NBER WORKING PAPER SERIES

EARLY LIFE SHOCKS, MARKET ADJUSTMENTS, AND BLACK-WHITE INEQUALITY

Karen Clay
Ethan J. Schmick

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020, Revised February 2025

This paper previously circulated under the title “The Impact of an Environmental Shock on Black-White Inequality: Evidence from the Boll Weevil.” We began this project with Werner Troesken, and he provided extremely helpful insights during the project’s early stages. We miss him greatly. We would also like to thank Ran Abramitzky, Daniel Fetter, James Fenske, Andy Ferrara, Daniel Jones, Jessica LaVoice, Paul Rhode, Allison Shertzer, Andrew Smyth, Randy Walsh, and Gavin Wright for helpful comments and suggestions. This paper also benefited from comments by seminar participants at Case Western University, Marquette University, Stanford University, University of Wisconsin-Madison, and Warwick University and conference participants at the ASSA, the Liberal Arts Colleges Development Economics Conference, the North American Regional Science Association Conference, and the All UC Economic History 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.

Early Life Shocks, Market Adjustments, and Black-White Inequality
Karen Clay and Ethan J. Schmick
NBER Working Paper No. 27101
May 2020, Revised February 2025
JEL No. I24, J10, J62, N32

ABSTRACT

This paper investigates the long run impacts of an early life agricultural shock on Black and White sons in the U.S. South. The boll weevil, one of the most destructive agricultural pests in American history, decreased cotton production and resulted in substantial changes to the Southern economy. The impact of this shock on children who were born before and after its arrival is not a priori obvious; it could be positive or negative depending on whether children born after the shock experienced better or worse early life conditions. To examine the empirical effects of this shock on Black and White fathers and sons, the analysis makes use of cross-census links from the Census Tree (Buckles et al., 2023) and race-specific difference-in-differences and triple difference empirical strategies. We find the arrival of the boll weevil benefited Black sons in the long run, as reflected in two 1940 measures of income – wages and imputed income – and did not harm White sons. These differential gains decreased inequality. We provide empirical and historical evidence on a range of mechanisms through which early life conditions may have improved for Black sons relative to White sons.

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

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

1 Introduction

Black-White inequality fell during much of the twentieth century as measured by wages and for some measures of intergenerational mobility (Margo, 2016; Bayer and Charles, 2018; Jácome, Kuziemko and Naidu, 2021). A variety of reasons for this have been advanced, most notably Black migration to the North and increases in Black education levels (Heckman, Lyons and Todd, 2000; Carruthers and Wanamaker, 2017; Collins, 2021). Migration can only be one part of the story, however, since up to 1940 relatively little migration had occurred.[1] The effects of education are unclear in the first half of the twentieth century. For example, Carruthers and Wanamaker (2017) suggest that human capital accounts for a large share of the Jim Crow wage gap in 1940. Yet, recent analysis by Mohammed and Mohnen (2025) suggests that Rosenwald schools had little effect on Black mens' occupational outcomes in 1940. Other factors were likely at play for the 1900-1940 period.

This paper investigates the long-run impact of a large negative agricultural shock on Black and White sons in the U.S. South. The boll weevil, an agricultural pest that destroys cotton crops, invaded the U.S. South during the early twentieth century. Starting in 1892 it began to gradually spread through the cotton growing region of the U.S. and by 1922 all cotton growing areas had been infested. The arrival of the boll weevil had a modest initial negative effect that grew over time as the infestation worsened. Within 5 years of its arrival in a county, total cotton production fell 39-50% (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017). While cotton continued to be produced, the mix of crops shifted towards local food crops. This shock affected approximately 22% of the U.S. population and 75% of the Black population.

The impact of this shock on early life conditions for children exposed to it is not a priori obvious. Before the shock, Black and White households faced different labor market frictions, as a result of anti-enticement laws and other state and local policies (Naidu, 2010; Hornbeck and Naidu, 2014; Ager, Brueckner and Herz, 2017). These

[1] In the 1940 census, 80% of Black individuals were living in the South.

labor market frictions made it particularly difficult for Black tenant farmers to work for other farm owners, obtain better contracts, or shift occupations. The boll weevil led to a shift in agricultural mix from cotton, which required large amounts of labor at specific times of the year, to a more diversified portfolio of agricultural products including food crops, which required labor on a year-round basis. As a result, many farm owners ended or changed existing tenancy contracts, reducing labor market frictions. In response, many Black and White households migrated out of their original counties and obtained new occupations or new tenancy contracts (Ager, Brueckner and Herz, 2017; Lange, Olmstead and Rhode, 2009). These changes may have positively or negatively impacted early life conditions for children in these households by changing household economic conditions, nutrition, schooling, fertility, and patterns of racial violence.

To examine the empirical effects of this shock on Black and White fathers and sons, the analysis makes use of cross-census links from the Census Tree (Buckles et al., 2023) and race-specific difference-in-differences and triple difference empirical strategies.[2] Our linking procedure starts with Black and White fathers who had a young son in the census prior to the boll weevil's arrival in their county of residence (1900 or 1910). They are linked to the next census (1910 or 1920) after the boll weevil had arrived in their initial county of residence. We observe fathers' characteristics and any changes, including whether they migrated to a new location, changed occupation, or had additional sons. Sons of these fathers are then linked to the 1940 Census. We observe their adult outcomes such as wage income, occupation, and migration from their father's initial location.[3] The race-specific difference-in-differences specifications leverage variation in fathers' initial county of residence and the timing of the boll weevil's arrival. The triple difference specifications pool Black and White sons and, therefore, additionally leverage variation in race, allowing us to examine inequality. The analysis controls for contemporaneous public health campaigns to eradicate

[2] The Census Tree uses FamilySearch's proprietary machine learning linking algorithm, combined with links from other sources, to generate a comprehensive and highly accurate set of links between censuses.

[3] Fewer than 10% of Black and White sons in our sample are observed outside the South in 1940.

hookworm and malaria and the establishment of Rosenwald schools (Bleakley, 2007, 2010; Aaronson and Mazumder, 2011).

The paper has three main findings. First, the arrival of the boll weevil benefited Black sons in the long run, as reflected in two 1940 measures of income – wages and imputed income – and did not harm White sons.[4] Black sons born after the shock experienced relative increases in wages compared to White sons of 11 percent and relative increases in imputed incomes of 5 percent. These gains were not driven by migration out of the South. Black sons who remained in the South also had increased wages and incomes.

Second, these differential gains decreased inequality as measured by wages and income rank. The increases in Black sons' wages and imputed incomes were substantial, accounting for 6-15% of the Black-White wage gap in 1940. Estimates from intergenerational mobility analyses show that being born after the boll weevil increased Black sons' imputed income rank by 1 and had a negative but insignificant effect on White sons' imputed income rank. This represents a 12% increase in the average income rank for Black sons. These increases are consistent with Margo (2016), which shows that Black-White wage ratios were rising from 1900-1940, and with Derenoncourt et al. (2024), which shows that White-Black wealth ratios were falling.[5]

Third, Black sons born after the boll weevil appear to have experienced improvements in early life conditions through a range of mechanisms that may have led to increased wages and imputed incomes. We present empirical and historical evidence on Black fathers migration and occupational upgrading; improvements in sons' nutrition; improvements in sons' schooling; increases in resources available to sons, including reductions in the number of children in the household; and reductions in racial violence.[6] The evidence suggests that Black sons born after the boll weevil may

[4]Wages were self reported by wage workers, which were less than half the population, and imputed income is based on occupation and is available for everyone in the labor force.

[5]The literature on intergenerational mobility is mixed. Collins and Wanamaker (2022) finds little variation in Black intergenerational mobility over 1880-2000, while Jácome, Kuziemko and Naidu (2021) find increases in Black intergenerational mobility from the 1910-1929 cohort to the 1940-1959 cohort.

[6]Our findings on reductions in fertility and increases in schooling are consistent with Ager, Herz and Brueckner (2020) and Baker, Blanchette and Eriksson (2020).

have benefited through all of these mechanisms.

Our paper contributes to four literatures. The first is the literature on Black-White inequality. The economic history literature on the Black-White wage gap is very large but has predominantly focused on the Great Migration or education and on the period after 1940 (Bayer and Charles, 2018; Collins, 2021; Carruthers and Wanamaker, 2017; Collins and Margo, 2006; Derenoncourt, 2022). Two exceptions are Margo (2016) and Derenoncourt et al. (2024). Margo (2016) provides new evidence on Black-White wage inequality, showing that it decreased continuously from 1870 to 1940. Derenoncourt et al. (2024) constructs new estimates of White-Black per capita wealth ratios from 1860-2020 and show that convergence stalled after 1950. A small but growing recent literature examines Black and White intergenerational mobility (Collins and Wanamaker, 2022; Saavedra and Twinam, 2020; Jácome, Kuziemko and Naidu, 2021; Ward, 2023). This paper presents new evidence on the effect of the boll weevil on Black-White wages and intergenerational mobility.

The second is the literature on labor market mobility, including the effects of coercion and restrictions in labor markets. Coercion has occurred in a wide range of historical contexts (Acemoglu and Wolitzky, 2011; Naidu and Yuchtman, 2013; Bobonis and Morrow, 2014; Buggle and Nafziger, 2021). One important context is the U.S. South (Engerman, 1992; Naidu, 2010; Hornbeck and Naidu, 2014; Ager, Brueckner and Herz, 2017). This paper provides new evidence on changes in Black fathers' labor market conditions relative to White fathers and the impact this had on their sons.

The third is the literature on early life conditions (Almond and Currie, 2011; Almond, Currie and Duque, 2018). Two papers suggest that markets may play a role in mitigating shocks on long run outcomes. Shah and Steinberg (2017) find that drought leads to higher childhood educational outcomes, because wages in local labor markets fall. Mulmi et al. (2016) find that food markets in Nepal mitigate the effects of climatic shocks on child height. This paper contributes new evidence on Black and White early life conditions during an important and understudied historical period. It highlights the role that a range of market mechanisms may have played in mitigating shocks.

The fourth is the literature on the boll weevil. This literature has almost exclusively examined the place-based effects of the boll weevil, documenting its effects at the county level on the production of cotton and other agricultural products, population, the price of land, farm wages, female labor force participation, and education (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017; Clay, Schmick and Troesken, 2019; Feigenbaum, Mazumder and Smith, 2020; Ager, Herz and Brueckner, 2020; Ferrara, Ha and Walsh, 2022). One important exception is Baker, Blanchette and Eriksson (2020), who use linked census data to examine the impact of the boll weevil on years of schooling. The boll weevil was likely to have affected the educational outcomes of school age children through reduced demand for child labor in cotton picking. Their paper compares men who were ages 4-18 when the boll weevil arrived in a county with others who were 19-30 and finds that men 4-9 years old at the time of the boll weevil's arrival had 0.24-0.36 greater years of schooling. Our paper finds a similar effect of the boll weevil on years of schooling, but only for Black sons that were wage workers. One reason for the lack of an effect for Black sons in a broader sample that includes non-wage workers may be because all of the men in our sample were at most 9 years old when the boll weevil arrived and so were treated according to the Baker, Blanchette and Eriksson (2020) definition. Our paper provides new individual-level evidence on the effects of the boll weevil on other economic outcomes, such as wages and income, for Black and White fathers and their sons born around the time of its arrival. It also provides new evidence on the mechanisms through which the boll weevil may have affected sons.

2 Historical Background

The arrival of the boll weevil in the cotton belt during the late 1800s and 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 over the next 30 years. By 1922, the entire cotton growing region of the United States had been infested. Appendix Figure A.1 shows the annual progression of the boll weevil through the cotton belt from 1892 to 1922.

The U.S. Department of Agriculture (1951), Ransom and Sutch (2001), Lange, Olmstead and Rhode (2009), Ager, Brueckner and Herz (2017), and Ferrara, Ha and Walsh (2022) all find that the arrival of the boll weevil had large negative effects on cotton yields and production. The effects emerged over time as the boll weevil infestation grew. In the first year, the amount of cotton ginned fell by 10% and within five years it had fallen by 39-50% (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017). This resulted in substantial disruptions to tenancy arrangements and affected local labor markets through decreased farm wages and female labor force participation (Ager, Brueckner and Herz, 2017; Bloome, Feigenbaum and Muller, 2017). Lange, Olmstead and Rhode (2009), Ager, Brueckner and Herz (2017), and Ferrara, Ha and Walsh (2022) all show that the boll weevil significantly reduced the value of land. These negative effects on tenancy, labor markets, and land values induced substantial migration throughout the South (Lange, Olmstead and Rhode, 2009; Ager, Brueckner and Herz, 2017; Feigenbaum, Mazumder and Smith, 2020).

Black and White families may have been differentially impacted by this negative shock to cotton production for two reasons. First, Black fathers were more likely to work in agriculture and were lower on the agricultural ladder than White fathers (Alston and Kauffman, 1997; Alston and Ferrie, 2005; Collins, Holtkamp and Wanamaker, 2024). In our sample, 83% of Black fathers and 73% of White fathers worked in agriculture; 28% of Black fathers and 18% of White fathers were farm laborers. Second, Black and White households differed in their ability to shift to other agricultural contracts, owners, or occupations. Anti-enticement laws imposed fines on planters who made offers to laborers already under contract (Naidu, 2010). At the state and local level, Black codes appear to have acted as constraints on Black households (Cohen, 1976; Roback, 1984). When the boll weevil led owners to voluntarily end tenancy contracts, Black and White households were able to re-optimize to other agricultural contracts, owners, occupations, and locations.

The boll weevil could affect the outcomes of Black and White children born after its arrival positively or negatively through a range of mechanisms, including migration and changes in household income; changes in resources available to children including conditions during pregnancy, the number of children in the household, and time spent on children; changes in diet; changes in access to schooling; and changes in the frequency of racial violence as measured by lynchings. We briefly speak to each of these mechanisms below.

The boll weevil led to substantial migration, which may have affected children in a variety of ways. In our sample, 22% of Black fathers and 37% of White fathers moved to a different county around the time of the weevil's arrival. The moves were generally from one rural area to another, but 5% of Black fathers and 7% of White fathers moved from a rural area to an urban area. Some fathers moved to other states – 6% of Black fathers and 12% of White fathers. A small number of fathers moved out of the South – 2% of Black fathers and 3% of White fathers. These moves likely changed incomes and access to food and may have affected resources available to children within and outside of the home.

Household income may have fallen for Black households relative to White households, as Black women and Black children reduced their labor force participation following the boll weevil's arrival (Ager, Brueckner and Herz, 2017; Baker, Blanchette and Eriksson, 2020). Women and children in White households were less likely to work initially, and so White households experienced smaller declines. Further, Black fathers in our sample saw small increases in imputed income and income rank after the weevil's arrival, while White fathers saw larger increases.[7]

The boll weevil may have led to changes in resources available to children, including conditions during pregnancy, the number of children in the household, and time spent

[7] We use a measure of imputed income from Collins and Wanamaker (2022), which is based on region-race-occupation cells and uses 1940 or 1960 income values – individuals who did not change their region-race-occupation cell between censuses are assigned the same imputed income. If the Black-White wage ratio in a region-occupation cell rose (or fell) over time, this would not be captured. Margo (2016) shows three series spanning 1900-1920. Two of the series show slight increases in average Black-White income ratios and one shows a slight decrease.

on children. Ager, Brueckner and Herz (2017) provide evidence that Black women reduced their labor force participation after the arrival of the boll weevil. Their post-boll weevil pregnancies may have been less stressful and led to better outcomes for their children. We show that Black fathers had fewer sons after the boll weevil relative to White fathers, which also may have led to better outcomes for sons. More broadly, lower female labor force participation may have led to greater time spent on children, benefiting sons born after the boll weevil.

Black and White households, including pregnant mothers and young children, may have eaten a somewhat better diet post-boll weevil. This could occur along a number of possible dimensions. One dimension would be eating more calories, holding the mix of foods in the diet constant. In their study of African American dietary patterns at the beginning of the twentieth century, Dirks and Duran (2001) highlight "frank undernutrition" in rural areas. If the boll weevil lowered food prices this may have increased the consumption of calories. A second dimension along which diet might have improved was replacing imported Midwestern corn meal, a staple of the Southern diet that was usually consumed daily, with corn meal from locally grown corn. Locally grown corn was more nutritious because the germ had not been removed (Clay, Schmick and Troesken, 2019). A third dimension along which diet might have improved was eating a greater variety of food as food prices changed. Dirks and Duran (2001) note that diets in urban areas were generally better than diets in rural areas in part due to differences in relative prices. Household food prices may have changed either because of changes in local production or, for families that migrated, because relative prices differed in the destination county. A fourth dimension was an increase in home production of food. Hawthorne, Montgomery and Dixon (1922), a USDA study that compared farms in Sumpter County, Georgia before and after the arrival of the boll weevil, reports (p. 16): "More croppers had gardens in 1918 [after the boll weevil] than in 1913 [before the boll weevil]. In fact, some farmers had come to require their croppers to grow a part or all of their own vegetables, in order to limit their need for credit and give them a better chance of breaking even should the boll weevil or other conditions cause heavy losses."[8]

After the arrival of the boll weevil, children may have been able start attending school, attend school for more days, or attend a Rosenwald schools. Sons were less likely to be employed in agriculture following the arrival of the boll weevil, and thus may have been able to start attending school or attend school for more days (Baker, 2015; Baker, Blanchette and Eriksson, 2020). Although unrelated to the boll weevil, children may also have been impacted by Rosenwald schools. In 1912, Booker T. Washington and Julius Rosenwald began a program to improve Black students' access to high quality eduction in the rural South (Aaronson and Mazumder, 2011). By 1920, about 6% of Black 7-13 year olds living in rural areas were educated in a Rosenwald school. By 1926, the percentage had risen to about 25%. Given the availability of data on the timing of opening of Rosenwald schools in counties, our empirical analysis directly controls for Rosenwald exposure.

To the extent that the arrival of the boll weevil reduced racial violence, Black families and children may have benefited in a variety of economic and non-economic ways. Feigenbaum, Mazumder and Smith (2020) find that the arrival of the boll weevil was associated with reductions in lynchings and the construction of Confederate monuments. The paper provides evidence that (p. 1) "the reductions in coercion were responses to African American out-migration." Increases in perceived safety may have changed the types of jobs and educational opportunities available to household members and children, reduced stress, and improved health.

In addition to the boll weevil, some sons in our sample were exposed to public health campaigns to reduce hookworm and malaria. The Rockefeller Sanitary Commission's hookworm eradication campaign began in 1910, and treatment started in earnest in most locations in 1912 (Rockefeller Sanitary Commission, 1910-1914; Bleakley, 2007). Many children and adults were treated and treatment continued over time. The malaria eradication campaign began around 1920 and malaria death rates

[8]Gardens were not necessarily common prior to the boll weevil. A Department of Labor study of Black migration (Leavell, Dillard and of Labor Division of Negro Economics, 1919) indicates one reason why (p. 87) "Often no active encouragement is given tenants and laborers to cultivate these gardens and sometimes the labor is pushed so hard by the landlord that no time is allowed for this work."

started falling in 1922 (Ferrell, 1931; Bleakley, 2010). For hookworm and malaria, it is not clear which birth cohorts were treated by the campaigns. As a result, in our empirical analysis, we flexibly control for exposure to these campaigns.[9]

3 Data

This section describes the data on the boll weevil's arrival, the construction of our linked sample of fathers and sons, our measures of income, controls, and summary statistics for the sample.

3.1 The Boll Weevil

Data on the year the boll weevil first arrived in a county were taken from Lange, Olmstead and Rhode (2009), which originally came from USDA boll weevil maps.[10] Appendix Figure A.1 shows the annual progression of the boll weevil through the cotton belt.[11] Our main estimating sample consists of sons born to fathers that were initially observed living in a county invaded by the boll weevil between 1901 and 1920.[12] Counties invaded by the boll weevil between 1901 and 1920 are located in Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, and Texas and contain more than 80% of the population that would eventually be invaded by the boll weevil.

3.2 Linking

To study the impact of the boll weevil on children born around the time of its arrival, we generated a linked sample of fathers and their sons. A linked sample of fathers

[9]The controls are described in greater detail in Section 3.4 and Appendix B.

[10]In a robustness check, we use a newspaper-based measure of the arrival of the boll weevil from Ferrara, Ha and Walsh (2022) to correct for possible measurement error in the USDA maps.

[11]Based on visual inspection of Appendix Figure A.1 we changed the date of arrival of the boll weevil for three counties. In Iredell County, North Carolina the boll weevil's arrival is coded as occurring in 1922 but appears to occur in 1921 on the map. Wake County, North Carolina is coded as 1922 but appears to be 1921, and Cherokee County, South Carolina is coded as 1921 but appears to be 1920.

[12]No counties were invaded by the boll weevil in 1900, which is why we start in 1901.

and sons is important because the boll weevil resulted in large changes to Southern labor and agricultural markets and led to substantial migration throughout the South. By linking both fathers and sons, we are able to observe and control for migration and changes in fathers' occupational status when analyzing sons' long-run outcomes.

Appendix B provides a detailed explanation of our linking method. The Census Tree (Buckles et al., 2023; Price et al., 2023*a*,*c*,*b*,*d*,*e*) is used for all linking. The Census Tree compiles links from seven different sources, including FamilySearch's proprietary machine learning linking algorithm.

Appendix Figure B.1 provides an example of our linking procedure. Fathers in the 1900 complete count census observed in a county invaded by the boll weevil between 1901 and 1910 are linked forward to 1910. Sons of these fathers, observed in either 1900 or 1910, are linked forward to 1940 to obtain their adult outcomes. We repeat this procedure with fathers in the 1910 census; fathers in the 1910 complete count census observed in a county invaded by the boll weevil between 1911 and 1920 are linked forward to 1920. We then link the sons of these fathers from either 1910 or 1920 to 1940.[13] The two sets of linked sons are stacked; only the earliest observation is kept for sons who are linked multiple times.[14] In our main analysis, we require that a link be identified in at least two of the seven sources contained in the Census Tree and that one of the sources is FamilySearch's proprietary machine learning linking algorithm. We demonstrate the robustness of our results to the number of linking sources.

We generate inverse propensity score weights to make our linked sample representative of the population. To do this, we combine our linked sample with the relevant population that was not linked and estimate a propensity of being linked. The inverse of this propensity score is used to weight the linked sample. Appendix Table B.2 compares the relevant population with the unweighted linked and weighted

[13]In the 1900, 1910, and 1920 complete count censuses about 20% of Black sons and 6% of White sons who were under age 10 and living in a county invaded by the boll weevil lived apart from their father. This is in line with Moehling (2004). Our results do not, necessarily, generalize to sons that lived apart from their father. However, differences in family structure will only bias our results if family structure systemically changed after the arrival of the boll weevil. We find no evidence of this. See Appendix B and Appendix Table B.1 for more details.

[14]For example, if a son is linked from 1900 to 1940 and 1910 to 1940 we only keep the 1900 to 1940 link.

linked sample. The weighting improves the representativeness of the sample. A more detailed discussion of weighting is provided in Appendix B.

3.3 Weekly Wages and Imputed Income

We use two 1940 measures of income as our main outcome variables: imputed income and weekly wages.

Our imputed income sample contains any son who was in the labor force, reported an occupation, and was not on work relief. Individuals working for public work relief programs had their occupation in the program recorded, which might not have corresponded to their usual occupation. 88% of all the sons in our linked sample are in the imputed income sample. For these sons, we assign an imputed income based on the method described in Collins and Wanamaker (2022).[15] Collins and Wanamaker (2022) construct imputed incomes within a region, race, and occupation cell (e.g., Black miners in the South) using the 1940 and 1960 censuses. This measure differentiates between farm owners and tenants/sharecroppers, both of whom constitute a large part of our sample. The measure is also useful for assigning income to self-employed individuals, such as farmers, who did not report their annual income in the 1940 census.

One weakness of our imputed income measure is that it does not vary within occupation-region-race cells. For example, all Black miners in the South are assigned the same imputed income. We believe there was substantial heterogeneity within occupations as many Black Southerners worked manual labor jobs that would either pay a piece rate or reward those that were more physically fit. Accordingly, we also use weekly wages as an outcome. We define weekly wages as an individual's yearly income in 1939 divided by the number of weeks they worked in 1939. The 1940 census was the first census to ask about income, although it only asked about wage or salary income earned as an *employee* in 1939. Thus, our wage worker sample takes the

[15] We are grateful to Collins and Wanamaker (2022) for providing us with their code to construct this measure of income.

imputed income sample and further restricts to only wage workers who worked over 30 weeks in 1939 and were not in the armed forces or unemployed. 42% of all the sons in our linked sample are in the weekly wage sample. The sensitivity of the results to a number of alternative wage worker restrictions is explored in Section 5.2. Appendix Table B.3 shows how these restrictions affect the sample size. Appendix Figure B.3 displays the cumulative distribution of both imputed income and wages for Black and White sons in our sample.

Appendix Table A.1 examines if selection into our wage worker and imputed income samples changed after the boll weevil's arrival. If Black sons born after the boll weevil's arrival were more or less likely to be wage workers (or be in the imputed income sample) it could bias our results on wages and imputed income depending on the direction of the selection. Panels A and B of Appendix Table A.1 regress an indicator for being in the wage worker or imputed income samples on fathers' initial county fixed effects, sons' birth year fixed effects, and other controls. There is no evidence that Black or White sons born after the boll weevil were more or less likely to be wage workers or in the imputed income sample. In Panel C, there is no evidence that Black sons born after the boll weevil were more or less likely to be wage workers or in our imputed income sample than White sons born after the boll weevil. We conclude that selection into our wage worker and imputed income samples is unlikely to bias our results.

3.4 Controls

We directly control for the Rosenwald rural school initiative, the malaria eradication campaign, and the hookworm eradication campaign in our empirical strategy. As mentioned in Section 2, these initiatives all occurred during the 1910s and 1920s. To control for the Roesnwald rural school initiative we use the same county-birth cohort measure of Rosenwald exposure as Aaronson and Mazumder (2011). To control for the malaria eradication campaign, we use county-level data on malaria ecology from Hong (2011). To control for the hookworm eradication campaign, we use county-level data on hookworm infection rates from Thoman (2009). The birth cohorts that were treated by the malaria and hookworm eradication campaigns are not well defined; or example, during the hookworm campaign many children *and* adults were treated *and* treatment continued over time. Accordingly, we flexibly control for these campaigns by interacting county-level malaria ecology or hookworm infection rates with birth year fixed effects. More details on each of these control variables is contained in Appendix B.

3.5 Summary Statistics

Appendix Table A.2 provides summary statistics for fathers (Panel A) and sons (Panel B) in our wage worker and imputed income samples. In Panel A there are large cross-sectional differences between Black and White fathers in most socioeconomic variables. For example, Black fathers have substantially lower imputed incomes than White fathers.[16] A few additional things are worth noting. First, Black fathers' imputed income ranks rose slightly, on average, between the two censuses; White fathers' imputed income ranks rose by much more. Second, as we highlighted in Section 2, the boll weevil led to significant migration throughout the South. Over 20% of Black fathers and over 35% of White fathers moved counties in the 10-year interval between the first and second census. This can be compared to migration rates for the entire United States of about 12%.[17] However, most fathers stayed in the same state and very few – about 2% of the Black fathers and 3% of White fathers – moved out of the South.

Panel B documents that there are also large cross-sectional differences in most socioeconomic variables for sons. It is worth noting that the migration patterns of Black and White sons are fairly similar. Approximately the same percentage of Black and White sons were living outside the South. Similarly, about the same percentage were living in a large city.

[16]We assign fathers an imputed income in the same way we do for sons; by following the method described in Collins and Wanamaker (2022).

[17]Calculated using the 1900 to 1910 Census Tree linked sample.

4 Empirical Strategy

The main analysis takes two approaches to estimating the effect of the boll weevil on weekly wages and imputed incomes: a race-specific difference-in-differences model and a triple difference model. The difference-in-differences model takes advantage of variation across counties and birth years, while the triple difference model takes advantage of variation across counties, birth years, and racial groups.

4.1 Race-specific Difference-in-Differences

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

$$\begin{aligned} Outcome_{ict} = \beta * \mathbb{I}[Born\ post\ boll\ weevil = 1]_{ct} + \theta_c + \alpha_t \\ + \delta_b + \xi_e + X_{ct}\zeta' + \epsilon_{ict} \end{aligned} \tag{1}$$

In the above specification, i indexes sons, c indexes the counties that sons' fathers were initially residing in (in either 1900 or 1910), and t indexes sons' birth years. Thus, $Outcome_{ict}$ is the adult outcome, as observed in 1940, of son i, whose father initially resided in county c, and who was born in year t.

$Outcome_{ict}$ is one of two 1940 measures of income: log of weekly wages or log of imputed income. The construction of both of these measures is detailed in Section 3. $\mathbb{I}[Born\ post\ boll\ weevil = 1]_{ct}$ is an indicator variable that takes a value of one for sons in birth cohorts born after the boll weevil arrived in their father's initial county of residence c. Sons born in the year the boll weevil first arrived in their father's initial county of residence are coded as not being treated (i.e. $\mathbb{I}[Born\ post\ boll\ weevil = 0]_{ct}$). The treatment variable is defined this way because the boll weevil usually did not become active and spread until the harvest season and infestation was often light in the first year. The effects on cotton production grew as the infestation worsened. Appendix Figure A.2 shows an event study of cotton production and supports this choice.

Specification (1) includes fixed effects to control for fathers initial county (θ_c), birth year (α_t), birth order (δ_b), and census enumeration year (ξ_e).[18] X_{ct} is a vector of controls for hookworm eradication, malaria eradication, and Rosenwald schools. Specification (1) is estimated separately by race, which controls for potential omitted variables that arise in models not run separately (Feigenberg, Ost and Qureshi, 2023).

Standard errors are clustered based on bins of longitude. Longitude is used, because much of the movement of the boll weevil was west to east (see Appendix Figure A.1). Our longitude bins correspond to a quarter degree of longitude (about 14 miles in the southern United States), which results in 96 bins in the sample.[19] We demonstrate robustness to other forms of spatial autocorrelation (Conley, 1999). Inverse propensity score re-weighting, as described in Section 3, is used in all regressions involving sons.

Both the date of the weevil's arrival and county fixed effects in Specification (1) are based on *father's initial county of residence*. We do this to account for migration that occurred after the weevil's arrival. Recall from Appendix Table A.2 that over 20% of Black fathers and over 35% of White fathers moved from their initial county. Appendix Table A.3 examines county characteristics for fathers who moved between the two censuses. Black and White fathers moved to counties that were more populous, more urbanized, had more manufacturing establishments, and had less cotton production. In short, they moved to places that were very different from their initial county of residence. By comparing sons born after the boll weevil whose fathers migrated to individuals in their father's initial county of residence (as opposed to their father's final county of residence) we answer the following counterfactual: how did sons who likely would have grown up in the same county in the absence of the boll weevil fare later in life?

Race-specific difference-in-differences event studies are estimated using the following specification:

[18] Birth order is determined based on the linked sons we observe with each father.

[19] Counties hit by the boll weevil between 1901 and 1920 run from about the 76th meridian west to the 100th meridian west.

$$Outcome_{ict} = \sum_{r=-10}^{r=-2} \beta_r + \sum_{r=0}^{r=9} \beta_r + \theta_c + \alpha_t + \delta_b + \xi_e + X_{ct}\zeta' + \epsilon_{ict} \quad (2)$$

In Specification (2), r indexes time relative to the arrival of the boll weevil in a father's initial county of residence; the year the weevil arrived is $r = 0$. The β_rs are coefficients on indicator variables that take a value of one if a son was born in year r relative to the weevil's arrival in their father's initial county of residence. All other variables are defined analogously to Specification (1). A complete set of leads and lags are estimated and the coefficient on the year -1 is omitted; all the β_rs are interpreted relative to event year -1. Specification (2) is estimated separately for Black and White sons, and standard errors are clustered based on 96 bins of longitude.

To evaluate pre-trends, we estimate a "parametric event study" as described in Dobkin et al. (2018). In this event study, the lead indicators (i.e. indicators for event time -10 through -2) are replaced with a linear pre-trend in event time r.[20] Everything else in the parametric event study remains the same as Specification (2).

The identifying assumption of the difference-in-differences approach is that the outcomes of sons of the same race born before and after the boll weevil would have trended similarly in the absence of the boll weevil. Because we include father's initial county fixed effects, this assumption need only hold for sons whose fathers initially resided in the same county. We present evidence that our results are robust to a number of possible concerns regarding identification including: alternative definitions of treated cohorts, spillovers to untreated cohorts from counties already treated, and issues regarding treatment effects that vary over time in staggered difference-in-differences designs.

[20] The precise specification is:

$$Outcome_{ict} = \gamma * r + \sum_{r=0}^{r=9} \beta_r + \theta_c + \alpha_t + \delta_b + \xi_e + X_{ct}\zeta' + \epsilon_{ict}$$

4.2 Triple Difference: Comparing Black Sons and White Sons

The second empirical approach involves triple differences and uses the following specification:

$$
\begin{aligned}
Outcome_{ict} = {} & \psi * \mathbb{I}[Black = 1]_i + \gamma * \mathbb{I}[Born\ post\ boll\ weevil = 1]_{ct} * \mathbb{I}[Black = 1]_i \\
& + \theta_c + \theta_c * \mathbb{I}[Black = 1]_i + \alpha_t + \alpha_t * \mathbb{I}[Black = 1]_i \\
& + \delta_b + \delta_b * \mathbb{I}[Black = 1]_i + \xi_e + \xi_e * \mathbb{I}[Black = 1]_i \\
& + X_{ct}\zeta' * \mathbb{I}[Black = 1]_i + \theta_c * \alpha_t + \epsilon_{ict} \qquad (3)
\end{aligned}
$$

The triple difference specification allows us to leverage all three dimensions of the data: birth years; fathers' initial county of residence; and race. The specification is similar to Specification (1), but interacts the post-boll weevil treatment variable with a dummy variable indicating if individual i is Black ($\mathbb{I}[Black = 1]_i$). Specification (3) also includes interactions of all fixed effects and control variables with the Black indicator ($\theta_c * \mathbb{I}[Black = 1]_i$, $\alpha_t * \mathbb{I}[Black = 1]_i$, $\delta_b * \mathbb{I}[Black = 1]_i$, $\xi_e * \mathbb{I}[Black = 1]_i$, and $X_{ct}\zeta' * \mathbb{I}[Black = 1]_i$). Finally, it includes birth year and father's initial county of residence fixed effects interacted with each other ($\theta_c * \alpha_t$).[21] Note that the stand-alone post-boll weevil indicator variable (i.e. $\mathbb{I}[Born\ post\ boll\ weevil = 1]_{ct}$) and the stand-alone controls (i.e. X_{ct}) are absorbed by the interaction of fathers' initial county of residence and birth year fixed effects (i.e. $\theta_c * \alpha_t$). Standard errors are clustered based on 96 longitude bins and we show robustness to other forms of spatial autocorrelation (Conley, 1999). Inverse propensity score re-weighting, as described in Section 3, is used in all regressions involving sons. γ in Specification (3) estimates the effect of being born after the boll weevil for Black sons relative to White sons whose fathers initially resided in the same county.

[21] These birth year-by-county fixed effects result in the estimation of over 9,000 coefficients, most of which are not significant. Accordingly, we demonstrate the robustness of this specification to using birth year and father's initial *state* of residence fixed effects interacted with each other (i.e. $\sigma_s * \alpha_t$ instead of $\theta_c * \alpha_t$).

In triple difference event studies, we plot the interaction between the event-time indicator variables and an indicator variable if a son is Black. The specification is:

$$\begin{aligned}
Outcome_{ict} = \psi * \mathbb{I}[Black = 1]_i + \sum_{r=-10}^{r=-2} \beta_r * \mathbb{I}[Black = 1]_i + \sum_{r=0}^{r=9} \beta_r * \mathbb{I}[Black = 1]_i \\
+ \theta_c + \theta_c * \mathbb{I}[Black = 1]_i + \alpha_t + \alpha_t * \mathbb{I}[Black = 1]_i \\
+ \delta_b + \delta_b * \mathbb{I}[Black = 1]_i + \xi_e + \xi_e * \mathbb{I}[Black = 1]_i \\
+ X_{ct}\zeta' * \mathbb{I}[Black = 1]_i + \theta_c * \alpha_t + \epsilon_{ict} \qquad (4)
\end{aligned}$$

Once again the β_rs are coefficients on indicator variables that take a value of one if a son was born in year r relative to the weevil's arrival.

5 Results

This section begins by presenting the main differences-in-differences and triple difference results. It then examines the robustness of the results to alternative measures of socioeconomic status, alternative sample restrictions, alternative treatment definitions, stricter linking criteria, alternative difference-in-differences estimators, and spatial standard errors. Finally, it examines intergenerational mobility.

5.1 Main Results

Figure 1 presents the results of the event studies. All the event studies plot the pre-trend from a parametric event study making it easy to identify the coefficients that are statistically different from the pre-trend. In Panels (a)-(d), which plot the difference-in-differences event studies from Specification (2), there are statistically positive effects of the boll weevil on Black sons' weekly wages and imputed income compared to the pre-trend but no effect for White sons. In Panels (e) and (f), which plot the triple difference event studies from Specification (4), Black sons born after the boll weevil experienced significant increases in wages and imputed income relative to White sons born after the boll weevil.

It is worth noting that the event studies show a gradual improvement in weekly wages and imputed income for Black sons born in the first few years after the arrival of the boll weevil. This is consistent with contemporary reports indicating that the boll weevil had a modest initial effect that grew over time as the infestation worsened (Lange, Olmstead and Rhode, 2009). This is in line with Appendix Figure A.2, which shows an event study for cotton production. Cotton production did not fall dramatically for the first two years of infestation.

Appendix Figure A.3 presents the coefficients from the event studies just shown alongside coefficients from event studies using the estimator proposed by Sun and Abraham (2021). The Sun and Abraham (2021) estimator eliminates biases that are potentially present in difference-in-differences and triple difference estimates; these biases emerge due to staggered treatment timing and evolving treatment effects. The estimates with the two methods are similar.

Table 1 presents our main difference-in-differences and triple difference results. Panels A and B report coefficients from the difference-in-differences Specification (1) for Black and White sons, respectively. Panels C and D report coefficients from the triple difference Specification (3) using both Black and White sons. Panel C uses father's initial *county of residence* fixed effects interacted with birth year fixed effects. The interaction of these fixed effects results in the estimation of over 9,000 coefficients, most of which are not significant. Accordingly, Panel D uses father's initial *state of residence* fixed effects interacted with birth year fixed effects. Columns 1-3 use the log of weekly wages as the dependent variable and columns 4-6 use the log of annual imputed income as the dependent variable. Columns 1 and 4 do not use inverse propensity score re-weighting. Columns 2 and 5 use inverse propensity score re-weighting, which is used throughout the rest of the analysis to make the linked sample more representative of the population. Columns 3 and 6 add controls for hookworm, malaria, and Rosenwald schools.[22] Columns 3 and 6 are our preferred

[22]Appendix Table A.4 presents the results from adding each of these controls individually. Appendix Figure A.4 plots the coefficients on hookworm infection rate interacted with birth year fixed effects and malaria ecology interacted with birth year fixed effects.

empirical specifications for the remainder of the paper and are the ones used in the event studies in Figure 1 and Appendix Figure A.3.

In the difference-in-differences specifications in Panel A of Table 1, Black sons born after the boll weevil had statistically significantly higher wages and imputed incomes than Black sons born before the boll weevil. In column 3, Black sons born after the boll weevil had wages that were 10 percent higher than Black sons born before the boll weevil. In column 6, Black sons born after the boll weevil had imputed incomes that were 4 percent higher than Black sons born before the boll weevil. In Panel B, White sons born after the boll weevil had wages and imputed incomes that were not statistically significantly different from White sons born before the boll weevil.

In the triple difference specifications in Panels C and D, Black sons born after the boll weevil had statistically significantly higher wages and imputed incomes than White sons born after its arrival. In column 3 Black sons born after the boll weevil had wages that were 11 to 12 percent higher. In column 6, Black sons born after the boll weevil had imputed incomes that were 5 to 6 percent higher.

Differential Black-White gains of 11 percent in wages and 5 percent in imputed income are large. In 1940, the Black-White wage gap was about 0.71 log points and the Black-White imputed income gap was about 0.85 log points.[23] The differential gains, therefore, account for 15% of the Black-White wage gap and 6% of the impute income gap.

Table 2 shows our results are similar if we restrict attention to Black and White sons of fathers who remained in the South or who themselves remained in the South.[24] The point estimates in the difference-in-differences specifications in Panel A and the triple difference specifications in Panel C for these sons are slightly smaller than, although not statistically significantly different from, our main estimates in Table 1. These results suggest that gains in wages and imputed income experienced by Black

[23]To get these numbers, we ran a regression of each dependent variable on county (in 1940) and birth year fixed effects and an indicator if a son was Black.

[24]A father moved out of the South if they were not in the South census region in the second census we observe them in. A son moved out of the South if they were not living in the South in the 1940 census.

sons born after the boll weevil were not driven by migration out of the South. In the difference-in-differences specification in Panel A, being born after the arrival of the boll weevil has a positive and statistically significant effect on the probability of migration out of the South for Black sons in the wage worker and imputed income samples. In the triple difference specification in Panel C, being born after the arrival of the boll weevil has a positive and statistically significant effect on the probability of migration out of the South for Black sons in the wage worker sample.

The results in Table 2 do not imply that those who migrated out of the South experienced similar or worse outcomes to those that did not migrate. In Appendix Table A.5, which shows the returns to migration and whether these returns varied for sons born after the boll weevil, we find large returns to moving out of the South.[25] In Panels A and C, two things are worth noting. First, the coefficients on "Born post BW" (Panel A) and "Born post BW x Black" (Panel C) continue to be positive and statistically significant. In contrast, the interaction terms with "Born post BW" and "Move out of South" are not statistically significant. Second, Black and White sons did benefit from moving out of the South. Specifically, the coefficients on "Moved Out of the South" and "Moved Out of the South x Black" are positive, statistically significant and large. For example, in Panel C the coefficients on "Moved Out of the South" range from 0.107 to 0.222 log points, and the coefficients on "Moved out of South x Black" range from 0.212 to 0.584 log points. Thus, while gains to moving out of the South for White sons were substantial, the gains for Black sons were much larger.

5.2 Robustness

Appendix Table A.6 demonstrates the robustness of our standard errors to other forms of spatial autocorrelation. In particular, we estimate spatial heteroskedasticity and autocorrelation-consistent (HAC) standard errors using the method proposed by Conley (1999). Estimating these standard errors is very data intensive; accordingly, in

[25] Moving to northern cities brought both higher wages and higher mortality for Black men (Black et al., 2015).

Panel C we use father's initial state of residence fixed effects interacted with birth year fixed effects rather than father's initial county of residence fixed effects interacted with birth year fixed effects. Columns 1 and 4 allow for spatial correlation within a 100 kilometer bandwidth; columns 2 and 5 use a 200 kilometer bandwidth and columns 3 and 6 use a 400 kilometer bandwidth. In all instances the standard errors are similar to, or smaller than, our baseline standard errors which are clustered based on 96 bins of longitude.

Appendix Table A.7 shows results using percentile in the national income distribution, log of annual income (as opposed to weekly wages), and weeks worked (in 1939) as the dependent variables. In panels A and C, Black sons born after the arrival of the boll weevil had income ranks that were about 2-3 percentiles higher. Black sons also had statistically significantly higher annual incomes. This implies that our weekly wage results are driven by increases in annual income, not decreases in weeks worked. In the difference-in-differences specifications in Panels A and B, being born after the arrival of the boll weevil did not affect Black sons' weeks worked, but did reduce White sons' weeks worked. As a result, in the triple difference specification in Panel C, being born after the arrival of the boll weevil increased Black sons' weeks worked.

Appendix Table A.8 demonstrates that the wage results are robust to alternative restrictions on the wage worker sample. The results across different wage restrictions and samples are very similar to the wage results in Table 1. Column 7 examines the results for a sample that only includes brothers and controls for father fixed effects. We define individuals as brothers if they reported having the same father in the censuses. The point estimates in column 7, Panel C, while not being statistically significant, are similar in magnitude to our baseline result in column 1.

Appendix Table A.9 shows that the results are robust to having the first year of treatment be event year 0 or to dropping various birth cohorts that might have been partially impacted by the boll weevil. Columns 1 and 6 repeat our baseline specification from Table 1. Columns 2 and 7 change our post boll weevil indicator to turn on in the year the boll weevil arrived, as opposed to the year after. Depending on the timing of the arrival of the boll weevil and the timing of a son's birth, the arrival of the boll weevil may have partially impacted the pregnancies of sons born in event year 0. It is possible that the pregnancies of some sons born in event year 1 were also impacted.[26] Columns 3 and 8 drop sons born the year the weevil arrived and columns 4 and 9 drop sons born the year the boll weevil arrived and the year after. When these years are dropped, the effects are larger than, but not statistically significantly different than, the baseline results.

Because the boll weevil advanced outward geographically and induced substantial migration, another concern is geographic spillovers to pre-event cohorts in counties not yet invaded. Spillovers could also occur if pre-event cohorts are hurt by or benefit from the effects of changing economic conditions after the weevil's arrival. If these changes in aggregate benefited (harmed) pre-boll weevil cohorts, they would bias down (up) estimated treatment effects. Columns 5 and 10 drop sons born one and two years prior to the weevil's arrival. These sons are the most likely to be impacted by spillovers resulting from the boll weevil. The results are similar to the baseline results.

Appendix Table A.10 documents that the results are not sensitive to alternative measures of the boll weevil's year of arrival. Ferrara, Ha and Walsh (2022) use a newspaper based measure of the boll weevil's arrival to correct for possible measurement error in the USDA maps used in this paper. One of their preferred methods of bias correction is to perform the estimation on the set of observations for which the post-boll weevil treatment variable is the same using the USDA maps and their newspaper measure. They refer to this as an agreement sample. Appendix Table A.10 compares the results using all three measures: USDA maps, newspapers, and agreement sample. The newspaper based measure of boll weevil's arrival, which likely has the largest measurement error, produces results with the smallest magnitude, followed by the USDA measure. The agreement sample, which likely has the least measurement error, produces results with the largest magnitude.

Appendix Tables A.11 and A.12 show that the results are robust to the number

[26] This would occur if the boll weevil arrived late in event year 0 and the son was born early in event year 1.

of linking sources in the Census Tree. We obtain similar weekly wage and imputed income results when a link is identified by one, two, three, or four sources in the Census Tree. Very few Black sons in our samples are identified by five or six sources in the Census Tree.

Appendix Table A.13 demonstrates that the results are robust to using alternative difference-in-differences estimators.[27] Due to the computationally intensive nature of using alternative difference-in-differences estimators, in this table we collapse our individual-level data to county-race-birth year cells and weight these by the number of observations used to generate the cell average (e.g. the average wage for Black sons born in 1903 whose father initially resided in Winston County, AL). We also do not control for hookworm, malaria, or Rosenwald exposure in these regressions because an assumption of the Callaway and Sant'Anna (2021) estimator used in this table is that control variables are stationary (i.e. not time varying). Columns 1 and 4 present our baseline two-way fixed effects difference-in-differences estimates with this collapsed data.[28] Columns 2 and 5 use the estimator from Callaway and Sant'Anna (2021) and columns 3 and 6 use the estimator from Sun and Abraham (2021). Both of these estimators eliminate bias in difference-in-differences settings with staggered treatment timing; the estimates are not statistically significantly different from the two-way fixed effects estimates in columns 1 and 4. The estimates in all columns of Appendix Table A.13 are similar to our baseline estimates in Table 1.

[27] This table focuses on two-way fixed effects difference-in-differences estimates. These estimates are subject to bias due to heterogeneous treatment effects over time. The estimators designed to deal with this bias have only been shown to work in difference-in-differences designs, not triple difference designs.

[28] These estimates should be compared to columns 1 and 4 of Table 1 because we do not use inverse propensity score re-weighting, but instead weight by the number of observations used to generate the cell average.

5.3 Intergenerational Mobility

Next, we explore the effect of the boll weevil on intergenerational mobility using the following specification:

$$\begin{aligned} \textit{Son income rank}_{ict} = \mu + \beta * \mathbb{I}[\textit{Born post boll weevil} = 1]_{ct} + \eta * \textit{Father income rank}_{ict} \\ + \theta_c + \alpha_t + \delta_b + \xi_e + X_{ct}\zeta' + \epsilon_{ict} \end{aligned} \quad (5)$$

This specification is similar to Specification (1), but now the outcome is sons' income rank and we control for fathers' income rank. Both sons' and fathers' income rank are based on their *imputed income* (Collins and Wanamaker, 2022). μ in Specification (5) is the intergenerational mobility intercept; a son's expected income rank if their father's income rank (and all other variables) are zero. η is the intergenerational mobility slope; for each one unit increase in father's rank, how much will be passed on to their son. In some specifications we let the intergenerational mobility slope change for sons born after the boll weevil's arrival by interacting father's income rank with the "Born post boll weevil" indicator.

It is worth noting that our analysis of intergenerational mobility differs from conventional analyses in a number ways. First, our sample is limited to fathers observed in counties in the South, prior to the arrival of the boll weevil, who had a son under the age of 10. Intergenerational mobility in the early twentieth century was much lower in the South than in other regions (Connor and Storper, 2020). Second, Black and White intergenerational mobility are examined separately by race.[29] Third, to align with the rest of the analysis in the paper, father's initial county fixed effects are included, as are controls for hookworm, malaria, and Rosenwald school exposure. Fourth, fathers are observed twice.[30] Fifth, some father-son linkages are between 1900 and 1910 and 1940 and some are between 1910 and 1920 and 1940. This reflects our

[29] Collins and Wanamaker (2022), Saavedra and Twinam (2020), and Jácome, Kuziemko and Naidu (2021) are the closest to our analysis in that they also analyze intergenerational mobility separately by race in some specifications. Ward (2023) includes Black and White fathers and sons but does not run Black-only specifications.

[30] Ward (2020) and Ward (2023) also observe fathers twice or even three times.

focus on the boll weevil.

Table 3 examines how Black fathers' income rank affected their sons' income rank. Columns 1 and 3 use fathers' initial income rank (in either 1900 or 1910) and columns 2 and 4 use fathers' final income rank (in either 1910 or 1920). Columns 1 and 2 include both our boll weevil treatment variable and fathers' income rank. Accordingly, they allow the intergenerational mobility intercept to change for sons born after the weevil's arrival. Columns 3 and 4 allow both the intercept and slope of intergenerational mobility to change after the weevil's arrival.

In all columns of Table 3, there is a statistically significant increase in the intergenerational mobility intercept for Black sons born after the arrival of the boll weevil. White sons born after the arrival of the boll weevil had an insignificant decrease in the intergenerational mobility intercept. Appendix Figure A.5 graphs the intergenerational mobility slope and intercept from column 2 for Black and White sons. The change in intercept for both Black and White sons is apparent. Black sons are only graphed for fathers' ranks from 0 to 30 because virtually no Black fathers had income ranks above 30.

Black sons born after the weevil's arrival experienced an increase in imputed income rank of about 1, which is substantial. This represents a 12% increase in the average income rank for Black sons. Jácome, Kuziemko and Naidu (2021) find that the intercept in the relationship between the rank of Black parental income to the rank of adult child income increased by 7 percentiles between the 1910s-1920s birth cohorts and the 1940s-1950s birth cohorts.[31] Along this dimension, it is substantial as well.

6 Mechanisms

We provide evidence on a variety of mechanisms through which early life conditions may have improved for Black sons relative to White sons including: fathers' migration and occupational change, improvements in some measures related to nutrition and

[31] It is worth noting that our birth cohorts are earlier, 1891-1920.

education, declines in the number of Black sons born after the boll weevil, and declines in racial violence after the boll weevil.

6.1 Migration and Changes in Household Income

Table 4 examines migration and changes in occupation between the first and second census (1900 to 1910 or 1910 to 1920) for Black fathers relative to White fathers. Panels A and B show that Black fathers were less likely to migrate along a range of dimensions than White fathers. Appendix Table A.2 shows that the baseline migration rates for Black fathers were substantial: 22% moved counties, 6% moved states, 2% moved out of the South, 16% moved from a rural area to another rural area, 5% moved from a rural area to an urban area, and 1% moved from a rural area to a city with 100,000 or more in population. The rates for White fathers were, however, larger. One exception is that Black fathers were more likely to migrate to large cities.

Table 4 Panels C and D show that while some Black fathers experienced improvements in their economic status relative to White fathers, on average Black fathers saw declines in their economic status following the boll weevil. Panel C shows that Black fathers were less likely to leave agriculture than White fathers. Black fathers were, however more likely to upgrade their agricultural status from farm laborer to tenant or owner. They were also more likely to move from being a laborer, whether farm laborer or another type of laborer, to having another occupation. Thus, some of the poorest Black fathers appear to have experienced improvements in their economic status. Panel D shows that on average, however, Black fathers saw declines in their imputed income score and their imputed income rank relative to White fathers.[32]

Appendix Table A.14 shows that one mechanism through which Black sons born after the boll weevil saw improved wages and income is through their father migrating or upgrading occupations. This table takes some of the changes in fathers' characteristics documented in Table 4 and examines outcomes for their sons.[33] For

[32]Note that if a father did not change their occupation-region-race cell between the first and second census they will have the same imputed income.

[33]Table 2 shows the results for sons whose father migrated out of the South.

the weekly wage sample, sons whose father moved from rural areas to other rural areas experienced increases that are greater than our baseline estimates. The same is true for sons whose father upgraded within agriculture from being a farm laborer to a tenant or farm owner and for sons whose father moved from being a laborer to another occupation. For the imputed income sample, sons whose father moved from a rural area to an urban area have imputed incomes that are greater than our baseline estimate. For both samples, we find substantial, although not statistically significant, effects for sons whose father did not move and did not change occupations.

6.2 Nutrition and Schooling

Given that we find substantial effects of being born after the boll weevil for sons whose father did not move and did not change occupations, a natural question is whether the positive effects of the boll weevil are the result of improved early childhood nutrition, increased educational opportunity, or some other factor. In this section, we present evidence that Black sons experienced improvements in some measures related to nutrition and education.

Table 5 shows that after the arrival of the boll weevil, cotton acreage, cotton's share of farm acreage, and cotton production fell while the acreage, share of acreage, and production of food-related crops rose.[34] Together with the historical evidence discussed earlier, this suggests that nutrition may have improved for sons born after the boll weevil. Further, Appendix Table A.15 shows evidence that pellagra mortality fell after the arrival of the boll weevil and fell more in counties with high Black-shares of the population.[35] Pellagra is a disease caused by insufficient niacin consumption that was widespread in the South during the early twentieth century. Pellagra mortality is

[34]Data are from the 1890, 1900, 1910, 1920, 1925, and 1930 Censuses of Agriculture (Haines, Fishback and Rhode, 2018). All crop variables in the Censuses of Agriculture are measured in the year prior to the census being taken (e.g. 1889 instead of 1890, 1899 instead of 1900, etc.) and we account for this when assigning the boll weevil's arrival date. PPML is used in Panels A and C and OLS is used in Panel B.

[35]High Black-share counties are counties with 41.7% or more of their population that are Black in 1900, which corresponds to the median percent Black in the sample of counties that we have pellagra data for. We have pellagra data both before and after the boll weevil's arrival for counties in North and South Carolina.

an indicator of very poor nutrition.

If sons experienced better nutrition, this may have translated into being taller and having better cognitive abilities. To explore this, we use data on heights and AGCT scores (a measure of cognitive ability) of Black and White male WWII enlistees born around the time of the arrival of the boll weevil from *U.S. World War II Army Enlistment Records, 1938-1946*.[36] The records include information on the state and county of residence at the time of enlistment as well as an enlistee's height (in inches) and in some cases AGCT score. The sample is restricted to men who lived in the same state they were born in, were born between 1915 and 1924, and were born within 5 years ($\pm$) of the boll weevil's arrival in their county of residence.[37] For the heights analysis we further restrict to enlistees with a valid height and weight.[38] Because the records do not include county of birth, we assume that men were living in their county of birth when they enlisted. This likely introduces measurement error, which would bias our coefficient estimates toward zero.

Panel A of Table 6 shows that Black recruits born after the boll weevil were statistically significantly taller than Black recruits born before the boll weevil. While an increase of 0.079 inches may seem small, Black sons born after the boll weevil were able to close about 12% of the height gap with White sons.[39] Being taller may have led to improved outcomes in the labor market, since most Black men were employed in manual labor. Black sons born after the boll weevil also had higher AGCT scores, but the effects were not statistically significant.

Table 7 shows that the boll weevil did have a statistically significant effect on average years of schooling, but only for Black sons who were wage workers. In Panels A and C of column 1, Black sons experienced an increase in schooling of about

[36]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 in this context. See Bodenhorn, Guinnane and Mroz (2017, 2019) and Komlos and A'Hearn (2019).

[37]Men born after 1924 could still have been growing when they enlisted.

[38]To serve in WWII an individual had to be between 5 and 6.5 feet tall and weigh over 105 pounds. We also exclude any enlistee that reported a weight over 400 pounds.

[39]0.079/0.649 where 0.649 is the difference in average height between Black and White enlistees in Table 6.

one-third of a year. Interestingly, exposure to Rosenwald schools did not increase years of schooling for Black sons in our sample. The Rosenwald exposure rate for most of our birth cohorts is generally low or zero since the Rosenwald school initiative did begin in earnest until the 1920s.[40] Columns 3 and 4 repeat our baseline weekly wage and imputed income analysis, but control for fixed effects for the number of years of schooling. The results are similar to our main results.

Our findings regarding years of schooling relate most closely to Baker, Blanchette and Eriksson (2020). Baker, Blanchette and Eriksson (2020) find that children who were young (4 to 9) when the boll weevil arrived experienced increases in schooling relative to young adults (19 to 30 year olds). Our paper finds an effect of the boll weevil on schooling in our weekly wage sample that is similar in magnitude to Baker, Blanchette and Eriksson (2020), but no effect in our broader imputed income sample. We speculate that the differences in findings across the two papers are driven by the fact that their empirical design has a much broader range of treatments. Baker, Blanchette and Eriksson (2020) have a large number of untreated individuals (e.g. 19-30 year olds) and individuals who were more intensively (e.g. 4-9 year olds) and less intensively treated (e.g. 15-18 year olds). In our sample, everyone was treated by their definition. That is, they were 9 or younger when the boll weevil arrived and reduced the demand for child labor.

Even given our mixed findings on years of schooling for Black sons born after the boll weevil, other dimensions of education may have improved. In particular, students may have been able to attend school more regularly as demand for child labor fell following the arrival of the boll weevil (Baker, 2015; Baker, Blanchette and Eriksson, 2020).

[40] Mohammed and Mohnen (2025) find increases of about 0.9 years of schooling for rural Black men in the 1910-1914 birth cohorts and about 0.3 years for the 1915-1919 birth cohorts. Our analysis spans the 1891-1920 birth cohorts and so includes a lot of years with very low or no treatment. Similar to how Mohammed and Mohnen (2025) split their sample into 1910-1914 and 1915-1919 birth cohorts, we split our sample into 1907-1913 and 1914-1920 birth cohorts and find significant effects of Rosenwald exposure on years of schooling for the 1907-1913 birth cohorts.

6.3 Fertility after the Boll Weevil and Lynchings

Table 8 documents that Black fathers had statistically significantly fewer male children born after the boll weevil than White fathers. The magnitude of the difference represents a 16% decline in Black fertility relative to the mean. These findings are consistent with Ager, Herz and Brueckner (2020)'s findings on fertility.[41] Black female labor force participation also decreased after the arrival of the boll weevil, which may have reduced stress on pregnant and nursing mothers and allowed mothers to spend more time with their children (Ager, Brueckner and Herz, 2017). At the margin, these changes may have allowed Black families to invest more in their children and so conferred benefits on Black sons born after the boll weevil relative to White sons born after the boll weevil.

Table 8 also shows that there may have been less racial violence, measured through lynchings, after the arrival of the boll weevil. This result is in-line with Feigenbaum, Mazumder and Smith (2020) who find that lynchings and the construction of Confederate monuments decreased in the South after the boll weevil's arrival. When we restrict to counties hit by the boll weevil between 1901 and 1920, the years that align with our analysis, the effect is negative and statistically insignificant. The magnitude of the effect is, however, sizable relative to the mean. The reduction in violence could have reduced a range of stressors on Black families and their children that translated into better long run outcomes. For example, reductions in violence may have allowed Black men, women, and children to travel longer distances safely to work or school or to buy food and change employers more easily.

In summary, the empirical evidence in this section and the historical evidence considered in Section 2 suggest five mechanisms through which Black sons born after the boll weevil might have experienced improvements in wages and imputed income: migration and occupational upgrading; changes in early childhood nutrition; changes in schooling; changes in resources available to children including conditions during

[41] Using repeated cross-sectional data and focusing on children under 5, Ager, Herz and Brueckner (2020) find that fertility of Black mothers was lower than White mothers after the arrival of the boll weevil, but the difference was small and not statistically significant.

pregnancy, the number of children in the household, and time spent on children; and changes in racial violence.

7 Conclusion

Drawing on a large newly linked data set of Black and White fathers and sons, we find a large negative agricultural shock, the boll weevil, benefited Black sons in the long run as measured by wages and imputed income in 1940 and did not harm White sons. Black sons born after the shock had wages that were 11 percent higher and imputed incomes that were 5 percent higher than White sons born after the shock. These gains were not driven by migration out of the South.

The boll weevil decreased racial inequality as measured by wages, imputed income, and income rank. The increases in Black sons' wages and imputed wages accounted for 6-15% of the Black-White wage gap in 1940. Increases in the intergenerational mobility intercept were 12% of the average income rank for Black sons.

These surprising improvements appear to have occurred through a range of mechanisms related to the boll weevil. The available evidence suggests that Black sons may have benefited from their fathers' migration and occupational upgrading; from improvements in nutrition and schooling; from increases in household resources available to sons, including reductions in the number of children in the household; and from reductions in racial violence.

The paper sheds new light on the improvement in the economic status of Black men during the first half of the twentieth century. The literature has primarily focused on education and migration to the North. This paper adds to the set of factors that led to improvement by showing that a large agricultural shock, the boll weevil, generated long run benefits for Black sons born after its arrival.

References

Aaronson, Daniel, and Bhashkar Mazumder. 2011. "The impact of Rosenwald schools on black achievement." *Journal of Political Economy*, 119(5): 821–888.

Abramitzky, Ran, Leah Boustan, Katherian Eriksson, Santiago Pérez, and Myera Rashid. 2020. "Census Linking Project: Version 2.0." https://censuslinkingproject.org.

Acemoglu, Daron, and Alexander Wolitzky. 2011. "The economics of labor coercion." *Econometrica*, 79(2): 555–600.

Ager, Philipp, Benedikt Herz, and Markus Brueckner. 2020. "Structural change and the fertility transition." *Review of Economics and Statistics*, 102(4): 806–822.

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.

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 Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

Alston, Lee J, and Joseph P Ferrie. 2005. "Time on the ladder: Career mobility in agriculture, 1890–1938." *The Journal of Economic History*, 65(4): 1058–1081.

Alston, Lee J, and Kyle D Kauffman. 1997. "Agricultural chutes and ladders: New estimates of sharecroppers and "true tenants" in the South, 1900–1920." *The Journal of Economic History*, 57(2): 464–475.

Bailey, Martha, Connor Cole, and Catherine Massey. 2020. "Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2): 80–93.

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.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent paths: A new perspective on earnings differences between black and white men since 1940." *The Quarterly Journal of Economics*, 133(3): 1459–1501.

Black, Dan A, Seth G Sanders, Evan J Taylor, and Lowell J Taylor. 2015. "The impact of the Great Migration on mortality of African Americans: Evidence from the Deep South." *American Economic Review*, 105(2): 477–503.

Bleakley, Hoyt. 2007. "Disease and development: evidence from hookworm eradication in the American South." *The quarterly journal of economics*, 122(1): 73–117.

Bleakley, Hoyt. 2010. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics*, 2(2): 1–45.

Bloome, Deirdre, James Feigenbaum, and Christopher Muller. 2017. "Tenancy, Marriage, and the Boll Weevil Infestation, 1892–1930." *Demography*, 54(3): 1029–1049.

Bobonis, Gustavo J, and Peter M Morrow. 2014. "Labor coercion and the accumulation of human capital." *Journal of Development Economics*, 108: 32–53.

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.

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.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley EB Wilbert. 2023. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." National Bureau of Economic Research.

Buggle, Johannes C, and Steven Nafziger. 2021. "The slow road from serfdom: labor coercion and long-run development in the former Russian Empire." *Review of Economics and Statistics*, 103(1): 1–17.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Carruthers, Celeste K, and Marianne H Wanamaker. 2017. "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.

Cohen, William. 1976. "Negro involuntary servitude in the South, 1865-1940: A preliminary analysis." *The Journal of Southern History*, 42(1): 31–60.

Collins, William J. 2021. "The Great Migration of Black Americans from the US South: A guide and interpretation." *Explorations in Economic History*, 80: 101382.

Collins, William J, and Marianne H Wanamaker. 2022. "African American intergenerational economic mobility since 1880." *American Economic Journal: Applied Economics*, 14(3): 84–117.

Collins, William J, and Robert A Margo. 2006. "Historical perspectives on racial differences in schooling in the United States." *Handbook of the Economics of Education*, 1: 107–154.

Collins, William J, Nicholas Holtkamp, and Marianne H Wanamaker. 2024. "Black Americans' Landholdings and Economic Mobility after Emancipation: Evidence from the Census of Agriculture and Linked Records." *The Journal of Economic History*, 1–34.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of econometrics*, 92(1): 1–45.

Connor, Dylan Shane, and Michael Storper. 2020. "The changing geography of social mobility in the United States." *Proceedings of the National Academy of Sciences*, 117(48): 30309–30317.

Derenoncourt, Ellora. 2022. "Can you move to opportunity? Evidence from the Great Migration." *American Economic Review*, 112(2): 369–408.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2024. "Wealth of two nations: The US racial wealth gap, 1860–2020." *The Quarterly Journal of Economics*, 139(2): 693–750.

Dirks, Robert T, and Nancy Duran. 2001. "African American dietary patterns at the beginning of the 20th century." *The Journal of nutrition*, 131(7): 1881–1889.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo. 2018. "The economic consequences of hospital admissions." *American Economic Review*, 108(2): 308–352.

Engerman, Stanley L. 1992. "Coerced and free labor: property rights and the development of the labor force." *Explorations in economic history*, 29(1): 1–29.

Feigenbaum, James, Soumyajit Mazumder, and Cory Smith. 2020. "When Coercive Economies Fail: The Political Economy of the US South After the Boll Weevil." *Working Paper*.

Feigenberg, Benjamin, Ben Ost, and Javaeria A Qureshi. 2023. "Omitted variable bias in interacted models: A cautionary tale." *Review of Economics and Statistics*, 1–47.

Ferrara, Andreas, Joung Yeob Ha, and Randall Walsh. 2022. "Using Digitized Newspapers to Refine Historical Measures: The Case of the Boll Weevil." National Bureau of Economic Research.

Ferrell, John A. 1931. "Challenge of Malaria in the South." *American Journal of Public Health and the Nations Health*, 21(4): 355–374.

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.

Haines, Michael, Price Fishback, and Paul Rhode. 2018. "United States Agriculture Data, 1840 - 2012." Inter-university Consortium for Political and Social Research.

Hawthorne, Harvey W, Frank Montgomery, and Harrison Morton Dixon. 1922. *Farm Management and Farm Organization in Sumter County, Ga: An Analysis of the Business of 534 Farms in 1913 and 550 Farms in 1918.* US Department of Agriculture.

Heckman, James J, Thomas M Lyons, and Petra E Todd. 2000. "Understanding black–white wage differentials: 1960–1990." *American Economic Review*, 90(2): 344–349.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. 2023. "IPUMS Multigenerational Longitudinal Panel: Version 1.1." https://doi.org/10.18128/D016.V1.1, IPUMS.

Hong, Sok Chul. 2011. "Malaria and economic productivity: a longitudinal analysis of the American case." *The Journal of Economic History*, 71(3): 654–671.

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.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2021. "Mobility for all: Representative intergenerational mobility estimates over the 20th century." National Bureau of Economic Research.

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.

Leavell, R. H., J. H. Dillard, and United States Department of Labor Division of Negro Economics. 1919. *The Negro migration in 1916-17.* U.S. Department of Labor, Office of the Secretary, Division of Negro Economics.

Margo, Robert A. 1995. "Explaining Black-white Wage Convergence, 1940–1950." *ILR Review*, 48(3): 470–481.

Margo, Robert A. 2016. "Obama, Katrina, and the persistence of racial inequality." *The Journal of Economic History*, 76(2): 301–341.

Moehling, Carolyn M. 2004. "Family structure, school attendance, and child labor in the American South in 1900 and 1910." *Explorations in Economic History*, 41(1): 73–100.

Mohammed, Shariq, and Paul Mohnen. 2025. "Black economic progress in the Jim Crow South: Evidence from Rosenwald schools." *Journal of Economic History*, Forthcoming.

Mulmi, Prajula, Steven A Block, Gerald E Shively, and William A Masters. 2016. "Climatic conditions and child height: Sex-specific vulnerability and the protective effects of sanitation and food markets in Nepal." *Economics & Human Biology*, 23: 63–75.

Naidu, Suresh. 2010. "Recruitment restrictions and labor markets: Evidence from the postbellum US South." *Journal of Labor Economics*, 28(2): 413–445.

Naidu, Suresh, and Noam Yuchtman. 2013. "Coercive contract enforcement: law and the labor market in nineteenth century industrial Britain." *American Economic Review*, 103(1): 107–144.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*a*. "The Census Tree, 1900-1910 [dataset]." Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*b*. "The Census Tree, 1900-1940 [dataset]." Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*c*. "The Census Tree, 1910-1920 [dataset]." Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*d*. "The Census Tree, 1910-1940 [dataset]." Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*e*. "The Census Tree, 1920-1940 [dataset]." Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. 2021. "Combining family history and machine learning to link historical records: The Census Tree data set." *Explorations in Economic History*, 80: 101391.

Ransom, Roger L, and Richard Sutch. 2001. *One Kind of Freedom: The Economic Consequences of Emancipation.* Cambridge University Press.

Roback, Jennifer. 1984. "Southern labor law in the Jim Crow era: exploitative or competitive?" *The University of Chicago Law Review*, 51(4): 1161–1192.

Rockefeller Sanitary Commission. 1910-1914. *Annual Report of Rockefeller Sanitary Commission for the Eradication of Hookworm Disease.* Rockefeller Foundation.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]." IPUMS.

Saavedra, Martin, and Tate Twinam. 2020. "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History*, 75: 101304.

Shah, Manisha, and Bryce Millett Steinberg. 2017. "Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital." *Journal of Political Economy*, 125(2): 527–561.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Thoman, Eric B. 2009. "Historic Hookworm Prevalence Rates and Distribution in the Southeastern United States: Selected Findings of the Rockefeller Sanitary Commission for the Eradication of Hookworm." *Rockefeller Arch Cent Res Rep Online*, 1–21.

U.S. Department of Agriculture, Bureau of Agricultural Economics. 1951. *Statistics on Cotton and Related Data. Statistical Bulletin No. 99.* Government Printing Office.

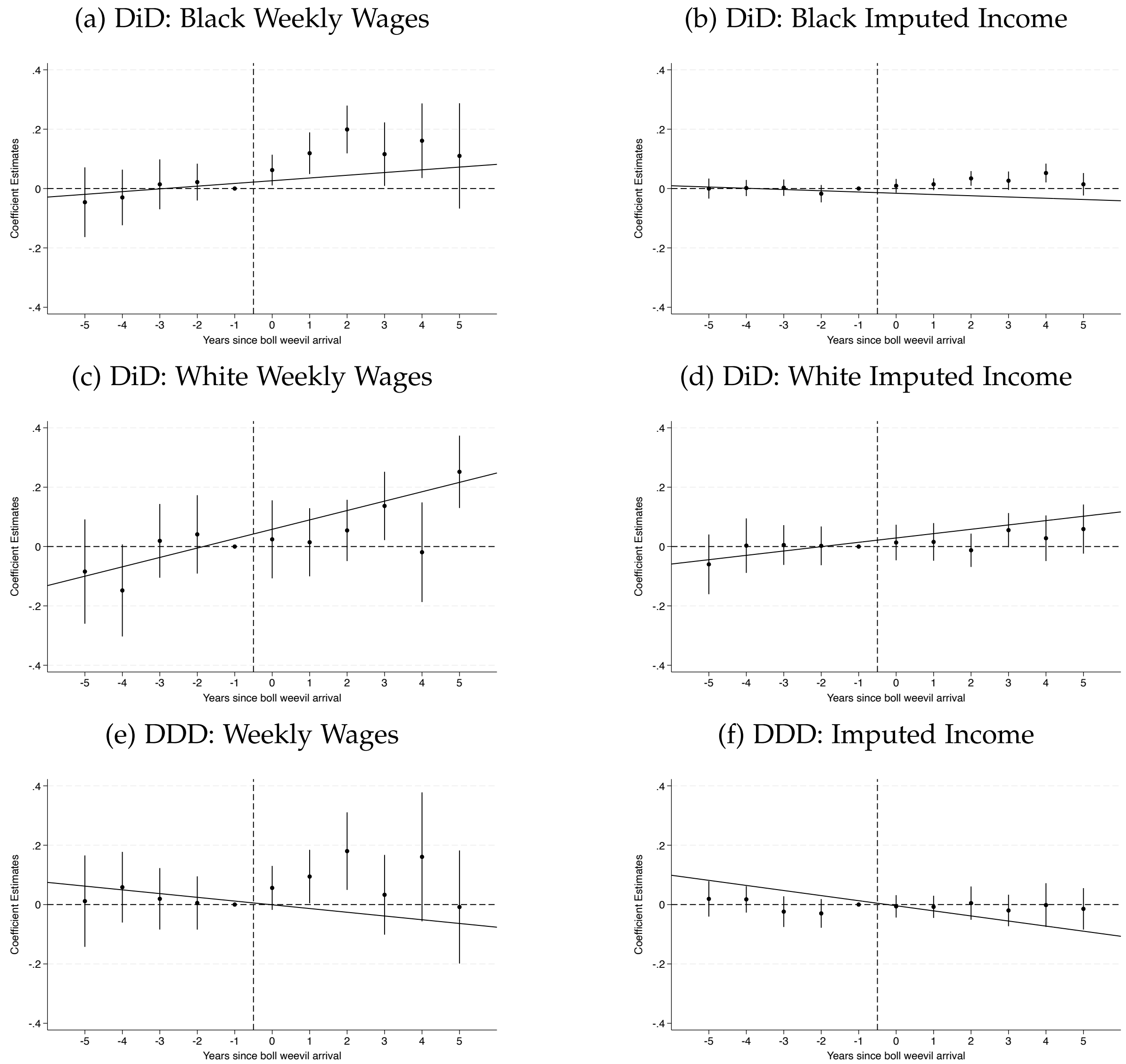
Ward, Zachary. 2020. "The not-so-hot melting pot: The persistence of outcomes for descendants of the age of mass migration." *American Economic Journal: Applied Economics*, 12(4): 73–102.

Ward, Zachary. 2023. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." *American Economic Review*, 113(12): 3213–48.

Williams, Jhacova. 2022. "Historical lynchings and the contemporary voting behavior of Blacks." *American Economic Journal: Applied Economics*, 14(3): 224–253.

Figures and Tables

Figure 1: Event studies - Weekly wages and imputed income



Notes: Panels (a), (b), (c), and (d) show estimates of Specification (2) in the text. Panels (e) and (f) show estimates of Specification (4). All panels graph the pre-trend from a parametric event study described in the text. The unit of observation is sons. The estimation is run on sons born -10 to +9 years relative to the boll weevil's arrival in their father's original county of residence. Only estimates for sons born -5 to +5 years relative to the boll weevil's arrival are shown. In all panels the event time indicator for the year -1 is omitted. All regressions use inverse propensity score re-weighting. Standard errors are clustered based on 96 bins of longitude. 90% confidence intervals are shown.

Table 1: The boll weevil, weekly wages, and imputed income

	Log(weekly wage)			Log(imputed income)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: DiD for Black sons						
Born post BW	0.070**	0.097***	0.096***	0.016*	0.021	0.034***
	(0.028)	(0.033)	(0.030)	(0.008)	(0.012)	(0.010)
Observations	11155	11155	11155	27077	27077	27077
Mean of dep. var.	2.095	2.095	2.095	5.936	5.936	5.936
Panel B: DiD for White sons						
Born post BW	-0.012*	-0.042	-0.043	0.001	-0.025	-0.021
	(0.007)	(0.049)	(0.053)	(0.003)	(0.024)	(0.024)
Observations	171064	171064	171064	352319	352319	352319
Mean of dep. var.	2.964	2.964	2.964	6.817	6.817	6.817
Panel C: Triple difference; county-by-year FE						
Born post BW * Black	0.078**	0.118***	0.114***	0.014	0.037**	0.045***
	(0.030)	(0.041)	(0.041)	(0.009)	(0.018)	(0.017)
Observations	181709	181709	181709	379237	379237	379237
Mean of dep. var.	2.911	2.911	2.911	6.754	6.754	6.754
Panel D: Triple difference; state-by-year FE						
Born post BW * Black	0.085***	0.118**	0.119**	0.012	0.055**	0.063**
	(0.028)	(0.052)	(0.052)	(0.009)	(0.026)	(0.024)
Observations	181709	181709	181709	379237	379237	379237
Mean of dep. var.	2.911	2.911	2.911	6.754	6.754	6.754
Two-way/DDD FE	X	X	X	X	X	X
Propensity score weighting		X	X		X	X
Hookworm, malaria, and Rosenwald controls			X			X

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panels C and D provide estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), and a full set of indicators for sons' birth order. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, and fathers' initial county-by-birth year fixed effects. Panel D uses the same controls as Panel C, but uses fathers' initial state-by-birth year fixed effects rather than fathers' initial county-by-birth year fixed effects. Columns 3 and 6 of Panels A and B include additional controls: county-level hookworm infection rates interacted with birth year fixed effects, county-level malaria ecology interacted with birth year fixed effects, and Rosenwald exposure. Columns 3 and 6 of Panels C and D include the following additional controls: county-level hookworm infection rates interacted with birth year-by-race fixed effects, county-level malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald exposure interacted with race. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 2: The boll weevil, weekly wages, and imputed income by migration status

	Log(weekly wage)		Log(imputed income)		Pr(Son not in South = 1)	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: DiD for Black sons			
Born post BW	0.075**	0.080**	0.027***	0.021**	0.048***	0.020*
	(0.031)	(0.035)	(0.010)	(0.010)	(0.016)	(0.010)
Observations	10936	9775	26646	24678	11155	27086
Mean of dep. var.	2.08	1.986	5.925	5.87	.123	.089
			Panel B: DiD for White sons			
Born post BW	-0.013	-0.061*	-0.000	-0.021	-0.018	-0.016
	(0.013)	(0.031)	(0.007)	(0.014)	(0.029)	(0.023)
Observations	165229	150048	342250	318798	171064	352341
Mean of dep. var.	2.955	2.923	6.81	6.791	.123	.095
			Panel C: Triple difference			
Born post BW * Black	0.084**	0.098**	0.024*	0.031**	0.046*	0.026
	(0.036)	(0.040)	(0.013)	(0.015)	(0.024)	(0.018)
Observations	175644	159232	368731	343279	181709	379237
Mean of dep. var.	2.901	2.865	6.746	6.725	.123	.095
Sample	Father remained in South	Son remained in South	Father remained in South	Son remained in South	Wage sample	Imputed income sample

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Fathers are considered to have remained in the South if we observe them in the South census region in both censuses (all fathers are initially living in the South census region). Sons are considered to have remained in the South if we observe them in the South census region in the 1940 Census. Standard errors are clustered based on 96 bins of longitude.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 3: Sons' imputed income rank on fathers' imputed income rank

	Panel A: Rank in imputed income distribution for Black sons			
	(1)	(2)	(3)	(4)
Born post BW	1.037***	1.006***	1.364***	1.166**
	(0.351)	(0.332)	(0.426)	(0.444)
Father income rank (initial)	0.292***		0.301***	
	(0.032)		(0.036)	
Father income rank (final)		0.347***		0.351***
		(0.027)		(0.029)
Father income rank (initial) * Born post BW			-0.048	
			(0.043)	
Father income rank (final) * Born post BW				-0.017
				(0.040)
Observations	24685	24685	24685	24685
Mean of dep. var.	8.280	8.280	8.280	8.280
	Panel B: Rank in imputed income distribution for White sons			
	(1)	(2)	(3)	(4)
Born post BW	-1.202	-1.396	-0.115	-0.864
	(1.545)	(1.407)	(1.954)	(2.215)
Father income rank (initial)	0.174***		0.184***	
	(0.014)		(0.021)	
Father income rank (final)		0.271***		0.274***
		(0.015)		(0.018)
Father income rank (initial) * Born post BW			-0.030	
			(0.035)	
Father income rank (final) * Born post BW				-0.009
				(0.037)
Observations	310704	310704	310704	310704
Mean of dep. var.	44.622	44.622	44.622	44.622

Notes: The unit of observation is sons. Both panels display estimates for Specification (5) in the text. All regressions control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 4: The boll weevil and fathers' outcomes

	Panel A: Father migration		
	Move county	Move state	Move out of South
	(1)	(2)	(3)
Black	-0.099***	-0.036***	-0.000
	(0.007)	(0.004)	(0.002)
Observations	316777	316777	316777
Mean of dep. var.	.356	.111	.028

	Panel B: Father urban migration		
	Move rural to rural	Move rural to urban	Move rural to ≥100,000 urban
	(1)	(2)	(3)
Black	-0.067***	-0.023***	0.003***
	(0.006)	(0.003)	(0.001)
Observations	316777	316777	316777
Mean of dep. var.	.268	.072	.009

	Panel C: Father labor market outcomes		
	Leave ag.	Upgrade ag.	Leave labor
	(1)	(2)	(3)
Black	-0.032***	0.045***	0.045***
	(0.003)	(0.004)	(0.004)
Observations	278924	278924	278924
Mean of dep. var.	.105	.136	.208

	Panel D: Imputed income	
	Δ imputed income score	Δ imputed income rank
	(1)	(2)
Black	-204.974***	-10.768***
	(5.114)	(0.342)
Observations	278924	278924
Mean of dep. var.	207.971	10.4

Notes: The unit of observation is fathers. In Panel A the dependent variable is whether a father migrated out of their initial county (column 1), state (column 2), or region (column 3; all fathers start in the South) of residence after the arrival of the boll weevil. In Panel B the dependent variable is whether a father migrated from a rural area to another rural area (column 1), to an urban area (column 2), or to a large urban area (column 3; population 100,000 or more). In Panel C the dependent variables takes a one if a father moved from an agricultural occupation to a non-agricultural occupation (column 1), upgraded in agriculture from being a farm laborer to a farmer (either tenant/sharecropper or owner-operator; column 2), or moved from being a laborer (farm or non-farm) to another occupation (column 3). In Panel D the dependent variable is the change in a father's imputed income score (column 1) or the change in a father's rank in the imputed income distribution (column 2). All columns control for fathers' initial county of residence fixed effects, initial census enumeration year fixed effects, and a full set of indicators for father's age at initial census. No weights are used. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 5: The boll weevil and agricultural outcomes

	Cotton	Corn	Wheat	Sweet potato	Peanut
	(1)	(2)	(3)	(4)	(5)
	Panel A: Acres				
Post BW	-0.205***	0.293**	0.302**	0.160***	0.077
	(0.039)	(0.149)	(0.117)	(0.046)	(0.213)
Observations	4114	3794	3725	4138	4004
Mean of dep. var.	37766	31319	3406	567	915
	Panel B: Share of farm acres				
Post BW	-0.042***	0.042***	0.003**	0.000***	0.002**
	(0.005)	(0.014)	(0.001)	(0.000)	(0.001)
Observations	4132	3794	4061	4138	4074
Mean of dep. var.	.144	.127	.008	.003	.004
	Panel C: Output (Bales or bushels)				
Post BW	-0.484***	-0.037	0.171	0.086	0.298
	(0.053)	(0.063)	(0.155)	(0.054)	(0.211)
Observations	4114	3794	3725	4138	4004
Mean of dep. var.	13130	410556	42308	47264	16306

Notes: This table presents estimates for the following Specification:

$$Outcome_{ct} = \beta[Post\ boll\ weevil_{ct} = 1] + \theta_c + \gamma_t + \epsilon_{ct}$$

where c indexes counties and t indexes years. θ_c are county fixed effects and γ_t are year fixed effects. Estimates in Panels A and C use PPML due to some dependent variables (such as peanuts) having a large number of zeroes; estimates in Panel B use OLS. Counties invaded by the boll weevil between 1901 and 1920 are included (to line-up with the individual-level analysis). The data come from the 1890, 1900, 1910, 1920, 1925, and 1930 Censuses of Agriculture (Haines, Fishback and Rhode, 2018). All crop variables in the Censuses of Agriculture are measured in the year prior to the census being taken (e.g. 1889 instead of 1890, 1899 instead of 1900, etc.) and we account for this when assigning the boll weevil's arrival date. Standard errors are clustered based on 96 bins of longitude.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 6: The boll weevil, height, and AGCT scores for WWII enlistees

	Height (inches)	AGCT score
	(1)	(2)
	Panel A: DiD for Black enlistees	
Born post BW	0.079**	0.604
	(0.038)	(1.090)
Observations	53815	5982
Mean of dep. var	68.234	63.468
	Panel B: DiD for White enlistees	
Born post BW	0.001	-0.598
	(0.024)	(0.886)
Observations	176520	16373
Mean of dep. var	68.883	89.512
	Panel C: Triple difference	
Born post BW * Black	0.038	0.614
	(0.049)	(1.571)
Observations	230335	21599
Mean of dep. var	68.731	82.499

Notes: The unit of observation is a World War II enlistee. The data used in this table comes from *U.S. World War II Army Enlistment Records, 1938-1946* from the National Archives and Records Administration. Panels A and B control for: county of residence at time of enlistment fixed effects, birth year fixed effects, year of enlistment fixed effects, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: county of residence at time of enlistment-by-race fixed effects, birth year-by-race fixed effects, year of enlistment-by-race fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. No weights are used. Standard errors are clustered based on 96 bins of longitude.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 7: The boll weevil and years of schooling

	Years of Schooling		Log (weekly wage)	Log(imputed income)
	(1)	(2)	(3)	(4)
Panel A: DiD for Black sons				
Born post BW	0.330**	0.057	0.078***	0.031***
	(0.156)	(0.097)	(0.028)	(0.009)
Rosenwald Exposure	0.052	-0.106	-0.017	-0.018
	(0.184)	(0.147)	(0.039)	(0.016)
Observations	11155	27077	11155	27077
Mean of dep. var.	6.645	6.293	2.095	5.936
Panel B: DiD for White sons				
Born post BW	0.029	0.217	-0.035	-0.031
	(0.208)	(0.139)	(0.050)	(0.021)
Rosenwald Exposure	-0.415***	-0.479***	-0.032	-0.098***
	(0.112)	(0.086)	(0.021)	(0.021)
Observations	171064	352319	171064	352319
Mean of dep. var.	10.065	9.474	2.964	6.817
Panel C: Triple difference				
Born post BW * Black	0.335*	0.048	0.087**	0.040***
	(0.195)	(0.152)	(0.038)	(0.014)
Rosenwald Exposure * Black	0.351	0.315	-0.011	0.045***
	(0.231)	(0.194)	(0.037)	(0.016)
Observations	181709	379237	181709	379237
Mean of dep. var.	9.857	9.247	2.911	6.754
Sample:	Wage workers	Imputed income	Wage workers	Imputed income

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure (displayed above). Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure fully interacted with race and an indicator if the father was in a rural area (the coefficient on the triple interaction is displayed above). Columns 3 and 4 control for a complete set of indicators for years of schooling (0 through 17; anyone with more than 17 years of schooling is assigned 17). All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table 8: Number of lynchings and male children

	Number of male children born after boll weevil	Pr(lynching=1)	
	(1)	(2)	(3)
Black	-0.065***		
	(0.012)		
Post BW		-0.027***	-0.015
		(0.010)	(0.015)
Observations	316777	34692	25921
Mean of dep. var.	.409	.07	.084
Sample:	Fathers	All counties	Counties invaded by BW between 1901-1920

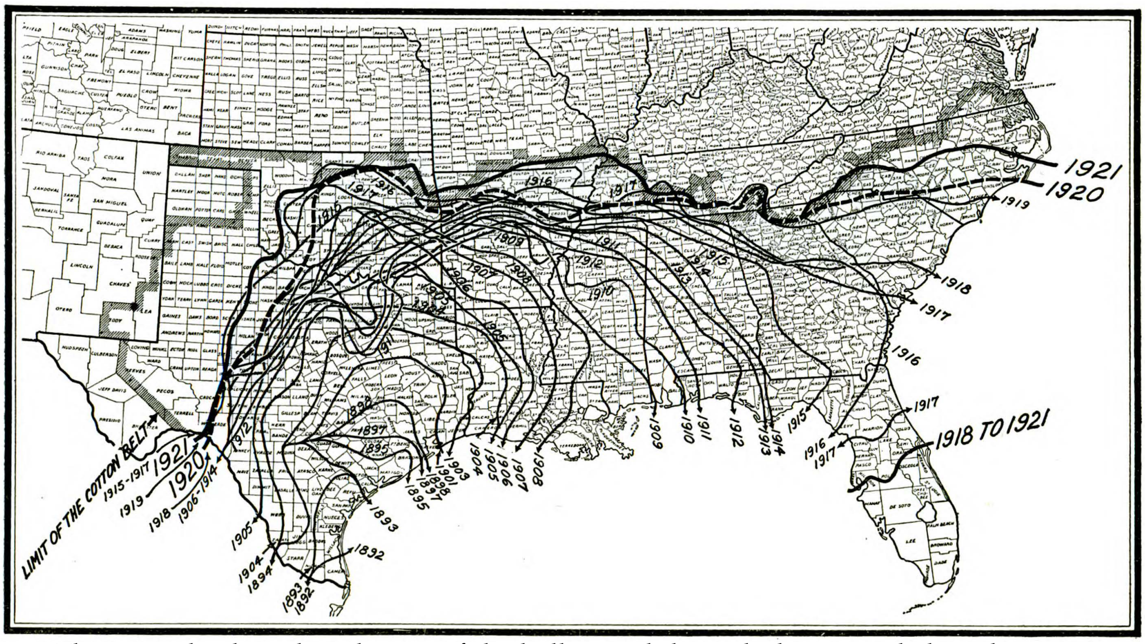
Notes: In column 1 the unit of observation is fathers. The dependent variable is the number of male children born to a father after the boll weevil. Controls include: fathers' initial county of residence fixed effects, initial census enumeration year fixed effects, and a full set of indicators for father's age at initial census. No weights are used. Column 1 controls for initial county of residence fixed effects, father's age at initial census, and initial census enumeration year fixed effects. In columns 2 and 3 the unit of observation is a county-year. Columns 2 and 3 present estimates for the following Specification:

$$Outcome_{ct} = \beta[Post\ boll\ weevil_{ct} = 1] + \theta_c + \gamma_t + \epsilon_{ct}$$

where c indexes counties and t indexes years. θ_c are county fixed effects and γ_t are year fixed effects. The dependent variable is the number of lynchings obtained from Williams (2022). These data are available for Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee. Columns 2 and 3 control for county and year fixed effects. Standard errors are clustered based on 96 bins of longitude in all regressions.
$*$ = $p < 0.10$; $**$ = $p < 0.05$; $***$ = $p < 0.01$

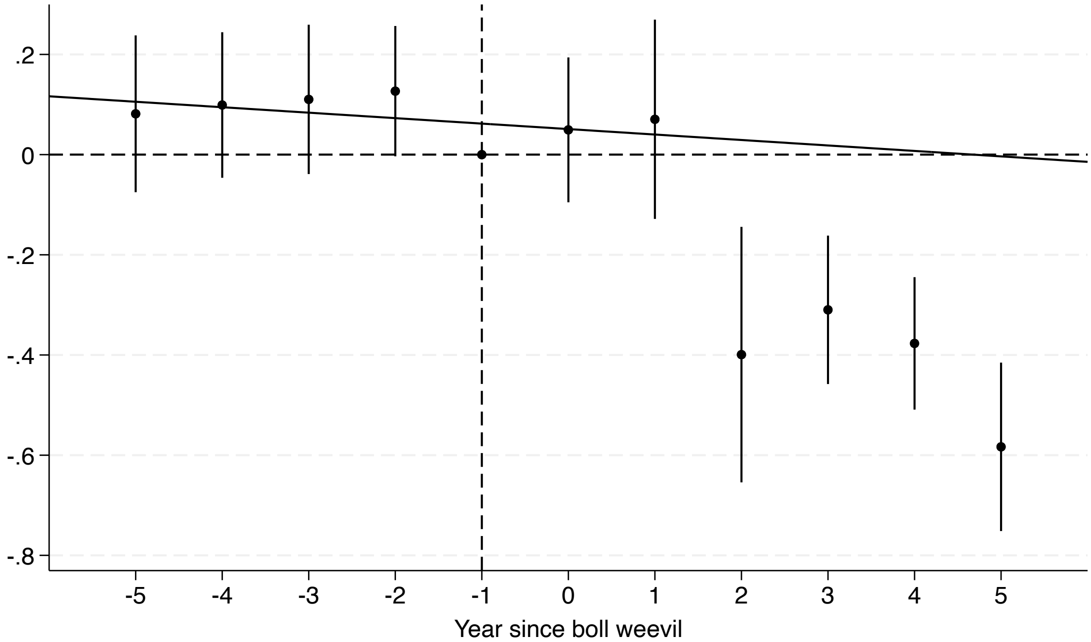
A Figure and Table Appendix

Figure A.1: The boll weevil's advance through the cotton belt



Notes: This map displays the advance of the boll weevil through the cotton belt and was originally published in Hunter and Coad (1923).

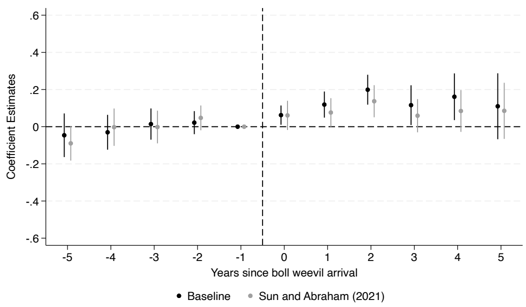
Figure A.2: Event study - Cotton bales



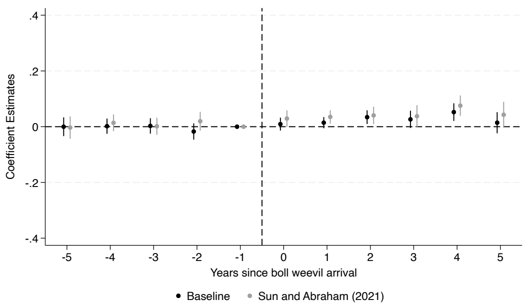
Notes: This figure presents estimates of the Specification in Table 5. The unit of observation is a county-year and the dependent variable is cotton bales. PPML is used for the estimation. The estimation is run for -10 to +10 years relative to the boll weevil's. Observations 10 or more years prior to the boll weevil's arrival in a county are binned into a "-10" bin; observations 10 or more years after the boll weevil's arrival in a county are binned into a "+10" bin. Only estimates for -5 to +5 years relative to the boll weevil's arrival in a county are shown. The coefficient on the year -1 is constrained to be zero and all coefficients are interpreted relative to this year. Counties invaded by the boll weevil between 1901 and 1920 are included (to line-up with the individual-level analysis). The data used in this figure comes from the 1890, 1900, 1910, 1920, 1925, and 1930 Censuses of Agriculture (Haines, Fishback and Rhode, 2018). All crop variables in the Censuses of Agriculture are measured in the year prior to the census being taken (e.g. 1889 instead of 1890, 1899 instead of 1900, etc.) and we account for this when assigning the boll weevil's arrival date. Standard errors are clustered based on 96 bins of longitude. 90% confidence intervals are shown.

Figure A.3: Event studies - Weekly wages and Imputed Income

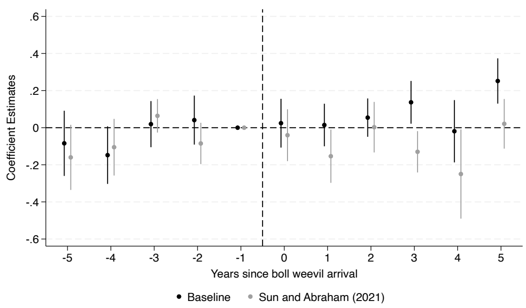
(a) DiD: Black Weekly Wages



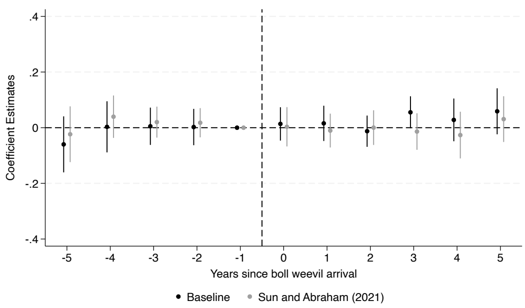
(b) DiD: Black Imputed Income



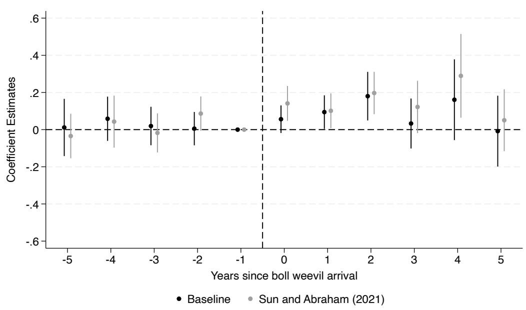
(c) DiD: White Weekly Wages



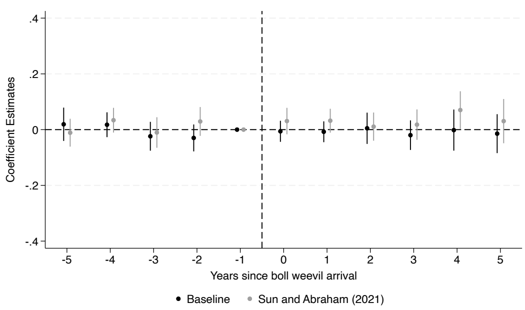
(d) DiD: White Imputed Income



(e) DDD: Weekly Wages



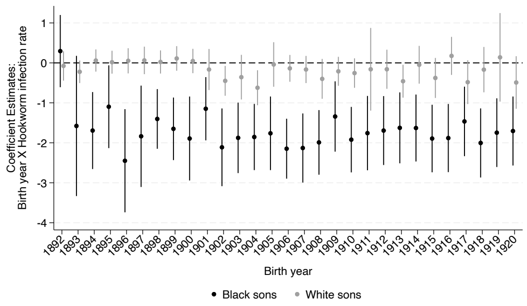
(f) DDD: Imputed Income



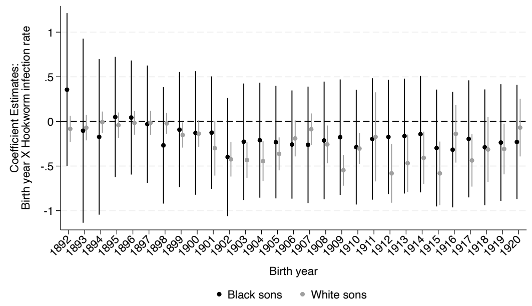
Notes: Panels (a), (b), (c), and (d) show estimates of Specification (2) in the text. Panels (e) and (f) show estimates of Specification (4). The unit of observation is sons. The estimation is run on sons born -10 to +9 years relative to the boll weevil's arrival in their father's original county of residence. Only estimates for sons born -5 to +5 years relative to the boll weevil's arrival are shown. In all panels the event time indicator for the year -1 is omitted. The Sun and Abraham (2021) estimator requires never treated units. For this estimator we define never treated counties as counties that had less than 10% of farm land in cotton acreage. All regressions use inverse propensity score re-weighting. Standard errors are clustered based on 96 bins of longitude. 90% confidence intervals are shown.

Figure A.4: Birth year effects interacted with hookworm infection rates and malaria ecology

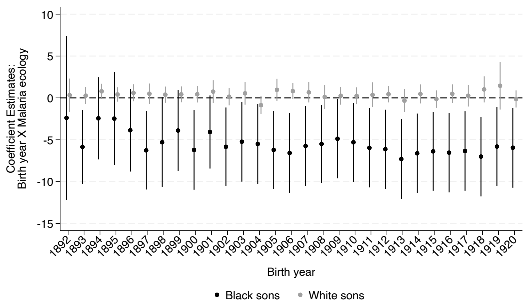
(a) Dependent variable: weekly wages; DiD controlling for hookworm infection rate



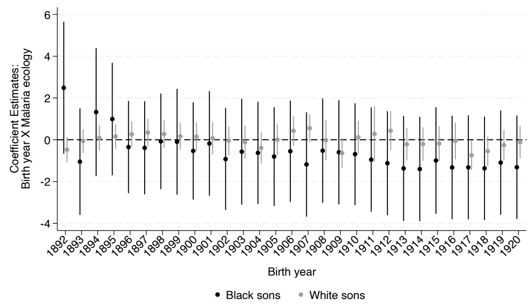
(b) Dependent variable: imputed income; DiD controlling for hookworm infection rate



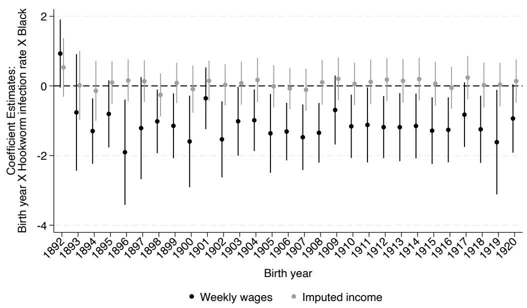
(c) Dependent variable: weekly wages; DiD controlling for malaria ecology



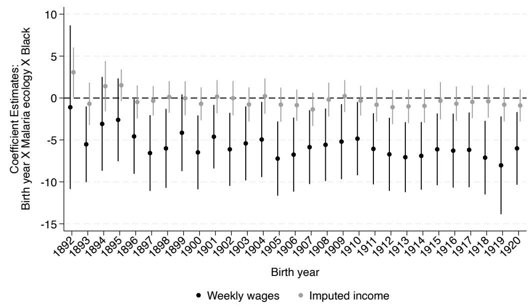
(d) Dependent variable: imputed income; DiD controlling for malaria ecology



(e) DDD controlling for hookworm infection rate



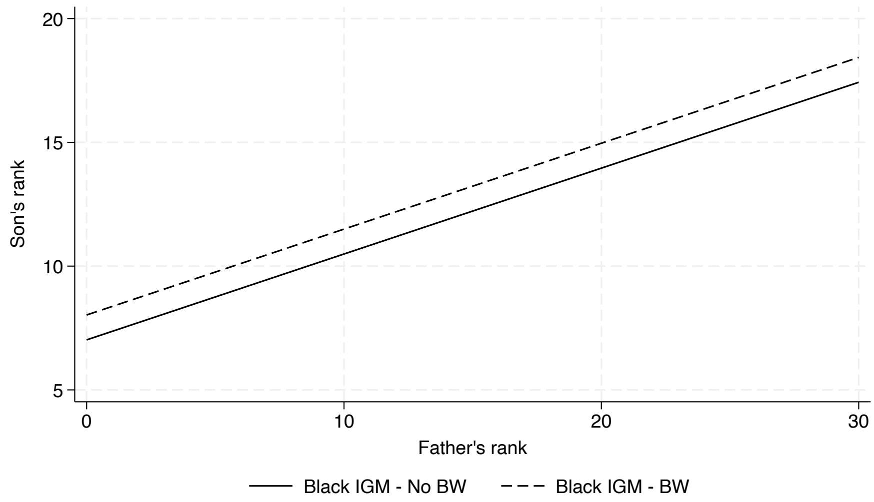
(f) DDD controlling for malaria ecology



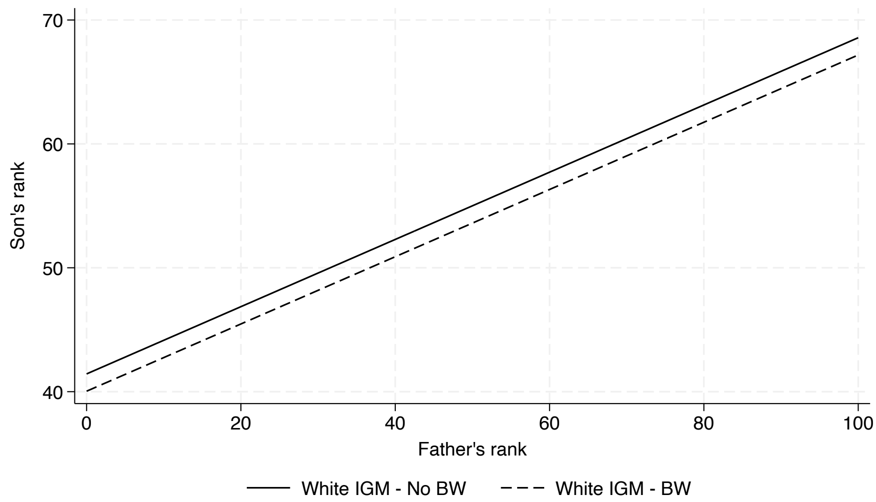
Notes: Panels (a) and (b) graph coefficient estimates for *county-level* hookworm infection rates interacted with birth year fixed effects, which are estimated in the regressions in Panel A and B, columns 1 and 5. Panels (c) and (d) graph coefficient estimates for *county-level* malaria ecology interacted with birth year fixed effects, which are estimated in the regressions in Panel A and B, columns 2 and 6. Panel (e) graphs coefficient estimates for *county-level* hookworm infection rates interacted with birth year-by-race fixed effects, which are estimated in the regressions in Panel C, columns 1 and 5. Panel (f) graphs coefficient estimates for *county-level* malaria ecology interacted with birth year-by-race fixed effects, which are estimated in the regressions in Panel C, columns 2 and 6. Standard errors are clustered based on 96 bins of longitude. 95% confidence intervals are shown.

Figure A.5: Intergenerational mobility by race

(a) Black



(b) White



Notes: Graphs are for estimates presented in column 2 of Table 3. For Black intergenerational mobility only fathers' income ranks from 0 to 30 are graphed because virtually no Black fathers had income ranks above 30.

Table A.1: The boll weevil and the likelihood of engaging in wage work

	Pr(Wage worker sample=1) (1)	Pr(Imputed income sample=1) (2)
	Panel A: DiD for Black sons	
Born post BW	0.007	-0.011
	(0.014)	(0.010)
Observations	31459	31459
Mean of dep. var.	.352	.352
	Panel B: DiD for White sons	
Born post BW	0.026	0.014
	(0.033)	(0.018)
Observations	397503	397503
Mean of dep. var.	.429	.429
	Panel C: Triple difference	
Born post BW * Black	0.017	-0.012
	(0.020)	(0.017)
Observations	428839	428839
Mean of dep. var.	.424	.424

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score reweighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.2: Summary statistics

	Black		White	
Sample:	Wage worker	Imputed income	Wage worker	Imputed income
	(1)	(2)	(3)	(4)
		Panel A: Fathers		
Works in ag. in first census	0.771	0.827	0.659	0.726
Works in ag. in second census	0.735	0.805	0.596	0.680
Farmer in first census	0.493	0.552	0.485	0.545
Farmer in second census	0.650	0.725	0.560	0.644
Farm laborer in first census	0.278	0.275	0.172	0.180
Farm laborer in second census	0.084	0.078	0.032	0.033
Laborer (all) in first census	0.408	0.377	0.234	0.234
Laborer (all) in second census	0.219	0.178	0.085	0.079
Imputed income in first census	345.032	332.821	760.757	719.727
Imputed income in second census	375.493	360.362	1002.999	946.509
Income rank in first census	7.488	6.620	33.931	31.708
Income rank in second census	8.103	7.213	46.292	43.210
Moved out of South	0.019	0.016	0.035	0.029
Moved out of state	0.061	0.056	0.125	0.115
Moved out of county	0.231	0.216	0.387	0.367
Rural location in first census	0.896	0.926	0.853	0.889
Moved rural to rural	0.159	0.159	0.272	0.277
Moved rural to urban area	0.068	0.054	0.094	0.074
Moved rural to large city (≥100,000)	0.017	0.014	0.011	0.008
Age in first census	28.491	28.540	29.385	29.605
Observations	10324	24107	153018	292672
		Panel B: Sons		
Weekly wage	10.512	10.531	24.475	23.037
Income rank	32.005	27.189	60.993	53.152
Imputed income	470.081	409.242	1120.419	1053.321
Imputed income rank	11.583	8.473	49.927	45.305
Years of schooling	6.645	6.293	10.065	9.474
Moved out of father's initial region (South)	0.123	0.089	0.123	0.095
Living in urban area	0.450	0.302	0.515	0.362
Living in large city (≥100,000)	0.219	0.146	0.194	0.130
Age in 1940	28.455	28.674	31.632	31.727
Observations	11155	27077	171064	352319

Notes: Weekly wages and imputed income are in 1939 dollars. Imputed incomes come from Collins and Wanamaker (2022).

Table A.3: County characteristics for fathers who migrated

	Black Fathers		White Fathers	
	First census	Second census	First census	Second census
	(1)	(2)	(3)	(4)
County population	32397	105814	31450	48853
Percent urban	0.123	0.250	0.136	0.214
Manufacturing establishment	127.803	458.165	131.283	157.174
Cotton acreage share of farm acreage (1899)	0.162	0.125	0.135	0.082
Observations	5198	5198	107434	107434

Notes: This table reports average county-level characteristics for fathers that moved counties between the first and second censuses they are observed in. Cotton acreage as a share on farm acreage is from 1899.

Table A.4: The boll weevil, weekly wages, and imputed income - controlling for hookworm and malaria eradication campaigns and Rosenwald schools

	Log(weekly wage)				Log(imputed income)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: DiD for Black sons								
Born post BW	0.097^{***}	0.093^{***}	0.098^{***}	0.096^{***}	0.019	0.033^{***}	0.022^{*}	0.034^{***}
	(0.033)	(0.029)	(0.033)	(0.030)	(0.013)	(0.010)	(0.012)	(0.010)
Rosenwald exposure	-0.127			-0.021	-0.131^{**}			-0.024
	(0.082)			(0.041)	(0.058)			(0.018)
Observations	11155	11155	11155	11155	27077	27077	27077	27077
Mean of dep. var.	2.095	2.095	2.095	2.095	5.936	5.936	5.936	5.936
Panel B: DiD for White sons								
Born post BW	-0.042	-0.038	-0.042	-0.043	-0.025	-0.019	-0.026	-0.021
	(0.049)	(0.052)	(0.050)	(0.053)	(0.024)	(0.025)	(0.024)	(0.024)
Rosenwald exposure	-0.095^{***}			-0.063^{***}	-0.216^{***}			-0.126^{***}
	(0.034)			(0.023)	(0.036)			(0.022)
Observations	171064	171064	171064	171064	352319	352319	352319	352319
Mean of dep. var.	2.964	2.964	2.964	2.964	6.817	6.817	6.817	6.817
Panel C: Triple difference								
Born post BW * Black	0.119^{***}	0.113^{***}	0.118^{***}	0.114^{***}	0.038^{**}	0.043^{**}	0.036^{**}	0.045^{***}
	(0.041)	(0.041)	(0.041)	(0.041)	(0.018)	(0.017)	(0.018)	(0.017)
Rosenwald exposure * Black			-0.075	0.012			0.008	0.062^{***}
			(0.082)	(0.036)			(0.041)	(0.018)
Observations	181709	181709	181709	181709	379237	379237	379237	379237
Mean of dep. var.	2.911	2.911	2.911	2.911	6.754	6.754	6.754	6.754
Rosenwald Schools	X			X	X			X
Malaria ecology		X		X		X		X
Hookworm infection rate			X	X			X	X

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), and a full set of indicators for sons' birth order. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, and fathers' initial county-by-birth year fixed effects. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.5: The boll weevil, weekly wages, and imputed income controlling for migration out of the South

†: Who moved?	Log (weekly wage) Father (1)	Son (2)	Log (imputed income) Father (3)	Son (4)
Panel A: DiD for Black sons				
Born post BW	0.088***	0.067**	0.025***	0.023**
	(0.032)	(0.034)	(0.009)	(0.009)
Moved out of South†	0.448***	0.705***	0.426***	0.667***
	(0.063)	(0.028)	(0.033)	(0.014)
Born post BW*Moved out of South†	-0.055	-0.028	0.051	-0.014
	(0.088)	(0.052)	(0.036)	(0.023)
Observations	11155	11155	27077	27077
Mean of dep. var.	2.095	2.095	5.936	5.936
Panel B: DiD for White sons				
Born post BW	-0.036	-0.061	0.015	-0.019
	(0.037)	(0.055)	(0.017)	(0.022)
Moved out of South†	0.232***	0.104**	0.253***	0.164***
	(0.029)	(0.049)	(0.020)	(0.022)
Born post BW*Moved out of South†	-0.008	0.029	-0.049**	0.002
	(0.042)	(0.039)	(0.022)	(0.027)
Observations	171064	171064	352319	352319
Mean of dep. var.	2.964	2.964	6.817	6.817
Panel C: Triple difference				
Born post BW * Black	0.104**	0.078*	0.035**	0.027
	(0.041)	(0.044)	(0.016)	(0.017)
Moved out of South†	0.189***	0.107**	0.222***	0.167***
	(0.034)	(0.050)	(0.019)	(0.020)
Born post BW * Moved out of South†	0.030	0.034	-0.011	-0.007
	(0.035)	(0.041)	(0.021)	(0.028)
Moved out of South†*Black	0.212***	0.584***	0.212***	0.500***
	(0.071)	(0.056)	(0.039)	(0.023)
Born post BW * Moved out of South † * Black	0.068	0.027	0.050	-0.004
	(0.105)	(0.065)	(0.042)	(0.031)
Observations	181709	181709	379237	379237
Mean of dep. var.	2.911	2.911	6.754	6.754

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.6: The boll weevil, weekly wages, and imputed income - alternative spatial autocorrelation standard errors

	Log(weekly wage)			Log(imputed income)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: DiD for Black sons					
Born post BW	0.096***	0.096***	0.096***	0.034***	0.034***	0.034***
	(0.026)	(0.015)	(0.015)	(0.010)	(0.009)	(0.009)
Observations	11155	11155	11155	27077	27077	27077
Mean of dep. var.	2.095	2.095	2.095	5.936	5.936	5.936
	Panel B: DiD for White sons					
Born post BW	-0.043	-0.043	-0.043	-0.021	-0.021	-0.021*
	(0.056)	(0.034)	(0.031)	(0.021)	(0.023)	(0.012)
Observations	171064	171064	171064	352319	352319	352319
Mean of dep. var.	2.964	2.964	2.964	6.817	6.817	6.817
	Panel C: Triple difference; state-by-year FE					
Born post BW * Black	0.119***	0.119***	0.119***	0.063***	0.063***	0.063***
	(0.043)	(0.035)	(0.034)	(0.019)	(0.023)	(0.007)
Observations	181709	181709	181709	379237	379237	379237
Mean of dep. var.	2.911	2.911	2.911	6.754	6.754	6.754
Spatial HAC bandwidth:	100km	200km	400km	100km	200km	400km

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial state-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. All columns report standard errors based on the Conley (1999) spatial heteroskedasticity and autocorrelation-consistent (HAC) estimation using bandwidths of 100, 200, or 400 kilometers.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.7: The boll weevil and other economic outcomes

	Percentile in national income distribution	Log(annual income)	Weeks worked
	(1)	(2)	(3)
	Panel A: DiD for Black sons		
Born post BW	2.401***	0.096***	0.044
	(0.750)	(0.031)	(0.311)
Observations	11155	11155	11155
Mean of dep. var.	32.005	5.93	46.914
	Panel B: DiD for White sons		
Born post BW	-1.500	-0.064	-0.858*
	(1.742)	(0.057)	(0.449)
Observations	171064	171064	171064
Mean of dep. var.	60.993	6.841	48.736
	Panel C: Triple difference		
Born post BW * Black	3.151***	0.130***	0.678*
	(1.149)	(0.042)	(0.372)
Observations	181709	181709	181709
Mean of dep. var.	59.223	6.785	48.624

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.8: Alternative weekly wage restrictions

	Log(weekly wage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: DiD for Black sons						
Born post BW	0.096***	0.085***	0.065**	0.075**	0.090***	0.096***	0.159**
	(0.030)	(0.028)	(0.027)	(0.029)	(0.028)	(0.030)	(0.067)
Observations	11155	11870	12527	10306	11155	11155	1406
Mean of dep. var.	2.095	2.098	2.103	2.094	2.095	2.095	2.014
	Panel B: DiD for White sons						
Born post BW	-0.043	-0.033	-0.029	-0.061	-0.037	-0.043	-0.203*
	(0.053)	(0.051)	(0.047)	(0.053)	(0.052)	(0.053)	(0.107)
Observations	171064	181143	188877	164179	171064	171064	33833
Mean of dep. var.	2.964	2.923	2.911	2.951	2.964	2.964	2.965
	Panel C: Triple difference						
Born post BW * Black	0.114***	0.092**	0.069*	0.125***	0.106***	0.114***	0.179
	(0.041)	(0.042)	(0.039)	(0.042)	(0.039)	(0.041)	(0.128)
Observations	181709	192552	200966	161486	181709	181709	31944
Mean of dep. var.	2.911	2.873	2.861	2.939	2.911	2.911	2.936
Restrictions	Baseline	Include work relief	Include unemployed and not in labor force	Work 40 or more weeks	Adjust farm laborer wages for perquisites	Wages censored at $100	Brothers

(Table A.8 continued)

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude. 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. 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 restricts the baseline sample to individuals that worked more than 40 weeks in the year. Goldin and Margo (1992) use this restriction as well as replacing incomes above $5,000 with 1.4 times $5,000, or $7,000. We chose not to do this winsorizing because we are comparing Black and White incomes and almost no Black incomes are above $5,000. Column 5 uses information from Collins and Wanamaker (2022) to adjust the wages of farm laborers for perquisites 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 perquisites for farm laborers in 1940. Column 6 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. Column 7 uses a sample of brothers and father fixed effects. We define brothers as male individuals who report having the same father in the censuses. The coefficient estimates in column 7 are identified by sets of brothers where at least one brother was born prior to the arrival of the boll weevil in their father's initial county and at least one brother was born after.

$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.9: The boll weevil, weekly wages, and imputed income - different treatment start date and dropping years close to arrival

	Log (weekly wage)					Log(imputed income)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: DiD for Black sons										
Born post BW	0.096***	0.092***	0.126***	0.134***	0.091**	0.034***	0.033***	0.042***	0.052***	0.032***
	(0.030)	(0.025)	(0.032)	(0.034)	(0.036)	(0.010)	(0.011)	(0.012)	(0.015)	(0.011)
Observations	11155	11155	10147	9190	9218	27077	27077	24661	22324	22565
Mean of dep. var.	2.095	2.095	2.102	2.116	2.106	5.936	5.936	5.938	5.941	5.938
Panel B: DiD for White sons										
Born post BW	-0.043	-0.048	-0.056	-0.038	-0.038	-0.021	-0.012	-0.025	-0.033	-0.016
	(0.053)	(0.059)	(0.064)	(0.066)	(0.049)	(0.024)	(0.029)	(0.031)	(0.032)	(0.028)
Observations	171064	171064	157288	144243	144869	352319	352319	324080	297339	299661
Mean of dep. var.	2.964	2.964	2.971	2.982	2.97	6.817	6.817	6.822	6.828	6.82
Panel C: Triple difference										
Born post BW * Black	0.114***	0.110***	0.153***	0.164***	0.099**	0.045***	0.047**	0.059***	0.073***	0.048**
	(0.041)	(0.038)	(0.045)	(0.049)	(0.045)	(0.017)	(0.019)	(0.020)	(0.023)	(0.019)
Observations	181709	181709	166932	152935	153613	379237	379237	348582	319507	322073
Mean of dep. var.	2.911	2.911	2.919	2.93	2.918	6.754	6.754	6.759	6.766	6.759
	Treated year after arrival (baseline)	Treated year of arrival	Drop year 0	Drop year 0 & 1	Drop year -2 & -1	Treated year after arrival (baseline)	Treated year of arrival	Drop year 0	Drop year 0 & 1	Drop year -2 & -1

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.10: Alternative measure of boll weevil arrival using Ferrara, Ha and Walsh (2022)

	Log(weekly wage)			Log(imputed income)		
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: DiD for Black sons			
Born post BW	0.096***	0.063**	0.112***	0.034***	0.024**	0.042***
	(0.030)	(0.031)	(0.040)	(0.010)	(0.010)	(0.013)
Observations	11155	11155	9035	27077	27077	21828
Mean of dep. var.	2.095	2.095	2.133	5.936	5.936	2.08
			Panel B: DiD for White sons			
Born post BW	-0.043	-0.058	-0.061	-0.021	-0.038	-0.051
	(0.053)	(0.052)	(0.079)	(0.024)	(0.029)	(0.035)
Observations	171064	171064	142646	352319	352319	291776
Mean of dep. var.	2.964	2.964	3.004	6.817	6.817	2.876
			Panel C: Triple difference			
Born post BW * Black	0.114***	0.055	0.112**	0.045***	0.034*	0.059***
	(0.041)	(0.038)	(0.050)	(0.017)	(0.020)	(0.022)
Observations	181709	181709	151225	379237	379237	313470
Mean of dep. var.	2.911	2.911	2.952	6.754	6.754	6.778
	USDA	Ferrara et al. (2024)	Agreement sample	USDA	Ferrara et al. (2024)	Agreement sample

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.11: Robustness of weekly wages to the number of linking sources in the Census Tree

	Log(weekly wage)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: DiD for Black sons					
Born post BW	0.087***	0.096***	0.097***	0.110**	0.104	0.463
	(0.029)	(0.030)	(0.033)	(0.043)	(0.076)	(0.314)
Observations	12005	11155	8978	5365	1948	219
Mean of dep. var.	2.085	2.095	2.111	2.151	2.201	2.042
	Panel B: DiD for White sons					
Born post BW	-0.048	-0.043	-0.031	-0.046	-0.046	-0.051
	(0.053)	(0.053)	(0.058)	(0.072)	(0.077)	(0.068)
Observations	174511	171064	152491	101665	59842	19320
Mean of dep. var.	2.961	2.964	2.968	2.954	2.961	2.933
	Panel C: Triple difference					
Born post BW * Black	0.115***	0.114***	0.100**	0.113**	0.011	1.025*
	(0.040)	(0.041)	(0.043)	(0.056)	(0.080)	(0.567)
Observations	186028	181709	160873	106116	60377	17182
Mean of dep. var.	2.905	2.911	2.92	2.914	2.94	2.932
Number of linking sources:	1	2	3	4	5	6

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude. $* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.12: Robustness of imputed income to the number of linking sources in the Census Tree

	Log(imputed income)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: DiD for Black sons					
Born post BW	0.030***	0.034***	0.038***	0.032**	0.016	0.086*
	(0.010)	(0.010)	(0.011)	(0.014)	(0.021)	(0.051)
Observations	29184	27077	21828	13246	5150	867
Mean of dep. var.	5.933	5.936	5.939	5.95	5.961	5.91
	Panel B: DiD for White sons					
Born post BW	-0.017	-0.021	-0.025	-0.020	-0.022	0.012
	(0.023)	(0.024)	(0.026)	(0.032)	(0.035)	(0.052)
Observations	359255	352319	315447	210566	124741	41784
Mean of dep. var.	6.815	6.817	6.819	6.809	6.814	6.794
	Panel C: Triple difference					
Born post BW * Black	0.039**	0.045***	0.054***	0.040**	0.031	0.063
	(0.017)	(0.017)	(0.018)	(0.020)	(0.034)	(0.107)
Observations	388287	379237	337074	223404	129082	40902
Mean of dep. var.	6.749	6.754	6.762	6.759	6.781	6.778
Number of linking sources	1	2	3	4	5	6

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude. $* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.13: Alternative DiD estimators

	Log(weekly wage)			Log(imputed income)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: DiD for Black sons					
Born post BW	0.071**	0.163***	0.039	0.016*	0.024	0.016**
	(0.028)	(0.059)	(0.028)	(0.008)	(0.015)	(0.008)
Observations	4760	3645	4760	6617	5707	6617
	Panel B: DiD for White sons					
Born post BW	-0.012*	-0.017	0.001	0.002	-0.011*	-0.002
	(0.007)	(0.012)	(0.008)	(0.003)	(0.006)	(0.004)
Observations	9373	8776	9373	9920	9403	9920
Method	TWFE	Callaway and Sant'Anna (2021)	Sun and Abraham (2021)	TWFE	Callaway and Sant'Anna (2021)	Sun and Abraham (2021)

Notes: The unit of observation is a birth year-county-race cell. Our individual-level data (used throughout the rest of the paper) are collapsed to the birth year-county-race level (e.g. averages for Black sons born in 1903 whose father initially resided in Winston County, AL). We do this to make it computationally easier to compute the Callaway and Sant'Anna (2021) estimator. All columns control for fathers' initial county fixed effects and birth year fixed effects. Columns 2 and 5 use the estimator proposed by Callaway and Sant'Anna (2021) without any control variables. Columns 3 and 6 use the estimator proposed by Sun and Abraham (2021) without any control variables. Standard errors in columns 1, 3, 4,and 6 are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.14: The boll weevil, weekly wages, and imputed income for changes in father's characteristics

	Log(weekly wage)						Log(imputed income)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: DiD for Black sons												
Born post BW	0.096***	0.181**	0.014	0.157**	0.110*	0.064	0.034***	0.029	0.102*	0.024	0.006	0.023*
	(0.030)	(0.079)	(0.115)	(0.074)	(0.061)	(0.049)	(0.010)	(0.023)	(0.056)	(0.017)	(0.020)	(0.014)
Observations	11155	1630	623	1825	2692	4384	27077	4171	1315	5071	6758	11906
Mean of dep. var.	2.095	2.028	2.444	1.866	1.986	2.046	5.936	5.925	6.283	5.839	5.899	5.885
Panel B: DiD for White sons												
Born post BW	-0.043	0.064	-0.020	-0.005	-0.038	-0.012	-0.021	-0.011	0.066	0.016	-0.020	-0.005
	(0.053)	(0.042)	(0.062)	(0.044)	(0.056)	(0.014)	(0.024)	(0.025)	(0.048)	(0.022)	(0.047)	(0.007)
Observations	171064	46592	15928	17028	29688	50926	352319	98069	24782	40495	62430	121010
Mean of dep. var.	2.964	2.929	3.196	2.646	2.787	2.891	6.817	6.798	7.06	6.596	6.685	6.761
Panel C: Triple difference												
Born post BW * Black	0.114***	0.183*	-0.012	0.277***	0.170*	0.047	0.045***	0.036	0.178**	0.026	0.047*	0.022
	(0.041)	(0.101)	(0.152)	(0.094)	(0.090)	(0.057)	(0.017)	(0.037)	(0.085)	(0.028)	(0.025)	(0.017)
Observations	181709	46471	13962	17018	30527	53838	379237	101191	23726	43973	67762	132251
Mean of dep. var.	2.911	2.904	3.178	2.57	2.72	2.829	6.754	6.763	7.025	6.511	6.608	6.683
Father:	Baseline	Move rural to rural	Move rural to urban	Upgrade ag.	Leave labor	No migration or change in occ.	Baseline	Move rural to rural	Move rural to urban	Upgrade ag.	Leave labor	No migration or change in occ.

Notes: The unit of observation is sons. Panels A and B display estimates for Specification (1) in the text. Panel C provides estimates for Specification (3). Panels A and B control for: birth year fixed effects, fathers' initial county of residence fixed effects (i.e. where the father resided in 1900 or 1910), sons' initial census enumeration year fixed effects (i.e. 1900, 1910, or 1920), a full set of indicators for sons' birth order, hookworm infection rates interacted with birth year fixed effects, malaria ecology interacted with birth year fixed effects, and Rosenwald school exposure. Panel C controls for: birth year-by-race fixed effects, fathers' initial county-by-race fixed effects, census enumeration year-by-race fixed effects, birth order-by-race fixed effects, fathers' initial county-by-birth year fixed effects, hookworm infection rates interacted with birth year-by-race fixed effects, malaria ecology interacted with birth year-by-race fixed effects, and Rosenwald school exposure interacted with race. All regressions use inverse propensity score re-weighting. Birth order is based on the linked sons we observe with each father. Standard errors are clustered based on 96 bins of longitude.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table A.15: Pellagra death rates

	Log(pellagra death rate)	
	(1)	(2)
Post BW	-0.177***	-0.093
	(0.046)	(0.057)
Post BW * High Black share		-0.144**
		(0.061)
Observations	1273	1273
Mean of dep. var.	.763	.763

Notes: The unit of observation is counties in North and South Carolina. Pellagra death rates are not available in other states prior to the arrival of the boll weevil. 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 years 1915-1925 and for counties in South Carolina for the years 1916-1925. High Black share counties have a Black share of the population above the 50th percentile in the 1900 census. This corresponds to a Black share above 41.7%. All columns control for county and year fixed effects. Standard errors are clustered at the county level.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

B Data and Linking Appendix

B.1 Linking

B.1.1 Linking data and methods

To construct this linked dataset, we use crosswalks provided by the Census Tree Project (Buckles et al., 2023; Price et al., 2021). The 1900-1910 (Price et al., 2023*a*), 1910-1920 (Price et al., 2023*c*), 1900-1940 (Price et al., 2023*b*), 1910-1940 (Price et al., 2023*d*), and 1920-1940 (Price et al., 2023*e*) crosswalks were used.

Our linking procedure is as follows. Starting with the 1900 complete count census we add in the 1900-1910 linked sample and the 1900-1940 linked sample. We then keep only sons whose father was linked from 1900 to 1910 and who themselves were linked from 1900 to 1940. We further restrict to only sons whose father was initially living in a county invaded by the boll weevil between 1901 and 1910. This provides us with information on fathers and their sons born prior to the weevil's arrival. To obtain information on sons born after the weevil's arrival we use the 1910 complete count census and add in the 1900-1910 linked sample and the 1910-1940 linked sample. Once again, we keep only sons whose father was linked between 1900 and 1910 and who themselves were linked between 1910 and 1940; we again restrict to males whose father was initially living in a county (in 1900) invaded by the weevil between 1901 and 1910. We stack the set of sons linked from 1900 to 1940 with the set of sons linked from 1910 to 1940. This provides us with sons born both before and after the weevil's arrival in the county their father initially resided in (in 1900).

We repeat this entire procedure using the 1910 and 1920 complete count censuses and restricting to sons whose father was living in a county invaded by the weevil between 1911 and 1920. If a son is linked from multiple censuses to 1940, we only keep the earliest link. For example if a son is linked from 1900 to 1940 and from 1910 to 1940 we only keep the 1900 to 1940 observation.

This linking procedure produces a rich data set containing information on adult outcomes for sons born around the time of the weevil's arrival. It also contains detailed information on these sons' fathers both before and after the boll weevil invaded the county they were residing in. For example, we observe whether fathers moved, where they moved to, whether they changed occupations, etc.

Figure B.1 provides an example of our linking method for two families. For the first family we observe a father in the 1900 complete count census living in Rapides Parish, LA, which was invaded by the boll weevil in 1904. We observe this father with his son who was born in 1898. We then link this father forward to the 1910 census using the Census Tree crosswalks and observe him with another son who was born in 1906. Both sons are then linked forward to 1940 to obtain their adult outcomes, which is what are used in our analysis. For the second family we observe a father in the 1910 complete count census living in Lowndes County, AL, which was invaded by the boll weevil in 1912. We observe this father with his son who was born in 1907. We then link him forward to the 1920 census and observe him with another son who was born in 1915. Both sons are then linked forward to 1940 to obtain their adult outcomes. Note that, despite the example, it is not always the case that a father has one son born before and one born after the boll weevil; sometimes all of a father's sons are born before the boll weevil and sometimes we only successfully link one son per father.

As just explained, a son is only in our linked sample if their father is also linked from either 1900 to 1910 or from 1910 to 1920. Moehling (2004) finds that Black children were less likely to live with one or both of their parents than White children. This level difference does not matter for our analysis since our analysis is conditional on living with your father. However, if family structure systematically changed after the boll weevil it might bias our estimates. To address this issue, we stacked the full count 1900, 1910, and 1920 censuses and included only children who were born within 10 years ($\pm$) of the boll weevil's arrival in the county they were living in.[B.1] We then regress a indicator for father being absent, mother being absent, or both parents being absent on an indicator for whether the boll weevil already invaded the county. The results are shown in Appendix Table B.1.[B.2] Similar to Moehling (2004), we find

[B.1]Children living in a county not invaded by the boll weevil are not included in the analysis.

[B.2]Panels A and B include county and birth year fixed effects. Panel C includes county-by-race birth that different shares of children in our sample – 20% of Black children and 6% of White children – were living without their father. Similarly, 10% of Black children and 4% of White children were living without their mother. However, the coefficients on father absent, mother absent, and both parents absent are all small and not statistically significant. We conclude that our estimates are not biased due to changes in family structure that occurred as a result of the boll weevil. Our results may not, however, generalize to Black or White sons who did not live with their father.

The Census Tree Project, which we use for all the linking in this paper, contains links from seven different methods. These are: (1) Family Tree links, (2) XGBoost algorithm links, (3) the Census Linking Project (CLP; Abramitzky et al. (2020)), (4) the Multigenerational Longitudinal Panel (MLP; Helgertz et al. (2023)), (5) FamilySearch "profile hints", (6) FamilySearch "direct hints", and (7) implied links. A brief description of each method follows; for more details see Buckles et al. (2023).

(1): Family Tree links are user created links from family trees on FamilySearch.org. Users have made over 317 million unique census links between 1850 and 1940. How FamilySearch users created these links is not known.

(2): Links made using the XGBoost algorithm were constructed specifically for the Census Tree Project and are an attempt to model how FamilySearch users made links. Buckles et al. (2023) start with Family Tree links made by FamilySearch users. After standardizing names, places, etc. they create a set of possible matches in each census by blocking the data based on first and last name (NYSIIS standardized), birthplace, birth year ($\pm$ 3 years), sex, and race. A subset of the "true" matches, as identified by FamilySearch users, are used to train the blocked data to identify true versus false matches. Over 70 variables/features are used in this training (including features such as the distance, in miles, between the two towns an individual lived in). After the data have been trained, a score can be assigned to every potential link in the blocked data, with this score being the predicted probability of a link being "true". A link is declared "true" if it has the highest probability score and has the highest sheet count

year-by-race and county-by-birth year fixed effects.

(i.e. total number of individual links between the census pages which contain the two records). Any remaining conflicts between two years (e.g. if two records are tied in probability score and both have the same sheet count) are removed. Over 98% of the links made by FamilySearch users satisfy these criteria. See Buckles et al. (2023) for complete details on the XGBoost algorithm.

(3): Links from the CLP are unsupervised (i.e. no training data is used) and use rules based on an individual's first name, last name, birth year, and birth place. The links provided by the CLP are not necessarily unique, but the Census Tree project excludes multiple links by requiring that names be unique within the birth year.

(4): The MLP links between adjacent censuses using a two step approach. First, a machine learning algorithm is used to obtain high-quality matches for men. Then other individuals in matched mens' households are linked provided they still reside in the household with the matched male. Since the MLP only links between adjacent census, there are no links for sons in our dataset that rely on MLP matches. If MLP links were "daisy-chained" together (i.e. combine links from 1910-1920, 1920-1930, and 1930-1940 to get links from 1910-1940) some of our links would almost surely be in the resulting dataset.

(5) and (6): FamilySearch has a proprietary machine learning algorithm that provides two types of "hints" for census records to users. The first, referred to by Buckles et al. (2023) as "profile hints", suggests to a FamilySearch user that a census record might belong in their family tree. When census records from two different years are both "hinted" at this creates a link between an individual in those two censuses. The second hint, referred to by Buckles et al. (2023) as "direct hints", directly identifies a possible link between two census records.

To create the Census Tree crosswalks, links from the six methods just described are combined. To handle discrepancies between the six methods, Buckles et al. (2023) calculate a sheet count for each link in all six methods and only keep the link with the highest count. Any remaining discrepancies (i.e. if two methods produce different links that have the same sheet count) are discarded.

(7): Finally, Buckles et al. (2023) create "implied" links by taking advantage of the fact that if a record is linked to two different censuses then a link can also be established between those two censuses. For example, if an individual is linked between 1900 and 1910 and is also linked between 1910 and 1920, then this individual can be linked directly from 1900 to 1920. These implied links are added to the six previous methods, discrepancies are once again removed (via sheet count), and the Census Tree crosswalks are complete.

The Census Tree sample is very accurate. Buckles et al. (2023) had research assistants (RAs) manually link 760 census records between 1900 and 1910. They find that the FamilySearch "hints" and the Family Tree links have the highest accuracy, with 95-97% of links made using these methods aligning with the RAs. They also find a steep increase in accuracy based on the number of methods by which a link is identified. If a link is only identified by one method, there is a 68 to 81% chance the link aligns with the RAs. However, if a link is identified by two methods there is an 86 to 94% chance the link aligns with the RAs. Link accuracy continues to increase with the number of methods, but less dramatically. In our estimating sample we make sure each link is identified in at least two sources and that at least one of the sources is the highly accurate FamilySearch "hints."

B.1.2 Representativeness of linked data and inverse propensity score re-weighting

A concern with any linked sample is whether it is representative of the entire population. If certain groups of individuals are more likely to be linked it could result in biased estimates, especially when studying intergenerational mobility. For example, Ward (2023) writes that if "children from low socioeconomic status families who remain poor in adulthood are less likely to be linked" than one would overestimate the amount of intergenerational mobility (Ward, 2023, p. 3222).

To make our linked sample more representative of the relevant population of interest, we generate inverse propensity weights, as described in Bailey, Cole and Massey (2020) and Appendix B of Ward (2023). We perform the following steps:

1. We pool each linked sample (1900-1940, 1910-1940 with father first observed in 1900, 1910-1940 with father first observed in 1910, and 1920-1940) with the sample of all children from the complete count censuses who satisfy similar restrictions. One challenge with defining the set of children from the complete count censuses who satisfy similar restrictions is that sons are only in our sample if their father was himself linked from either 1900 to 1910 or 1910 to 1920. We obviously cannot require these same restrictions of the population in the 1900, 1910, or 1920 complete count censuses. Accordingly, when generating weights we use liberal definitions for the population of interest. For example, we pool all sons in the 1900-1940 linked sample who were living in a state invaded by the boll weevil (regardless of whether their father was living in a county invaded by the boll weevil in the next ten years) with individuals in the 1900 complete count census who were: male, under the age of 10, and living in a state invaded by the boll weevil. We pool all sons in the 1910-1940 linked sample (with father first observed in 1900), who could have been living anywhere in the country, with individuals in the 1910 complete count census who were: male and under the age of 10. We no longer limit the comparison to individuals living in a state invaded by the boll weevil because the linked sons might be living anywhere in the United States, especially if their father moved out of the South. We use similar comparisons for the 1910-1940 (father first observed in 1910) and 1920-1940 linked samples.

2. We estimate a probit model to predict who will be in the linked sample. We use the following variables to predict who will be linked: a Black indicator, dummy variables for each age and their interaction with the Black indicator, state of residence and its interaction with the Black indicator, and farm status (living on a farm) and its interaction with the Black indicator.

3. Using the estimates from the probit model, we calculate $\hat{p}$, the probability that an individual is linked. Figure B.2 shows the distribution of probabilities for linked and unlinked individuals in each of our four linked samples. The figures show a large amount of overlap in the probability of being linked meaning we are not just linking certain types of individuals.

4. We re-weight the sample using an inverse propensity weight: $(\frac{1-\hat{p}}{\hat{p}}) * (\frac{1}{1-q})$. $\hat{p}$ is the predicated probability that an individual is linked and q is the share of the population of interest that is linked.

Table B.2 compares our linked sample to the population of interest and finds that, prior to weighting, the linked sample appears representative along some dimensions, such as age, but not others, such as race. For example, in 1900 (Panel A) the average age of the relevant population was 4.43, 37% were Black, and 59% lived on a farm (column 1). In our 1900-1940 linked sample, the average age was 4.275, 16% were Black, and 67% lived on a farm (column 2). We link about 32% of the relevant population. Once the inverse propensity weights are applied, our linked sample appears more similar to the relevant population (column 3). Inverse propensity re-weighting is used throughout our analysis except in columns 1 and 4 of Table 1, which show our baseline Specification without re-weighting.

The Census Tree claims to be able to match 72% of men from 1900 to 1940, 75% of men from 1910 to 1940, and 78% of men from 1920 to 1940. There are several reasons why our linking rates in Table B.2 are lower. First, for a son to be included in our linked sample, their father must also be linked. Second, we require that each link is identified by at least two of the seven linking sources. Finally, we discard duplicate links; if an individual is linked from 1900-1940, 1910-1940, and 1920-1940 we only keep the 1900-1940 link. This mechanically lowers the link rate for 1910-1940 and 1920-1940.

B.2 Sample Restrictions

We make several restrictions on who is included in our imputed income and wage worker samples. Appendix Table B.3 shows how these restrictions impact the number of observations in our sample.

We start with all sons who were born within 10 years of the boll weevil's arrival in their father's initial county of residence (either in 1900 or 1910). There are 31,479 Black sons and 397,505 White sons in our linked sample that meet this criteria.

To construct the imputed income sample, we impose two restrictions. First, we exclude individuals that did not report an occupation. Second, we exclude individuals who were not in the labor force, but with a reported occupation, and individuals that worked for public work relief programs, such as the CCC or WPA. Census enumerators were instructed to record occupations for "each person who was classified as at work; as at work on, or assigned to public emergency work; as seeking work; or as with a job." Thus, individuals not in the labor force should not have a reported occupation. In addition, individuals working for public work relief programs had their occupation in the program recorded, which might not be their usual occupation. There are additional reasons to not include individuals on work relief in the sample when examining wages. 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). [B.3] This is the sample we use to estimate our results for imputed income.

To construct the wage worker sample, we impose three additional restrictions on the imputed income sample. First, we exclude workers who were self employed or unclassified. Census enumerators were only supposed to record the wage and salary income earned as an *employee*. The exact instructions say that income should be recorded "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., 2021). There are also a very small number of workers who are not classified as either "self-employed" or "work for wages." We exclude these unclassified workers from the sample.

Second, we exclude individuals in the armed forces, who were unemployed, or who worked less than 30 weeks in the prior year (i.e. in 1939). If an individual was unemployed in 1940 (when the questions was asked), it is possible they also were unemployed for a time in 1939. Thus, any income that is reported from 1939 might not be representative of the amount they would have earned if they had been fully

[B.3] We show the robustness of our results to the inclusion of individuals that were on work relief in Appendix Table A.8.

employed.[B.4] In addition, individuals that worked less than 30 weeks in a year likely did not have a steady job.

Third, we exclude individuals who satisfy all of the above criteria, but for some reason do not have a reported income. This is the sample we use to estimate our results for weekly wages.

Other papers that examine weekly wages from the censuses use similar approaches to determine who is included in the sample. 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. Appendix Figure B.3 displays the cumulative distribution of both imputed income and wages for Black and White sons in our sample.

B.3 Controls

As mentioned in Section 2, the hookworm eradication campaign, malaria eradication campaign, and Rosenwald rural school initiative all occurred during the 1910s and 1920s.

To control for Rosenwald schools, we use the Rosenwald exposure measure from Aaronson and Mazumder (2011). This measure is the the share of each county's Black school-age population (7-17 year olds) that would have had a seat in a Rosenwald school during any given year averaged over the years a birth cohort was 7 to 13 years old.[B.5] See Aaronson and Mazumder (2011) for more details. We assign Rosenwald exposure to sons based on the county their father resided in during the second census we observe them in (either 1910 or 1920; the Rosenwald rural school initiative did not begin in earnest until the 1920s). These data are available for every county in the United States from 1919 through 1931; if a county did not have a Rosenwald school, they had zero Rosenwald teachers.

[B.4]We show the robustness of our results to the inclusion of unemployed workers in Appendix Table A.8.

[B.5]The number of seats is determined by the number of teachers; each teacher is assumed to teach 45 students. Aaronson and Mazumder (2011) define exposure over the years 7 to 13 because they cannot identify which schools built after 1926 were high schools.

To control for the malaria eradication campaign, we use county-level data on malaria ecology from Hong (2011). We assign malaria ecology data to sons based on the county their father resided in during the second census we observe them in (either 1910 or 1920; the malaria eradication campaign did not begin until the 1920s). Although the Hong (2011) data covers most of the United States, it is missing for about 9% of counties that we observe fathers in during the second census.[B.6] We impute missing malaria ecology using the average malaria ecology for other counties in the same tenth of a degree of latitude.[B.7]

To control for the hookworm eradication campaign, we use county-level data on hookworm infection rates from Thoman (2009). These data were originally reported by the Rockefeller Sanitary Commission (RSC) (Rockefeller Sanitary Commission, 1910-1914) and are only available for counties the RSC operated in. We assign hookworm infection rates to sons based on the county their father resided in during the 1910 census (since the RSC began to operate in 1910). This measure is missing for about 50% of counties in the Southern United States due to the RSC not operating in these counties.[B.8] For these counties, we impute missing hookworm rates using the average hookworm rate for other counties in the same tenth of a degree of latitude.[B.9] By 1910, some fathers and sons were living outside the South. Since hookworm was not prevalent outside the South, we assign a hookworm infection rate of 0 for counties outside of the South.

[B.6] It is missing for 211 of the 2457 counties.

[B.7] e.g. Houston County, AL is missing malaria ecology data and its latitude is 31.14. We impute malaria ecology for Houston County using the average malaria ecology for counties with latitudes between 31.10 and 31.20.

[B.8] It is missing for 530 of the 1096 counties in the South that we observe fathers in during 1910.

[B.9] e.g. Benton County, TN is missing hookworm infection data and its latitude is 36.03. We impute hookworm infection rates for Benton County using the average hookworm infection rate for other counties with latitudes between 36.00 and 36.10.

Figure B.1: Linking procedure example

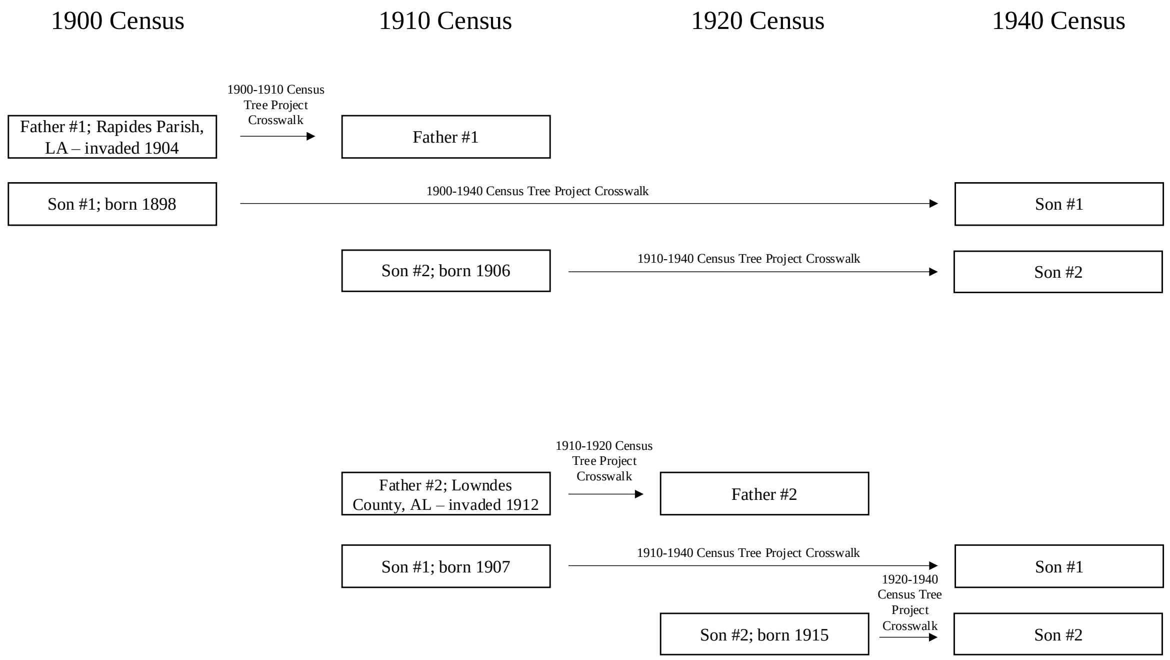
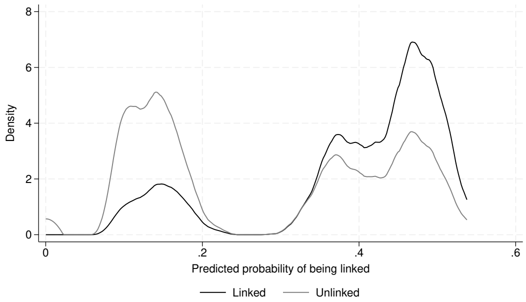
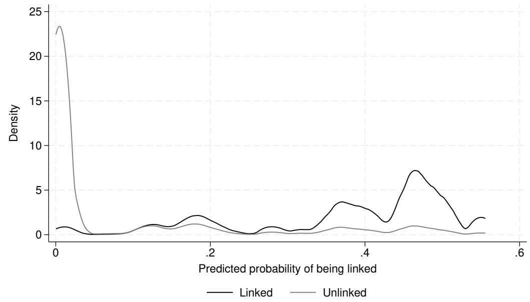


Figure B.2: Predicted probability of being linked

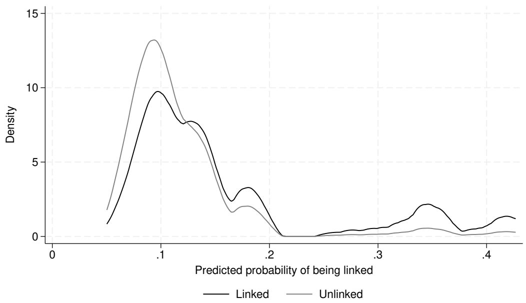
(a) 1900-1940



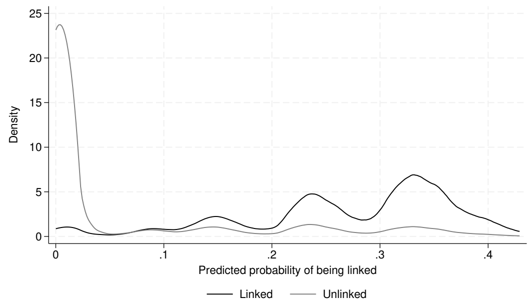
(b) 1910-1940; Father first observed in 1900



(c) 1910-1940; Father first observed in 1910



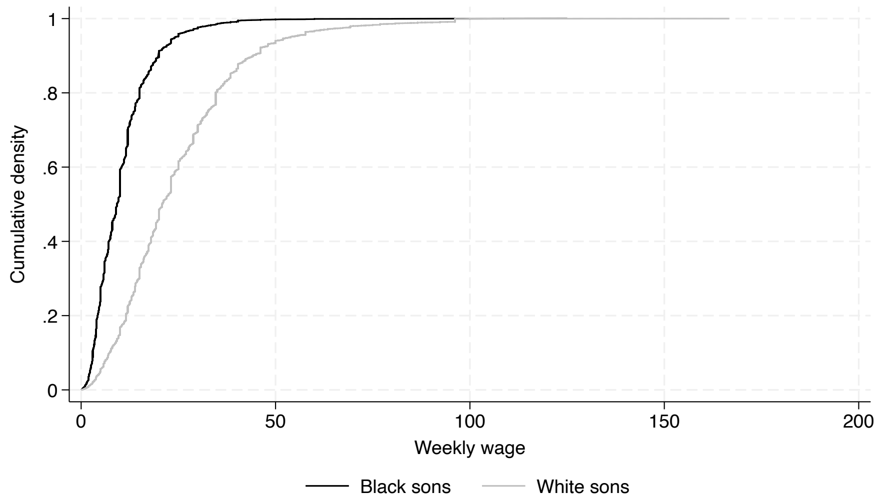
(d) 1920-1940



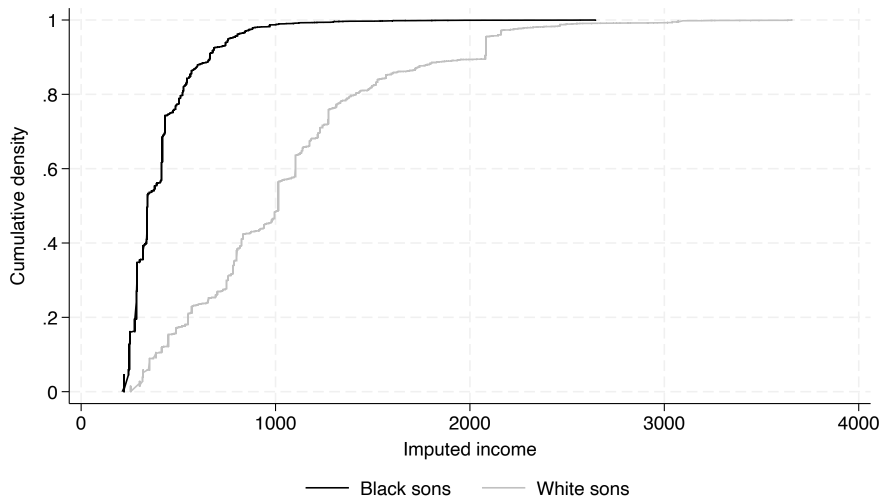
Notes: These figures present kernel density estimates of the predicated probability of being linked for both linked and unlinked individuals. The predicated probability of being linked was obtained from a probit regression where the dependent variable was a dummy if the individual was linked and the independent variables were: a Black indicator, indiactor variables for each age and their interaction with the Black indicator, state of residence indicators and their interaction with the Black indicator, and farm status (living on a farm) and its interaction with the Black indicator. In panel A, the linked sample of individuals from 1900-1940 are pooled with all individuals in the 1900 census who were male, under the age of 10, and living in the South census region. In panel B, the linked sample of individuals from 1910-1940 (fathers first observed in 1900) are pooled with all individuals in the 1910 census who were male and under the age of 10. In panel C, the linked sample of individuals from 1910-1940 (fathers first observed in 1910) are pooled with all individuals in the 1910 census who were male, under the age of 10, and living in the South census region. In panel D, the linked sample of individuals from 1920-1940 are pooled with all individuals in the 1920 census who were male and under the age of 10. The kernel density estimates use an Epanechnikov kernel function with a bandwidth of 0.01.

Figure B.3: CDF of weekly wages and imputed income by race

(a) Weekly wage



(b) Imputed income



Notes: Panel A shows a CDF of weekly wages using our wage worker sample. Panel B shows a CDF of imputed income using the imputed income sample.

Table B.1: The boll weevil and family structure

	Father absent (1)	Mother absent (2)	Both absent (3)
	Panel A: DiD for Black sons		
Born post BW	-0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)
Observations	1546918	1546918	1546918
Mean of dep. var.	.195	.098	.068
	Panel B: DiD for White sons		
Born post BW	-0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.001)
Observations	2791648	2791648	2791648
Mean of dep. var.	.062	.039	.017
	Panel C: Triple difference		
Born post BW * Black	0.001	0.001	-0.001
	(0.002)	(0.001)	(0.001)
Observations	4338566	4338566	4338566
Mean of dep. var.	.109	.06	.035

Notes: The unit of observation is a child in the 1900, 1910, or 1920 census. Children born within 10 years (±) of the boll weevil's arrival in their county are included in the sample. Children living in a county not invaded by the boll weevil are not included in the sample. Panels A and B include county and birth year fixed effects. Panel C includes county-by-race birth year-by-race and county-by-birth year fixed effects. Standard errors are clustered at the county level.
$* = p < 0.10$; $** = p < 0.05$; $*** = p < 0.01$

Table B.2: Comparison of linked to unlinked individuals - main sample

	Population	Linked unweighted	Linked weighted
	Panel A: 1900-1940 linked sample		
Age	4.430	4.275	4.484
Black	0.372	0.159	0.471
Live on farm	0.591	0.670	0.551
Observations	2819027	901980	901980
	Panel B: 1910-1940 linked sample; Father first observed in 1900		
Age	4.356	4.280	4.202
Black	0.124	0.158	0.111
Live on farm	0.374	0.648	0.239
Observations	10274579	1129887	1129887
	Panel C: 1910-1940 linked sample; Father first observed in 1910		
Age	4.364	3.997	4.402
Black	0.337	0.241	0.351
Live on farm	0.588	0.530	0.593
Observations	3300836	408920	408920
	Panel D: 1920-1940 linked sample		
Age	4.463	4.249	4.510
Black	0.105	0.152	0.093
Live on farm	0.345	0.657	0.224
Observations	11572158	874558	874558

Notes: This table presents a comparison of means between individuals that are part of the relevant population and individuals in our linked sample. Panel A compares men linked from 1900 to 1940 with the population of men in the 1900 complete count census who were under the age of 10 and living in the South census region. Panel B compares men linked from 1910 to 1940 (father first observed in 1900) with the population of men in the 1910 complete count census who were under the age of 10. Panel C compares men linked from 1910 to 1940 (father first observed in 1910) with the population of men in the 1910 complete count census who were under the age of 10 and living in the South census region. Panel D compares men linked from 1902 to 1940 with the population of men in the 1920 complete count census who were under the age of 10. Column (1) shows the mean for the population, column (2) shows the mean for linked individuals, and column (3) re-weights the mean for linked individuals using inverse propensity score weights so they appear more representative of the population. The weight applied to individuals is given by the formula: $(\frac{1-\hat{p}}{\hat{p}}) * (\frac{1}{1-q})$, where $\hat{p}$ is the predicated probability that an individual is linked and q is the share of the population that is linked. In Panel A, 32% of the population is linked. In Panel B, 11% of the population is linked. In Panel C, 12% of the population is linked. In Panel D, X% of the population is linked.

Table B.3: Wage worker sample restrictions

	Black sons (1)	White sons (2)	Total (3)
Born within 10 years of boll weevil arrival	31,479	397,505	428,984
Minus individuals with no occupation	29,022	375,701	404,723
Minus individuals not in labor force or on work relief (**Imputed income sample**)	27,077	352,319	379,396
Minus self employed and unclassified workers	19,873	239,483	259,356
Minus Armed forces, unemployed, or worked less than 30 weeks	14,363	190,181	204,544
Minus no income reported (**Wage worker sample**)	11,155	171,064	182,219

Notes: This table shows the sample sizes for the imputed income and wage worker samples. It also shows the change in the number of observations for various restrictions we make.