven exclusively by migration.

### 5.3 Other Measures of Income and Selection

Table 4 compares our main results for reported weekly wage income with results using two alternative measures of income – OCC score and constructed income. Columns 1, 2, and 4 use our sample of wage workers, while columns 3 and 5 expand the sample to all workers that we were able to construct an income score for. In Panel A we find that Black sons born after the boll weevil, generally, experienced significant improvements in their occupational standing. The fact that the coefficient estimates are smaller in magnitude than in column 1 is not surprising. In columns 2-5 individual-level variation is lost and a single average is applied to an entire occupational group or region, race, occupation, sex, and worker class cell. This means that the averages may differ between 1940 wages, the OCC score, and constructed income.<sup>13</sup> In addition, Carruthers and Wanamaker (2017*a*) and Carruthers and Wanamaker (2017*b*) use both weekly wages and a constructed occupation score from the 1940 census and their coefficient estimates are 30-60% smaller when using the occupation score. Columns 3 and 5 of Panel B show that some white sons might have benefited from being born after the boll weevil. Finally, the estimates in Panel D are noisy with the exception of column 3. Again, this is not entirely surprising since there is no individual variation and Panel D contains a large number of fixed effects.

The fact that the sample size nearly doubles when we move from wage workers to all workers raises questions about who engages in wage work. Figure 2 plots the percentage of sons in our sample who were wage workers by race and birth cohort. The share of all workers engaged in wage work trended up across birth cohorts. After the 1900 birth cohort, Black and white shares of workers engaged in wage work are very similar. Column 6 of Table 4 shows that the boll weevil did not have a statistically significant impact on the probability that a Black or white son was a wage worker.

---

<sup>13</sup>Constructed income uses 1940 income and 1960 income-ratios between wage workers and self-employed workers. In the 1960 census both self-employed and wage workers reported income. Thus, we calculate the ratio of income between self-employed individuals and wage workers in each region, race, occupation, and sex cell. The 1940 income for wage workers in this cell is multiplied by the ratio to get a constructed income score for self-employed workers in the cell. See Collins and Wanamaker (2017) for more detail.

While we do not have direct evidence on selection, evidence from years of schooling suggests that Black and white wage workers may have been, on average, positively selected. Appendix Figure B.4 shows schooling for Black and white sons who were in our baseline sample of wage workers (wage workers in the figure) and sons who were not in our main sample but were in the larger sample used in columns 3, 5, and 6 of Table 4 (non-wage workers in the figure).

Since the share of wage workers is increasing across birth cohorts it is worth considering how selection into wage work might affect our estimates. Ignoring controls, if selection into wage work were positive, then the average quality of the wage workers would decline across birth cohorts (since sons born later are more likely to become wage workers), biasing down estimates of wage increases associated with the boll weevil in panels A and B of Tables 2 and 3. Similarly, if selection into wage work were negative, then the average quality of wage workers would increase across birth cohorts, biasing up estimates of wage increases associated with the boll weevil in panels A and B. The inclusion of birth year fixed effects may address this, if selection is affecting all counties similarly. The inclusion of county specific time trends is likely to further mitigate selection concerns. In Table 2, the difference-in-differences estimates increase with the inclusion of county-specific time trends, which is consistent with positive selection of Black sons into wage work.

The situation becomes more complicated in panel D, because it depends on relative selection for Black and white sons. Ignoring controls, triple difference estimates will be biased downward if Black sons were relatively more positively selected into wage work than white sons. Conversely, the triple difference estimates would be biased upward if Black sons were relatively less positively selected into wage work than white sons. Recall that the triple difference specification in equation 3 includes a variety of controls – county by race, birth year by race, and county by birth year fixed effects – that are likely to mitigate selection. Further, Appendix Table B.7 highlights that the wage gains experienced by Black sons born after the boll weevil are relatively constant over time. Column 1 reproduces column 3 from Table 2. In column 1, the 1891-1920 birth cohorts are used. Column 2 restricts attention to sons whose father lived in a county invaded

by the boll weevil in 1905 and later, which translates to the 1895-1920 birth cohorts. Column 3 restricts attention to counties invaded in 1910 and later, which would be the 1900-1920 birth cohorts. The results across all panels are similar to the results in column 1. Restricting attention to counties where the boll weevil invaded after 1905 or 1910 reduces the difference in wage work over the sample period, which may mitigate selection into wage work. It also tests the identifying assumption that the treatment effect does not vary over time.

## 5.4 Schooling

Table 5 indicates that the boll weevil did not have a statistically significant effect on reported years of schooling for Black or white sons who were wage workers. This also holds in the full sample, which includes linked sons who were not wage workers (see Appendix Table B.9). Appendix Figure B.4 shows that years of schooling are trending upward slowly over time for both Black sons and white sons.

Our findings regarding schooling relate most closely to Baker, Blanchette and Eriksson (2020), who also use data from the 1940 census. Using linked data for men aged 23-58 in 1940, they find effects of the boll weevil on children who were 4-9 at the time of arrival when compared to ages 19-30. Our sample and comparison groups are different than theirs. We examine the effects of the boll weevil on children born within 10 years of the weevil's arrival comparing children born just after its arrival to those born just prior. Further, the county fixed effects differ – where the child was residing in their paper versus where the father was initially observed in ours. Notwithstanding these differences, Baker, Blanchette and Eriksson (2020) also find that the boll weevil did not have differential schooling impacts by race.

## 5.5 Robustness

Appendix Table B.10 investigates the robustness of our results to a variety of alternative samples. Column 1 reproduces the result from Table 2 column 3 for reference. As

explained in Appendix A, our baseline sample in column 1 includes wage workers that were in the labor force, not on work relief, not unemployed, and worked more than 30 weeks in the year. We also truncate our baseline sample of wage workers at the 97th percentile of weekly wages. We then examine whether our results are affected by: including workers on work relief (column 2); including workers that were either unemployed or not in the labor force, but reported a wage for the previous year (column 3); truncating weekly wages at the 96th percentile (column 4); truncating weekly wages at the 98th percentile (column 5); restricting the sample to individuals who worked more than 40 weeks (column 6).<sup>14</sup> Column 7 uses the restrictions from Goldin and Margo (1992), which replaces incomes above \$5,000 with 1.4 times \$5,000, or \$7,000 when computing weekly wages.<sup>15</sup> To address the fact that many farm laborers received perquisite compensation we increase the weekly wages of farm laborers by 26% in column 8.<sup>16</sup> Finally, column 9 censors, rather than truncates, weekly wages. In particular, we replace weekly wages above \$100 with \$100. We selected this amount because an individual who worked 50 weeks in a year and earned \$100 a week would have an annual income of \$5,000, which is the top-coded income in the 1940 Census.

Appendix Table B.10 demonstrates that our results are very similar across different samples. In particular, Black sons born after the boll weevil have weekly wages that are 7-8% higher than Black sons born prior the its arrival (Panel A); white sons born after its arrival do not have significantly higher wages (Panel B); and Black sons born after its arrival have weekly wages that are 5-7% higher than white sons born after its arrival (Panel D).

---

<sup>14</sup>The 96th percentile weekly wage is \$53.84 and the 98th percentile weekly wage is \$67.23.

<sup>15</sup>Goldin and Margo (1992) also restrict to individuals that worked more than 40 weeks in the year and we do this as well in column 7. We, also, continue to restrict to only wage workers that were in the labor force, not on work relief, and not unemployed. Finally, Goldin and Margo (1992) restrict to individuals that had a weekly wage higher than \$6 a week. We do not impose this restriction, because laborers and farm laborers (who make up over 20% of our sample in column 1) often had weekly wages less than \$6 a week, but may have received perquisite compensation in the form of room and board. See Weiss (1989) for an argument that, at least in the nineteenth century, many workers that were reported as “laborers” in the census were actually farm laborers. Indeed, we find that over 17% of workers that report being “laborers” in our sample were living on a farm.

<sup>16</sup>This 26% adjustment is the amount used by Collins and Wanamaker (2017), which they based off of USDA reports.



We conclude this section by exploring the sensitivity of our main results to different linking procedures. We use three different linking criteria that are each more strict (i.e. will result in fewer false positive matches) than our baseline linking procedure. The first restricts our sample to individuals that match exactly on first name (not phonetically cleaned), last name (not phonetically cleaned), birthplace, birth year, and race. The second keeps individuals from our sample whose first and last names (phonetically cleaned) are unique within a five-year age band. The third restricts our sample to individuals that both match exactly and are unique within a five-year age band. Appendix Table B.11 estimates our main empirical result with these alternative, more conservative linking criteria. Column 1 of Table B.11 displays our main empirical result from column 3 of Table 2 for comparison with the other linking methods. Column 2 of Table B.11 uses only those who match exactly on first name (not phonetically cleaned), last name (not phonetically cleaned), birthplace, birth year, and race. Column 3 uses individuals whose names are unique within a five-year age band and column 4 uses individuals that both match exactly and are unique within a five-year age band. Note that using these stricter linking criteria results in smaller sample sizes. Nevertheless, our point estimates remain similar to column 1 even if they are sometimes estimated less precisely.

## 5.6 Mechanisms

As we discussed in section 2.1, the observed wage increases for Black sons born after the boll weevil in Tables 2 and 3 may have been driven by changes in fertility, female labor force participation, or nutrition. In this section, we present new evidence on fertility and two measures related to nutrition: pellagra and height.

Table 6 shows that the number of male children born after the boll weevil to Black fathers was significantly lower than for white fathers, but the magnitude was small. Our initial sample, by design, includes fathers with at least one son under the age of 10. In column 1, Black fathers had about 0.017 fewer male children after the boll weevil than white fathers. When we restrict attention to fathers who remained in the South or the

same state, the effects are similar in magnitude. These numbers are small relative to both the mean (0.169) and the standard deviation (0.375) of the number of male children born after the boll weevil. One possibility is that fathers who moved out of agriculture had lower demand for sons. Note, however, that only 6% of Black fathers and 5% of white fathers moved out of agriculture between the first and second censuses (Table 1). Only 3% of Black fathers and 0% of white fathers moved into manufacturing between the first and second censuses. Table 6 only uses fathers whose sons are included in our main results in Tables 2 and 3. Appendix Table B.12 expands this to include the fathers of all sons that we were able to link. Once again, Black fathers are significantly less likely to have children after the boll weevil relative to white fathers.<sup>17</sup>

Next we turn to pellagra, a disease caused by insufficient niacin consumption and an indicator of poor nutrition.<sup>18</sup> Using county data from North and South Carolina and a difference-in-differences estimation strategy, Clay, Schmick and Troesken (2019) show that after the arrival of the boll weevil pellagra rates fell.<sup>19</sup> Column 1 of Appendix Table B.13 shows the result from Table 3, column 5 of Clay, Schmick and Troesken (2019), which uses the log of the pellagra death rate as the dependent variable. Pellagra death rates significantly decreased by over 25% after the arrival of the boll weevil. In column 2 of Appendix Table B.13, we extend the Clay, Schmick and Troesken (2019) analysis to show that pellagra fell more in counties with higher shares of the population that were Black. Column 2 interacts the post boll weevil variable with the share of a county that was Black in 1910, the census prior to the weevil’s invasion of North and South Carolina. The share of a county’s population that was Black was standardized to have a mean of zero and a standard deviation of one. Counties with an average share Black population saw pellagra death rates decrease by about 19% after the arrival of the boll weevil; a one standard deviation increase in the share of the population that is Black resulted in an additional

---

<sup>17</sup>These findings on fertility relate most closely to Ager, Brueckner and Herz (2020). Using repeated cross sectional data and focusing on children under 5, they find small, statistically significant reductions in fertility after the boll weevil. The differential effect for Black mothers was negative but not statistically significant. Their negative point estimate is consistent with our finding of negative effects of fertility for Black fathers.

<sup>18</sup>It was not known that insufficient niacin was the cause of pellagra until 1937.

<sup>19</sup>North and South Carolina are the only states that reported pellagra deaths at the county level both before and after the boll weevil invasion.

decrease in the pellagra death rate by 8%.

We next analyze adult height, which is a marker for early life conditions including nutrition. Data on adult height is taken from *U.S. World War II Army Enlistment Records, 1938-1946*, provided by the National Archives and Records Administration. These records contain enlistment information for over 8.5 million individuals who served in World War II. The records contain information on the state and county of residence at the time of enlistment as well as the individual's height (in inches). We restrict these records to men who lived in the same state they were born in, were drafted, were born between 1915 and 1924, had a valid height and weight and were living in a county that was invaded by the boll weevil after 1914.<sup>20</sup> The counties in the sample are shown in Appendix Figure B.5.

A few points are worth explicitly noting. First, the birth cohorts considered in the height data set (1915-1924) differ from the birth cohorts in the wage regressions (1891-1920).<sup>21</sup> Second, because the records do not include county of birth, we assume that individuals are living in their county of birth when they enlisted. While the assumption clearly does not hold in practice, any movements made within the state of birth (since we are restricting to men who were living in the state they were born in) will likely introduce measurement error and bias our coefficient estimates towards zero. Third, there is selection into the enlistee sample, since individuals were screened along many dimensions as part of the enlistment process. Selection and differences in selection by race may impact the analysis of heights.<sup>22</sup> Thus the results should be interpreted with these factors in mind.

The evidence on enlistee heights in Table 7 is consistent with early life conditions

---

<sup>20</sup>Individuals born after 1924 may have still been growing when they enlisted. To serve in WWII an individual had to be between 5 and 6.5 feet tall and weigh over 105 pounds. Thus, a valid height is between 60 and 78 inches and a valid weight is 105 pounds and above.

<sup>21</sup>In principal it would be interesting to examine wage results for this group. Constraining the sample to sons who were born in 1915 or later and are living in their state of birth limits the sample size to 14,600 observations. When we do this the coefficient from the triple difference specification for Black sons is still positive and significant at the 10% level.

<sup>22</sup>The use of nineteenth century heights has been actively debated by economic historians because of selection issues. Less has been said about twentieth century heights, but similar issues are likely to apply. See Bodenhorn, Guinnane and Mroz (2017, 2019) and Komlos and A'Hearn (2019)

having improved for Black sons born after the boll weevil, both relative to Black sons born before and to white sons born after its arrival. Black enlistees born after the boll weevil were 0.09-0.11 inches taller than Black enlistees born before the boll weevil; white enlistees born after the boll weevil were the same height as white enlistees born before the boll weevil; and Black enlistees born after the boll weevil experienced height gains of 0.06-0.12 inches relative to white enlistees born after its arrival.

Taken together with the historical evidence discussed in section 2.1, the fertility, pellagra, and height findings support the idea that early life conditions improved after the arrival of the boll weevil.

## 6 Conclusion

A large negative agricultural shock, the boll weevil, spread across the American South from the early 1890s to the early 1920s. The literature has focused on the widespread economic disruption induced by the boll weevil. Total cotton production fell dramatically, which led to changes in the mix of crops, agricultural tenancy arrangements, occupations, and induced substantial migration out of affected counties. Given the magnitude of the disruption, one might have expected the boll weevil to have had adverse effects on children born after its arrival.

Strikingly, we find that Black sons who were born after the shock experienced significant relative wage gains. Drawing on a large newly linked sample of Black and white fathers and sons, we find that Black sons born immediately after the arrival of the boll weevil had wages that were 8% higher than Black sons born before its arrival, while white sons born and after its arrival had similar wages. In triple difference specifications, Black sons born after the arrival of the boll weevil saw a 6% increase in their wages relative to white sons born after its arrival. The increases are similar if we restrict attention to the 80 percent of sons who remained in the South. This narrowing of the Black-white wage gap is sizeable compared to declines in the gap over the next decade (1940-1950) in the South (11%) and in the US as a whole (19%) .

How did this negative shock lead to benefits for Black sons? Fertility, female labor force participation, and nutrition all appear to have contributed to relative improvements in early life conditions. We find Black fathers had slightly fewer children than white fathers after the boll weevil. This may have allowed greater investment in child quality. Black female labor force participation was lower after the arrival of the boll weevil (Ager, Brueckner and Herz, 2017), which may have improved child outcomes. Increases in the production of high nutritive value foods following the arrival of the boll weevil may have differentially benefited Black households. Using height data from World War II enlistees, we show that Black enlistees born after the boll weevil experience relative height gains compared to white enlistees.

What do the effects of this negative shock tell us about Black-white inequality? Our findings suggest that improvements in early life conditions in the late nineteenth and early twentieth centuries played an important and underappreciated role in the improvement of the economic status of Black men by 1940.

## References

- Abramitzky, Ran, Leah Boustan, and Katherine Eriksson.** 2019. “To the New World and Back Again: Return Migrants in the Age of Mass Migration.” *ILR Review*, 72(2): 300–322.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson.** 2012. “Europe’s Tired, Poor, Huddled Masses: Self-selection and Economic Outcomes in the Age of Mass Migration.” *American Economic Review*, 102(5): 1832–56.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson.** 2014. “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” *Journal of Political Economy*, 122(3): 467–506.
- Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, and Santiago Pérez.** 2021. “Intergenerational Mobility of Immigrants in the US over Two Centuries.” *American Economic Review*, Forthcoming.
- Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J Feigenbaum, and Santiago Pérez.** 2019. “Automated Linking of Historical Data.” *Journal of Economic Literature*, Forthcoming.
- Abramitzky, Ran, Roy Mill, and Santiago Pérez.** 2020. “Linking Individuals Across Historical Sources: A Fully Automated Approach.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2): 94–111.
- Acemoglu, Daron, and Joshua Angrist.** 2000. “How Large are Human-capital Externalities? Evidence from Compulsory Schooling Laws.” *NBER Macroeconomics Annual*, 15: 9–59.
- Ager, Philipp, Markus Brueckner, and Benedikt Herz.** 2017. “The Boll Weevil Plague and its Effect on the Southern Agricultural Sector, 1889–1929.” *Explorations in Economic History*, 65: 94–105.

- Ager, Philipp, Markus Brueckner, and Benedikt Herz.** 2020. "Structural Change and the Fertility Transition in the American South." *Review of Economics and Statistics*, 102(4): 806–822.
- Almond, Douglas, and Janet Currie.** 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives*, 25(3): 153–72.
- Almond, Douglas, Janet Currie, and Mariesa Herrmann.** 2012. "From Infant to Mother: Early Disease Environment and Future Maternal Health." *Labour Economics*, 19(4): 475–483.
- Almond, Douglas, Janet Currie, and Valentina Duque.** 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.
- Arthi, Vellore.** 2018. "'The Dust was Long in Settling': Human Capital and the Lasting Impact of the American Dust Bowl." *The Journal of Economic History*, 78(1): 196–230.
- Bailey, Martha J, Connor Cole, Morgan Henderson, and Catherine Massey.** 2020. "How Well do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature*, 58(4): 997–1044.
- Baker, Richard B.** 2015. "From the Field to the Classroom: the Boll Weevil's Impact on Education in Rural Georgia." *The Journal of Economic History*, 75(4): 1128–1160.
- Baker, Richard B, John Blanchette, and Katherine Eriksson.** 2020. "Long-run Impacts of Agricultural Shocks on Educational Attainment: Evidence from the Boll Weevil." *The Journal of Economic History*, 80(1): 136–174.
- Bhalotra, Sonia R, and Atheendar Venkataramani.** 2015. "Shadows of the Captain of the Men of Death: Early Life Health Interventions, Human Capital Investments, and Institutions." *SSRN Working Paper*.
- Bloome, Deirdre, James Feigenbaum, and Christopher Muller.** 2017. "Tenancy, Marriage, and the Boll Weevil Infestation, 1892–1930." *Demography*, 54(3): 1029–1049.

- Bodenhorn, Howard, Timothy W Guinnane, and Thomas A Mroz.** 2017. “Sample-selection Biases and the Industrialization Puzzle.” *The Journal of Economic History*, 77(1): 171–207.
- Bodenhorn, Howard, Timothy W Guinnane, and Thomas A Mroz.** 2019. “Diagnosing Sample-Selection Bias in Historical Heights: A Reply to Komlos and A’Hearn.” *The Journal of Economic History*, 79(4): 1154–1175.
- Boustan, Leah Platt.** 2009. “Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970.” *The Journal of Economic History*, 69(3): 755–782.
- Bremer, William W.** 1975. “Along the “American way”: The New Deal’s Work Relief Programs for the Unemployed.” *The Journal of American History*, 62(3): 636–652.
- Carruthers, Celeste K, and Marianne H Wanamaker.** 2017*a*. “Returns to School Resources in the Jim Crow South.” *Explorations in Economic History*, 64: 104–110.
- Carruthers, Celeste K, and Marianne H Wanamaker.** 2017*b*. “Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap.” *Journal of Labor Economics*, 35(3): 655–696.
- Clay, Karen, Ethan Schmick, and Werner Troesken.** 2019. “The Rise and Fall of Pellagra in the American South.” *The Journal of Economic History*, 79(1): 32–62.
- Collins, William J, and Marianne H Wanamaker.** 2014. “Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data.” *American Economic Journal: Applied Economics*, 6(1): 220–52.
- Collins, William J, and Marianne H Wanamaker.** 2017. “African American Inter-generational Economic Mobility Since 1880.” *National Bureau of Economic Research Working Paper*, #23395.
- Currie, Janet, and Douglas Almond.** 2011. “Human Capital Development Before Age Five.” In *Handbook of Labor Economics* Vol. 4, 1315–1486. Elsevier.

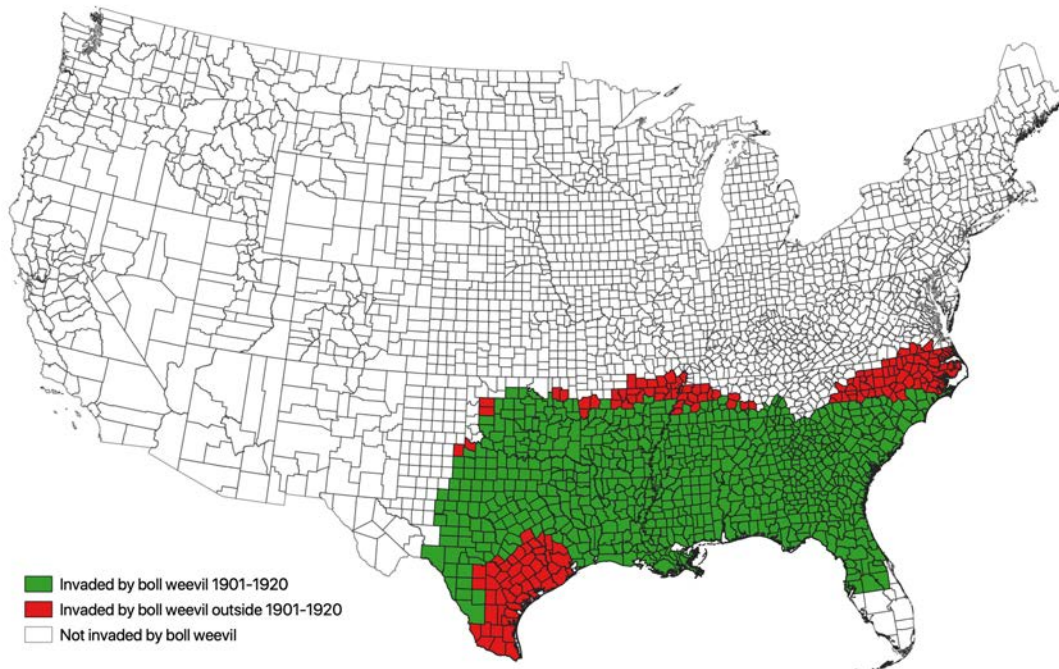


- Derenoncourt, Ellora.** 2019. “Can you Move to Opportunity? Evidence from the Great Migration.” *Working Paper*.
- Feigenbaum, James.** 2016. “A Machine Learning Approach to Census Record Linking.” *Working Paper*.
- Feigenbaum, James, Soumyajit Mazumder, and Cory Smith.** 2019. “When Coercive Economies Fail: The Political Economy of the US South After the Boll Weevil.” *Working Paper*.
- Ferrie, Joseph P.** 1996. “A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 29(4): 141–156.
- Goldin, Claudia, and Robert A Margo.** 1992. “The Great Compression: The Wage Structure in the United States at Mid-century.” *The Quarterly Journal of Economics*, 107(1): 1–34.
- Goodman-Bacon, Andrew.** 2018. “Difference-in-differences with Variation in Treatment Timing.” *National Bureau of Economic Research Working Paper*, #25018.
- Hornbeck, Richard.** 2012. “The Enduring Impact of the American Dust Bowl: Short- and Long-run Adjustments to Environmental Catastrophe.” *American Economic Review*, 102(4): 1477–1507.
- Hornbeck, Richard, and Suresh Naidu.** 2014. “When the Levee Breaks: Black Migration and Economic Development in the American South.” *American Economic Review*, 104(3): 963–90.
- Hunter, Walter David, and Bert Raymond Coad.** 1923. *The Boll Weevil Problem*. US Dept. of Agriculture.

- Komlos, John, and Brian A’Hearn.** 2019. “Clarifications of a Puzzle: The Decline in Nutritional Status at the Onset of Modern Economic Growth in the United States.” *The Journal of Economic History*, 79(4): 1129–1153.
- Lange, Fabian, Alan L Olmstead, and Paul W Rhode.** 2009. “The Impact of the Boll Weevil, 1892–1932.” *The Journal of Economic History*, 69(3): 685–718.
- Long, Jason, and Joseph Ferrie.** 2013. “Intergenerational Occupational Mobility in Great Britain and the United States Since 1850.” *American Economic Review*, 103(4): 1109–37.
- Margo, Robert A.** 1995. “Explaining Black-white Wage Convergence, 1940–1950.” *ILR Review*, 48(3): 470–481.
- Officer, Lawrence, and Samuel H Williamson.** 2021. “Measuring Worth.” <https://www.measuringworth.com>.
- Ransom, Roger L, and Richard Sutch.** 2001. *One Kind of Freedom: The Economic Consequences of Emancipation*. Cambridge University Press.
- Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek.** 2020. “IPUMS USA: Version 10.0 [dataset].”
- U.S. Department of Agriculture, Bureau of Agricultural Economics.** 1951. *Statistics on Cotton and Related Data. Statistical Bulletin No. 99*. GPO.
- Weiss, Thomas J.** 1989. “Economic Growth Before 1860: Revised Conjectures.” *National Bureau of Economic Research Working Paper*, #7.

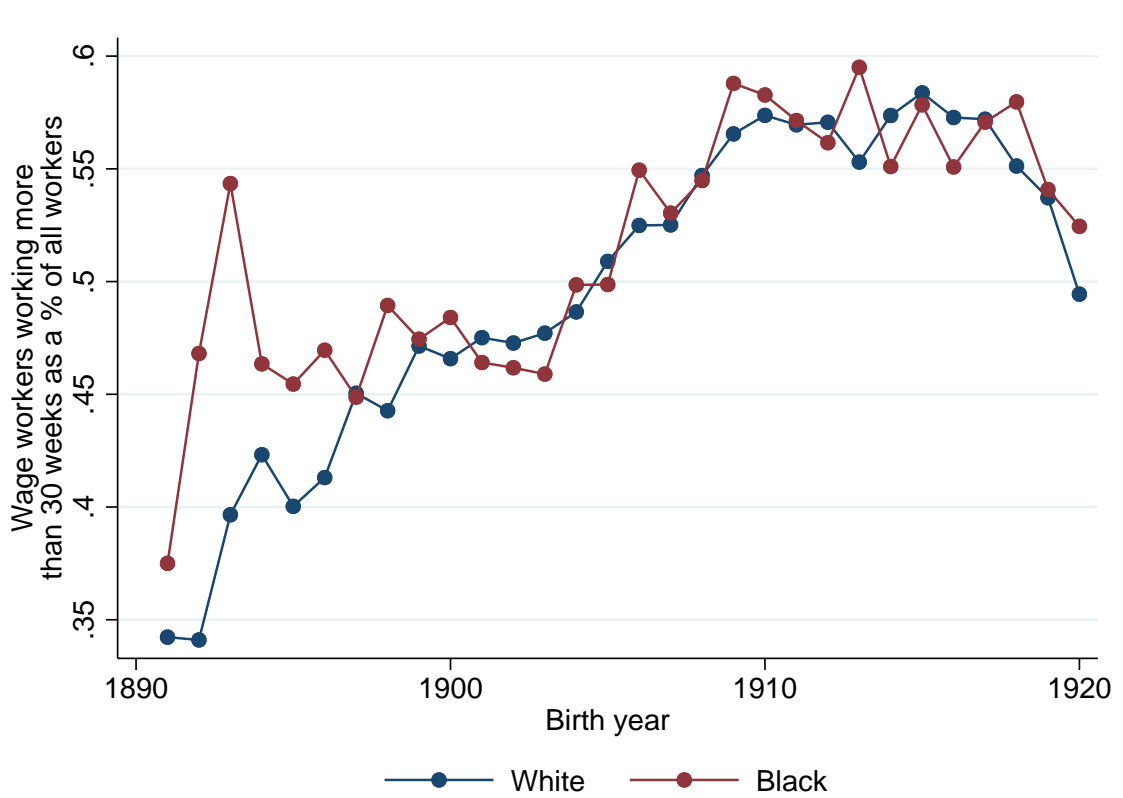
# Figures and Tables

Figure 1: Counties invaded by boll weevil



*Notes:* This map displays counties that were invaded by the boll weevil. Counties in green were invaded by the boll weevil sometime between 1901 and 1920. Counties in red were invaded by the boll weevil, but not within the 1901-1920 time frame. We use fathers living in a county invaded by the boll weevil between 1901-1920 to construct our sample.

Figure 2: Wage workers by birth cohort and race as a % of all workers



*Notes:* This figure shows the percent of individuals in our sample that were wages workers as a percentage of all workers. More precisely, the numerator is the number of wags workers that worked more than 30 weeks in the previous year, were not on work relief, were employed, and were in the labor force in each birth year (the sample in Table 2, Panels C and D). The denominator is the number of workers that were not work relief, were employed, were in the labor force, and that we could construct income scores for (the sample in Table 4; columns 3, 5, and 6; Panels C and D).

Table 1: Summary statistics

	Black		White	
	Wage workers (1)	All workers (2)	Wage workers (3)	All workers (4)
<i>Panel A: Fathers</i>				
Works in ag. in first census	0.72	0.75	0.61	0.65
Works in ag. in second census	0.66	0.69	0.56	0.61
Works in manufac. in first census	0.04	0.04	0.06	0.05
Works in manufac. in second census	0.07	0.06	0.06	0.05
Age in first census	34.00	33.98	34.66	34.84
Moved out of South	0.05	0.04	0.10	0.09
Moved out of state	0.24	0.23	0.30	0.30
Moved out of county	0.63	0.61	0.59	0.58
Observations	10502	19005	35268	65938
<i>Panel B: Sons</i>				
Weekly wage	12.31	12.34	23.28	24.57
OCC Score	18.62	17.56	25.57	23.86
Constructed income	513.72	468.49	1113.13	1143.11
Years of schooling	5.83	5.44	9.36	8.96
Age in 1940	32.22	32.66	32.57	33.18
Moved out of father's initial region (South)	0.19	0.15	0.18	0.15
Moved out of father's initial state	0.43	0.35	0.42	0.37
Moved out of father's initial county	0.87	0.82	0.84	0.79
Observations	12902	23505	46199	87385

*Notes:* Weekly wages and constructed income score are in 1939 dollars. OCC Score was originally in hundreds of 1950 dollars, but we converted it to 1939 dollars for this table using the CPI from Officer and Williamson (2021).

Table 2: The boll weevil and weekly wages

	Log(weekly wage)			
	(1)	(2)	(3)	(4)
<i>Panel A: Difference-in-differences for Black sons</i>				
Born post BW	0.0463*	0.0488**	0.0771**	0.0751**
	(0.0246)	(0.0245)	(0.0326)	(0.0303)
Observations	12902	12901	12901	12901
R-sq.	0.0957	0.0974	0.135	0.237
<i>Panel B: Difference-in-differences for white sons</i>				
Born post BW	-0.00656	-0.00684	0.0000625	-0.00241
	(0.0116)	(0.0117)	(0.0144)	(0.0137)
Observations	46199	46199	46199	46199
R-sq.	0.121	0.122	0.136	0.261
<i>Panel C: Difference-in-differences complete interaction</i>				
Born post BW * Black	0.0529**	0.0557**	0.0773**	0.0777**
	(0.0264)	(0.0263)	(0.0348)	(0.0327)
Observations	59101	59100	59100	59100
R-sq.	0.242	0.243	0.260	0.362
<i>Panel D: Triple differences</i>				
Born post BW * Black	0.0369**	0.0376**	0.0625**	0.0597**
	(0.0183)	(0.0182)	(0.0314)	(0.0303)
Observations	59101	59101	59101	59101
R-sq.	0.231	0.232	0.335	0.425
County and birth year FE	X	X	X	X
Birth order and census year FE		X	X	X
County TT/DDD interactions			X	X
Years of schooling FE				X

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. Columns (3) and (4) of Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Columns (3) and (4) of Panel D control for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where “county” is the father’s initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table 3: The boll weevil, weekly wages, and father and son migration status

	Log(weekly wage)				
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Difference-in-differences</i>					
<i>for Black sons</i>					
Born post BW	0.0771** (0.0326)	0.0712** (0.0326)	0.0712** (0.0349)	0.0740** (0.0352)	0.0945** (0.0423)
Observations	12901	12307	10374	9847	7359
R-sq.	0.135	0.140	0.148	0.164	0.216
<i>Panel B: Difference-in-differences</i>					
<i>for white sons</i>					
Born post BW	0.0000625 (0.0144)	0.00192 (0.0152)	0.00856 (0.0169)	0.00856 (0.0171)	0.0212 (0.0196)
Observations	46199	41856	38081	32037	26827
R-sq.	0.136	0.138	0.142	0.155	0.177
<i>Panel C: Difference-in-differences</i>					
<i>complete interaction</i>					
Born post BW * Black	0.0773** (0.0348)	0.0693** (0.0348)	0.0626* (0.0376)	0.0655* (0.0381)	0.0732 (0.0454)
Observations	59100	54163	48455	41884	34186
R-sq.	0.260	0.260	0.293	0.276	0.334
<i>Panel D: Triple differences</i>					
Born post BW * Black	0.0625** (0.0314)	0.0604* (0.0320)	0.0719** (0.0354)	0.0549 (0.0368)	0.0422 (0.0447)
Observations	59101	54021	48076	41366	33204
R-sq.	0.335	0.340	0.375	0.372	0.430
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table 4: The boll weevil and other long-run outcomes

	Log(weekly wage)	Log(OCC Score)	Log(Constructed income)		P(Wage worker = 1)	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Diff-in-diff for Black sons</i>						
Born post BW	0.0771** (0.0326)	0.0186 (0.0224)	0.0248* (0.0142)	0.0392** (0.0196)	0.0300** (0.0132)	-0.0193 (0.0202)
Observations	12901	12901	23503	12901	23503	23503
R-sq.	0.135	0.111	0.0688	0.118	0.0838	0.0989
Only Wage Workers	X	X		X		
<i>Panel B: Diff-in-diff for white sons</i>						
Born post BW	0.0000625 (0.0144)	0.00512 (0.00994)	0.0154** (0.00774)	0.0143 (0.0120)	0.0252*** (0.00894)	0.00184 (0.00900)
Observations	46199	46199	87385	46199	87385	87385
R-sq.	0.136	0.0706	0.0590	0.0859	0.0996	0.0479
Only Wage Workers	X	X		X		
<i>Panel C: Diff-in-diff complete interaction</i>						
Born post BW * Black	0.0773** (0.0348)	0.0135 (0.0234)	0.00945 (0.0160)	0.0248 (0.0218)	0.00477 (0.0155)	-0.0211 (0.0224)
Observations	59100	59100	110888	59100	110888	110888
R-sq.	0.260	0.179	0.124	0.351	0.418	0.0589
Only Wage Workers	X	X		X		
<i>Panel D: Triple differences</i>						
Born post BW * Black	0.0625** (0.0314)	0.0225 (0.0211)	0.0311** (0.0151)	0.0148 (0.0208)	0.0129 (0.0141)	-0.0187 (0.0198)
Observations	59101	59101	110890	59101	110890	110890
R-sq.	0.335	0.261	0.175	0.416	0.452	0.128
Only Wage Workers	X	X		X		

*Notes:* The unit of observation is sons. In columns 1, 2, and 4 we use the sons that were used in the estimates in Table 2 and 3. In columns 3, 5, and 6 we expand the sample to use the entire set of sons that we have both an OCC score and a constructed income score for. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$



Table 5: The boll weevil, years of schooling, and father and son migration status - only wage workers

	Years of Schooling				
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Difference-in-differences for Black sons</i>					
Born post BW	-0.0160 (0.186)	-0.0391 (0.187)	-0.0468 (0.192)	0.00524 (0.206)	0.114 (0.225)
Observations	12901	12307	10374	9847	7359
R-sq.	0.122	0.124	0.153	0.152	0.207
<i>Panel B: Difference-in-differences for white sons</i>					
Born post BW	0.0304 (0.0871)	0.0126 (0.0914)	0.0388 (0.101)	0.0718 (0.101)	0.0845 (0.123)
Observations	46199	41856	38081	32037	26827
R-sq.	0.0647	0.0672	0.0733	0.0853	0.102
<i>Panel C: Difference-in-differences complete interaction</i>					
Born post BW * Black	-0.0464 (0.204)	-0.0519 (0.206)	-0.0866 (0.215)	-0.0670 (0.230)	0.0294 (0.258)
Observations	59100	54163	48455	41884	34186
R-sq.	0.214	0.219	0.246	0.243	0.280
<i>Panel D: Triple differences</i>					
Born post BW * Black	-0.0856 (0.177)	-0.0374 (0.177)	0.0446 (0.201)	-0.00946 (0.212)	0.0571 (0.262)
Observations	59101	54021	48076	41366	33204
R-sq.	0.291	0.300	0.330	0.336	0.378
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table 6: Number of siblings born after the boll weevil by race and father migration status

	Number of male children born after boll weevil		
	(1)	(2)	(3)
Black	-0.0171*** (0.00440)	-0.0182*** (0.00462)	-0.0187*** (0.00538)
Observations	45350	41555	32247
R-sq.	0.148	0.154	0.178
Sample	Full	Father remained in South	Father remained in same state
Mean of dep. var.	0.169	0.175	0.188
SD of dep. var.	0.375	0.380	0.390

*Notes:* The unit of observation is fathers. This table presents results from estimating a regression where the dependent variable is the number of male children born to a father after the boll weevil. All columns control for county fixed effects, father's age at initial census, and initial census enumeration year fixed effects.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table 7: The boll weevil’s effect on the height of WWII enlistees

	Height (inches)	
	(1)	(2)
<i>Panel A: Difference-in-differences</i>		
<i>for Black sons</i>		
Born post BW	0.112*** (0.0433)	0.0876* (0.0491)
Observations	51658	51658
R-sq.	0.0317	0.0417
Mean of dep. var.	68.22	68.22
SD of dep. var.	2.63	2.63
<i>Panel B: Difference-in-differences</i>		
<i>for white sons</i>		
Born post BW	0.00664 (0.0257)	0.0243 (0.0289)
Observations	122128	122128
R-sq.	0.0296	0.0341
Mean of dep. var.	68.76	68.76
SD of dep. var.	2.53	2.53
<i>Panel C: Difference-in-differences</i>		
<i>complete interaction</i>		
Born post BW * Black	0.105** (0.0492)	0.0632 (0.0558)
Observations	173786	173786
R-sq.	0.0390	0.0451
Mean of dep. var.	68.60	68.60
SD of dep. var.	2.57	2.57
<i>Panel D: Triple differences</i>		
Born post BW * Black	0.0619* (0.0327)	0.115** (0.0548)
Observations	173786	173786
R-sq.	0.0343	0.0584
Mean of dep. var.	68.60	68.60
SD of dep. var.	2.57	2.57
County, birth year, enlistment year FE	X	X
County TT/DDD controls		X

*Notes:* The unit of observation is a World War II enlistee. In particular, we use the *U.S. World War II Army Enlistment Records, 1938-1946* from the National Archives and Records Administration (see text for more details). Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and panel D provides estimates for Equation 3. Column (1) controls for county of residence (at time of enlistment), birth year, and year of enlistment fixed effects. Column (2) of Panels A-C control for county of residence time trends, while Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year fixed effects where “county” is the county of residence at time of enlistment.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

# A Data Appendix

## A.1 Linking

To perform all of the linking we use the ABE algorithm, which is commonly used in economics and was developed by Abramitzky, Boustan, and Eriksson (Abramitzky, Boustan and Eriksson, 2012, 2014, 2019). This algorithm is similar to the algorithm used in Ferrie (1996) and Long and Ferrie (2013). We begin by adjusting first names for common nicknames and then standardize each first and surname using the NYSIIS algorithm, which transforms a name into a phonetic code. We then restrict our sample to individuals who are unique by NYSIIS first name, NYSIIS surname, birthplace, birth year, and race. Using these variables we search for the individual in the census we want to link them to. If we find a unique match we declare this observation to be a match. If we find multiple matches the observation is discarded. If we do not find a unique match we continue to search for individuals who match exactly on NYSIIS first name, NYSIIS surname, birthplace, and race, but we now allow birth year to differ by up to one year (e.g. if an individual in the 1910 Census is reported as being born in 1902 we will search for individuals in the 1940 Census with a birth year of 1901 and 1903). If still no unique match is found we continue to search for individuals who match exactly on NYSIIS first name, NYSIIS surname, birthplace, and race but we now allow birth year to differ by up to two years. The ABE algorithm is one of many algorithms currently used to link individuals across censuses. Other approaches include the Expectation-Maximization (EM) algorithm (Abramitzky, Mill and Pérez, 2020), machine learning approaches (Feigenbaum, 2016), and combinations of hand matched samples and computer programming (Bailey et al., 2020). Despite the variety of approaches, Abramitzky et al. (2019) show that automated approaches, including the ABE algorithm used in this paper, result in low false positive rates and similar coefficient estimates to a hand linked sample. Nevertheless, in section 5.5 we perform robustness checks on our results using alternative, and more conservative, linking approaches.

The results from this linking procedure are displayed in Appendix Table B.2. We begin with 425,317 Black fathers and 784,194 white fathers with sons under the age of 10 who were observed in either the 1900 or 1910 Censuses living in a county that would be invaded by the boll weevil in the next ten years. We were able to successfully link 24-25% of Black fathers and 32-33% of white fathers to the next census. We then located sons of successfully linked fathers that were born within 10 years of the boll weevil arriving in their father's initial county of residence and linked them to the 1940 Census. We were able to successfully link 18-25% of Black sons and 26-34% of white sons to the 1940 Census.<sup>23</sup> The overall link rate for Black sons was 20.5% (30,724 out of 149,875), while it was 29.3% for white sons (106,680 out of 363,531). Our overall link rates are slightly higher than the literature for two reasons. First, as mentioned by Baker, Blanchette and Eriksson (2020), we have the chance to match some sons twice (e.g. a son born in 1905 could be match in the 1900-1910 sample and again in the 1910-1920 sample). Even though we only keep the match from the earliest census, having multiple chances to match an individual mechanically increases the match rate. Second, unlike Baker, Blanchette and Eriksson (2020), we do not link both forward (from the 1900, 1910 and 1920 censuses to the 1940 census) and backward (from the 1940 census to the 1900, 1910, and 1920 censuses) and then take the intersection of the two linked sets.<sup>24</sup> This, again, results in mechanically higher link rates. Nevertheless, our link rates are fairly comparable to the literature. For example, Baker, Blanchette and Eriksson (2020) are able to link 18.6% of Black children (versus 20.5% for us) and 27.4% of white children (versus 29.3% for us). Abramitzky et al. (2021) are able to link 29.4% of white children, aged 0-16, from 1910 to 1940, which is almost identical to our link rate.

---

<sup>23</sup>Note that the link rates for sons whose fathers were successfully linked from 1910 to 1920 are much lower than the link rates for sons whose fathers were successfully linked from 1900 to 1910 due to some sons being observed in both the 1900-1910 sample and the 1910-1920 sample. In particular, it is possible for us to link a father from 1900 to 1910 and observe him with a son in 1910 that we link to 1940. We then observe the same father in 1910 and link him to 1920 and we observe his same son in 1910 and link him to 1940 again. In order to prevent this son from showing up twice in our dataset, we only keep the first instance in which he is observed, which was when we linked his father from 1900 to 1910. Thus, our link rate is lower for sons from fathers that were successfully linked between 1910 and 1920 because we discard these duplicated sons.

<sup>24</sup>Linking both forward and backward and taking the intersection is most applicable when one is not trying to link specific individuals. In our case, we have a specific set of sons that we are attempting to link.

## A.2 Sample Restrictions

We make several restrictions on who is included in our baseline sample of sons when weekly wages are the outcome. Appendix Table B.3 shows how our sample is reduced with each subsequent restriction. First, we exclude any individual that did not report wage or salary income. This could include individuals that were not employed or individuals that were self-employed. Census enumerators in 1940 were only supposed to record the wage and salary income earned as an employee.<sup>25</sup> Second, we exclude workers that reported a wage or salary income, but also reported being self-employed. As just explained, census enumerators were not supposed to report income for self-employed individuals. Income might have been reported for some self-employed individuals because they had a second job that paid them a wage or salary as an employee. Alternatively, some census enumerators might have made mistakes and recorded income for individuals that they were not supposed to. Regardless, it is difficult to know how to interpret income for self-employed individuals so we exclude them from our analysis. Third, we exclude individuals that worked for public work relief programs, such as the CCC or the WPA, were unemployed or not in the labor force. Individuals on work relief did receive a wage, but these wages were set by strict formulas and 75-80% of workers received the lowest wage on the scale (Bremer, 1975).<sup>26</sup>

Finally, we exclude individuals that worked less than 30 weeks in a year or that were in the top 3% of the weekly wage distribution. Individuals that worked less than 30 weeks in a year likely did not have a steady job. In addition, census enumerators were supposed to top-code any individual with an annual income over \$5,000 a year as having an income of \$5,000. This practice was not universally followed as there are several individuals for whom income is over \$5,000. We address these outliers by excluding individuals in the top 3% of the weekly wage distribution.<sup>27</sup> Appendix Figure B.6 displays a histogram of

---

<sup>25</sup>The instructions to census enumerators say to record wage or salary income “for work done as an employee, including public emergency project work, in 1939. Do not include the earning of businessmen, farmers, or professional persons derived from business profits, sale of corps, or fees” (Ruggles et al., 2020).

<sup>26</sup>We show the robustness of our results to the inclusion of both individuals that were on work relief and those that were unemployed or not in the labor force in Appendix Table B.10.

<sup>27</sup>For our sample, the 97th percentile of weekly wages is \$57.69, so individuals with wages above this

our weekly wage variable for our baseline sample, but it also includes the top 3 percentile of the weekly wage distribution. The vertical line is drawn at \$57.69, which represents the 97th percentile weekly wages. Observations above this are excluded in our final baseline sample.

Other papers that examine weekly wages from the censuses use similar approaches to determine who is included in the sample and how to address outliers. For example, Goldin and Margo (1992) and Margo (1995) compare wages across the 1940 and 1950 censuses. For the 1940 census, they include only wage or salary workers that worked more than 40 weeks in a year in their sample and impute incomes for those whose income was top-coded with 1.4 times the top code, which was \$5,000 in 1940.<sup>28</sup> They also truncate the bottom of their distribution by omitting individuals whose weekly wage was less than \$6. It is important to note that we compare wages within the 1940 census and, therefore, do not have to worry about wages being comparable across censuses.

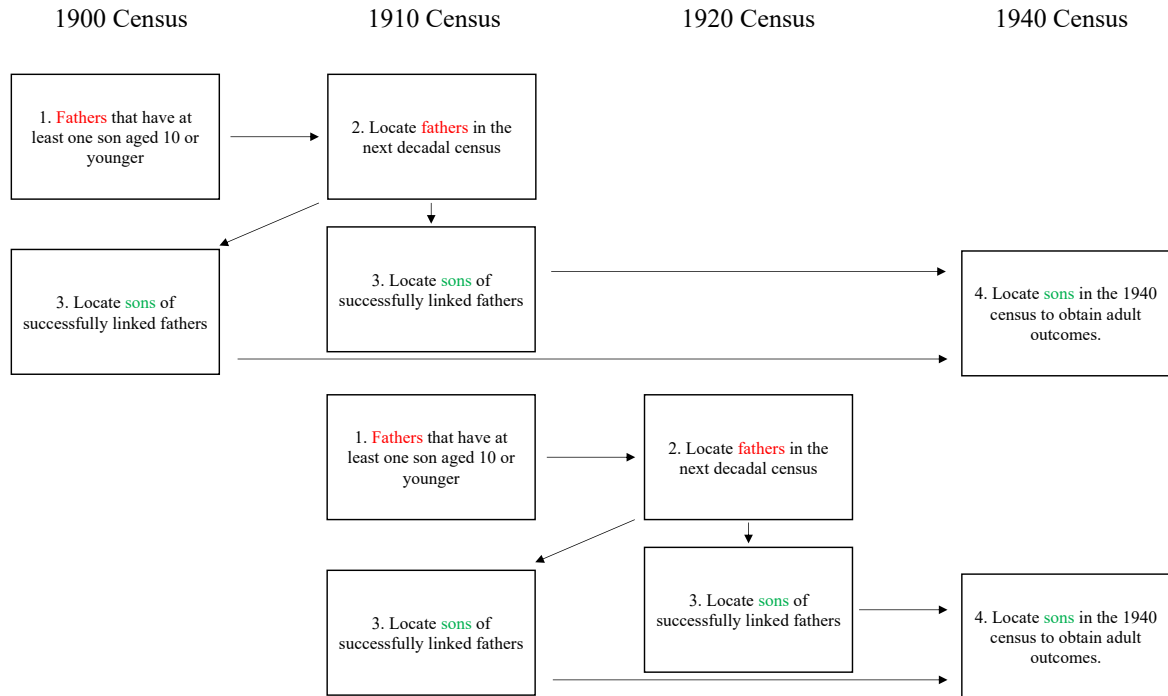
---

level are dropped from our analysis. Our results are not sensitive to changing this criteria to drop individuals with incomes above the 96th or 98th percentile or to censoring wages above \$100 per week at \$100. See Appendix Table B.10.

<sup>28</sup>Acemoglu and Angrist (2000) take a similar approach when using the 1960-1980 censuses. They censor weekly wages at the 98th percentile and replace wages above this with 1.5 times the 98th percentile wage. One reason we prefer truncation to censoring is that censoring takes outliers in the weekly wage distribution, which are basically all white sons, and makes them even bigger outliers. Nevertheless we show the robustness of our results to censoring in Appendix Table B.10.

## B Figure and Table Appendix

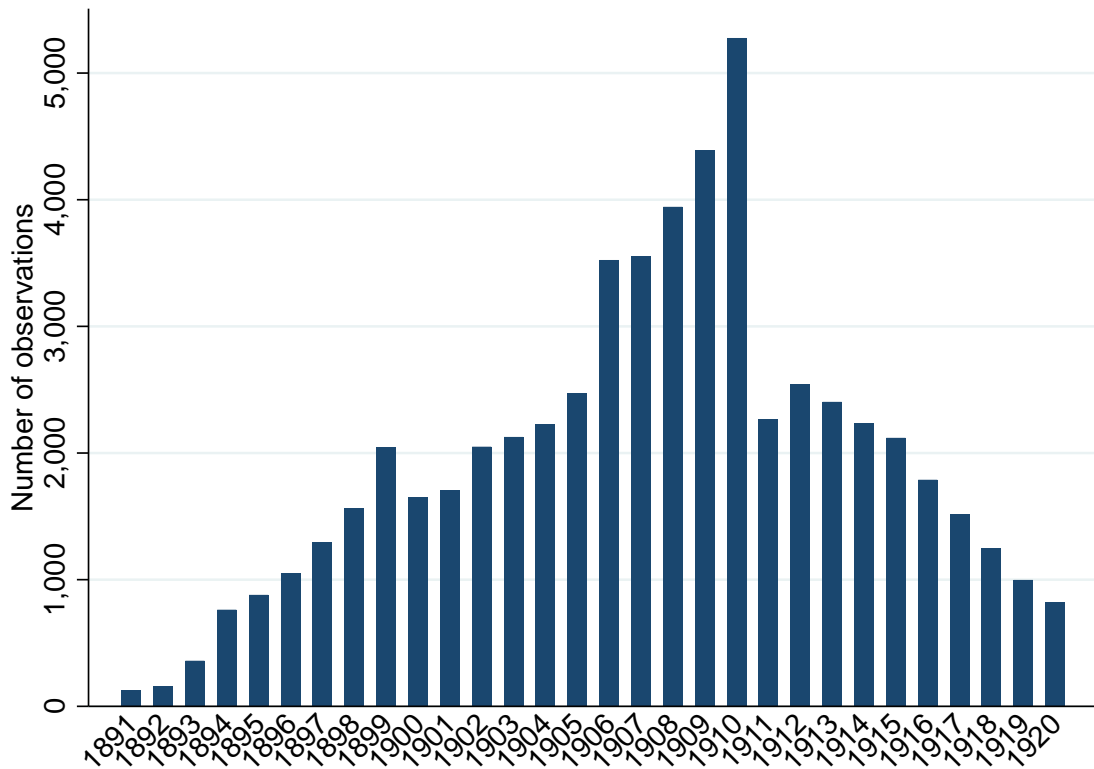
Figure B.1: Multi-step linking procedure



*Notes:* This figure displays our multi-step linking procedure.

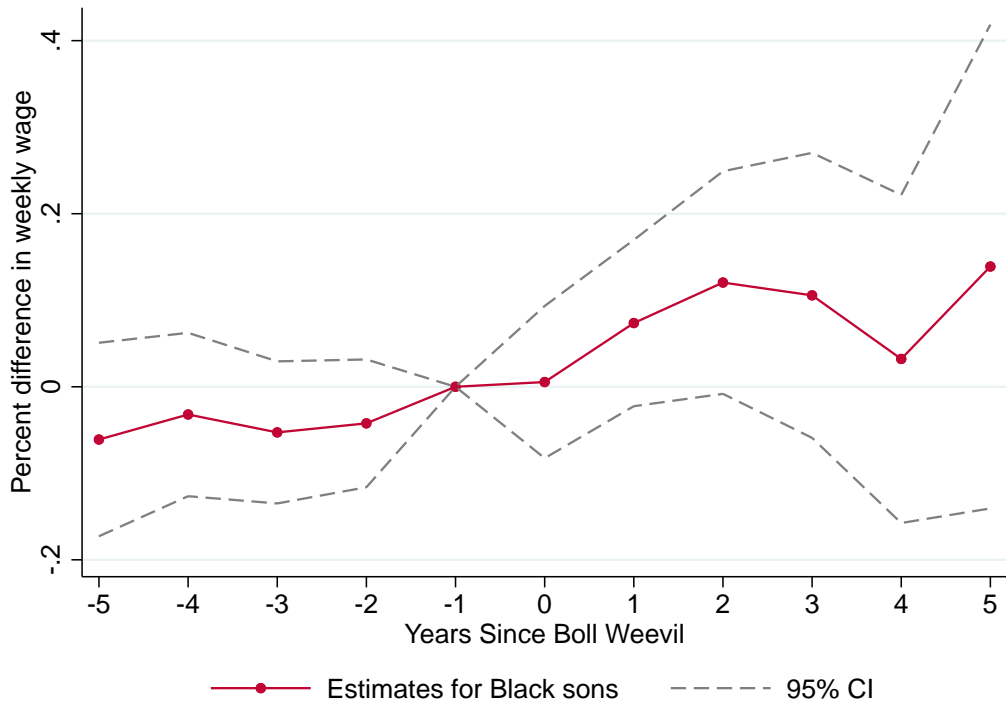


Figure B.2: Number of observations by birth cohort

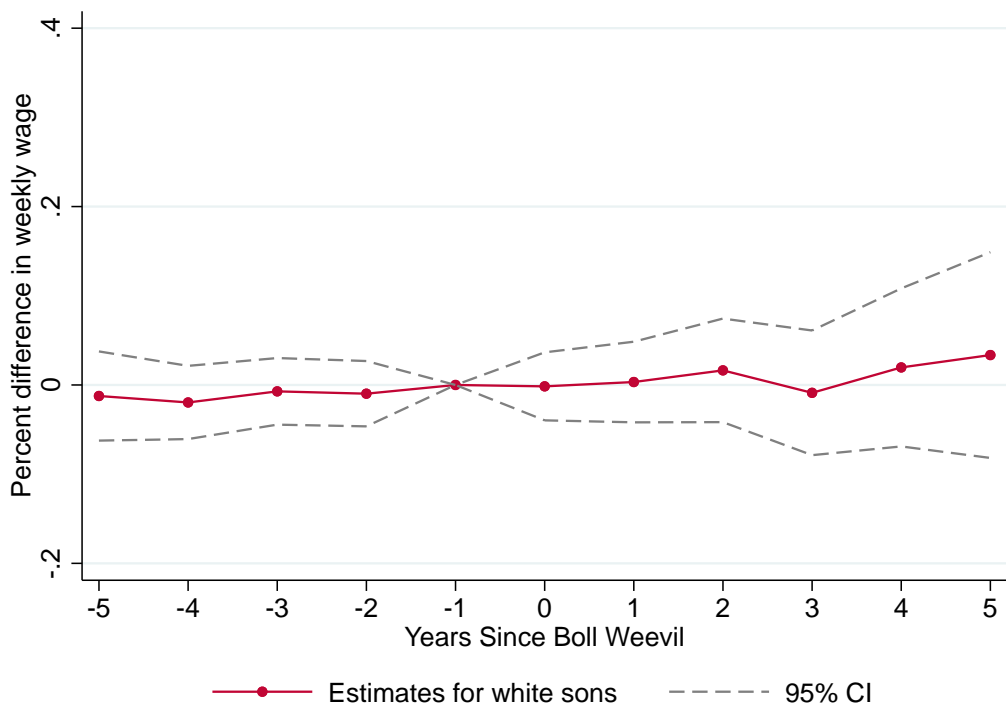


*Notes:* The graph shows the number of observations by birth cohort in our main sample, which is the sample used in Table 2 Panels C and D. Fathers in the 1900 census that we link to the 1910 census can have sons born between 1891 and 1910. Fathers in the 1910 census that we link to the 1920 census can have sons born between 1900 and 1920. Thus the years from 1900-1910 have sons from both the 1900 fathers and the 1910 fathers. The years prior to 1901 only have sons from the 1900 fathers and the years after 1910 only have sons from the 1910 fathers.

Figure B.3: Event studies  
 Panel A: Black sons

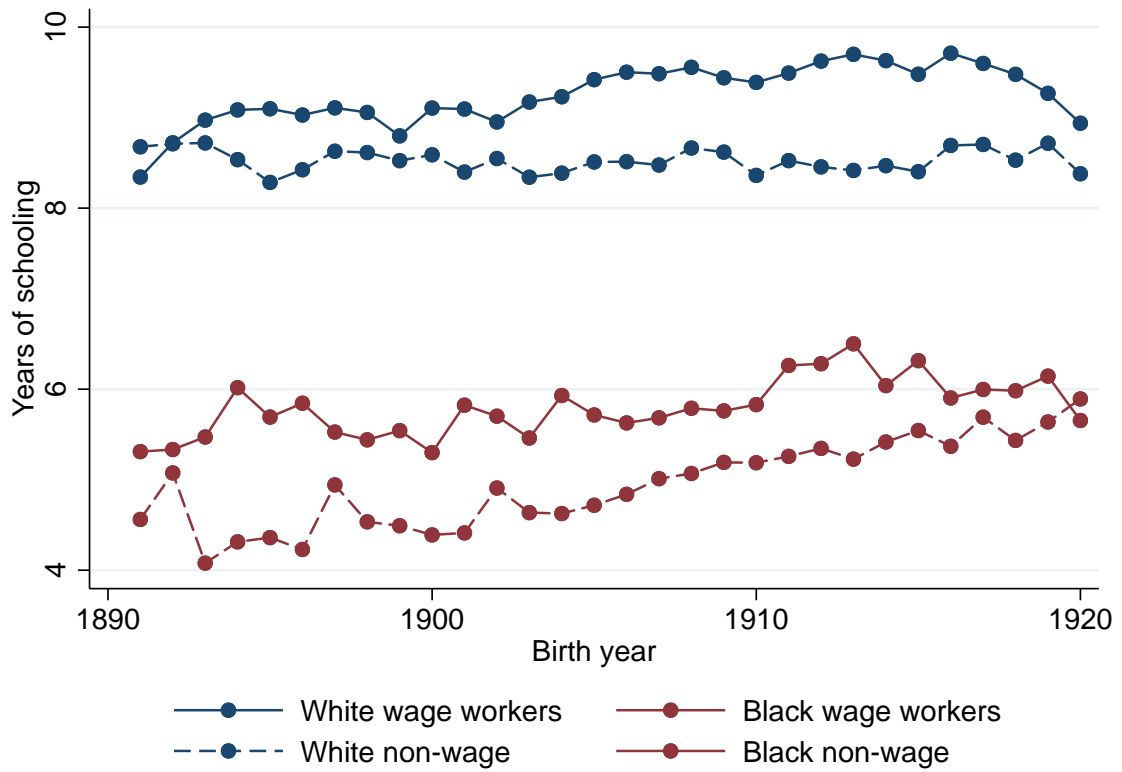


Panel B: White sons



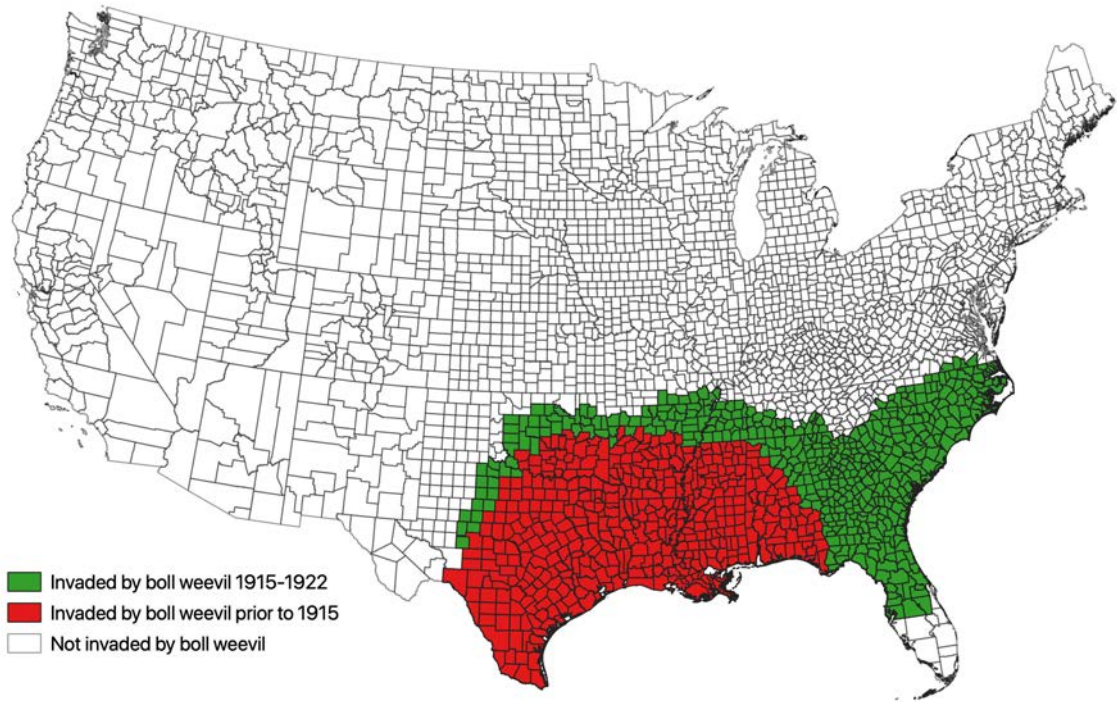
*Notes:* These figures show estimates of the event study equivalent to Equation 1 in the text. Panel A is estimated on Black sons, while Panel B is estimated on white sons. The coefficient -5 contains individuals born 5 to 10 years prior to the weevil’s arrival in their father’s original county of residence. The coefficient 5 contains individuals born 5 to 10 years after the weevil’s arrival in their father’s original county. The coefficient for individuals born 1 year prior to the weevil’s arrival is omitted, so all other coefficients are relative to those born 1 year prior to arrival.

Figure B.4: Years of schooling by birth cohort by race



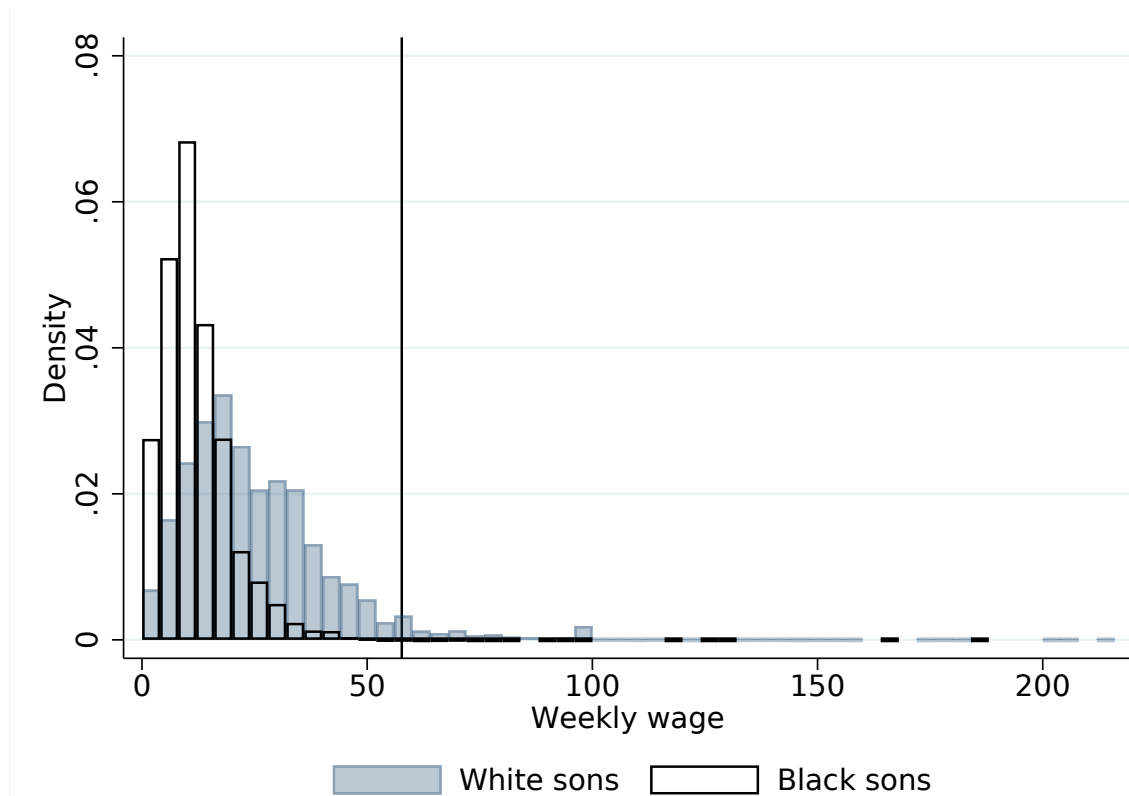
*Notes:* This figure compares years of schooling for Black wage workers (Table 2, Panel A) versus Black non-wage workers (Table 4, Panel A, columns 3, 5, and 6). It also compares years of schooling for white wage workers versus white non-wage workers.

Figure B.5: Counties invaded by boll weevil 1915-1922



*Notes:* This map displays counties that were invaded by the boll weevil. Counties in green were invaded by the boll weevil sometime between 1915 and 1922. Counties in red were invaded by the boll weevil prior to 1915. We use World War II enlistees born between 1915 and 1924 in our height analysis.

Figure B.6: Histogram of Weekly Wages by Race



*Notes:* The graph shows a histogram of our weekly wage variable using our baseline sample from Tables 2 and 3, but without truncating the top three percentiles of the distribution. The vertical black line in the figure is drawn at \$57.69, which represents the 97th percentile of the weekly wage distribution and the point at which we truncate our data in the baseline sample.

Table B.1: Comparison of linked sons to sons attempted to link

	Black sons		White sons	
	Linked	Attempted to Link	Linked	Attempted to Link
Age at first census	5.355**	5.410	5.744	5.733
In owner occupied housing <sup>†</sup>	0.270***	0.239	0.551***	0.518
In urban area	0.118***	0.111	0.146***	0.140
Father moved states	0.232**	0.237	0.306**	0.310
Father moved region	0.040**	0.043	0.089***	0.082
Father initially farmer	0.750**	0.756	0.665***	0.670
Born before 1905 (WWI Cohorts)	0.309***	0.287	0.345***	0.326
<i>N</i>	30724	149875	106680	363531

*Notes:* The stars report significance from a test of equality of means. All tests were conducted relative to the base sample that we attempted to link for each race.

†: Owner occupied housing status is available for fewer observations than is reported in the table. There are 30,711 linked Black sons and 149,802 Black sons that we attempted to link that have owner occupied housing status. There are 106,585 linked white sons and 363,186 white sons that we attempted to link that have owner occupied housing status.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.2: Linking results by race

	1900-1910 Black (1)	1900-1910 White (2)	1910-1920 Black (3)	1910-1920 White (4)
Fathers with at least one son under 10	147,081	296,583	278,236	487,611
Linked fathers (Match rate)	36,982 25%	96,624 33%	66,547 24%	158,381 32%
Sons of linked fathers	49,914	142,642	99,961	220,889
Linked sons of linked fathers* (Match Rate)	12,303 25%	48,206 34%	18,421 18%	58,474 26%

\* To be included in this sample a son has to be from a family with under 10 sons.

Table B.3: Sample sizes

	Black sons (1)	White sons (2)	Total (3)
All linked sons born within 10 years of boll weevil	30,724	106,680	137,404
Minus individuals with no reported wage	20,514	70,530	91,044
Minus self-employed workers	19,223	64,414	83,637
Minus individuals on work relief, unemployed, or not in labor force	15,910	55,411	71,321
Minus worked less than 30 weeks and top 3% of wage distribution	12,902	46,199	59,101

This table shows how we moved from the entire sample that we linked to the sub-sample used in Tables 2 and 3. The final row does not include individuals for whom we are missing OCC score, constructed income score, or both.



Table B.4: Cotton intensity and weekly wages

	Log(weekly wage)				
	(1)	(2)	(3)	(4)	(5)
<i>Difference-in-differences</i>					
<i>for Blacks</i>					
Born post BW	-0.00698 (0.0484)	-0.00473 (0.0503)	0.00694 (0.0521)	0.00508 (0.0497)	0.0493 (0.0584)
Born post BW * Above med. cotton	0.127** (0.0551)	0.115** (0.0572)	0.131** (0.0612)	0.104* (0.0593)	0.0952 (0.0689)
Observations	12901	12307	9337	9847	7449
R-sq.	0.136	0.140	0.162	0.165	0.206
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. This table display estimates for Equation 1 in the text, but we now interact the post boll weevil dummy variable with a high cotton intensity dummy variable. We define high cotton intensity as having above median cotton acres per capita (the median is 1.57 acres per capita). All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, census enumeration year dummies, and county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.5: The boll weevil, annual income, and father and son migration status

	Log(annual income)				
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Diff-in-diff</i>					
<i>for Black sons</i>					
Born post BW	0.0881** (0.0345)	0.0819** (0.0344)	0.0825** (0.0372)	0.0896** (0.0371)	0.107** (0.0447)
Observations	12901	12307	10374	9847	7359
R-sq.	0.135	0.140	0.149	0.165	0.214
<i>Panel B: Diff-in-diff</i>					
<i>for white sons</i>					
Born post BW	-0.00258 (0.0152)	0.000678 (0.0162)	0.00665 (0.0181)	0.00771 (0.0180)	0.0202 (0.0210)
Observations	46199	41856	38081	32037	26827
R-sq.	0.137	0.139	0.145	0.157	0.180
<i>Panel C: DiD complete</i>					
<i>interaction</i>					
Born post BW * Black	0.0908** (0.0364)	0.0812** (0.0366)	0.0758* (0.0396)	0.0819** (0.0399)	0.0869* (0.0479)
Observations	59100	54163	48455	41884	34186
R-sq.	0.258	0.258	0.292	0.275	0.332
<i>Panel D: Triple</i>					
<i>differences</i>					
Born post BW * Black	0.0698** (0.0332)	0.0653* (0.0337)	0.0775** (0.0372)	0.0689* (0.0388)	0.0539 (0.0464)
Observations	59101	54021	48076	41366	33204
R-sq.	0.333	0.338	0.374	0.371	0.428
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father’s initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where “county” is the father’s initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.6: The boll weevil, weeks worked, and father and son migration status

	Weeks worked				
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Diff-in-diff</i>					
<i>for Black sons</i>					
Born post BW	0.470 (0.358)	0.456 (0.359)	0.460 (0.407)	0.657 (0.404)	0.522 (0.481)
Observations	12901	12307	10374	9847	7359
R-sq.	0.0993	0.102	0.117	0.119	0.152
<i>Panel B: Diff-in-diff</i>					
<i>for white sons</i>					
Born post BW	-0.121 (0.164)	-0.0581 (0.171)	-0.0912 (0.179)	-0.0507 (0.187)	-0.0539 (0.216)
Observations	46199	41856	38081	32037	26827
R-sq.	0.0459	0.0505	0.0546	0.0641	0.0770
<i>Panel C: DiD complete</i>					
<i>interaction</i>					
Born post BW * Black	0.592 (0.385)	0.514 (0.386)	0.554 (0.428)	0.707 (0.431)	0.576 (0.511)
Observations	59100	54163	48455	41884	34186
R-sq.	0.0657	0.0703	0.0801	0.0858	0.105
<i>Panel D: Triple</i>					
<i>differences</i>					
Born post BW * Black	0.333 (0.336)	0.231 (0.338)	0.240 (0.380)	0.623 (0.388)	0.518 (0.467)
Observations	59101	54021	48076	41366	33204
R-sq.	0.159	0.169	0.187	0.204	0.235
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.7: The boll weevil timing, weekly wages, and sons living out of South

	Log(weekly wage)			Pr(Son not in South = 1)
	(1)	(2)	(3)	(4)
<i>Panel A: Difference-in-differences for Black sons</i>				
Born post BW	0.0771** (0.0326)	0.0682* (0.0393)	0.0705 (0.0449)	0.0115 (0.0172)
Observations	12901	11024	8522	12901
R-sq.	0.135	0.144	0.143	0.124
<i>Panel B: Difference-in-differences for white sons</i>				
Born post BW	0.0000625 (0.0144)	-0.00676 (0.0172)	0.00896 (0.0201)	0.00481 (0.00912)
Observations	46199	35608	27114	46199
R-sq.	0.136	0.141	0.128	0.118
<i>Panel C: Difference-in-differences complete interaction</i>				
Born post BW * Black	0.0773** (0.0348)	0.0751* (0.0412)	0.0616 (0.0469)	0.00674 (0.0194)
Observations	59100	46632	35636	59100
R-sq.	0.260	0.256	0.239	0.120
<i>Panel D: Triple differences</i>				
Born post BW * Black	0.0625** (0.0314)	0.0822** (0.0372)	0.107** (0.0444)	-0.0116 (0.0185)
Observations	59101	46632	35636	59101
R-sq.	0.335	0.331	0.315	0.205
Sample	Full	BW invaded 1905 or later	BW invaded 1910 or later	Full

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. In column 2 we restrict to sons whose fathers were originally residing in a county invaded by the boll weevil in 1905 or later. In column 3 we restrict to sons whose fathers were originally residing in a county invaded by the boll weevil in 1910 or later. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.8: The returns to migrating out of the South

†: Who moved?	Log(weekly wage)		
	(1)	Son (2)	Father (3)
<i>Panel A: Difference-in-differences for Black sons</i>			
Born post BW	0.0771** (0.0326)	0.0794** (0.0318)	0.0693** (0.0320)
Moved out of South <sup>†</sup>		0.683*** (0.0142)	0.175*** (0.0341)
Born post BW * Moved out of South <sup>†</sup>		-0.0630 (0.0425)	0.367*** (0.0962)
Observations	12901	12901	12901
R-sq.	0.135	0.273	0.139
<i>Panel B: Difference-in-differences for white sons</i>			
Born post BW	0.0000625 (0.0144)	0.000357 (0.0151)	-0.00434 (0.0146)
Moved out of South <sup>†</sup>		0.234*** (0.0107)	0.118*** (0.0112)
Born post BW * Moved out of South <sup>†</sup>		-0.00814 (0.0185)	0.0608** (0.0285)
Observations	46199	46199	46199
R-sq.	0.136	0.152	0.139
<i>Panel C: Difference-in-differences complete interaction</i>			
Born post BW * Black	0.0773** (0.0348)	0.0791** (0.0345)	0.0736** (0.0342)
Moved out of South <sup>†</sup>		0.234*** (0.0108)	0.118*** (0.0113)
Moved out of South <sup>†</sup> * Black		0.448*** (0.0170)	0.0571 (0.0360)
Born post BW * Moved out of South <sup>†</sup>		-0.00813 (0.0186)	0.0608** (0.0286)
Born post BW * Moved out of South <sup>†</sup> * Black		-0.0549 (0.0471)	0.306*** (0.0979)
Observations	59100	59100	59100
R-sq.	0.260	0.296	0.262

<i>Panel D: Triple differences</i>			
Born post BW * Black	0.0625** (0.0314)	0.0654** (0.0318)	0.0571* (0.0313)
Moved out of South <sup>†</sup>		0.231*** (0.0115)	0.116*** (0.0116)
Moved out of South <sup>†</sup> * Black		0.441*** (0.0180)	0.0282 (0.0373)
Born post BW * Moved out of South <sup>†</sup>		0.00594 (0.0204)	0.0650** (0.0296)
Born post BW * Moved out of South <sup>†</sup> * Black		-0.0425 (0.0491)	0.332*** (0.104)
Observations	59101	59101	59101
R-sq.	0.335	0.368	0.337

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text, but include an interaction with a dummy variable for a son moving out of the South in column 2 and a father moving out of the South in column 3. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3 but both panels include the complete triple interaction of born after the boll weevil, moving out of the South (son or father), and Black. A son is considered to have moved out of the South if they are not observed in the South census region in 1940. A father is considered to have moved out of the South if he is observed outside the South census region in the second census we observe him in (i.e. the 1910 or 1920 census). All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.9: The boll weevil, years of schooling, and father and son migration status -  
Full sample

	Years of Schooling				
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Difference-in-differences for Black sons</i>					
Born post BW	-0.000791 (0.106)	-0.0355 (0.108)	0.0134 (0.113)	-0.00279 (0.114)	0.0819 (0.125)
Observations	30723	29486	25249	23602	18976
R-sq.	0.0825	0.0813	0.0970	0.0981	0.130
<i>Panel B: Difference-in-differences for white sons</i>					
Born post BW	0.0363 (0.0555)	0.0131 (0.0567)	0.00779 (0.0590)	0.0367 (0.0646)	0.00483 (0.0693)
Observations	106680	97149	90929	74032	67159
R-sq.	0.0510	0.0511	0.0552	0.0671	0.0769
<i>Panel C: Difference-in-differences complete interaction</i>					
Born post BW * Black	-0.0365 (0.123)	-0.0483 (0.124)	0.00590 (0.129)	-0.0395 (0.131)	0.0771 (0.143)
Observations	137403	126635	116178	97634	86135
R-sq.	0.174	0.177	0.195	0.197	0.221
<i>Panel D: Triple differences</i>					
Born post BW * Black	-0.0163 (0.115)	0.0301 (0.114)	0.110 (0.124)	-0.0378 (0.126)	0.121 (0.144)
Observations	137404	126562	116038	97483	85835
R-sq.	0.214	0.219	0.240	0.246	0.274
Sample	Full	Father remained in South	Son remained in South	Father remained in same state	Son remained in same state

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.10: The boll weevil and weekly wages - Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log(weekly wage)								
<i>Panel A: Diff-in-diff for Black sons</i>									
Born post BW	0.0771** (0.0326)	0.0644** (0.0301)	0.0726** (0.0302)	0.0768** (0.0327)	0.0788** (0.0329)	0.0672* (0.0361)	0.0696* (0.0411)	0.0741** (0.0304)	0.0789** (0.0331)
Observations	12901	13827	13798	12898	12907	10978	11111	12901	12934
R-sq.	0.135	0.129	0.128	0.135	0.135	0.150	0.141	0.133	0.135
<i>Panel B: Diff-in-diff for white sons</i>									
Born post BW	0.0000625 (0.0144)	-0.00151 (0.0140)	-0.00332 (0.0146)	-0.00268 (0.0147)	0.000422 (0.0145)	0.0132 (0.0155)	0.0250 (0.0162)	-0.000344 (0.0138)	0.00860 (0.0146)
Observations	46199	49131	48340	45790	47103	41112	43663	46199	48411
R-sq.	0.136	0.131	0.131	0.134	0.141	0.142	0.142	0.139	0.151
<i>Panel C: DiD complete interaction</i>									
Born post BW * Black	0.0773** (0.0348)	0.0661** (0.0324)	0.0761** (0.0328)	0.0795** (0.0349)	0.0783** (0.0348)	0.0540 (0.0379)	0.0447 (0.0425)	0.0745** (0.0326)	0.0705** (0.0353)
Observations	59100	62958	62138	58688	60010	52090	54774	59100	61345
R-sq.	0.260	0.241	0.251	0.257	0.265	0.272	0.265	0.262	0.274
<i>Panel D: Triple differences</i>									
Born post BW * Black	0.0625** (0.0314)	0.0579* (0.0306)	0.0520* (0.0304)	0.0650** (0.0317)	0.0582* (0.0310)	0.0654* (0.0345)	0.0534 (0.0366)	0.0541* (0.0300)	0.0521* (0.0315)
Observations	59101	62959	62140	58689	60011	52091	54775	59101	61346
R-sq.	0.335	0.314	0.325	0.333	0.339	0.353	0.344	0.338	0.345
Restrictions	Baseline	Include work relief	Include unemployed and not in labor	Truncate at 96 pct.	Truncate at 98 pct.	Work 40 or more weeks	Goldin and Margo (1992)	Adjust farm laborer wages for prerequisites	Wages censored at 100



(Table B.10 continued)

[ht] *Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. Column 1 uses our baseline sample, which includes wage workers that were in the labor force, not on work relief, not unemployed, and worked more than 30 weeks in the year. We truncate our baseline sample of wage workers at the 97th percentile of weekly wages. Column 2 includes workers on work relief. Column 3 includes workers that were unemployed or not in the labor force, but reported a wage for the previous year. Column 4 changes the truncation from the 97th percentile to the 96th percentile and column 5 changes it to the 98th percentile. Column 6 restricts the baseline sample to individuals that worked more than 40 weeks in the year. Column 7 uses the restrictions mentioned in Goldin and Margo (1992). In particular, when computing weekly wages we replace incomes above \$5,000 with 1.4 times \$5,000, or \$7,000. We also restrict to individuals that worked more than 40 weeks in the year. We continue to restrict to only wage workers that were in the labor force, not on work relief, and not unemployed. Column 8 uses information from Collins and Wanamaker (2017) to adjust the wages of farm laborers for prerequisites that they received in the form of room and board. In particular, we scale farm laborers weekly wages up by 26%, which is the amount of prerequisites for farm laborers in 1940. Column 9 censors weekly wages at \$100 and replaces wages about \$100 per week with \$100. We selected \$100, because a worker who worked 50 weeks in a year earning \$100 a week, would have an annual income of \$5,000 a year, which is the top coded annual income in the 1940 census. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.11: Sensitivity of main empirical results to alternative linking algorithms

	Log(weekly wage)			
	(1)	(2)	(3)	(4)
<i>Panel A: Difference-in-differences</i>				
<i>for Black sons</i>				
Born post BW	0.0771** (0.0326)	0.0975 (0.0716)	0.0708* (0.0371)	0.109 (0.0808)
Observations	12901	3651	10081	2607
R-sq.	0.135	0.292	0.162	0.360
<i>Panel B: Difference-in-differences</i>				
<i>for white sons</i>				
Born post BW	0.0000625 (0.0144)	0.0170 (0.0240)	0.000200 (0.0154)	0.0168 (0.0253)
Observations	46199	19146	40697	16718
R-sq.	0.136	0.178	0.147	0.194
<i>Panel C: Difference-in-differences</i>				
<i>complete interaction</i>				
Born post BW * Black	0.0771** (0.0347)	0.0805 (0.0733)	0.0705* (0.0390)	0.0926 (0.0807)
Observations	59100	22797	50778	19325
R-sq.	0.260	0.294	0.265	0.302
<i>Panel D: Triple differences</i>				
Born post BW * Black	0.0625** (0.0314)	0.0414 (0.0719)	0.0601* (0.0359)	0.147* (0.0877)
Observations	59101	21102	50503	17365
R-sq.	0.335	0.418	0.350	0.430
Sample	Full	Exact Name and Age Matches	Unique Name and Age Matches	Exact and Unique Name and Age Matches

*Notes:* The unit of observation is sons. Panels A and B of this table display estimates for Equation 1 in the text. Panel C provides estimates for Equation 2 and Panel D provides estimates for Equation 3. All columns control for: father's initial county fixed effects, birth year dummies, birth order dummies, and census enumeration year dummies. Panels A-C control for county specific time trends using the county that we first observe the father in (i.e. the county they resided in when the 1900 or 1910 census was taken). Panel D controls for the triple difference interactions: county-by-race, birth year-by-race, and county-by-birth year where "county" is the father's initial county of residence. Birth order is determined by the age of the sons who have the same father in the censuses. Thus, it does not take into account older sons who either moved out of the house or died before the censuses were taken. It also does not take into account daughters. Standard errors are clustered at the fathers-initial-county level.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.12: Number of siblings born after the boll weevil by race and father migration status - Full sample

	Number of male children born after boll weevil		
	(1)	(2)	(3)
Black	-0.0183*** (0.00347)	-0.0195*** (0.00360)	-0.0180*** (0.00420)
Observations	94194	86916	67303
R-sq.	0.345	0.348	0.364
Sample	Full	Father remained in South	Father remained in same state
Mean of dep. var.	0.262	0.271	0.288
SD of dep. var.	0.508	0.514	0.528

*Notes:* The unit of observation is fathers. This table presents results from estimating a regression where the dependent variable is the number of male children born to a father after the boll weevil. All columns control for county fixed effects, age at initial census, and initial census enumeration year fixed effects.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$

Table B.13: The boll weevil and pellagra death rates in North and South Carolina counties

	Log(pellagra death rate)	
	(1)	(2)
Post BW	-0.235*** (0.0490)	-0.181*** (0.0566)
Post BW * Percent black		-0.0758* (0.0400)
Observations	1312	1312
R-sq.	0.580	0.582

*Notes:* The unit of observation is counties in North and South Carolina. This table displays estimates for a regression of the pellagra death rate on a post boll weevil dummy variable for counties in North Carolina for the year 1915-1925 and for counties in South Carolina for the years 1916-1925. The share of a county's population that was Black was taken from the 1910 census and standardized to have a mean of zero and a standard deviation of one. All columns control for county and year fixed effects.

\* =  $p < 0.10$

\*\* =  $p < 0.05$

\*\*\* =  $p < 0.01$