Abstract

This paper studies the long-run impact of racially oppressive institutions, finding that Black Americans’ socioeconomic status today is lower than that of white Americans in large part due to the US’s history of slavery and Jim Crow. We overcome the challenge of measuring each Black family’s historical exposure to slavery and Jim Crow by tracing their records from 1850 to 2000. First, we document that Black families who were enslaved until 1865 continue have considerably lower education, income, and wealth. Second, we show this persistence is entirely driven by post-slavery oppression under Jim Crow. We use a regression discontinuity design that compares the outcomes of families who were freed across state borders with more or less stringent Jim Crow laws, finding that states with more oppressive regimes sharply reduced Black economic progress in the long run. Using quasi-experimental variation in access to schools, we show that the limited access to human capital under Jim Crow was key in perpetuating racial disparities after slavery.

*We thank Leah Boustan, Ellora Derenoncourt, Jeremiah Dittmar, Jonathon Hazell, Ethan Ilzetzki, Ilyana Kuziemko, Camille Landais, David Lee, Ben Moll, Suresh Naidu, Steve Redding, Ricardo Reis, Maarten de Ridder, Tianyi Wang, and Zach Ward for insightful comments. We thank all seminar and conference participants for sharing their suggestions. Tre’ McMillan, Cynthia Nwankwo, and Bracklinn Williams provided excellent research assistance. This work was supported by Princeton’s Program for Research on Inequality.
†Department of Economics, Princeton University. lalthoff@princeton.edu
‡Department of Economics, London School of Economics. h.a.reichardt@lse.ac.uk
1. Introduction

The socioeconomic gaps between Black and white Americans are one of the most persistent features of US society. While racial disparities have narrowed considerably over the past two centuries, the progress has been far from perfect. For example, the average wealth of Black Americans remains below 20 percent of that owned by white Americans today (Derenoncourt et al., 2022).

One explanation for the lower socioeconomic status of Black Americans is the US’s particular history of institutionalized racial oppression. Throughout America’s colonial history, slavery was practiced—until around 1800 in the North and until 1865 in the South of the US. However, slavery was not the end of institutionalized oppression. Soon after the Civil War (1861–1865) led to the abolition of slavery in the South, state governments there started passing a mounting number of Jim Crow laws. Jim Crow (1877–1964) was an institution designed to limit the economic progress of the newly freed Black people and comprised laws of racial segregation (e.g., of schools and public transport), disenfranchisement (e.g., literacy requirements and poll taxes), and restricted geographic mobility (e.g., vagrancy laws and enticement laws). After almost 100 years, the Civil Rights legislation of the 1960s outlawed racial discrimination and ended Jim Crow, making it “one of the most significant legislative achievements in American history” (U.S. Senate, 2019).

This paper studies whether and to what extent Black families’ historical exposure to slavery and Jim Crow continues to shape US racial inequality today. In sum, our results show that the economic progress a Black family made after slavery critically depended on the location in which they were freed. The majority of families who were enslaved until the Civil War were freed in the most southern states. After slavery ended, those states implemented the most severe forms of Jim Crow institutions. Had it not been for Jim Crow, Black economic progress after slavery would have been substantially faster. With its critical importance for intergenerational mobility, the denial of equal access to education through Jim Crow was a key factor that made the institution so detrimental to Black economic progress in the long run.

We overcome the challenge of measuring a family’s historical exposure to slavery and Jim Crow by tracing their records from 1850 to 2000 using automated record-linking methods (Abramitzky et al., 2019). To measure a family’s exposure to slavery, we leverage the fact that the 1850 and 1860 censuses did not record enslaved people. We identify

1Throughout this paper, we use the term “Jim Crow” to refer to laws that limited Black Americans’ civil rights. Extralegal factors—such as lynchings or employer discrimination that often went above and beyond the letter of the law—do not fall under our definition of Jim Crow. In terms of measurement, we focus on state laws, ignoring less prevalent instances of local Jim Crow-like ordinances.

2Using their name, year and place of birth, and race, we follow individuals across full-count decennial census records from 1850 to 1940. We use information on the family interrelationships for individuals in the same household from the census to build family trees based on those linked records. We also link families to administrative records that allow us to measure socioeconomic outcomes at the neighborhood level.
families who were freed before the Civil War as those having ancestors recorded in the
1850 or 1860 census; others are classified as enslaved until 1865. We validate this method
using a name-based approach to determine how likely a family was to have been en-
slaved until 1865 (Ager et al., 2021). To measure a family’s exposure to Jim Crow, we use
that record linkage allows us to observe where a family was freed from slavery. Where a
family was freed reflects their exposure to state-level Jim Crow institutions over the sub-
sequent 75 years well because geographic mobility was low before 1940, especially for
those under intense Jim Crow regimes. We measure a state’s Jim Crow intensity using
both a composite index of state-level racial oppression and measures based on a newly
constructed dataset on over 900 Jim Crow laws.

Being enslaved until 1865 and being exposed to intense Jim Crow regimes after slavery
often went hand in hand. Because of the rapid southern-expansion of the US plantation
economy, the longer a family was enslaved, the more likely they were to be concentrated
in the most southern states. Those states would soon after slavery become the epicenter
of Jim Crow. We divide our sample in two: Black families who had higher exposure to
both slavery and Jim Crow because they were enslaved until 1865; and families who had
less exposure to both slavery and Jim Crow because they had been freed an average of 50
years earlier.

We document that Black families who were enslaved until 1865 have considerably
lower education, income, and wealth today than Black families freed before the Civil
War. Those within-race gaps have magnitudes equal to almost half of the corresponding
Black-white gaps. To understand the institutional forces that shape these gaps, we de-
compose them into variation within and across states. If the main driver is slavery—an
institution that transcended state borders within the South—then large gaps within states
would persist for families enslaved around 50 years longer. If the state-led institution
of Jim Crow was the main driver, then a family’s own history of enslavement should matter
less. We find that while slavery drastically lowered the socioeconomic status of Black
families, Jim Crow was the key force that perpetuated this disparity. Conditional on liv-

3Linking the historical records of women remains difficult, allowing us to follow only the paternal line
of ancestry. We estimate that intermarriage between families freed before 1865 and families freed in 1865
likely attenuates our estimates of the socioeconomic gaps between them in 1940 by one-third.

4This approach leverages changes in the distribution of last names in the census from 1860 to 1870—
before and after the inclusion of the newly freed Black families—to assign a probability of having been
enslaved until 1865 to each last name. For example, the last name “Freedman” did not exist in 1860 but
was chosen by many families who were newly freed in 1865. In contrast, the last name “Du Bois” became
10 times less frequent in the census after the formerly Enslaved were included in 1870.

5As a family’s location of freedom, we use their ancestor’s state of birth or county of residence as ob-
served in the 1870 census. We only use this information for families who were enslaved until 1865.

6Mobility was low due to both institutional factors—such as Jim Crow laws that limited labor mobil-
ity (Roback, 1984)—and economic factors—such as high migration costs (Carrington et al., 1996) or the
elusiveness of opportunities in potential destinations (Akbar et al., 2020; Derenoncourt, 2022).

7This composite index is based on a state’s population share that was enslaved in 1860; its share of share-
croppers who were Black in 1930; its number of disfranchisement devices; and its share of congressional
delegates that signed the Southern Manifesto (Baker, 2022).
ing in the same state, the initially large socioeconomic gaps faced by Black families who were enslaved an average of 50 years longer vanished by 1940. Across states, however, Black families freed at the same time fared very differently. Those state differences in Black economic progress fully explain the persistently low socioeconomic status of Black families with high exposure to pre-Civil Rights oppression.

To identify the effect of Jim Crow, we use a regression discontinuity design that compares the socioeconomic outcomes of Black families who were freed across state borders with more or less stringent Jim Crow laws. We validate our empirical strategy by showing that our estimates are consistent with the beginning of Jim Crow in 1877, consistent across space with the intensity of Jim Crow institutions, and consistent with the race-specific nature of Jim Crow to not affect white Americans directly. By focusing on border counties, we isolate the role of institutions from factors that transcend borders, such as culture, climate, or the economy. Our main identifying assumption is that an enslaved person’s birthplace is exogenous to future generations’ potential socioeconomic outcomes. This assumption is supported by historical evidence. Specifically, enslaved people had no freedom of movement before 1865, leaving no room for self-selection into location. Selection could only have occurred through forced migration, to which slaveholder migration and the domestic slave trade equally contributed. Slaveholders were generally non-selective in moving all their enslaved people with them (Fogel and Engerman, 1974; Pritchett, 2001; Tadman, 2008; Pritchett, 2019). Selective slave trade is only evident in the small sugar cultivation areas. By the nature of the work required, enslaved people there tended to be physically stronger and more likely to be male (Phillips, 1918). While having able-bodied men to clear the new highly fertile land is the kind of selection that slave owners cared about, it is not the kind of selection that would influence human capital 100 years later.

We find that Black families who were freed in states with more oppressive regimes experienced sharply lower rates of economic progress starting in the Jim Crow era (1877–1964). The resulting differences in socioeconomic status are increasing in the differences in Jim Crow intensity across a border. For example, consistent with Louisiana’s Jim Crow legislation being far more extensive than Texas’s, we find that families freed in Louisiana attained 1.2 years less in education by 1940 than families freed only a few miles away in Texas. Analyzing the content of specific laws, we find that the largest number pertain to education, restricting Black Americans’ access to schools. Our analysis of the content of these laws motivates our exploration of access to human capital as a key mechanism in

---

8 We limit our sample to families who were enslaved until 1865.
9 In principle, selection could also arise through differences in the slaveholders who choose to migrate. However, for selection to arise, the slaveholder’s decision would need to be correlated with the potential outcomes of their enslaved people—a scenario we cannot rule out but deem unlikely.
10 Sugar cultivation accounted for only 6 percent of the rural enslaved population (Tadman, 1977, 1979).
11 Contrary to the most plausible scenario for positive selection, we find that families freed in the Louisiana sugar areas achieved lower socioeconomic status by 1940 than families freed in other areas.
the persistent effect of Jim Crow.

To understand how access to human capital mediated the effect of Jim Crow on outcomes in the long run, we leverage a natural experiment in access to schools for Black children in the early 20th century. Specifically, we compare the education of children depending on whether their ancestors were freed in a county that would receive one of 5,000 schools built by the Rosenwald program (1914–1931) by the time they were in school age (Aaronson and Mazumder, 2011). We find that the Rosenwald school program had persistent positive effects on the economic progress of Black families, especially in the most oppressive states. A Black child who could attend a Rosenwald school attained 0.3 years more education than a child who could not attend. In the most oppressive states, the effect was more than twice as large. We find that school access not only increased the education of the benefiting individuals but also improved the economic conditions of their children today.

This paper contributes to our understanding of whether and how historical institutions affect economic outcomes in the long run. Acemoglu et al. (2002), Dell (2010), Donaldson (2018), and Dell and Olken (2019) show that regions can be lastingly transformed by temporary policies and institutions. In this paper, we develop innovative methods to study the impact of institutions on individual families, rather than on regions, and apply them in the context of US historical racial oppression. Such individual-level evidence allows for the geographic mobility of families (attenuating regional differences) and can generate novel insights into the mechanisms that drive various forms of persistence. For example, we leverage quasi-experimental variation in how likely a Black family was have access to schools to show that one of Jim Crow’s key mechanisms was to limit their access to human capital.12

This paper also enhances our understanding of geographic disparities in intergenerational mobility. Upward mobility has been lower in the South than in any other region, both historically (Olivetti and Paserman, 2015) and in recent decades (Chetty et al., 2014). Consistent with this evidence, our results suggest that Black economic progress after slavery was slowest in the South. Chetty and Hendren (2018) analyze children whose families move across places to show that locations caused low upward mobility in the South rather than being a result of selection. Our empirical strategy relies on the immobility of a specific population rather than their mobility, and the results confirm the importance of exposure to location-specific factors for the intergenerational accumulation of human and physical capital. Indeed, a state’s capacity to generate upward mobility is highly persistent: our estimates have a strong correlation ($0.617, p = 0.000$) with the causal effect that states have on the intergenerational mobility of low-income groups in recent decades.

---

12 This finding is consistent with evidence from development economics showing that the provision of educational resources is most effective in areas where, before the intervention, children were most deprived of opportunities to access such resources (e.g., see Duflo, 2001; Glewwe and Kremer, 2006).
(Chetty and Hendren, 2018). The high persistence of a location’s capacity to generate upward mobility is consistent with persisting cultural, economic, and institutional factors that set boundaries to economic opportunity.

This paper further contributes to the evidence on the long-run effects that oppressive institutions can have on racial inequality. Regions—within and outside of the US—that relied on slave labor continue to have lower and more unequally distributed incomes (Nunn, 2008), lower upward mobility (Berger, 2018), larger racial disparities (Bertocchi and Dimico, 2014), and higher levels of racial resentment against Black Americans (Acharya et al., 2018). Evidence of how racially oppressive institutions affect individual families in the long run are scarcer. Sacerdote (2005) uses Southern place of birth as a proxy for being enslaved until 1865 and shows that Black descendants of this group continued to have lower socioeconomic status than those who were freed earlier. By combining linked records, exogenous variation in ancestor location, and new details on state institutions, we assess why Black families whose ancestors were enslaved until 1865 still experience lower socioeconomic outcomes. We show that after 1940, the single reason why those differences persist is the fact that the most severe Jim Crow regimes arose in the states where Black families enslaved until 1865 were concentrated.

Lastly, this paper contributes to a historical literature on the evolution of Black economic progress after slavery. Margo (1991) argues that beyond market forces such as the supply of and demand for educated Black workers, one of the main reasons for persistence of educational gaps between Black and white children was the barriers faced by Black parents, both in the labor market and in obtaining literacy. Our results show that this “intergenerational drag” of slavery was quantitatively important for around three generations. We extend Margo’s model of Black economic progress to encompass the racially oppressive institutions after slavery which we show to be decisive in the long run. The dependence of Black economic progress on institutional factors is consistent with the seminal work of Du Bois (1935), Woodward (1955), and Ransom and Sutch (2001), who highlight that when and where their environment allowed for it, Black families did make rapid progress—such as in the Reconstruction era (1865–1877). Consistent with that, the Rosenwald program being particularly effective in the most oppressive states suggests that it was not a lack of demand for education among Black children there but rather a lack of access to education that slowed their human capital accumulation (see also Aaronson and Mazumder, 2011).

Those locations also have lower productivity (Mitchener and McLean, 2003). Fujiwara et al. (2019) find that in Brazil, a location’s past reliance on slavery caused weaker institutions and higher inequality until today.
2. Historical Context

In this section, we provide some historical context for the evolution of racially oppressive institutions in the US—from slavery to Jim Crow and beyond.

2.1 Free Black Americans before 1865

In 1860, just before the Civil War (1861–1865) that led to the abolition of slavery, 4 million enslaved and 0.4 million free Black people lived in America. Enslaved people existed on American soil from the country’s colonial origins in the 16th century. The roots of the free Black population trace back to 1619, when the first 20 Black people were purchased by settlers in Virginia. Little is known about their fate, but it is likely that at least some of them were treated as servants who had to work for a fixed term and gained freedom after (Frazier, 1949). Around 1660, both law and practice had changed, implying that virtually all Black individuals who arrived in the colonies were enslaved for life. From 1662 onwards, the law also mandated that a child would inherit their legal (i.e., free or enslaved) status from their mother regardless of race.

For those enslaved, the Revolutionary War (1775–1783) provided a road to freedom. Responding to a need for troops and laborers, the British governor promised freedom to all enslaved willing and able to serve the British. It is estimated that up to 100,000 enslaved people ran away from plantations to do so (Schama, 2006). After the war, many of them remained in the US as free persons. As a result, the free Black population in some states increased dramatically.

The Revolutionary War also brought with it a spirit of egalitarianism, challenging the institution of slavery in some regions. In the North, the abolitionist movement spread quickly after the war. While only few Black people lived free of slavery before the Revolutionary War, most Northern states adopted gradual Emancipation laws after the war. New Jersey was the last Northern state to do so in 1804.

In the South, the path to freedom was narrow, especially in the Lower South. All Southern states except North Carolina allowed masters to free (“manumit”) their slaves by 1790, but the practice was employed to different degrees across regions. In the Upper South, the first wave of manumissions occurred between 1783 and 1793, the first decade after the Revolutionary War. Motivated by anti-slavery beliefs, the vast majority of manumitters freed all their slaves at once. However, manumission gradually became more selective and turned into a reward system designed to uphold slavery (Wolf, 2006). By 1860, 0.2 million of the 1.8 million Black Americans in the Upper South were free (11.1

---

14The Lower South comprises Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, South Carolina, and Texas. The Upper South comprises Delaware, Washington, D.C., Kentucky, Maryland, Missouri, North Carolina, Tennessee, Virginia, and West Virginia. The North comprises all other states.
percent). The Lower South did not see a similar manumission wave after the war as manumissions there were usually limited to masters’ “illicit offspring, special favorites, or least productive slaves” (Berlin, 1974). The free Black population of the Lower South mostly originated from refugees who fled from the French colony Saint-Domingue (now Haiti) and the purchase of Louisiana from France that had a sizable free Black population. By 1860, 40,000 of the 2.5 million Black Americans in the Lower South were free (1.6 percent).

The legal and socioeconomic status of the Free varied greatly across locations and over time before 1865. In most states, free Black Americans were deprived of the right to vote and to hold political office. But their legally protected property rights were respected in most cases. With the limited freedom they enjoyed, some free Black families could accumulate modest wealth and social status. Most of them, however, lived in poverty “under conditions barely distinguishable from those of the mass of slaves” (Berlin, 1974). Their economic status varied greatly across the country and, perhaps surprisingly, tended to be better further South (Berlin, 1976). In the North, free Black families were concentrated in cities where they suffered from competition with and hostility from white laborers (Frazier, 1949). Most free Black families in the South lived in rural areas, working as farmhands and casual laborers (Berlin, 1974).

2.2 Freedom of All Black Americans after 1865

By the beginning of the Civil War (1861–1865), the enslaved population was concentrated in the Lower South (see Figure 1). The free Black population, in contrast, was concentrated in the North and the Upper South. These differences in geographic location exposed them to different institutional regimes after slavery.

The Civil War led to the emancipation of enslaved families, giving all Black Americans the same legal status. The average free Black family had likely already been free for around 50 years. For the first 12 years after the Civil War—the Reconstruction era (1865–1877)—the Union Army occupied the South. Reconstruction allowed Black Americans to experience unprecedented socioeconomic progress. New schools and universities were built to educate Black Americans throughout the South. Black men participated politically, casting their votes in high numbers and serving in public office (Logan, 2020).

Throughout Reconstruction, Black economic and political progress was met with violent opposition from white Southerners (Du Bois, 1935; Foner, 1963). In 1877, the Union troops left the South, abandoning the project of Reconstruction. What followed was the disenfranchisement of Black people through informal and legal means, leading to massive declines in Black political participation (Kousser, 1974; Naidu, 2012). Many free

---

1550 years is an estimate derived from assessing free Black population counts and assuming constant population growth before 1865.
Black Americans lost their higher social status and left the South (Woodson, 1918).

Black Americans who remained in the South after Reconstruction faced increasing oppression through the rise of Jim Crow (1877–1964). Jim Crow laws governed almost every aspect of Black life. Schools, work places, public transport, medical facilities, and parks were racially segregated. Poll taxes, literacy tests, and other means limited Black suffrage (Naidu, 2012; Walton et al., 2012). Enticement laws, contract enforcement laws, and emigrant-agent laws prevented Black workers from seeking economic opportunities with new employers or in states outside the South (Roback, 1984). Vagrancy laws criminalized the unemployment of Black people (Blackmon, 2009). In addition to institutionalized segregation, various informal means of excluding Black Americans spread through the South and beyond.

From 1910 to 1940, many Black Americans started to leave the (Upper) South in the first wave of the Great Migration. Black families from the Lower South only participated in this migration in small numbers before 1940, both because Jim Crow limited their geographic mobility and because migration was more costly for them (Roback, 1984; Carrington et al., 1996). While the Civil Rights Movement successfully fought oppression starting in the mid-1950s, the Great Migration continued until the end of the movement in the late 1960s. By then, six million Black Americans had left the South. However, opportunities in the North proved elusive to Black families (Akbar et al., 2020; Derenoncourt, 2022). In addition, even after the achievements of the 1960s, old forms of racial oppression have persisted and new forms—such as mass incarceration and “color-blind”
voter suppression—have arisen since (Western, 2006; Alexander, 2010; Darity et al., 2016). Progress in narrowing racial gaps in socioeconomic status have largely stalled since the 1960s (Bayer and Charles, 2018; Althoff, 2021; Derenoncourt et al., 2022).

3. Data and a New Method to Measure a Family’s Exposure to Slavery and Jim Crow

A key empirical challenge we overcome in this paper is to measure a Black family’s exposure to slavery and Jim Crow. We construct family histories for Black Americans in the census between 1850 to 2000 and develop new methods to measure the two key components of a family’s exposure to pre-Civil Rights oppression: how long a family was enslaved and where they were freed, determining the intensity of the Jim Crow regime they likely lived under.

3.1 Measuring How Long a Family Was Enslaved

To measure how long a family was enslaved, we leverage that the 1850 and 1860 censuses did not record enslaved people.\textsuperscript{16}

\textbf{Main method.} We identify Black Americans who were free before 1865 (“the Free”) as those whose were (1) recorded in the 1850 or 1860 census or (2) born in a state that had already abolished slavery; others are classified as enslaved until 1865 (“the Enslaved”).\textsuperscript{17} We then carry this information forward to their descendants. To do so, we build family trees using information on family interrelationships for members of the same household from the census and by linking individuals’ census and administrative records from 1850 and 2000.

This classification strategy identifies high precision whether a Black family’s ancestors were enslaved until 1865 or not. In principle, if a family cannot be linked back to the 1850 or 1860 census, this could either mean that they were enslaved until 1865 or that they could not be linked using automated methods—for example because their name was misspelled in one census. Hence, in the South, we inevitably mis-classify some Black families who were free before 1865. However, census records show that only 6 percent of the Southern Black population were free in 1860, many of whom we correctly classify as such. As a result, we compare a group that was free with almost certainty in 1860 with a

\textsuperscript{16}These are the only pre-1865 census decades with individual-level data.

\textsuperscript{17}We refer to Black families who were free before 1865 as “the Free” even though they or their ancestors may have been enslaved at some point earlier than 1865. We refer to those enslaved until 1865 as “the (formerly) Enslaved.” We choose this terminology to avoid confusion engendered by the sometimes-used terms “Freemen” (Free) and “Freedmen” (formerly Enslaved). We avoid the term “slave” and capitalize “Free” and “Enslaved” when used as nouns to be respectful of the people we study.
group of which at least 94 percent were enslaved. The potential for attenuation bias due to imperfect linking rates is therefore minimal.

Our classification method has two important advantages over previous research, which typically relied on birthplaces to identify how long a family was likely enslaved. First, because the census only provides information on birthplaces for a person and their parents, the intergenerational effects of slavery beyond the second generation cannot be studied in the census cross-section. Our panels allow us to follow families until today.\footnote{This extension also lets us use the rich data on education, income, and wealth in the 1940 census.} Second and more importantly, relying on a person’s birthplace can only identify free Black families born in the North. However, 50 percent of all Black families freed before 1865 lived in the South. Our method correctly identifies a large number of those families. Measuring both how long a family was enslaved and where it was freed is crucial to determine what role slavery, Jim Crow, and their interaction play in shaping the persistent effects of pre-Civil Rights oppression.

Because women tended to change their last names upon marriage, automated linking methods only allow us to follow a family’s male ancestry line. An advantage of considering only male ancestry is that it avoids bias through selective marriage. A disadvantage is that we do not know which fraction of one’s ancestors were enslaved until 1865 and which fraction was freed earlier. Any amount of intermarriage will attenuate the group differences we measure towards zero.\footnote{In Appendix E, we use spouses’ state of birth (and their mothers’ state of birth) to show that while intermarriage was rare in absolute numbers due to the differences in the two groups’ geographic concentration, it likely attenuates our estimates of the socioeconomic gaps between them by one-third, making them appear smaller than they are in truth. This is because the population of free Black Americans is relatively small, so that even low absolute occurrences of intermarriage have a relatively large impact on this group.}

**Alternative method.** We develop a second strategy to identify descendants of the Free and the Enslaved based solely on last names. We use the change in the distribution over last names before 1865, when the census included only free Black Americans, to after 1865, when the it included all Black Americans. This approach allows us to use the full (rather than only the linked) sample of Black Americans in the census. The two approaches yield Free-Enslaved classifications that are highly correlated.\footnote{See Appendix Figure A.3.}

While some last names were common both among the Black families freed in 1865 or before, others were characteristic of one group. For example, the last name “Du Bois” was relatively frequent among free Black families in the 1860 census. However, with the inclusion of the families newly freed in 1865 in the 1870 census, Du Bois became 10 times less frequent—an indication that having this last name meant a person likely descended from the Free. In contrast, the last name “Freedman” did not exist in the 1860 census but appears in the 1870 census after some newly freed families chose it as their new last name. Thus, Black families called Freedman likely have ancestors who were enslaved.
This alternative classification method trades off accuracy in favor of coverage. Based on this probabilistic measure, we expect the resulting Free-Enslaved gaps to be subject to attenuation bias, making them appear smaller than they are in truth. In contrast to our main method, however, this classification can be applied to the entire population of Black Americans rather than being limited to those who we can link back in time.

This classification also allows us to identify descendants of the Enslaved in non-census data that include last names. We are collaborating with one of the main credit score providers in the US to extend our results to 2022.

3.2 Measuring the Exposure to State-Led Oppression During Jim Crow

Black families’ exposure to slavery and Jim Crow is highly correlated. Families who were enslaved until 1865 were also geographically concentrated in states that would become the epicenter of Jim Crow. In contrast, families who had been freed an average of 50 years earlier also were concentrated in states that would adopt less intensive Jim Crow regimes. These different geographic distributions are the result of the rapid South-expansion of the US plantation economy. The longer a family was enslaved, the more likely it was to be freed in the Lower South.

To measure a family’s exposure to Jim Crow, we use that record linkage allows us to observe the birthplace of its formerly enslaved ancestors. The state in which a family was freed is good proxy for exposure to state-level Jim Crow institutions over the subsequent 75 years because their geographic mobility across states was low before 1940, especially for those in the most oppressive Jim Crow states.

To use two different measures of a state’s Jim Crow intensity. Our first measure is a composite index of state-level racial oppression—the Historical Racial Regime (HRR) score. This score was developed as data-driven proxy of a state’s intensity of racial oppression from slavery to Jim Crow (Baker, 2022). The measure is a factor extracted from four components: a state’s population share that was enslaved in 1860; its share of sharecroppers who were Black in 1930; its number of disfranchisement devices; and its share of congressional delegates that signed the Southern Manifesto.

Our second measure is the number of Jim Crow laws that each state passed until 1950. For this measure, we collected data on over 900 Jim Crow laws. Specifically, we digitized the laws recorded in “States’ Laws on Race and Color” (Murray, 1950)—a resource aiming to document all state laws that regulated race and color in 1950. We classify each law as discriminatory, not discriminatory, or anti-discriminatory based on the legislative text and sometimes context provided by the author. We use discriminatory laws to proxy

---

21 In an effort independent of ours, Cook et al. (2022) have also collected this data from (Murray, 1950).
for Jim Crow laws. We also determine the domain which each law pertains to, including education, marriage, or public transport. We complete our dataset on Jim Crow laws by digitizing two important types of laws that were largely omitted by Murray (1950): Laws on employment from Roback (1984) and laws on suffrage from Walton et al. (2012).

3.3 Linked Data

We use full-count census data for all available decades between 1850 and 1940 (Ruggles et al., 2020) and link observations across adjacent and non-adjacent decades using the automated linking methodology provided by Abramitzky et al. (2020). A person is linked from one census to another if their name, year of birth, and state of birth match and, importantly, if the match is unique conditional on race. We use a method that allows for misspellings by matching names based on their phonetic sound (NYSIIS). Because women tend to change their last name upon marriage, only men can be linked over time.

The census also contains information on the relationship between individuals in the same household. Most importantly, by observing a person in their parents’ household during childhood, we can build family trees based on this information. These family trees allow us to study the evolution of a family’s social, economic, and geographic mobility across generations. Our main outcomes include education, income, and wealth. Over time, the census data tend to provide increasingly rich information on those outcomes. We study families’ outcomes in census records between 1870 (the first census to include all Black Americans) and 1940 (currently the most recent full-count census available), before focusing particular attention on the rich information included in the 1940 census.

To extend our analysis to the 21st century, we link the 1940 census to administrative death records between 1988 and 2005 (Goldstein et al., 2021). These records cover the near-universe of deaths among American citizens and contain the nine-digit ZIP code of the decedent’s residence at their time of death. Nine-digit ZIP codes are highly granular indicators of location, which refer to a “segment or one side of a street” (USPS, 2021), allowing us to obtain rich information on the socioeconomic characteristics of a person’s neighborhood.22 We use National Historical Geographic Information System (NHGIS) data on the distribution of education, income, and wealth by race within these areas to extend our analysis to the present day. We also examine broader measures of well-being such as health, mortality, residential segregation, evictions, and fatal police encounters. Last, we build a database of historical location characteristics including agriculture, demographics, lynchings, segregation, railroad networks, and transportation costs.23

To extend our results even further, we are collaborating with one of the main credit

---

22We map nine-digit ZIP codes into statistical areas, such as census blocks, which are small and designed to have socioeconomically homogeneous populations (Census Bureau, 2017). In Appendix A.8, we lay out the procedure to link nine-digit ZIP codes to statistical areas.

23We describe this dataset in detail in Appendices A.5–A.7.
score providers in the US. We use our name-based method to identify Black individuals whose ancestors were enslaved until 1865. This data allows us to extend our estimates of the degree to which Black Americans continue to be affected by their ancestors’ pre-Civil Rights oppression to the present day.

3.4 Sample

For our main analysis, we focus on Black men aged 20 to 54 years. For two reasons we also limit our sample to individuals who can be linked to their ancestors in 1880 or earlier. First, to identify a family who gained freedom before 1865 in a state that had not abolished slavery, it must be linkable to their ancestors in 1850 or 1860. Restricting the sample to Black Americans linkable to 1880 or earlier minimizes the bias that may result from comparing families who can be linked back in time easily (e.g., because they have unique names) with those who cannot. Second, this restriction excludes families who immigrated to the US after 1880 who might have experienced very different sets of oppressive institutions historically.

For 1940, the latest year available, our sample of Black prime-age men consists of 155,813 descendants of families who were enslaved until 1865 and 9,325 descendants of families who were freed before 1865. We achieve a linking rate of 10 percent from 1870 to 1940—an important benchmark because those links allow us to observe the state in which a Black family’s ancestors were freed from slavery via their birthplace in the 1870 census. Despite the imperfect linking rate, our sample is highly balanced on observables. For example, the literacy rate of formerly enslaved families in our linked sample matches that of the 1870 census population: 20.4%. For free Black families in our linked sample, literacy is very close to that of the 1860 census population: 65.1% compared to 66.8%. From the 1940 census to administrative records in 2000, we can link 21,059 descendants of enslaved families and 1,591 descendants of free Black families.

Potential Linking Bias. In constructing our main sample, we rely on linking families across census records. One may be concerned that linking procedures introduce mechanical differences in families who were enslaved until 1865 and those freed earlier. The most plausible concern is that a person’s socioeconomic status depends on how many generations or decades they can be linked backward in time.

To examine the quantitative importance of this concern, we group Black Americans in 1940 by the earliest decade in which we can link them back to one of their ancestors, and

\[\text{Linking rates are lower for Black than for white individuals. For example, Ager et al. (2021) achieve a linking rate of around 20 percent from the 1860 to either 1870 or 1900 for white men. Our average linking rates are 12 percent between adjacent census decades and 20 percent from one census to any other census.}\]

\[\text{See Appendix Table A.1 for a balance check for the two groups.}\]

\[\text{For example, children of single mothers typically cannot be linked to their grandparents in earlier censuses because existing methods exclude women.}\]
plot their average outcomes by group (see Figure 2). In 1870, Black families who were enslaved until 1865 are included in the census for the first time. Consistent with that change in sample composition, we observe a large drop in average income and education for people who can be linked to ancestors in 1870 but not 1860 or 1850. Aside from this drop, there are no trends in income or education, suggesting that individuals who can be linked back further do not have mechanically higher socioeconomic status. The evidence suggests that the group differences estimated here are not affected by such a mechanical bias. To err on the side of caution, we limit our sample to individuals who can be linked back to 1880 or earlier throughout this paper.

4. Socioeconomic Gaps between Descendants of Free and Enslaved Families

In this section, we document the gaps in education, income, and wealth from 1870 to 2000 between descendants of families who were enslaved until 1865 and those freed earlier. We then decompose this gap to assess the relative importance of state-specific factors and factors that transcended state borders.
4.1 Evolution of the Free-Enslaved Gap until 1940

We first estimate the Free-Enslaved gap (β_t) in socioeconomic outcomes (y_{i,t}) separately for each decade t from 1870 to 1940:

\[ y_{i,t} = \alpha_t + \beta_t s_i + \gamma' X_{i,t} + \varepsilon_{i,t}, \]  

(1)

where \(s_i\) is equal to one if person \(i\) is classified as a descendant of the Enslaved and is zero otherwise. \(X_{i,t}\) is a vector of controls that includes a quadratic term of age in our baseline specification. We cluster standard errors at the family level.\(^{27}\)

**Figure 3: Free-Enslaved Gap (1870–1940)**

\(<A>\) Literacy

\(<B>\) Occupational Skill

Notes: This figure shows the gaps in literacy and occupation skill (HISCLASS) among prime-age (20-54) male descendants of free vs. enslaved Black Americans in each census decade. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix A for details on the sample and data.

We find that the socioeconomic differences between descendants of the Free and the Enslaved are large and persistent. In 1870, the formerly Enslaved were 2 times (over 40 percentage points) more likely to be illiterate than free Black Americans (see Figure 3). By 1940, the gap was still 1.8 times (5 percentage points). Descendants of the Enslaved work in less skill-intensive occupations than descendants of the Free throughout 1870 to 1940. Consistent with this skill gap, descendants of the Enslaved earn substantially lower incomes and are significantly less likely to own their home.\(^{28}\)

The rich information on education, income, and wealth provided by the 1940 census allows us to get a detailed picture the Free-Enslaved gap 75 years after slavery ended. Consistent with our previous results, we find that descendants of the Enslaved are substantially less educated, earn lower incomes, and have accumulated less wealth than

---

\(^{27}\)We define a family to be the group of individuals with a common 1870 ancestor.

\(^{28}\)See Appendix Figure I.19.
descendants of free Black Americans (see Table 1). The gap in education amounts to 1.6 years—more than one-quarter of the average years of education among Black men in 1940.\textsuperscript{29} The likelihood that a descendant of the Enslaved earned a high school or college degree was only half compared to descendants of the Free; for graduate degrees it was one-third.\textsuperscript{30} Consistent with the educational gap, the income and wealth of the Enslaved are substantially lower.\textsuperscript{31}

\begin{table}[h]
\centering
\caption{Free-Enslaved Gap (1940)}
\begin{tabular}{lcccc}
\hline
 & Education (Years) & Wage Income (USD) & Homeownership (%) & House Value (USD) \\
 & Mean: 5.99 & Mean: 381.20 & Mean: 29.25 & Mean: 1,371.95 \\
\hline
Ancestor Enslaved & -1.59*** & -145.92*** & -7.24*** & -694.69*** \\
(0.05) & (6.13) & (0.62) & (65.85) \\
\hline
Controls (age, age\(^2\)) & Y & Y & Y & Y \\
Adjusted R\(^2\) & 0.04 & 0.05 & 0.01 & 0.01 \\
Observations & 163,549 & 154,463 & 164,357 & 46,971 \\
Ancestor Free & 9,078 & 8,551 & 9,070 & 3,227 \\
\hline
\end{tabular}
\end{table}

Notes: This table shows the gap in years of education, wage income, homeownership, and house value among prime-age (20–54) male descendants of free vs. enslaved Black Americans in 1940. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).

The narrowing of the Free-Enslaved gap from 1870 to 1940 is slow. To benchmark the speed of convergence, we estimate socioeconomic gaps between white families whose ancestors had no measurable physical or human capital in 1870 and all other white families between 1870 and 1940.\textsuperscript{32} In only 30 years, the gap in literacy between those two groups of white Americans rapidly shrunk from over 90 percentage points to less than 10 (from twice the Free-Enslaved gap in 1870 to half the Free-Enslaved gap in 1900). The homeownership gap for the two groups of white Americans was similar to the respective Free-Enslaved gap in 1870 but closed by 1900—while the Free-Enslaved changed very little until then. Thus, consistent with racial disparities in intergenerational mobility throughout US history (Collins and Wanamaker, 2022; Chetty et al., 2020), we find that the Free-Enslaved gap narrowed far more slowly over this period than comparable gaps among white Americans.

While our estimates are consistent with prior evidence, our estimation strategy allows

\textsuperscript{29}This gap is at the lower end of the range that Sacerdote (2005) estimates based on comparing Southern- and Northern-born Black men older than 54.

\textsuperscript{30}See Appendix Table I.8.

\textsuperscript{31}Appendix Table I.7 compares the Free-Enslaved gap across different measures of income. The relative gap in (imputed) total income is similar to the relative gap in years of education: 25 percent. In Appendix I.2, we document the Free-Enslaved gap in intergenerational mobility that has contributed to this persistence.

\textsuperscript{32}See Appendix Figure H.18.
for a broad reinterpretation of the results. The literacy gaps we estimate are comparable to those estimated by Sacerdote (2005). For 1880, he finds that Black people born in the South before 1865 are 47 percentage points less likely to be literate than those born in the North. As expected, our estimated Free-Enslaved gap of 40 percentage points in 1880 is somewhat smaller because our classification can correctly identify free Black Americans in the South. The Southern Free have lower literacy rates than free Black Americans in the North. Similarly, Sacerdote finds gaps of around 16 percentage points for descendants in 1920, whereas we find gaps of around 11 percentage points.

We validate the robustness of those estimates using our alternative measure of how likely a family was to be enslaved until 1865 based on the distribution of last names. Those estimates apply to the full census population, rather than the subset of linked families, confirming that large Free-Enslaved gaps persist until 1940. As expected, the results based on this probabilistic classification yields somewhat attenuated gaps but are still large and highly significant. The implied gaps correspond to 9 percent of the average education among Black prime-age men, 5 percent for income, 11 percent in homeownership rates, and 61 percent in house values. Without record linkage, we cannot assure that all Black families in the sample were present in the US during both slavery and Jim Crow. However, when re-weighting observations in 1940 to hold the distribution of last names constant at the 1870 level, the Free-Enslaved gap in years of education is estimated to be $-1.40^{**} (0.09)$, very close to the Free-Enslaved gap of $-1.59^{***} (0.04)$ based on our preferred classification.

We also conduct an array of placebo exercises that give additional support to our empirical strategy. First, we use 1875 as a placebo year of emancipation. Specifically, we classify Black families as descending from the Free or the Enslaved based on whether or not we can link them back to ancestors in 1870 (rather than 1860) or not. This placebo exercise yields no economically meaningful gaps. For example, a small gap of less than 1 percent in education emerges (compared to 25 percent in our baseline). Second, we use white Americans as a placebo group. Specifically, we divide white families into two groups depending on whether or not we can link them back to ancestors in the 1860 census. Again, this placebo exercise yields no economically (and rarely statistically) significant estimates. Lastly, adding all of those placebo groups as controls to our baseline specification leaves the overall patterns of persistence in the Free-Enslaved gap unaffected.

---

33See Appendix Table I.9.
34See Appendix Tables H.4 and H.5.
4.2 The Free-Enslaved Gap in the 21st Century

We extend our results beyond 1940 by link Black families from the 1940 census to administrative death records around 2000. Those records include the nine-digit ZIP code of a person’s residence at the time of death from which we derive neighborhood-level information on the distribution of income, wealth, and education. It is reasonable to suspect that the narrowing of the Free-Enslaved gap accelerated over the course of the Civil Rights Movement (1954–1968). Existing evidence suggests that the social mobility of Black Americans temporarily increased around 1970 (Clark, 2014).

<table>
<thead>
<tr>
<th>Table 2: Free-Enslaved Gap (2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ancestor Enslaved</strong></td>
</tr>
<tr>
<td>Mean: 68.85</td>
</tr>
<tr>
<td><strong>-3.02</strong>*</td>
</tr>
<tr>
<td>(0.51)</td>
</tr>
<tr>
<td><strong>Level of outcome</strong></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
</tr>
<tr>
<td><strong>Ancestor Free</strong></td>
</tr>
</tbody>
</table>

Notes: This table shows the gap in neighborhood-level outcomes (high school and college degree, median income, and median house value) among descendants of free vs. enslaved Black Americans in 2000. A neighborhood is defined as a census tract. Each person is assigned the respective value of the census tract in which they last lived according to administrative death records. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

We find that in 2000, descendants of the Enslaved reside in neighborhoods that have substantially lower education, income, and wealth than those of the Free’s descendants (see Table 2). Descendants of the Enslaved are 3.9 percentage points less likely to hold a high school degree and 2.6 percentage points less likely to hold a college degree. Their expected income is lower by $5,100—around 17 percent of the Black median income. Conditional on owning their home, their houses are worth around $17,500 less—around 19 percent of the median house value among Black Americans. We also show that descendants of the Enslaved live in neighborhoods associated with one year less in life expectancy and with a 16 percent higher chance of a fatal police encounter. Because those estimates ignore within-neighborhood differences, they should be seen as an underestimate of the true Free-Enslaved gap today.

---

35Our sample primarily covers the 1910–1940 birth cohorts.
36Clark (2014) uses the last names of “underclass” Black Americans to study their representation among licensed physicians and attorneys, two “elite” occupations.
37See Appendix Table I.13.
These Free-Enslaved gaps are equal in magnitude to almost half of the corresponding Black-white gaps today. This comparison yields a lower bound for the importance of historical racial oppression for modern racial disparities. Both groups, the Free and the Enslaved, were affected by both slavery and Jim Crow. It is the mere difference in intensity of their experiences that yields socioeconomic gaps of such large magnitude. Next, we turn to the drivers of this persistence.

4.3 A Simple Model of Black Economic Progress after Slavery

We propose a simple model to guide the interpretation of the Free-Enslaved gap and to analyze the channels that drive its persistence. The framework incorporates intergenerational mobility, the effects of exposure to location-specific factors, (potentially selective) migration, and the effect of delayed freedom. We discuss the following questions: Is the persisting disadvantage faced by descendants of the Enslaved a causal effect of slavery or Jim Crow? Could selection into being a free Black American drive a large part of the gap? What role did the differential exposure to location-specific factors among the Enslaved and the Free have in shaping the gap?

Let \( y_{i,t} \) denote the human capital—or any other outcome of interest—for family \( i \) at time \( t \). For simplicity, let there be two time periods, \( t \in \{0, 1\} \); the model is easily extendable to more time periods. We think of \( t = 0 \) as reflecting 1865, the year of Emancipation, and \( t = 1 \) as reflecting 1940, the last census year to which we can link families. We model \( y_{i,t} \) to be determined by

\[
y_{i,t} = \alpha_{i,t} + \gamma_{l(i,t)} + \rho_{l(i,t-1)}y_{i,t-1} + \epsilon_{i,t},
\]

such that it depends on five factors: a factor capturing innate ability \( \alpha_{i,t} \) with c.d.f. \( F(\cdot) \), the family’s previous human capital \( y_{i,t-1} \), their location \( l(i,t) \in L \), and a random error term \( \epsilon_{i,t} \) that satisfies \( \mathbb{E}[\epsilon_{i,t} | s_i, \alpha_{i,t}, l(i,t)] = 0 \). Last, we define \( \gamma_{l}^1 \) as the effect of being exposed to factors specific to location \( l \) at time \( t \). The mobility parameter \( \rho_{l} \) can vary across locations. We model \( y_{i,0} \) (the starting condition) as

\[
y_{i,0} = \alpha_{i,0} + \gamma_{l(i,0)}^0 - \delta s_i + \epsilon_{i,0},
\]

where \( s_i \) is an indicator for whether the family was enslaved until 1865. That is, in 1865, the outcomes depend on ability, location, and whether a person had been free before the Civil War. The parameter \( \delta \geq 0 \) may capture any direct advantage that free Black Americans had relative to the enslaved population, such as access to education during slavery. At time \( t = 1 \), the outcomes then become

\[
y_{i,1} = \left( \lambda + \rho_{l(i,0)} \right) \alpha_{i,0} + \rho_{l(i,0)} \gamma_{l(i,0)}^0 + \gamma_{l(i,1)}^1 - s_i \rho_{l(i,0)} \delta + \rho_{l(i,0)} \epsilon_{i,0} + \epsilon_{i,1},
\]
where \( \alpha_{i,1} = \lambda \alpha_{i,0} \) allows for intergenerational transmission of ability over multiple generations. Thus, outcomes over time are determined by ability of the initial generation through direct inheritance of ability (\( \lambda \)) and through intergenerational advantage derived from ability in previous generations (\( \rho_{l(i,0)} \)). The current location (\( \gamma_{l(i,1)}^1 \)) shifts the level of a person’s human capital. Through intergenerational transmission, human capital is also affected by (1) how previous generations were affected by where they lived (\( \gamma_{l(i,0)}^0 \)), (2) whether their ancestors were enslaved until 1865 (\( \delta \)), and (3) their ancestors’ idiosyncratic human capital shocks (\( \epsilon_{i,0} \)).

We think of the effect of being enslaved until 1865 as the expected difference between the two groups holding constant any confounding factors (i.e., ability). That is, we define the average “treatment” effect (ATE) at \( t = 1 \) as

\[
ATE \equiv \int (\mathbb{E}[y_{i,1} | s_i = 1, \alpha_{i,0}] - \mathbb{E}[y_{i,1} | s_i = 0, \alpha_{i,0}]) \, dF(\alpha_{i,0}) = \\
= \int \mathbb{E} \left[ \rho_{l(i,0)} (\gamma_{l(i,0)}^0 - \delta) + \gamma_{l(i,1)}^1 | s_i = 1, \alpha_{i,0} \right] \, dF(\alpha_{i,0}) - \int \mathbb{E} \left[ \rho_{l(i,0)} \gamma_{l(i,0)}^0 + \gamma_{l(i,1)}^1 | s_i = 0, \alpha_{i,0} \right] \, dF(\alpha_{i,0}).
\]

(5)

Importantly, in our definition, the effect of descending from an enslaved person includes not just the effect of delayed freedom but also any potential effect operating through differential exposure to location-specific factors. Combining (3), (4), and (5), the observed Free-Enslaved gap is equal to

\[
\mathbb{E}[y_{i,1} | s_i = 1] - \mathbb{E}[y_{i,1} | s_i = 0] = ATE - B,
\]

where the (negative of) the selection bias \( B \), arising from (1) potential selection into being free, (2) potential selection into location by (descendants of) the Free, and (3) potential selection into location by (descendants of) the Enslaved.\(^{38}\) Naturally, if there was no selection into the status of being a free Black American, the observed Free-Enslaved gap

\[
B = \mathbb{E} \left[ \left( \lambda + \rho_{l(i,0)} \right) \alpha_{i,0} | s_i = 0 \right] - \mathbb{E} \left[ \left( \lambda + \rho_{l(i,0)} \right) \alpha_{i,0} | s_i = 1 \right] + \\
\left( \mathbb{E} \left[ \rho_{l(i,0)} \gamma_{l(i,1)}^0 + \gamma_{l(i,1)}^1 | s_i = 0 \right] - \int \mathbb{E} \left[ \rho_{l(i,0)} \gamma_{l(i,1)}^0 + \gamma_{l(i,1)}^1 | s_i = 0, \alpha_{i,0} \right] dF(\alpha_{i,0}) \right) - \\
\left( \mathbb{E} \left[ \rho_{l(i,0)} (\gamma_{l(i,0)}^0 - \delta) + \gamma_{l(i,1)}^1 | s_i = 1 \right] - \int \mathbb{E} \left[ \rho_{l(i,0)} (\gamma_{l(i,0)}^0 - \delta) + \gamma_{l(i,1)}^1 | s_i = 1, \alpha_{i,0} \right] dF(\alpha_{i,0}) \right).
\]

\(^{38}\) Naturally, if there was no selection into the status of being a free Black American, the observed Free-Enslaved gap...
becomes an unbiased estimate of the ATE in (5):

\[ \mathbb{E}[y_{i,1} \mid s_i = 1] - \mathbb{E}[y_{i,1} \mid s_i = 0] = ATE \quad \text{if } \alpha_{i,0} \perp \perp s_i. \] (7)

We discuss the historical evidence on the plausibility of this assumption and provide bounds for the bias \( B \) under weaker assumptions in Appendix J. In summary, whether someone was enslaved until 1865 was in many cases a result of factors largely exogenous to their “ability.” For instance, half of the free population became free through general emancipation in the North. In the South, freedom was attained in different ways, but a large fraction of Black families who gained freedom before 1865 there did so through non-selective emancipation, such as manumission upon the death of their slave master or manumissions of slave masters who freed all of their enslaved people at once.

To assess the relative importance of being enslaved longer, but facing the same set of Jim Crow institutions after, we next examine the Free-Enslaved gap conditional on the location of the ancestor living in 1870.\(^{39}\)

### 4.4 Regional Variation in the Free-Enslaved Gap

While most of the enslaved population was freed in the Lower South, most free Black families lived in the North or in the Upper South. Before and after slavery ended, these regions differed in terms of their economies, institutions, and politics. We explore to what extent regional differences account for the persistence of the Free-Enslaved gap.

To analyze the importance of regional differences, we control for the location that \( i \)'s ancestors lived in around 1865. Formally, we add location fixed effects (\( \eta \)) to our initial specifications for outcomes (1870–2000),

\[ y_{i,t} = \alpha + \beta_1 s_i + \eta_{l(i,1870)} + \gamma' X_{i,t} + \epsilon_{i,t}, \] (8)

where \( l(i,1870) \) is the location of \( i \)'s ancestors in 1865. We use three different levels of granularity of location: (1) the region (South or North) and (2) the state in which families’ 1870 ancestors were born during slavery; and (3) the county in which families’ 1870 ancestors lived. The estimate of \( \beta_1 \) reflects the socioeconomic gap between descendents of the Free and the Enslaved conditional on their families originating from the same region, state, or county.

We find that differences across ancestor states fully account for the Free-Enslaved gap after 1940 (see Figure 4). The 1940 gaps in literacy, income scores, skill, and homeown-

\(^{39}\)This analysis also defines a tight upper bound on the bias \( B \) in Equation (6). Our results show that this upper bound is essentially zero, suggesting that positive selection into freedom before 1865 does not affect the Free-Enslaved gap after 1940.
**Figure 4:** Free-Enslaved Gap in Literacy Conditional on Ancestor Location (1940)

<table>
<thead>
<tr>
<th>Baseline</th>
<th>+ Region\textsubscript{1865-FE}</th>
<th>+ State\textsubscript{1860-FE}</th>
<th>+ County\textsubscript{1870-FE}</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.2 ppt.</td>
<td>-3.2 ppt.</td>
<td>-0.2 ppt.</td>
<td>-0.2 ppt.</td>
</tr>
</tbody>
</table>

Notes: This figure shows the 1940 Free-Enslaved gap in literacy before and after including different levels of origin location fixed effects. We successively add fixed effects for the region (South or North) and state a family’s 1870 ancestor were born, and the county in which their 1870 ancestors lived. The sample includes only Black prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix A for details on the sample and data.

The importance of state differences stands out from differences at other levels of granularity. First, the Free-Enslaved gap does not merely reflect a North-South disparity but also exists for Black families whose ancestors lived within the South in 1870. Second, accounting for differences between counties of the same state of origin does not help to further account for the Free-Enslaved gap. In the next section, we explore the critical role of states in shaping the persistence of disparities that existed after slavery.

**Implications for interpreting the Free-Enslaved gap.** The result of convergence in the conditional Free-Enslaved gap rules out the importance of selection into freedom before 1865, suggesting a causal interpretation of the unconditional Free-Enslaved gap as the long-run effect of being enslaved until the Civil War. This effect captures both that the Enslaved were freed later and that they were more likely to be freed in areas where Black people faced more intensive Jim Crow institutions after slavery. With virtually all of the variation in the Free-Enslaved gap accounted for, there is little potential for positive

---

\(^{40}\)See Appendix Figure I.20 and Appendix Tables I.11 and I.12.

\(^{41}\)See Appendix Figure I.22.
selection into being a free Black American to affect the Free-Enslaved gap.\footnote{See Appendix J.}

In methodological and conceptual contrast to prior work (e.g., Sacerdote, 2005), we argue that the average treatment effect of slavery should not be thought of merely as an effect \textit{conditional on location}. Descending from an enslaved person made a person much more likely to come from (and still live in) environments that were relatively harmful to their economic progress. The location of enslavement was directly caused by their own enslavement status, and its impact should thus be included in the treatment effect. From an econometric perspective, geographic location can then be interpreted as a \textit{bad control} since it is a mediating variable through which slave status affected future descendants.

### 4.5 States’ Effects on Black Economic Progress After Slavery

Motivated by the fact that state variation in Black families’ long-run economic progress fully explains the Free-Enslaved gap after 1940, we analyze this state variation in detail. We limit our sample to families who were enslaved until 1865, for whom we observe the plausibly exogenous state of birth of families enslaved ancestors in the 1870 census. Using this variation, we can interpret each state fixed effect as the causal long-run effect of being freed in that state on socioeconomic outcomes.\footnote{Given the model introduced in Section 4.3 and under the assumption that a person’s enslavement location is exogenous to their unobserved ability \((s_{i,0} \perp \perp l(i,0) \text{ if } s_i = 1)\), \(\hat{\eta}_c - \hat{\eta}_c'\) is a consistent estimate of
\[
\mathbb{E}[y_{i,1} | s_i = 1, l(i,0) = c, X_{i,t}] - \mathbb{E}[y_{i,1} | s_i = 1, l(i,0) = c', X_{i,t}] = \\
\rho_{l(i,0)}\gamma_c^0 + \mathbb{E}[^1_{l(i,1)} | s_i = 1, l(i,0) = c, X_{i,t}] - \left(\rho_{l(i,0)}\gamma_c^0 + \mathbb{E}[^1_{l(i,1)} | s_i = 1, l(i,0) = c', X_{i,t}]\right). \tag{9}
\]
That is, the estimate reflects both the (inherited) effect the state of birth \(c\) had on the ancestor in 1870 (relative to state of birth \(c'\)) and the expected effects of future locations given the 1870 location. One can interpret the object in (9) as an intent-to-treat (ITT) effect since the initial location is plausibly randomly assigned, but the later location (e.g., in 1940) is a result of endogenous (and potentially selective) migration decisions.}

The effect of being freed in each state in 1865. We find that there is a distinct geography of Black economic progress after slavery (see Figure 5). Being freed in a state further south had substantial negative effects on Black families’ socioeconomic outcomes in the long run. For example, a family who was freed in Louisiana would have received over two years more education had they instead been freed in, say, Kentucky. States affect other outcomes, such as literacy, income, or wealth, with similar magnitudes.\footnote{See Appendix Figure I.23.} This remains true until 2000 when, for example, families who had been freed in Louisiana have average incomes that are lower by more than one-quarter of the average income among Black Americans compared to those rooted in states along the border to the North.\footnote{See Appendix Figure I.24.}

The effect of living in each state between 1865 and 1940. We formally account for mi-
Notes: This figure shows the 1870 ancestor state of birth fixed effect (FE) estimates on 1940 years of education for descendants of the Enslaved. A state’s FE is the deviation from the population-weighted average across all states (baseline mean) after controlling for a quadratic function of age. The sample includes only Black prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix A for details on the sample and data.

Migration and find that the effect of being freed in a given state is a good proxy for the effect of living there from 1865 to 1940. The effect of being freed in a location is an intent-to-treat effect that reflects an average of treatment effects by destination locations, weighted by the probability of migrating there. Under mild assumptions, we can recover the effect that living in each location between 1865 and 1940 had on the economic progress of Black families. Our framework flexibly allows for selection into migration. Due to low geographic mobility, the recovered estimates are almost identical to the original ones.

With Black families freed in the Lower South faring so much worse that families freed elsewhere, it may seem puzzling why so few of them left the region, while large fractions of those freed in the Upper South migrated to the North. Institutional and economic factors explain this puzzle. First, Jim Crow directly targeted the geographic mobility of Black people (Roback, 1984): Enticement laws and contract enforcement laws limited Black workers’ ability to terminate their employment contracts; Vagrancy laws es-

46 We formalize this decomposition in Appendix F.

47 The assumption required is that the experiences that a family made in location c during slavery cease to directly affect their descendants by 1940 (in our model: \( \rho_{c(i,0)} \gamma_{c}^{i} = 0 \)). Equation 9 then suggests that ITT effect of initial location \( c, \eta_{c}, \) is the average over all potential future locations ATE, \( \gamma_{c}^{f} \), weighted by the probability of migrating from \( c \) to \( c' \). We invert the migration probability matrix to recover the ATE of living in each state, which itself is not affected by selective migration on average under the assumption that the average innate ability of Black Americans did not differ across enslavement locations.

48 See Appendix Figure C.7.
sentially criminalized not having a job; Emigrant-agent laws prevented employers from seeking workers from other states. Second, moving to the North was costly, especially from the Lower South. Their geographic distance to the North limited the potential for social networks that would lower the cost of migration (Carrington et al., 1996). Even for those Black families who did migrate to the North, opportunities there were elusive (Akbar et al., 2020; Derenoncourt, 2022).

**States’ Jim Crow regimes and the effect of living there.** We find that the vast majority of variation in state effects can be explained by variation in Jim Crow intensity (see Figure 6). For example, Louisiana is both the state that passed the highest number of Jim Crow laws—almost 100—and the state where Black families made the least progress in terms of education by 1940. Motivated by this suggestive evidence, we next identify the causal role that states’ Jim Crow institutions played in shaping Black economic progress.

**Figure 6: State Effects and Jim Crow Instensity**

- **Left Panel:** Effect on Years of Education (1940) vs. Historical Racial Regime Score (HRR score)
  - $R^2 = 0.87$

- **Right Panel:** Effect on Years of Education (1940) vs. Number of Jim Crow Laws
  - $R^2 = 0.68$

**Notes:** This figure relates the causal effect of living in each state between 1865 and 1940 on 1940 years of education to each state’s Jim Crow intensity. Jim Crow intensity is measured as the HRR score (left panel) or the number of Jim Crow laws (right panel). The sample includes only prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix A for details on the sample and data.

5. **The Jim Crow Effect**

To identify the effect of Jim Crow, we use a regression discontinuity design that compares the socioeconomic outcomes of Black families who were freed across state borders with more or less stringent Jim Crow laws.
5.1 Border Discontinuity Design

Our border discontinuity design compares the socioeconomic status of families in 1940 whose ancestors were freed on different sides of (but in close proximity to) state borders within the South in 1865 (see Figure 7). We thereby isolate the effect of institutions from other factors that may affect Black economic progress, such as culture, climate, or the economy—factors which typically do not adhere to state borders.

**FIGURE 7: Southern Counties’ Distance to State Borders**

Notes: This map shows each county’s distance to the closest state border within the South. Darker shades correspond to closer proximity to a border. Distances are measured from a county’s centroid to the border. We limit our analysis to counties within 100 kilometers (62 miles) of any border but show that our results are robust to other cutoffs.

The border discontinuity design takes the following form:

\[ y_{i,b}^{1940} = \alpha_b + \beta_b \cdot \text{JimCrow}_{i,b}^{1870} + \nu_b \cdot \text{dist}_{i,b}^{1870} + \psi_b \cdot \text{dist}_{i,b}^{1870} \cdot \text{JimCrow}_{i,b}^{1870} + \epsilon_{i,b}, \]  

(10)

separately for each border \( b \in B^{\text{South}} \), where \( y_{i,b}^{1940} \) is the years of education of Black person \( i \) in 1940 whose ancestors were freed close to state-border \( b \), JimCrow\(_{i,b}^{1870}\) indicates whether \( i \)'s 1870 ancestors lived on the side of border \( b \) that had a more intensive Jim Crow regime than the state on the other side of the border, and dist\(_{i,b}^{1870}\) is the distance between border \( b \) and the county’s centroid in which \( i \)'s ancestors lived in 1870.\(^{49}\) The main coefficient of interest, \( \beta_b \), captures the long-run effect of being freed on the more oppressive side of border \( b \) on a Black family’s human capital.

\(^{49}\)We thereby allow the slope coefficients of the distance to the border to flexibly vary for each state-border pair.
We next relate each border discontinuity estimate $\hat{\beta}_b$ to states’ difference in Jim Crow intensity across the border because those institutions differ more drastically across some state borders than others. We measure those differences in terms of the HRR index as well as the number of Jim Crow laws. These measures are very different in nature, approaching the challenge of quantifying institutional differences from different angles.

We find that Black families who were freed in states with more oppressive regimes experienced sharply lower rates of economic progress starting in the Jim Crow era (1877–1964). The resulting differences in socioeconomic status are increasing in the differences in Jim Crow intensity across a border (see Figure 8). For example, consistent with Louisiana’s Jim Crow legislation being far more extensive than Texas’s, we find that families freed in Louisiana attained 1.2 years less education by 1940 than families freed only a few miles away in Texas. The relationship between RD estimates and differences in Jim Crow intensity are very similar for the HRR score and the number of Jim Crow laws.

There are at least two reasons why—in principle—Jim Crow may have also harmed white Americans. First, some Jim Crow laws are likely to harm some poor white Americans directly. For example, poll taxes aimed at disenfranchising Black voters but also disenfranchised some poor white voters. Second, the negative effect that institutions had on Black Americans may have spillover effects. For example, slowing Black economic progress may hamper a region’s general economic development.

We find that in contrast to Black families, the socioeconomic status of white families was not affected by the Jim Crow intensity of the state in which their ancestors lived in 1870. It appears that some states did affect the educational attainment of both Black and white Americans. For example, both groups fared substantially worse if their ancestors lived in Delaware in 1870 compared to families whose ancestors lived on the other side of the border in Maryland. These differences may point to differences in their school system or other state institutions that affected families of both races. However, there is no systematic relationship between how racially oppressive state institutions were and how state institutions affected white Americans there.

### 5.2 Validation of the Empirical Strategy

To validate our empirical strategy, we pool all borders, rather than estimating discontinuities for each border separately. The pooled regression equation closely follows Equation (10). We equally divide our sample into two types of borders: “policy borders” between...
Notes: This figure shows each separate RD estimate in 1940 years of education for Black families who were freed on different sides of state borders in 1865. Negative estimates reflect lower education in the more oppressive state. Lines show the best linear fit between RD estimates and the differences in Jim Crow intensity, weighted by the inverse of each estimate’s standard error. Shaded areas represent robust 95 percent confidence bands. See Data Appendix A for details on the sample and data.
states that strongly differ in their Jim Crow intensity (more than the median border difference); and “placebo borders” between states that differ little in their Jim Crow intensity (less than the median border difference). Our validation exercises will focus on policy borders.

**Figure 9: Pooled Regression Discontinuity (RD) Estimate**

(A) Policy borders

Discontinuity: -0.60*** (0.13)

(B) Placebo borders

Discontinuity: -0.08 (0.10)

Notes: This figure shows the RD estimate in 1940 years of education for Black families who were freed across state borders with different Jim Crow intensity in 1865. Panel (A) shows “policy borders” where Jim Crow intensity differed more than across the median border; Panel (B) shows “placebo borders” where they differed less than the median. The left half of each figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix A for details on the sample and data.

Consistent with our main RD estimates, sharp differences in education only arose for Black families freed across borders where institutions differed substantially (see Figure 9).\(^{51}\) Being freed on the more oppressive side of such a policy border reduced the years of education in 1940 sharply by 0.6 years—10 percent of the average education among Black men.

First, we confirm that differences across policy borders only arose after the onset of Jim Crow (see Figure 10). We find that before Jim Crow, there were no differences in literacy among families freed in states that would become substantially more oppressive during Jim Crow. In 1880, three years after the start of Jim Crow, the literacy rates of families were still equal. Only starting in 1900, Black families attained substantially lower rates of literacy in more oppressive states. These differences grew over time in absolute terms, but even more so in relative terms. By 1930, while almost 90 percent of all Southern Black men were literate, families freed in more oppressive states were still 4.6 percentage points less likely to read and write.

Second, we confirm that during slavery, location characteristics evolved smoothly

---

\(^{51}\) Appendix Figure 1.28 shows the pooled RD estimate for all borders—both policy and placebo.
Figure 10: RD in Literacy

(A) Pre-period: 1870
Discontinuity: -2.60 (3.77)

(B) 1880
Discontinuity: -0.40 (1.56)

(C) 1900
Discontinuity: -4.31*** (1.26)

(D) 1910
Discontinuity: -5.61*** (1.35)

(E) 1920
Discontinuity: -6.87*** (1.09)

(F) 1930
Discontinuity: -4.59*** (0.95)

Notes: This figure shows the RD estimate in literacy for Black families who were freed across state borders with different Jim Crow intensity in 1865. The sample is restricted to policy borders. The left half of each figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix A for details on the sample and data.
FIGURE 11: No Border Discontinuities in Location Characteristics During Slavery

(A) Slaves Per-Capita
Discontinuity: 0.05 (0.03)

(B) Share of Black Population
Discontinuity: 4.85 (3.55)

(C) Cotton Output
Discontinuity: 0.01 (0.03)

(D) Farm Value
Discontinuity: -0.66 (0.77)

Notes: This figure shows the RD estimate in counties’ characteristics in 1860 across state borders with different Jim Crow intensity in 1865. The sample is restricted to policy borders. The left half of each figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix A for details on the sample and data.
across state borders before Jim Crow (see Figure 11). In 1860, neither the number of slaves relative to a county’s overall population, the share of its Black population, its cotton output per capita, or its average farm value differed across state borders in the South. The same holds true for various other location characteristics such as population density, incomes, the age of enslaved people, and migration costs.52

Third, we show that our results are robust to choosing different cutoffs for the distance between a county’s centroid and a state border: 100, 150, 200, or 250 kilometers.53 For example, the pooled RD estimates across policy borders (as shown in panel (A) of Figure 9) for those cutoffs all range between -0.58 and -0.63 and are all highly significant. We choose 100 kilometers as our baseline, because it is close to the mean squared error (MSE)-optimal bandwidth.

5.3 Location of Freedom and the Question of Exogeneity

In any study that aims to estimate the effect of places on outcomes, the key assumption for causal interpretation is for location to be orthogonal to potential outcomes. In the literature, researchers have typically relied on “mover designs” (e.g., Chetty et al., 2016). In those studies, the effects of places is estimated from the outcomes of families who move between them. The causal interpretation derives from assumptions on the nature of their moves.

Our empirical strategy relies on the immobility of a specific population rather than their mobility. In particular, we build on the circumstance that the Enslaved did not enjoy freedom of movement before 1865, leaving no room for self-selection into location. This circumstance lends plausibility to the key identifying assumption of an enslaved person’s birthplace to be orthogonal to potential outcomes of their (third-generation) descendants. The main threat to our identification assumption is the possibility of selective slave migration. Even though the Enslaved did not choose where they lived, it is possible that decisions by their owners or traders induced selection into locations of enslavement.

Slaveholder migration and the domestic slave trade equally contributed to the forced migration that occurred before 1865 (Fogel and Engerman, 1974; Tadman, 1979; Pritchett, 2001). Slaveholders were generally non-selective in moving all their enslaved people with them (Fogel and Engerman, 1974; Pritchett, 2001; Tadman, 2008; Pritchett, 2019). In principle, selection could also arise through differences in the slaveholders who choose to migrate. However, for selection to arise, the slaveholder’s decision would need to be correlated with the potential outcomes of their enslaved people—a scenario we cannot rule out but deem unlikely. The domestic slave trade accounts for the remaining inter-regional slave mobility. Selective slave trade is only evident in the small sugar cultivation

52See Appendix Figure I.29.
53See Appendix Figure I.30.
areas. Sugar cultivation accounted for only 6 percent of the rural enslaved population \( (\text{Tadman, 1977, 1979}) \). By the nature of the work required, enslaved people there tended to be physically stronger and more likely to be male \( (\text{Phillips, 1918}) \). While having able-bodied men to clear the new highly fertile land is the kind of selection that slave owners cared about, it is not the kind of selection that would influence human capital 100 years later.

Overall, selective migration among the Enslaved may, if anything, have an attenuating effect on our estimates of place effects. If anything, one might hypothesize that the selection into location based on physical traits might have biased upward the estimates of states that supposedly selected positively on height and strength. In contrast, we find that such states—those in the Lower South in general and those in the sugar region of Louisiana in particular—were especially detrimental to Black economic progress.

Lastly, for our county level evidence, we rely on a family’s county of residence in 1870, around five years after Emancipation. It is possible that those who were freed in 1865 had already sorted into specific locations by 1870. Census data show that in 1870, 96 percent of children under the age of 10 lived in their state of birth (75 percent of the population lived in a household with at least one such child). Therefore, it seems that across-state migration was limited. Within-state migration is no threat to our estimates unless selection was explicitly towards or away from border regions—a scenario we deem implausible, especially given that states did not differ across those borders.

5.4 Summary: Jim Crow and Black Economic Progress After Slavery

In sum, our evidence suggests that racially oppressive institutions played a critical role in shaping the South’s detrimental effect on Black economic progress. The estimates are a lower bound to the importance of Jim Crow institutions for two reasons. First, all Southern states adopted Jim Crow laws. Our estimates only isolate the additional effect of more oppressive institutions, rather than the overall effects of Jim Crow. Second, “laws are not an adequate index of the extent and prevalence of segregation and discriminatory practices in the South. The practices often anticipated and sometimes exceeded the laws”

---

54In contrast to the sugar industry, the cotton and tobacco industries (accounting for around 87 percent of enslaved agricultural workers) was generally non-selective on age and sex \( (\text{Tadman, 1977}) \).

55Traded enslaved people were found to be disproportionately likely to be young adults \( (\text{e.g., Pritchett, 2019}) \) and more likely to be male \( (\text{Fogel and Engerman, 1974}) \), but some of this evidence is nuanced by \( \text{Tadman (1977, 1979)} \). Moreover, \( \text{Pritchett (2001)} \) finds that traded slaves were, conditional on age and sex, marginally taller than the average enslaved population—although this finding is contested by \( \text{Steckel and Ziebarth (2016)} \). Moreover, such physical characteristics were co-determined by environmental influences such as nutrition, illness, the type of work, and stress \( (\text{Steckel, 1979; Carson, 2008}) \). There is no evidence that traders selected enslaved people on anything other than such basic physical characteristics. This is consistent with the dehumanization and commodification of Black people that characterized the slave trade, which “reduced people to the sum of their biological parts” \( (\text{Smallwood, 2008, p. 43}) \).

56We compute this statistic for households with children to avoid picking up mobility through forced migration that may have occurred long before 1865.
Our empirical strategy captures only institutions that change discontinuously at the border—such as de jure segregation at the state level—leaving some institutional features unaccounted for. The aggregate importance of institutions is almost certainly larger.

## 6. The Mechanism of Limited Access to Education

In this section, we assess restrictions to human capital as a mechanism of how Jim Crow reduced Black economic progress in the long run.

### 6.1 Jim Crow Regulated Black Family’s Access to Human Capital

To gauge the relative importance of different domains that Jim Crow laws pertained to, we use our newly built database on laws and their content. We document that the largest number of laws pertains to education, mainly to segregate schools and educational resources (see Figure 12).

For example, some laws established the provision of resources for new schools or colleges for white Americans only. Others required the racial segregation of existing schools or required local school boards to comprise only white people. Even the use of school books was regulated, stipulating that once a Black or white child had used a book, children of the other race were forbidden to use the same book. While school quality measures are scarce in this period, it seems likely that those laws created drastic differences in the educational resources available to Black and white children.

**Figure 12: Jim Crow Laws by Type**

![Figure 12: Jim Crow Laws by Type](image)

Notes: This figure shows the number of Jim Crow laws across Southern states that pertain to each category.

The importance of Jim Crow in limiting Black families’ human capital accumulation in the long run has been pointed out by leading scholars. Booker T. Washington writes that
“few people [have an] idea of the intensive desire which [Black people] showed for education. It was a whole race trying to go to school” (Washington, 1907). However, Black people’s desire for education was met with resistance. “[Black Americans’] attempts at education provoked the most intense and bitter hostilities as evincing a desire to render themselves equal to the whites” (American Freedmen’s Commission Report cited in Du Bois, 1935, p. 645). Robert Higgs argues that governments were the leading force of this resistance:

“Most damaging of all [racial discrimination after slavery] was the discriminatory behavior of the southern state and local governments. By providing only scant resources for black education, public school boards helped to perpetuate illiteracy […], and they thereby set in motion a variety of adverse effects.” (Higgs, 1989, p. 25)


Empirically, the intergenerational importance of education is consistent with Margo (1991), who shows that literacy and education among Black Southerners was linked to improved educational outcomes of their children, an early transition out of the farm economy, and an increased likelihood of moving to the North. Similarly, Collins and Wanamaker (2022) point to education as the key factor in shaping Black-white disparities in intergenerational mobility throughout the 20th century.

6.2 Access to Schools Mediated Jim Crow’s Negative Effects

To identify the lack of resources to accumulate human capital as a mechanism through which Jim Crow slowed the long-run economic progress of Black families, we leverage quasi-experimental variation from the Rosenwald school program. This program built 5,000 new schools for Black children throughout rural communities in South between 1914 and 1931 (see Figure 13). Aaronson and Mazumder (2011) document that there was substantial quasi-random variation in when and where Rosenwald schools were built and that the receiving counties did not systematically differ from other rural counties before the program. Our research design, however, is robust to unobserved differences between treated and untreated counties.
In particular, we estimate

\[ Y_{i,1940} = Rosenwald_{c_{1870},y} + \eta_{c_{1870}} + \eta_{s_{1870},y} + urban_{i,1870} + \varepsilon_{i,1940}, \]

where \( Y_{i,1940} \) is the 1940-outcome of child \( i \) born in year \( y \), \( c_{1870} \) is the county in which \( i \)'s ancestors lived in 1870, \( Rosenwald_{c_{1870},y} \) is the fraction of Black children born in \( y \) who could attend a Rosenwald school in their ancestors' county \( c_{1870} \), \( \eta_{c_{1870}} \) and \( \eta_{s_{1870},y} \) are county and state-cohort fixed effects, and \( urban_{i,1870} \) is a dummy for whether \( i \)'s ancestors lived in an urban part of county \( c_{1870} \).\(^{57}\) Our main outcome is years of education, but we also consider literacy, income, and wealth.

We make three empirical innovations. First, we use the county in which a Black child’s enslaved ancestors were freed, not their own county of birth, to determine the likelihood it attended a Rosenwald school. Holding the place of treatment constant at the ancestor level addresses the concern that parents may sort into locations based on educational opp-

\[ Rosenwald_{c_{1870},y} = \frac{1}{10} \sum_{t=y+6}^{y+16} \frac{RosenwaldTeachers_{c_{1870},t} \times 45}{BlackChildren_{c_{1870},t}}, \]

where \( y \) is the year in which a child was born, \( c_{1870} \) is the county in which \( i \)'s ancestors lived in 1870, \( RosenwaldTeachers_{c_{1870},t} \) and \( BlackChildren_{c_{1870},t} \) are the numbers of Rosenwald teachers and Black children in school ages (6–16) in county \( c_{1870} \) at time \( t \); 45 approximates the average class size at the time.

---

\(^{57}\)Our continuous treatment measure reflects the fraction of Black children in county \( c \) who were able to attend a Rosenwald school in the years in which child \( i \) from \( c \) was in school ages (6–16):
opportunities there. Second, by following the children of potential Rosenwald school students from the 1940 census to administrative records in 2000, we analyze the program’s intergenerational effects. Last, we use our measures of Jim Crow intensity to explore the heterogeneity of treatment effects across different institutional environments.

It is theoretically ambiguous whether Rosenwald schools are more or less beneficial to Black children in racially oppressive places. On the one hand, the demand for human capital may be lower among Black families in oppressive areas, potentially reducing the impact of Rosenwald schools. For example, Jim Crow institutions may have disincentivized human capital accumulation by limiting Black children’s return to education. On the other hand, oppressive institutions may directly restrict the supply of educational resources, potentially increasing the impact of Rosenwald schools.

We find that the Rosenwald school program had persistent positive effects on the economic progress of Black families, especially in the most oppressive states (see Table 3). A Black child able to attend a Rosenwald school attained 0.3 years more in education than a child not able to attend. In the most oppressive states, the effect was more than twice as large. We confirm that only families who stayed close to where their ancestors were freed until at least 1920 benefited from new Rosenwald schools built there after 1919. For families who had moved out of the area by 1920, Rosenwald school in their place of origin had, as expected, no effect on their education.

---

58 We thereby also broaden the sample to all men, rather than men who served in WWII—Aaronson and Mazumder (2011) link the census to enlistment records which contain enlistees’ county of birth.
59 An increase of the cohort × county specific Rosenwald exposure measure from 0 to 1 yields an increase in the cohort × county specific share of Black children in school by 0.65 (F-statistic: 1,736.2).
Table 3: Rosenwald Schools and Education

<table>
<thead>
<tr>
<th>Rosenwald exposure (%)</th>
<th>Baseline</th>
<th>Placebo</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stayers</td>
<td>Movers</td>
<td>Few Laws</td>
</tr>
<tr>
<td></td>
<td>Sample mean: 5.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.31**</td>
<td>0.38***</td>
<td>0.09</td>
<td>0.22</td>
</tr>
<tr>
<td>(0.10)</td>
<td>(0.12)</td>
<td>(0.23)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Controls</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>Observations</td>
<td>113,107</td>
<td>90,527</td>
<td>22,520</td>
</tr>
</tbody>
</table>

Notes: This table shows the effect of Rosenwald school exposure (0 to 1) on a person’s years of education in 1940. All regressions control for ancestor county FEs, year of birth \times ancestor state FEs, and the within-county rural/urban status of each family in 1870. Columns 2 and 3 compare individuals who remained in the state of their ancestors until at least 1920 (“Stayers”) and those who did not (“Leavers”). “Few Laws” considers the bottom quartile of states in terms of their number of educational Jim Crow laws; “Many Laws” considers the top quartile. “Low HRR” considers the bottom quartile of states in terms of their HRR score (least oppressive); “High HRR” considers the top quartile (most oppressive). Standard errors are clustered at both the county and cohort level. See Data Appendix A for details on the sample and data. ** p < 0.01, * p < 0.1.

In addition, we find that having access to a Rosenwald school not only increased the education of the benefiting individuals, but also improved the economic conditions of their children today (see Table 4). We find that the children of Rosenwald school attendees live in neighborhoods where Black people tend to be more likely to have a college degree, earn higher incomes, and live in more valuable homes. Thus, the Rosenwald school program’s positive effects persist beyond the Jim Crow era.

Table 4: Intergenerational Effects of the Rosenwald Schools

<table>
<thead>
<tr>
<th>Father’s Rosenwald exposure</th>
<th>Children’s Neighborhood Level Outcomes in 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HS Degree (%)</td>
</tr>
<tr>
<td></td>
<td>Mean: 69.33</td>
</tr>
<tr>
<td>2.78</td>
<td>4.95**</td>
</tr>
<tr>
<td>(3.06)</td>
<td>(2.10)</td>
</tr>
</tbody>
</table>

Notes: This table shows the effect of a father’s Rosenwald school exposure (0 to 1) on the residential neighborhood quality of the their children in 2000. All regressions control for ancestor county FEs, father’s year of birth \times ancestor state FEs, and the within-county rural/urban status of each family in 1870. Standard errors are clustered at both the county and father’s cohort level. ** p < 0.01, * p < 0.1.

38
The end of one racially oppressive institution is not the end of all racially oppressive institutions. After slavery ended, Jim Crow emerged in the states that concentrated the formerly enslaved population. Since the end of Jim Crow, new oppressive institutions have arisen—for example, mass incarceration (the “new Jim Crow”, Alexander, 2010) or “color-blind” voter suppression. Has the geography of Black economic progress changed since the end of Jim Crow?

We compare each state’s causal effect on Black economic progress after slavery to the same location’s causal effect on intergenerational mobility in recent decades as estimated by Chetty and Hendren (2018) (see Figure 14). We find that these estimates are highly positively correlated ($\rho = 0.679$). This tight link between historical and modern estimates suggests that a state’s ability to generate upward mobility is highly persistent over time: locations that spurred faster Black economic progress after slavery also increased the intergenerational mobility of American families in recent decades.

The strength of this correlation may appear surprising given that drastic changes in the institutional and economic environment across the two periods. Specifically, throughout 1870 to 1940 the economic progress of Black families in the South was institutionally limited by Jim Crow. In addition, the Southern economy was largely agricultural with 40 percent of Black men working in this sector. By the 2000s, Jim Crow had been abolished for more than 30 years and less than two percent of Black men worked in agriculture. However, while Jim Crow ended and the Southern economy evolved, racial oppression did not end. Acharya et al. (2018) show that counties most reliant on slavery until 1865 continue to have high rates of racial resentment against Black Americans and low political support for policies that could promote Black progress. These factors likely continue to limit Black economic progress in those locations (see also Berger, 2018).

While the effects of places estimated in this paper is specific to the formerly Enslaved, those estimated by Chetty and Hendren (2018) are based on children with parents at the 25th percentile of the income distribution regardless of race. Even if not specifically conditioning on race, the large Black-white disparities in family incomes lead to a disproportionate contribution of Black families to their estimates. We find that the correlation shrinks substantially ($\rho = 0.204$) when considering the mobility of children with parents at the 75th percentile. See Appendix Figure 1.27. We interpret the effect of an ancestor’s location of emancipation on their descendant’s outcomes in 1940 as a measure of locations’ effect on upward mobility. While our estimates do not condition on ancestors’ outcomes in 1870 explicitly because of data limitations, they do so implicitly by conditioning on ancestors being enslaved up until 1865.
**Figure 14: Persistence of a County’s Capacity to Generate Upward Mobility**

Notes: This figure shows a binned scatter plot relating the effect of ancestor’s county of residence in 1870 on 1940 years of education among descendants of the Enslaved (as shown in Figure 5) to the causal effects those same counties had on intergenerational mobility in recent decades as estimated by Chetty and Hendren (2018). We use their estimates of children’s mean percentile rank in the national household income distribution at age 26 conditional on growing up with parents at the 25th percentile of the national household income distribution. Appendix Figure I.26 repeats this figure at the county level. See Data Appendix A for details on the sample and data.

8. Conclusion

This paper provides new evidence on the historical roots of modern racial disparities, finding that Black Americans’ socioeconomic status today is lower than that of white Americans in large part due to the US’s history of slavery and Jim Crow. First, we document that Black families who were enslaved until 1865 continue have considerably lower socioeconomic status today. Second, we show this persistence is entirely driven by post-slavery oppression under Jim Crow and highlight the limited access to human capital as a key mechanism.

Our findings have important implications for policies that aim to reduce the disadvantage faced by descendants of the Enslaved. First, our results highlight the importance of disparities within racial groups that are not necessarily addressed by race-specific policies. College affirmative action is a prime example. Massey et al. (2007) show that the more selective a college, the less likely their Black students descend from the Enslaved. While only 13 percent of the country’s 18- to 19-year-old Black Americans have an immigration background, 41 percent of Black Ivy League students do. Affirmative action increases racial diversity on campuses but is less effective in alleviating disadvantages faced by descendants of the Enslaved.
Second, our results highlight that increasing the supply of educational resources to Black children can have long-lasting positive effects on their families’ economic progress. We find that externally providing such resources is particularly effective in areas where Black children are most deprived of them. Similar evidence has been documented by development economists (e.g., Duflo, 2001; Glewwe and Kremer, 2006). Our evidence also suggests that those policies can have large intergenerational spillovers. Ignoring those effects may lead to the design of policies that are below their optimal scale.

Third, there has been renewed interest in the specific policy of reparations, i.e., wealth transfers to descendants of the Enslaved (e.g., Darity, 2008; Craemer et al., 2020; Boerma and Karabarbounis, 2021). We argue that any assessment of the legacy of slavery should incorporate both when and where a family was freed—i.e., how long they were enslaved and how intensively they were exposed to Jim Crow after slavery. Our empirical evidence suggests that, indeed, Black families continue to be impacted drastically by when and where their ancestors were freed. For that matter, we must stress again that we only quantify the additional disadvantage faced by those whose ancestors were enslaved until 1865 and concentrated in the Lower South compared to those who gained freedom earlier, mostly in the Upper South and North. Many in the group of the Free were enslaved in earlier periods, and all Black Americans faced discrimination as a result of slavery regardless of their specific family history. Note that while some argue that reparations should only be received by those who can prove their ancestors were enslaved, our results suggest that post-slavery institutions also harmed Black Americans who descended from the Free—a group that may find it harder to prove their ancestors had been enslaved decades before the Civil War.

This paper has limitations that future work may be able to overcome. First, we limit our analysis to men because automated census-linking methods are either not available or have poor coverage for women. Women have historically tended to change their last names upon marriage, making it impossible for conventional methods to link them across census records. While not within the scope of this paper, in a separate project we link millions of women across census records by using information on their maiden and married names from social security applications (Althoff et al., 2022). Second, while we highlight Jim Crow laws that regulate Black families’ education, laws pertaining to other domains may have limited Black economic progress. Some of those laws have been studied more (e.g., limits to Black suffrage—see Naidu, 2012) than others (e.g., limits to interracial marriage).
REFERENCES


42


APPENDIX

A Data Appendix
A.1 Census Data ................................................. 50
A.2 Census Linking ............................................. 50
A.3 Identifying Descendants of the Free and the Enslaved .......... 50
A.4 Main Sample ................................................. 58
A.5 Individual-Level Outcome Variables ............................. 58
A.6 Neighborhood-Level Outcome Variables ....................... 60
A.7 County Characteristics ...................................... 62
A.8 Nine-Digit ZIP to Census 2000 Crosswalks .................... 63

B Socioeconomic Status among Descendants of the Free and the Enslaved 68

C Migration among Descendants of the Free and the Enslaved 69

D Adjusting Estimates for Misclassification Bias .......................... 73

E Adjusting Estimates for Intermarriage ................................ 74

F The Direct Effect of Locations After Accounting for Migration .......... 76

G Empirical Bayes Shrinkage of Place Effects .......................... 77

H Placebo Exercises .............................................. 80

I Additional Results .............................................. 85
I.1 The Free-Enslaved Gap in Alternative Measures .................. 85
I.2 The Free-Enslaved Gap in Intergenerational Mobility ................ 87
I.3 The Free-Enslaved Gap Based on the Distribution of Last Names .... 89
I.4 The Free-Enslaved Gap for Free Without Physical and Human Capital ... 89
I.5 The Free-Enslaved Gap between and within Ancestor’s Birthplace .... 90
I.6 The Free-Enslaved Gap in Mortality ................................ 91
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.7 The Intergenerational Effect of Exposure to State-Specific Factors</td>
<td>92</td>
</tr>
<tr>
<td>I.8 The Persistence of Place Effects</td>
<td>95</td>
</tr>
<tr>
<td>I.9 Additional Results from the Border Discontinuity Design</td>
<td>96</td>
</tr>
<tr>
<td>J Model Appendix</td>
<td>98</td>
</tr>
<tr>
<td>J.1 Selection into Being Free</td>
<td>99</td>
</tr>
<tr>
<td>J.2 Bounding the Selection Bias</td>
<td>101</td>
</tr>
</tbody>
</table>
A. DATA APPENDIX

A.1 Census Data

We use the publicly available full-count US Census Bureau data at the individual level for Black Americans for all decades between 1850 and 1940. The 1850 census is the first to include individual-level data; the 1940 census is the latest for which restricted access to the names of individuals is available. We use linking methods that rely on those names. The data from 1890 were destroyed by a fire and are therefore not included.

A.2 Census Linking

We use crosswalks between observations in different census decades provided by Abramitzky et al. (2020), publicly available here at https://censuslinkingproject.org. The crosswalks can be merged into the public version of the census data using the histid identifier. We do so, linking all adjacent and non-adjacent census decades in our sample period.

There are multiple crosswalks available, each based on different linking techniques. Our main results use the “abe_race_nysiis_standard” link, which matches observations based on first name, last name, and age. It requires each name to be unique within a five-year window for each race but allows some names to be matched even if their spelling differs. We also provide results for “abe_nysiis_conservative,” which requires names to be unique within and across races and allows for fewer deviations in matching characteristics.

To study intergenerational dynamics, we inherit parents’ characteristics to the children in their household. Certain characteristics, such as occupation or education, are only inherited from prime-age male (ages 20–54) ancestors to ensure comparability over time.

A.3 Identifying Descendants of the Free and the Enslaved

Main Method: Linking Historical Census Records

Figure A.1 illustrates our new method to identify descendants of the Free and descendants of the Enslaved in census records between 1870 and 1940. It mainly relies on census-linking methods (Abramitzky et al., 2019) but also uses information on place and year of birth.

The method consists of three steps. First, we identify the Free themselves before identifying their descendants. In 1850 and 1860, the enslaved population was excluded from the individual-level censuses. By definition, every Black American included in the census was therefore free before 1865. We link the 1850 and 1860 censuses forward to all
census decades between 1870 and 1940 and then classify every Black American who can be linked to 1850 or 1860 as free.

**Figure A.1: Illustration of Our Free-Enslaved Classification Algorithm**

<table>
<thead>
<tr>
<th>1860 Census</th>
<th>1870 Census</th>
<th>1880 Census</th>
<th>1940 Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5m obs.</td>
<td>4.8m obs.</td>
<td>6.6m obs.</td>
<td>12.7m obs.</td>
</tr>
<tr>
<td>(+4.0m Enslaved)</td>
<td></td>
<td></td>
<td>3.0m prime-age men</td>
</tr>
<tr>
<td>Isaac Smith</td>
<td>Jonah Smith</td>
<td>Jonah Smith</td>
<td>Free</td>
</tr>
<tr>
<td>Ida Smith</td>
<td>Jemima Smith</td>
<td>Jemima Smith</td>
<td>K</td>
</tr>
<tr>
<td>Jonah Smith</td>
<td>King Smith</td>
<td>6.9k</td>
<td></td>
</tr>
<tr>
<td>Moses Brown</td>
<td>Moses Brown</td>
<td>Mina Brown</td>
<td>+</td>
</tr>
<tr>
<td>1860, NJ</td>
<td></td>
<td></td>
<td>2.5k</td>
</tr>
<tr>
<td>Abe Williams</td>
<td></td>
<td></td>
<td>158.8k</td>
</tr>
<tr>
<td>Ann Williams</td>
<td>Titus Williams</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** This figure illustrates our new method to identify descendants of the Free and the Enslaved in census records 1870-1940. The names are chosen are arbitrary examples and do not reflect real data. Jonah Smith is identified as a descendant of the Free because he can be linked back to the 1860 census; Moses Brown because he was born in a state (New Jersey) that had abolished slavery by the time of his birth (1860). Abe Williams does not fall into either category and is therefore classified as formerly enslaved or a descendant of the Enslaved. The Free-Enslaved status is assigned to descendants based on their male ancestor. In 1940, the final year of our sample, we identify 9,400 descendants of the Free (6,800 through direct linking to 1850–1860 and 2,600 through their ancestor’s birthplace) and 155,800 descendants of the Enslaved. While not comprehensively illustrated here, we do link across all adjacent and non-adjacent census records of 1850, 1860, 1870, 1880, 1900, 1910, 1920, 1930, and 1940.

In addition to linking, we use information on place and year of birth in our classification algorithm. All Northern states had begun banning or restricting slavery by 1804—some of them decades earlier. Any Black person born in those states was either free upon birth or would be emancipated by a certain age (typically in their 20s). While the latter case opens up the possibility of a Northern-born Black person being sold into slavery in other states before their emancipation, this possibility was ruled out by law.

In Appendix Table A.2, we compare the de jure to the de facto status of slavery in the North. As a de facto measure, we show the number of slaves in the state in absolute numbers and as a fraction of the state’s Black population. Based on this evidence, we classify any Black American born outside of the slave states after 1804 and before 1865 as Free. In addition, we use the state-specific years in which slavery was abolished or restricted in non-slave states to go even further back in time.
Second, we identify the descendants of the Free by using information on the relationship between individuals within census households. Specifically, we classify Black people with a free Black American ancestor as being descendants of the Free. Any person without a free ancestor is classified as a descendant of the Enslaved. In 1940, the final year of our sample, we identify 9,400 descendants of the Free and 155,800 descendants of the Enslaved.

**Attenuation bias.** Because we can only link men, the descendant classification is determined exclusively through male ancestors. This data limitation prevents us from assessing inter-marriage between the Free and the formerly Enslaved or their descendants. The fact that some Black Americans will be both descendants of the Free and descendants of the Enslaved potentially biases our estimates of the Free-Enslaved gap toward zero. This attenuation bias is one reason why our results should be interpreted as a lower bound for the true Free-Enslaved gap.

**Linking bias.** Any study that uses automated linking methods faces the problem that individuals who can be linked across decades may not represent the overall population. For example, families with a high socioeconomic status may choose more unique names for their children, making it easier to create a unique match across census records. A socioeconomic gap between two sub-populations is only biased if the linking procedure differentially selects them into the sample. Table A.1 shows that, if anything, the linking procedure biases the Free-Enslaved gap toward zero.

**Table A.1: Assessing Linking Bias**

<table>
<thead>
<tr>
<th></th>
<th>Free (1860)</th>
<th>Enslaved (1870)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linked</td>
<td>Population Δ</td>
</tr>
<tr>
<td>Literacy (%)</td>
<td>65.1</td>
<td>66.8 -3%</td>
</tr>
<tr>
<td>Occupation Score</td>
<td>6.0</td>
<td>6.1 -1%</td>
</tr>
<tr>
<td>Real property ($)</td>
<td>1,217</td>
<td>1,230 -1%</td>
</tr>
<tr>
<td>Personal property ($)</td>
<td>312</td>
<td>316 -1%</td>
</tr>
<tr>
<td>Lives in North (%)</td>
<td>45.1</td>
<td>52.1 -13%</td>
</tr>
<tr>
<td>Lives on Farm (%)</td>
<td>21.2</td>
<td>18.2 17%</td>
</tr>
<tr>
<td>Observations</td>
<td>20,994</td>
<td>79,374</td>
</tr>
</tbody>
</table>

Notes: This table suggests that sample selection is small and negative among the Free and more positive among the formerly Enslaved. If anything, this kind of selection biases the Free-Enslaved gap toward zero. The left panel compares the Free who can be linked to any future decade to the entire population (which only consists of free Black Americans) in 1860. The right panel compares the formerly Enslaved (anyone not classified as Free by our algorithm) in 1870 who can be linked to any future decade to the 1860 population of the formerly Enslaved.

In addition, a family’s socioeconomic status may affect not only whether they can be linked across decades but also over how many decades they can be linked. For example, children who grow up with single mothers can typically not be linked to their grandparents because women cannot be linked by these methods due to name changes at mar-
riage. Our classification algorithm identifies descendants of the Free mainly through whether they can be linked back to 1850 or 1860, which could lead to an almost mechanically higher socioeconomic status. We addressed this concern in Section 3.4 (see Figure 2).

Last, one may be concerned that the effect of place in 1870 on outcomes in 1940 may be biased by differences in linking rates across those locations. In particular, areas with large Black populations may have lower linking rates because the linking relies on the uniqueness of a person’s name, state of birth, and age. Lower linking rates may imply that only individuals with particularly rare names—and therefore potentially different socioeconomic statuses—are selected into the sample. Figure A.2 addresses this concern by showing counties’ average likelihood of a resident in 1870 being linkable to the 1940 census. Linking rates are similar across the country except for the most sparsely populated counties in the North (which do not contribute to most of our main results).

**Figure A.2: Linking Rates by County from 1870 to 1940**

![Map of linking rates by county](image)

**Notes:** This figure shows the average linking rate for Black prime-age (20–54) men in 1870 to 1940. Only counties with a Black population of at least 50 prime-age men in 1870 are included.
<table>
<thead>
<tr>
<th>Year</th>
<th>State</th>
<th>Abolition of Slavery</th>
<th>Number of Slaves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>De Jure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>De Facto</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
<tr>
<td>1777</td>
<td>Vermont</td>
<td>Slavery was banned immediately upon founding of Vermont <em>(Constitution of Vermont, 1777)</em>.</td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
<tr>
<td>1780</td>
<td>Pennsylvania</td>
<td>Law of gradual emancipation passed in 1780 <em>(Pennsylvania General Assembly, 1780)</em>. Black Americans born to enslaved mothers after 1780 would be freed at age 28. Slavery was ended in 1847.</td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
<tr>
<td>1781</td>
<td>Maine</td>
<td>Slavery was abolished by Supreme Judicial Court rulings in three related court cases, collectively known as the &quot;Quock Walker case&quot; <em>(Cushing, 1961; Zilversmit, 1968)</em>. Slavery was ruled incompatible with the new state constitution of 1780.</td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td>Massachusetts</td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
<tr>
<td>1783</td>
<td>New Hampshire</td>
<td>Similar to Massachusetts, New Hampshire’s constitution essentially abolished slavery by stating &quot;all men are born equal and independent&quot; <em>(Constitution of the State of New Hampshire, 1783)</em>. However, it is not clear whether court rulings indeed interpreted the constitution as being at odds with slavery or not.</td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
<tr>
<td>1784</td>
<td>Rhode Island</td>
<td>Law for gradual emancipation passed in 1784 <em>(General Assembly of Rhode Island, 1784)</em>. Black Americans born to enslaved mothers after 1784 would be freed at age 18 (women) or 21 (men).</td>
<td>1790</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1850</td>
</tr>
</tbody>
</table>
Table A.2: Abolition of Slavery in the North

<table>
<thead>
<tr>
<th>Year</th>
<th>State</th>
<th>De Jure Abolition of Slavery</th>
<th>De Facto Number of Slaves</th>
<th>Year</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1784</td>
<td>Connecticut</td>
<td>Law for gradual emancipation passed in 1784 (Connecticut General Assembly, 1784). Black Americans born to enslaved mothers after 1784 would be freed at age 25. This age was lowered to 21 in 1797. Slavery was abolished in 1848.</td>
<td>1790  2,759 (50%)</td>
<td>1800  951 (15%)</td>
<td>1810  310 (5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1787</td>
<td>Ohio</td>
<td>The Confederation Congress’s Northwest Ordinance of 1787 both banned and enforced slavery (Confederation Congress, 1787). A clause allowed Northerners to capture and enslave runaway slaves. Slavery was abolished by Ohio in 1802, Indiana in 1816, and Illinois in 1818.</td>
<td>1790  –</td>
<td>1800  135 (21%)</td>
<td>1810  429 (28%)</td>
</tr>
<tr>
<td></td>
<td>Indiana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Illinois</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Michigan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wisconsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minnesota</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1799</td>
<td>New York</td>
<td>Law for gradual emancipation passed in 1799 (New York State Legislature, 1799). Black Americans born to enslaved mothers after 1799 would be freed at age 25 (women) or 28 (men). In 1817, state decided to free all slaves born before 1799 (but not their children) in 1827 (New York State Legislature, 1817).</td>
<td>1790  21,324 (82%)</td>
<td>1800  20,343 (66%)</td>
<td>1810  15,017 (37%)</td>
</tr>
<tr>
<td>1804</td>
<td>New Jersey</td>
<td>Law for gradual emancipation passed in 1804 (New Jersey State Legislature, 1804). While not freeing living slaves, Black Americans born to enslaved mothers after 1804 would be freed at age 21 (women) or 25 (men).</td>
<td>1790  11,423 (81%)</td>
<td>1800  12,422 (74%)</td>
<td>1810  10,851 (58%)</td>
</tr>
</tbody>
</table>

Notes: This table provides a timeline for the abolition of slavery in the North. The first column indicates the year which we choose as the states’ final year of slavery. We classify any Black American born in the state after this cutoff as free. The third column shows the laws that abolished slavery. In many cases, slavery was not abolished outright, but rather it was restricted in ways that would imply a person is free before 1865 in all likelihood. The final column shows the actual number of slaves who reside in the state and the percentage of the state’s Black population being enslaved in parentheses. The number of slaves is taken from aggregate counts in census records (1790–1850).

62 While the 1790 census states that 16 slaves were in Vermont that year, this is likely an error.

63 There is some evidence that after 1804, some Black Americans were sold to slave states before they
Alternative Method of Free-Enslaved Classification: Distribution of Last Names

While our main method provides a high-accuracy classification of descendants of the Free and the Enslaved, accuracy comes at the cost of reduced sample sizes due to imperfect linking rates across the decades. To use the full census sample of Black Americans after 1870, rather than a linked sub-sample thereof, we develop an additional strategy for identifying descendants of the Free and the Enslaved based on last names. Figure A.3 shows that the name-based measures are highly correlated with the free status based on our preferred measure, though they are attenuated as expected.

**Figure A.3: Comparing Name-Based and Linking-Based Measures**

(A) Exact Last Names

(B) Phonetic Version of Last Names

Notes: This figure compares the probabilistic measures of descending from free Black Americans with our preferred measure based mainly on census linking. This binned scatter plot shows that among Black prime-age men in the 1940 census, the fraction of people classified as Free closely coincides with the predicted probability based on the people’s last names.

Our alternative classification algorithm uses changes in the distribution over last names from 1850–1860 to 1870–1880. Before 1865, the census only includes free Black Americans—after, it also includes the formerly Enslaved and their descendants.

We compute the relative frequency of each last name before and after 1865. We then create a measure of how likely a person is to descend from the Free by dividing their last name’s relative frequency before 1865 by its relative frequency after 1865. For example, the last name Du Bois appears with relatively high frequency in the 1850 and 1860 censuses, while Freedman does not appear at all. After the four million formerly enslaved individuals enter the census sample in 1870 and 1880, the name Du Bois is far less (one-tenth) frequent, whereas a substantial number of individuals entered the sample with the last name Freedman for the first time. These changes suggest that anyone named Du Bois after 1865 likely descends from the Free, whereas anyone named Freedman likely descends from the Enslaved. Note that not all names give us a good idea of whether a reached the age to be emancipated (Armstead et al., 2016, p.104).
person descends from the Enslaved or not. Especially names very common among Black Americans before 1865, such as Johnson, Brown, or Smith, remain very common after 1865. Other names such as Washington did exist among Black Americans before 1865, but became substantially more common after many newly freed enslaved people chose this name in honor of the country’s first president.

Formally, using the example of the last name Du Bois, we estimate the name-specific likelihood of descending from free Black Americans defined as

\[
P(\text{Free}_{it} = 1|\text{Name}_{it} = \text{DuBois}_t) = \frac{P(\text{Free}_{i,1860} = 1, \text{Name}_{i,1860} = \text{DuBois}_t)}{P(\text{Name}_{i,1860} = \text{DuBois}_t)}
\]

\[
= \frac{P(\text{Name}_{i,1870} = \text{DuBois}_t)}{P(\text{Name}_{i,1870} = \text{DuBois}_t)}
\]

where the second equation follows from assuming that a last name conveys a constant probability of descending from free Black Americans. The last equation follows from the fact that the 1860 census only contained free Black Americans. This equation can be approximated by

\[
\hat{P}(\text{Free}_{it} = 1|\text{Name}_{it} = \text{DuBois}_t) = \frac{\#(\text{DuBois}_t)_{1860}/\text{BlackPop}_{1860}}{\#(\text{DuBois}_t)_{1870}/\text{BlackPop}_{1870}},
\]

where $\#\text{DuBois}_t$ is the number of individuals with the last name Du Bois in a given year and $\text{BlackPop}_t$ is the population of all Black Americans (free and enslaved). To reduce noise, we combine the names from the 1850 and 1860 censuses as a pre-1865 count and the 1870 and 1880 censuses as a post-1865 count. Before 1865, we compute the population by adding up the census sample size (the Free) and the number of the Enslaved (Berlin, 1974). We truncate our estimated probability by 0 and 1. Names that only appear pre-1865 but not post-1865 are assigned probability 1; those that only appear post-1865 are assigned probability 0.

To allow for misspellings, we also compute this measure based on the phonetics of last names. Specifically, we transform last names using the New York State Identification and Intelligence System (NYSIIS) phonetic code. For example, the last names “Browne” and “Brown” both become “Bran.” For placebo exercises, we also compute the above measure as a pseudo-probability of being free for white Americans as well as for 1875 as a time placebo for Emancipation.
A.4 Main Sample

For our main sample, we focus on prime-age (20–54) Black men who can be linked to ancestors in 1880 or before. Our focus on prime-age individuals provides a certain form of comparability, limiting the possibility that an individual is in school or retired. We focus on men because we rely on automated census-linking techniques that are either unavailable or have notoriously low coverage for women. We restrict the sample to individuals who can be linked back to 1880 or before for two reasons. First, this requirement excludes families who migrated to the US after 1880. Any comparison made in our analysis will thus be for individuals whose families have lived in the country for at least 60 years. Second, it reduces the potential for linking bias as discussed in Section A.3.

A.5 Individual-Level Outcome Variables

Our main outcomes variables can be categorized as (proxies of) income, education, or wealth. Most individual-level data draw on census records provided through IPUMS (Ruggles et al., 2020).

Income

- **Occupational income scores, 1850–1940 (census)**. Because the census does not include any continuous measure of income before 1940, researchers have instead relied on occupational income scores. The most popular version, “occscore,” reflects the median total income of a person in that occupation in 1950.

- **Lido income scores, 1850–1940 (Saavedra and Twinam, 2020)**. Occupational income scores do not contain any age-, sex-, or race-specific information. The recent literature has used regression and machine learning techniques to improve on the traditional occupational income score (e.g., Saavedra and Twinam, 2020; Abramitzky et al., 2021). We use the Lido score constructed by Saavedra and Twinam (2020). The authors constructed it using machine learning techniques using 1950 and 2000 census data to validate their results against occscore in the 1915 Iowa census. According to Abramitzky et al. (2021), the Lido score has a correlation of 0.99 with their own measure.

- **Occupational skill, 1850–1940 (Leeuwen and Maas, 2011)**. We use HISCLASS, a classification to compare occupations based on the skill they typically required. The classification ranges from “higher managers” to “unskilled farm workers.” We coarsen this classification by assigning “skilled” to every occupation classified as “medium skilled workers” or above and “unskilled” to everyone else.
• **Wage income, 1940 (census).** We use wage income for 1940, the only year it is available for in our sample period.

• **Predicted total income, 1940 (census).** Total income, the sum of wage and non-wage income, only becomes available in the census after 1940. Following the prior literature, we use a rich set of covariates to predict total income in 1940 (see, e.g., Abramitzky et al., 2021). In particular, we make use of the fact that in 1960, the census includes both wage income and total income. We regress 1960 total income on 1960 wage income, a quadratic function in age, and all interactions of state fixed effects $\times$ self-employment status $\times$ occupation scores. To avoid over-fitting, we shrink the coefficients using LASSO with tenfold cross-validation. We then apply the estimated equation to 1940, yielding a prediction of 1940 total income.

Note that we use 1960 data for the prediction because 1950 is affected by the “sample-line problem,” reducing the number of observations that include information on incomes.

**Education**

• **Literacy, 1850–1940 (census).** We use literacy for all years. In 1940, literacy becomes unavailable, and instead the census starts to include educational attainment. We proxy for literacy by having completed at least the second grade.

• **Years of education, 1940 (census).** We impute years of education from the highest educational level attained (“educd”).

• **High school, 1940 (census).** We impute whether a person holds a high school degree based on whether they completed at least 12 years of schooling (“educd”).

• **College, 1940 (census).** We impute whether a person holds a college degree based on whether they completed at least 16 years of schooling (“educd”).

• **Graduate, 1940 (census).** We impute whether a person holds a graduate degree based on whether they completed at least 17 years of schooling (“educd”).

**Wealth**

• **Personal property, 1860–1870 (census).** Measures “the contemporary dollar value of all stocks, bonds, mortgages, notes, livestock, plate, jewels, and furniture” as reported to the census. It is not clear whether zeros indicate missing values or true zero personal property, and therefore we replace zeros with “missing.”

• **Real property, 1850–1870 (census).** Measures “the contemporary dollar value of any real estate owned by the respondent” as reported to the census. It is not clear
whether zeros indicate missing values or true zero personal property, and therefore we replace zeros with “missing.”

- **Homeownership, 1850–1940 (census).** Measures whether the individual rents or owns their home. For 1900 to 1940, the census reports homeownership directly. For 1850 to 1870, we follow Collins and Margo (2011) in imputing homeownership status using information on wealth, where every household with positive real property is classified as owner-occupied. Collins and Margo (2011) exempt households who live in multi-family homes from this classification but the information necessary to follow them in doing so is not included in the full-count version of the census we use. However, creating homeownership proxies using their and our method yields a correlation of 0.9733 in the 1 percent sample.

- **House value, 1930–1940 (census).** Measures the house value conditional on owning the house.

### A.6 Neighborhood-Level Outcome Variables

While we cannot link our data to censuses after 1940, we can link the 1940 census to administrative death records from 1988 and 2005 using the CenSoc-Numident file (Goldstein et al., 2021). Importantly, the death records contain the nine-digit ZIP codes of residence at the time of death. We link these codes to statistical census geographic areas, i.e., census tracts, block groups, and blocks (see Section A.8 for more detail on the procedure). Census tracts contain between 1,200 and 8,000 people and are designed to be “relatively homogeneous units with respect to population characteristics, economic status, and living conditions” (Census Bureau, 2017). Block groups (between 600 and 3,000 people) and blocks are subdivisions of a census tract.

We assigned to each decedent various socioeconomic characteristics based on these statistical areas at the time of death. Since the sample is about evenly split between deaths before 2000 and deaths after 2000, we used the aggregated census data for the year 2000 from the NHGIS database. For variables from other sources, we selected the data to refer to a period as close to 2000 as availability allowed.

#### Income

- **Income, 2000 (NHGIS).** The median household income by race of householder. Available by ZCTA, census tracts, and block groups.
Wealth

- **House value, 2000 (NHGIS).** The median value of owner-occupied housing units by race of householder. Available by ZCTA and census tracts.

- **Homeownership, 2000 (NHGIS).** The share of occupied housing units that is occupied by the owner (relative to a renter) by race. Available by ZCTA, census tracts, block groups, and blocks.

Education

- **High school degree, 2000 (NHGIS).** The share of the population over 25 years old by race and sex who hold a high school degree. Available by ZCTA, census tracts, and block groups.

- **College degree, 2000 (NHGIS).** The share of the population over 25 years old by race and sex who hold a college degree. Available by ZCTA, census tracts, and block groups.

Health

- **Age at death, 1988–2005 (BUNMD, Goldstein et al., 2021).** The median age at death by race and sex. Available by five-digit ZIP code, census tracts, block groups, and block.


- **Physical health, 2017–2018 (CDC, 2020).** Prevalence of poor physical health in the last 14 days among individuals over 18 years. Available by ZCTA and census tracts.

- **Mental health, 2017–2018 (CDC, 2020).** Prevalence of poor mental health in the last 14 days among individuals over 18 years. Available by ZCTA and census tracts.

Miscellaneous

- **Evictions, 2000–2005 (Desmond et al., 2018).** Number of eviction filings and evictions per 100 renter-occupied households. Available by census tracts and block groups.

- **Fatal police encounters, 2000–2021 (fatalencounters.org).** “Fatal Encounters documents non-police deaths that occur when police are present or are precipitated by police action or presence. Officer deaths are included when caused by another
officer, including friendly fire incidents, and criminal actions—like domestic violence—and suicides that occur when other officers are present. Officer vehicle-related deaths are included when they are caused by another officer. Homicides of officers by felons or deaths in the regular course of duties are not generally documented in the database.” The data cover the entire US from 2000 to September 09, 2021, but the database is continuously updated by journalist D. Brian Burghart. Available by five-digit ZIP code.

- **Racial segregation, 2000 (NHGIS).** The Theil Index of racial segregation using the racial composition of a census tract or block group relative to the block that it contains (Theil, 1972). For the exact formula, see Equation (4) in Chetty et al. (2014). Whereas Chetty et al. (2014) compute the measure on the level of the commuting zone relative to the census tracts it contains, we compute the measure on the level of a census tract and block group relative to the block. Available by census tracts and block groups.

- **Percentage Black, 2000 (NHGIS).** The share of the population that is Black. Available by ZCTA, census tracts, block groups, and blocks.

### A.7 County Characteristics

To assess the characteristics of “good” and “bad” counties and the persistence over time, we compile a dataset on the county level. (Manson et al., 2021).

- **Distance to the North, East (NHGIS).** A county’s distance to the North and the East is proxied by its centroid’s latitude and longitude.

- **Free, 1860 (NHGIS).** Measures the percentage of a county’s 1860 Black population that is free.

- **Black, 1860 (NHGIS).** Measures the percentage of a county’s 1860 population that is Black.

- **Tobacco, cotton, rice, and sugar, 1860 (NHGIS).** Measures the value of a county’s tobacco, cotton, rice, or sugar output as a percentage of the total agricultural output in 1860.

- **Population density, 1870 (NHGIS).** Measures a county’s 1870 population per square kilometer area.

- **School, 1870 (NHGIS).** Measures the fraction of a county’s Black children (ages 6–16) attending school in 1870.
• **Farm, 1870 (NHGIS)**. Measures the fraction of a county’s population living on a farm in 1870.

• **Migration cost North, 1870 (Donaldson and Hornbeck, 2016)**. Measures the transportation cost through land and water ways from a given county to the Northern cities that were the main destinations of the Great Migration: Chicago, Detroit, Pittsburgh, and New York. The migration cost estimates are based on the 1870 railroad network.

• **Lynchings, 1883–1941 (Seguin and Rigby, 2019)**. Measures the number of lynchings that occurred in a county between 1883 and 1941.

• **Racial segregation, 1880 and 1940 (Logan and Parman, 2017)**. Measures racial segregation based on a comparison of the probability of different-race neighbors in a county relative to the counterfactual probability had the population been randomly distributed across the county.

• **Slaves per slaveholder, 1860 (NHGIS)**. The average number of enslaved people per slaveholder.

• **Intergenerational mobility, 1996–2012 (Chetty and Hendren, 2018)**. Measures the causal effect of a county on the expected rank in the national income distribution conditional on one’s parents’ income ranking at the 25th percentile during childhood.

### A.8 Nine-Digit ZIP to Census 2000 Crosswalks

The administrative death records contain nine-digit ZIP codes (“ZIP9”) of the place of residence at the time of death. We use the Census Bureau’s TIGER/Line ASCII files (1994, 1995, 1997, 1998, 1999, 2000, 2002, 2003, 2005, and 2006) to link ZIP9s to 2000 census statistical areas (i.e., census blocks, block groups, and census tracts). A ZIP9 is a characteristic of a range of addresses, usually a side or segment of a street. The relevant records in the TIGER/Line files for our purpose are record types 1, 6, and Z. Each entry in record type 1 represents a complete chain (a street segment) and contains the five-digit ZIP ("ZIP5") for the main address range of the complete chain. It also contains the census block number of the polygon on either side of the complete chain.

Record type 6 provides remaining address ranges and their ZIP5s in case the relevant segment of a street is associated with multiple address ranges. Record type Z provides ZIP+4 add-on codes for each address range in record types 1 and 6. Merging the three record types, we obtain a database of ZIP9s and corresponding blocks. The TIGER/Line 63

---

63Extensive documentation is available online for each TIGER/Line version.
versions before 2000 linked ZIP9s to 1990 census areas. For those versions, we extract the ZIP9 and longitude and latitude of the beginning and end of the street segment that the complete chain corresponds with. Then, we map the street segment’s midpoint onto the 2000 census shape files.

In most cases, a ZIP9 maps into a unique block (and hence maps into a unique block group and census tract). For instance, in 2000, 81 percent of ZIP9s were matched to a unique block. For block groups and census tracts, 96 percent and 97 percent of the ZIP9 matches were unique, respectively. In cases where a ZIP9 occurs in more than one statistical area, we assign the area that has the largest number of matches with the relevant ZIP9. This yields a one-to-one mapping of ZIP9s to blocks for each TIGER/Line year between 2000 and 2006. However, not all ZIP9s in the Censoc-Numident death records occur in the TIGER/Line files. To improve the coverage, we sort the data by ZIP9 for each version and interpolate the census statistical areas in case the next non-missing census area is exactly equal to the previous non-missing area (using that the ZIP9s are ordered geographically).

Last, for each decedent, we assign the census area corresponding to their ZIP9 derived from a TIGER/Line version before and after their year of death (if available). For instance, if someone was born in 1996, we first try to assign the census area based on the TIGER/Line in 1995 and in 1997. If either of them is not available, we try to match using the next proximate version. Using this procedure, we link around 84 percent of the decedents with ZIP9s to a census tract, 82 percent to a block group, and 77 percent to a block. For decedents for which we can find the census area corresponding to their ZIP9 both before and after their death, the agreement rate between the different versions is high (98 percent for census tracts, 96 percent for block groups, and 88 percent for blocks).

**References**


Slavery and Dispossession in Rutgers History, ed. by M. J. Fuentes and D. G. White, Rutgers University Press, 91–122.


CONSTITUTION OF VERMONT (1777): https://avalon.law.yale.edu/18th_century/ vt01.asp.


B. Socioeconomic Status among Descendants of the Free and the Enslaved

Figure B.4: Socioeconomic Characteristics of Family by Region of Origin (1870-1940)

Notes: This figure shows the averages of characteristics in the cross-section of prime-age male descendants of the Free and the Enslaved by their ancestor’s region (family’s residence pre-1880). Incomes Score uses the Lido score developed by Saavedra and Twinam (2020). Skill level of occupations is inferred from the classification by Leeuwen and Maas (2011). See Data Appendix A for details on the sample and data.
C. MIGRATION AMONG DESCENDANTS OF THE FREE AND THE ENSLAVED

Figure C.5: Population of the Free and the Enslaved by Region (1790-1940)

(A) The Enslaved and Their Descendants

(B) The Free and Their Descendants

Notes: This figure shows the number of enslaved (Panel A) and free (Panel B) Black Americans and their descendants across regions. Dashed lines represent aggregate counts as reported in historical records (Berlin, 1974). Solid lines represent population counts inferred from our data. We compute the fraction of the Free and the Enslaved living in each region and multiply by the total number of Black Americans who descend from the Free and the Enslaved respectively. This procedure assumes that both population grow at a similar rate, implying that each accounts for a fraction of the Black population that is constant at the 1860 level.
Figure C.6: Probability of Participating in the Great Migration

Notes: This figure shows the fraction of descendants of the Free and the Enslaved who live in the North in 1900 (before the Great Migration), 1940 (after the first wave of the Great Migration), and 2000 (around 30 years after the end of the Great Migration), conditional on their family residing in the South in 1870. The error bar indicates the 95 percent confidence interval for the difference between the two groups. Descendants of free Black Americans are 50 percent more likely to participate in the Great Migration until 1940 than descendants of the Enslaved. The gap remains constant even after controlling for their socioeconomic status. Controlling for their family’s county of residence in 1870, however, shrinks this gap by half. The geographic location of enslaved ancestors substantially delayed the participation in the Great Migration of their descendants, potentially also delaying their economic progress after slavery.

Figure C.7: Black Families Leaving the Slave States by 1870 State of Origin

(A) 1870-1880
(B) 1870-1900
(C) 1870-1920
(D) 1870-1940
(E) 1870-2000
FIGURE C.8: County Population of Enslaved and Free (1790 & 1860)

(A) The Enslaved in 1790

(B) The Free in 1790
(C) The Enslaved in 1860

(D) The Free in 1860
D. ADJUSTING ESTIMATES FOR MISCLASSIFICATION BIAS

We may incorrectly classify a substantial number of Black families as descendants of the Enslaved if they are born in a slave state where a large share of Black Americans was free before 1860. For example, among Black Americans from Maryland, around 50 percent are free in 1860. Due to imperfect linking rates, many of those free Black families will be misclassified as Enslaved by our method.

We adjust our estimates for bias that may arise from this type of misclassification. We make use that our original estimates are a weighted average of the (unknown) unbiased estimate and the non-causal estimate for free Black Americans:

\[
\hat{\beta}_{\text{original}} = \frac{\text{Enslaved}_{s,\text{links}}}{\text{Enslaved}_{s,1860}} \cdot \hat{\beta}_{\text{unbiased}} + \left(1 - \frac{\text{Enslaved}_{s,\text{links}}}{\text{Enslaved}_{s,1860}}\right) \cdot \hat{\beta}_{\text{free}},
\]

where \(\text{Enslaved}_{s,\text{links}}\) is the share of Black Americans who descend from the Enslaved of state \(s\) according to our classification in 1940, \(\text{Enslaved}_{s,1860}\) is the true share of Black Americans who descend from the Enslaved of state \(s\) according to the 1860 census, and \(\hat{\beta}_{\text{free}}\) is the non-causal estimate for outcomes of free Black families with ancestors from state \(s\).

In our sample, for instance, 70 percent of Black Americans with ancestors from Maryland are classified as descendants of the Enslaved in 1940—almost 20 percentage points more than the true share. Comparing the two groups suggests, unsurprisingly, that the free families of Maryland achieved somewhat higher levels of education by 1940 than descendants of the Enslaved. As a consequence, our original estimates (that incorrectly include some free families) are upward biased estimates of the true effect living in Maryland had on descendants of the Enslaved. Using the equation above, we can adjust our estimates.

Adjusting our estimates for bias that may result from this form of misclassification turns out not to quantitatively matter. Figure D.9 shows that the share of Black Americans that descend from the Enslaved only deviates substantially from our classification for three small slave states. Accordingly, adjusting our original estimates of the causal effect of each state barely affects our estimates. Panel (A) of Figure D.10 shows that the original estimates tend to almost perfectly coincide with the adjusted estimates. Panel (B) shows the map of states’ effects before and after the adjustment.
Notes: This figure assesses on misclassification of the Free-Enslaved status and the impact misclassification has on our estimates. Panel (A) shows the extent of misclassification as descendants of the Enslaved or the Free among Black Americans in 1940 with ancestors born in a given state before 1870. Panel (B) shows our causal estimates of living in each state before and after adjusting for misclassification bias. See Data Appendix A for details on the sample and data.

Notes: This figure compares the geographic patterns in our original estimates of states’ effect on Black economic progress after slavery with the adjusted estimates. See Data Appendix A for details on the sample and data.

E. ADJUSTING ESTIMATES FOR INTERMARRIAGE

Our empirical strategy identifies the gap between Black individuals whose ancestors along the male line were enslaved until the Civil War and those whose ancestors along the male line were free before the Civil War. Women tended to change their last names upon marriage, making it
impossible to assess the Free-Enslaved status of female ancestors. Therefore, some individuals who are classified as descendants of one group, may in truth also partly descend from the other group. This type of misclassification biases our estimate of the Free-Enslaved gap towards zero.

We use information on a person’s birth place, their parents’ birth places, and their spouse’s birth place to assess the degree of intermarriage. In particular, we can assess the degree to which Black Americans from slave states and free states intermarried to proxy for how common intermarriage was. Note that the Free and the Enslaved were distributed very differently across America.

**Figure E.11: Black Intermarriages between Descendants of the Free and the Enslaved**

![Graph showing Black Intermarriages between Descendants of the Free and the Enslaved](image)

Figure E.11 shows the fraction of Black marriages that constituted intermarriage between descendants of the Free (mother born in free state prior to 1865) and the Enslaved (mother born in slave state prior to 1865). By 1940, less than 2 percent of all Black marriages constituted such intermarriage. Therefore, those who we classify as descendants of the Enslaved likely have no free ancestors. However, because the number of the Free is small, even this small degree of intermarriage makes it likely that some fraction of the people we classify as Free in 1940 have some ancestor who was enslaved.

We can use the aggregate proxy for intermarriage to correct the original estimate of the Free-Enslaved gap as follows:

\[
\hat{\beta}_t^\text{true} = \overline{Y}_t^\text{Ensl} - \overline{Y}_t^\text{Free}
\]

\[
\hat{\beta}_t^\text{original} = \left( s_{t,\text{Ensl}}^\text{IM} \cdot \overline{Y}_t^\text{Ensl} + [1 - s_{t,\text{Ensl}}^\text{IM}] \cdot \overline{Y}_t^\text{Free} \right) - \left( s_{t,\text{Free}}^\text{IM} \cdot \overline{Y}_t^\text{Free} + [1 - s_{t,\text{Free}}^\text{IM}] \cdot \overline{Y}_t^\text{Ensl} \right)
\]

where \(\hat{\beta}_t^\text{true}\) is the (unknown) gap between a person who descends only from free Black Americans and a person who only descends from enslaved people, \(\overline{Y}_t^\text{Ensl}\) and \(\overline{Y}_t^\text{Free}\) are the average outcomes of descendants of the Enslaved and Free, and \(s_{t,\text{Ensl}}^\text{IM}\) and \(s_{t,\text{Free}}^\text{IM}\) are the share of En-
slaved and Free who engage in intermarriage. Using our estimates for 1940, we have

\[
\hat{\beta}_{1940}^{\text{original}} \approx (0.01 \cdot Y_t^{\text{Ensl}} + 0.99 \cdot Y_t^{\text{Free}}) - (0.33 \cdot Y_t^{\text{Free}} + 0.67 \cdot Y_t^{\text{Ensl}}) \\
\approx 0.66 \cdot (Y_t^{\text{Free}} - Y_t^{\text{Ensl}})
\]

Thus, the true Free-Enslaved gap is around one-and-a-half times as large in 1940 as our original estimates would suggest. For earlier decades, the adjustment factor is somewhat smaller.

F. THE DIRECT EFFECT OF LOCATIONS AFTER ACCOUNTING FOR MIGRATION

Our estimates of how being freed in a given location affected the economic progress of Black families reflects both the effect the original location and the expected effects of future locations conditional on the 1870 location. We decompose this intent-to-treat (ITT) effect into (1) the direct effect of the location of origin and (2) the potential indirect effect of destinations that Black families may migrate to. To do so, we invert the location-to-location migration probability matrix to back out the ATE of living in each state from the estimated ITT effect.

Figure F.12: ITT Effect and ATE of Living in Each Southern State (1870–1940) on Years of Education in 1940

Notes: This figure compares our original (ITT) estimates of how being freed in a given state affected a Black family’s economic progress to the direct (ATE) effect that living in that state had. The estimates are in years of education in 1940. See Data Appendix A for details on the sample and data.
G. Empirical Bayes Shrinkage of Place Effects

When estimating the place effects with many geographic units (counties), a common problem is that some estimates may be very noisy. While these estimates are unbiased, they are on average further from the truth—in a total squared error sense—than optimal (Efron, 2010). Shrinkage techniques address this problem.

Empirical Bayes methods have become a popular means to shrink noisy estimates (e.g., Angrist et al., 2017; Chetty and Hendren, 2018). The method is motivated by the fact that under the assumption of place effects resulting from a common (unknown) distribution, the optimal point estimator has the form of a Bayesian posterior mean (Armstrong et al., 2021). One does not need to make any assumptions on the specific distribution that the place effects result from.

Figure G.13: Example of Empirical Bayes Shrinkage With a Single Covariate

Panel (A) shows the preliminary estimates and 95 percent confidence bands clustered at the family level. Panel (B) shows the estimates after shrinking them toward the regression line based on the county’s Black population share and robust 95 percent empirical Bayes confidence intervals (Armstrong et al., 2021). County FEs based on ten observations or less are discarded. See Data Appendix A for details on the sample and data.

Figure G.13 provides a simple example of empirical Bayes shrinkage with a single covariate. Panel (A) shows the negative correlation between the preliminary estimates of a county’s causal effect on years of education in 1940 and county’s share of the population that is Black in 1860. Panel (B) shows the empirical Bayesian shrinkage estimates. Noisier preliminary estimates are pulled toward the regression line more drastically than more precise preliminary estimates.

We employ a empirical Bayes shrinkage to our baseline county-effects. We provide two forms of shrinkage estimates. The first set does not use covariates, shrinking the baseline estimates toward a common mean. The second set includes covariates, shrinking the baseline estimates toward the place effect predicted by the covariates. We prefer the second set of estimates because they leverage a larger set of information in the shrinkage process.
The empirical Bayes estimate for county $c$ including covariates takes the form

$$
\hat{\theta}_c = X'_c \hat{\delta} + \frac{s}{s + \hat{\sigma}_c^2} (\hat{\eta}_c - X'_c \hat{\delta}),
$$

(13)

where $\hat{\eta}_c$ is the preliminary estimate of county $c$’s effect, $\hat{\delta} = \left( \sum_{c=1}^N \hat{\sigma}_c^{-2} X_c X'_c \right)^{-1} \sum_{c=1}^N \hat{\sigma}_c^{-2} X_c \hat{\eta}_c$ is the ordinary least squares estimate of $\hat{\eta}_c$ on the county covariates $X_c$, $\hat{\sigma}^2_c$ is the standard error of $\hat{\eta}_c$, and $s = \max \left\{ \frac{-N+\sum_{c=1}^N \hat{\sigma}_c^{-2} \hat{\epsilon}_c^2}{\sum_{c=1}^N \hat{\sigma}_c^{-2} \hat{\epsilon}_c^2} \right\}$ with $\hat{\epsilon}_c = \hat{\eta}_c - X'_c \hat{\delta}$. The shrinkage estimate is therefore a weighted average of the preliminary county fixed effect and the predicted effect based on the county’s characteristics, with greater weight assigned to a preliminary fixed effects when it is more precisely estimated.

Figure G.14 shows a comparison of the place effects before and after the empirical Bayes shrinkage. While the general pattern of negative effects being concentrated in the Lower South holds before and after, the shrunk estimates are far more spatially correlated.

**Figure G.14: Causal Place Effects on 1940 Years of Education**

(A) Preliminary Estimates

(B) Shrinkage Estimates (No Covariates)

(C) Shrinkage Estimates (Covariates)

(D) Shrinkage Estimates (Covariates & State FEs)

Notes: This figure shows the 1870 ancestor county fixed effect (FE) estimates on 1940 years of education for descendants of the Enslaved. Panel (A) shows the preliminary estimates. Panel (B) shows the estimates after shrinking them to their common mean. Panel (C) shows the estimates after shrinking them to the regression line based on various covariates. Panel (D) repeats Panel (C) adding state FEs. County FEs based on ten observations or less are discarded. See Data Appendix A for details on the sample and data.
Figure G.15: Place Effects Across Groups Before and After Shrinkage

(A) Preliminary Estimates

(B) Shrinkage Estimates

(C) Preliminary Estimates

(D) Shrinkage Estimates

Notes: This figure compares the 1870 ancestor county fixed effect estimates on years of education in 1940 for descendants of the Enslaved (causal) with those of white Americans and descendants of free Black Americans (both non-causal). Panels (A) and (C) show the estimates before shrinkage, Panels (B) and (D) show the estimates after shrinkage. The shrinkage does not preserve a county’s original rank. County fixed effects based on ten observations or less are discarded. See Data Appendix A for details on the sample and data.

References


H. Placebo Exercises

In two types of placebo exercises, we test our method of quantifying the Free-Enslaved gap. First, we estimate the placebo Free-Enslaved gap for white Americans. White families that cannot be linked to the 1850 or 1860 censuses are classified as (placebo) descendants of the Enslaved. This exercise may not yield pure placebo estimates because families immigrating after 1860 may be different from those who immigrated earlier. Table H.3 shows that the composition of white Americans indeed experienced some changes after 1860. White Americans grew more likely to be first-generation immigrants and among those immigrants fewer came from the United Kingdom and more from Northern, Central, and Eastern Europe. When evaluating the placebo estimates, we should bare in mind that these changes contaminate a pure placebo.

Second, we estimate the Free-Enslaved gap using 1875 as the (placebo) end of slavery. Figure 2 already suggests that there is no gap between Black Americans who can be linked back to 1880 (but not 1870 or earlier) and those who can be linked back to 1870 or earlier. In this section, we also estimate the placebo Free-Enslaved gap based on the change in the distribution of last names from 1870 to 1880.

Beyond the two placebo exercises, we also estimate the effect that the 1870 ancestor birthplace has on white Americans’ socioeconomic status in 1940. Because we do not have exogenous variation in the ancestor birthplace for free people such as white Americans, these estimates are non-causal. We also trace the speed of convergence in the socioeconomic status of white families whose 1870 ancestors did or did not have physical or human capital. This exercise yields a benchmark for the speed of convergence between descendants of the Enslaved and the Free. Lastly, we estimate the intergenerational effect that the Lower South’s institutions had on white Americans using our border discontinuity design.

Table H.3: White Americans Before and After 1865

<table>
<thead>
<tr>
<th>Region of Birth</th>
<th>Share in 1860 (%)</th>
<th>Share in 1870 (%)</th>
<th>Change (Ppts.)</th>
<th>Change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>86.5</td>
<td>85.3</td>
<td>-1.2</td>
<td>-1.4</td>
</tr>
<tr>
<td>Central America</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>25.3</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>0.2</td>
<td>1.0</td>
<td>0.8</td>
<td>337.3</td>
</tr>
<tr>
<td>UK</td>
<td>8.1</td>
<td>7.0</td>
<td>-1.0</td>
<td>-12.8</td>
</tr>
<tr>
<td>Western Europe</td>
<td>0.7</td>
<td>0.8</td>
<td>0.1</td>
<td>10.8</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>76.6</td>
</tr>
<tr>
<td>Central/Eastern Europe</td>
<td>4.2</td>
<td>5.4</td>
<td>1.3</td>
<td>30.7</td>
</tr>
</tbody>
</table>

Notes: This table shows the share of white American men born in each region in the 1850–1860 and the 1870–1880 censuses. Regions with shares below 0.1 percent in any year are omitted.
FIGURE H.16: Free-Enslaved Gap (1870-1940) vs. Placebo for White Americans

Notes: This figure shows the true and placebo gaps in literacy rates, income scores (OCCSCORE), occupation skill levels (HISCLASS), and homeownership rates among prime-age (20-54) male descendants of free vs. enslaved Black Americans in each census decade. The placebo applies the exact same procedure to the sample of white Americans. The comparison shows that some linking bias may affect results in early periods, but all of it vanishes over time. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. All estimates control for a quadratic function in age and include 95 percent confidence bands that are clustered at the family level. While literacy is reported directly in the census decades only until 1930, we use schooling in second grade as a proxy in 1940. See Data Appendix A for details on the sample and data.
### Table H.4: Placebo Free-Enslaved Gap (1940) for White Americans

<table>
<thead>
<tr>
<th>Placebo</th>
<th>Education (Years)</th>
<th>Wage Income (USD)</th>
<th>Home Ownership (%)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean: 9.76</td>
<td>Mean: 892.68</td>
<td>Mean: 49.74</td>
<td>Mean: 3,284.56</td>
</tr>
<tr>
<td>-0.17***</td>
<td>(0.00)</td>
<td>-1.68</td>
<td>0.09</td>
<td>12.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls (age, age&lt;sup&gt;2&lt;/sup&gt;)</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.03</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Observations</td>
<td>5,015,270</td>
<td>4,770,969</td>
<td>5,012,884</td>
<td>2,425,204</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>3,158,604</td>
<td>3,001,138</td>
<td>3,155,980</td>
<td>1,536,909</td>
</tr>
</tbody>
</table>

Notes: This table shows the placebo gaps in years of education, total income, homeownership, and house value among prime-age (20-54) male Black Americans in 1940. The placebo applies the exact same procedure to the sample of white Americans. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

### Table H.5: Placebo vs. Free-Enslaved Gap Based on the Distribution of Last Names (1940)

<table>
<thead>
<tr>
<th>P(Ancestor Enslaved)</th>
<th>Education (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.47***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>Placebo: 1875</td>
<td>-0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Placebo: White</td>
<td>-0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Placebo: 1875 × White</td>
<td>-0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name-measure</th>
<th>Exact</th>
<th>Exact</th>
<th>Exact</th>
<th>Exact</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls (race, race × age, race × age&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Observations</td>
<td>2,859,747</td>
<td>24,267,079</td>
<td>2,859,747</td>
<td>27,126,826</td>
<td>27,126,826</td>
</tr>
</tbody>
</table>

Notes: This table repeats Table 1 showing the gap in years of education, total income, homeownership, and house value among prime-age (20-54) male descendants of free vs. enslaved Black Americans in 1940. The sample includes the entire universe of prime-age Black men, not just those linkable. The coefficients can be interpreted as a 100 percentage point increase in the likelihood of descending from the Enslaved based on their (exact) last name. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Robust standard errors are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
**TABLE H.6: Placebo Free-Enslaved Gap (2000) at Census Tract Level**

<table>
<thead>
<tr>
<th></th>
<th>HS Degree (%)</th>
<th>College Degree (%)</th>
<th>Income (USD)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean:</td>
<td>Mean: 83.52</td>
<td>Mean: 25.52</td>
<td>Mean: 46,123.73</td>
<td>Mean: 133,616.06</td>
</tr>
<tr>
<td><strong>Ancestor Enslaved</strong></td>
<td>-0.11**</td>
<td>0.22***</td>
<td>133.37</td>
<td>908.63*</td>
</tr>
<tr>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(91.00)</td>
<td>(499.87)</td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls (age, age²)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Observations</td>
<td>437,099</td>
<td>437,099</td>
<td>437,076</td>
<td>435,809</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>383,221</td>
<td>383,221</td>
<td>383,213</td>
<td>382,049</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the placebo gap in neighborhood-level outcomes among male white Americans around 2000. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. We estimate the gap in terms of the fraction of people who hold a high school or college degree, the median income earned, and the median house value. House values are measured conditional on ownership. Each person is assigned the respective value of the census tract in which they lived at the time of death. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**FIGURE H.17: Place Effects Across Poor Groups**

(A) Immigrants from Ireland, Italy, and Poland

(B) White with no Physical or Human Capital

**Notes:** This figure compares the 1870 ancestor county fixed effect estimates on years of education in 1940 for descendants of the Enslaved (causal) with those of white Americans who descend from Irish, Italian, or Polish immigrants ($\rho = -0.041$, Panel A) and white Americans whose ancestors were illiterate and had zero wealth in 1870 ($\rho = 0.151$, Panel B). A county’s fixed effect is the deviation from the race-specific population-weighted average across all Southern counties after controlling for a quadratic function of age. The sample includes only prime-age (20–54) men whose ancestors can be located in 1870. All county fixed effects based on ten observations or less are discarded. See Data Appendix A for details on the sample and data.
Figure H.18: Benchmark for Speed of Convergence—White Americans Whose Ancestors Did vs. Did Not Have Any Physical or Human Capital

Notes: This figure shows the gaps in literacy, income, skill, and homeownership among white prime-age (20-54) male descendants of ancestors with vs. without any physical or human capital in 1870. Physical capital is measured in terms of real and personal property; Human capital is measured in terms of literacy. The comparison yields a benchmark for the convergence of large socioeconomic gaps from 1870 to 1940. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. All estimates control for a quadratic function in age and include 95 percent confidence bands that are clustered at the family level. While literacy is reported directly in the census decades only until 1930, we use schooling in second grade as a proxy in 1940. See Data Appendix A for details on the sample and data.
I. ADDITIONAL RESULTS

I.1 The Free-Enslaved Gap in Alternative Measures

**FIGURE I.19: Free-Enslaved Gap (1870–1940)**

(A) Income

(B) Homeownership

Notes: This figure shows the gaps in income (occupational income score) and homeownership among prime-age (20-54) male descendants of free vs. enslaved Black Americans in each census decade. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix A for details on the sample and data.
**Figure I.20: Free-Enslaved Gap Conditional on Ancestor State (1870-1940)**

(A) Literacy

(B) Income Score

(C) Occupational Skill

(D) Homeownership

**Notes:** This figure shows the gaps in literacy, income, skill, and homeownership before (light) and after (dark) including fixed effects for 1870 ancestor state of birth. The comparison is made between prime-age (20-54 years) male descendants of free vs. enslaved Black Americans in each census decade. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included, minimizing bias due to the fact that the Free by definition have a link to 1850 or 1860. Both panels control for age and include 95 percent confidence bands clustered at the family level. See Data Appendix A for details on the sample and data.

**Table I.7: Free-Enslaved Gap (1940) in Different Income Measures**

<table>
<thead>
<tr>
<th></th>
<th>OCCSCORE (1950-USD)</th>
<th>LIDO Score (1950-USD)</th>
<th>Wage Income (1940-USD)</th>
<th>Total Income (1940-USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean: 1,604.09</td>
<td>Mean: 1,161.69</td>
<td>Mean: 381.20</td>
<td>Mean: 793.47</td>
</tr>
<tr>
<td>Ancestor Enslaved</td>
<td>-148.39***</td>
<td>-279.00***</td>
<td>-145.92***</td>
<td>-204.29***</td>
</tr>
<tr>
<td></td>
<td>(10.86)</td>
<td>(8.59)</td>
<td>(6.13)</td>
<td>(10.29)</td>
</tr>
<tr>
<td>Controls (age, age²)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Observations</td>
<td>168,138</td>
<td>142,743</td>
<td>154,463</td>
<td>146,871</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>9,325</td>
<td>7,517</td>
<td>8,551</td>
<td>8,100</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the Free-Enslaved gap in income across different measures: Occupational income score (OCCSCORE), a refined occupational income score (Lido), wage income, and total predicted income. All estimates are for Black prime-age men in 1940. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
TABLE I.8: Free-Enslaved Gap (1940) in Different Income Measures

<table>
<thead>
<tr>
<th>Ancestor</th>
<th>Literacy (%)</th>
<th>Education (Years)</th>
<th>High School (%)</th>
<th>College (%)</th>
<th>Graduate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enslaved</td>
<td>Mean: 91.49</td>
<td>Mean: 5.99</td>
<td>Mean: 9.28</td>
<td>Mean: 1.70</td>
<td>Mean: 0.46</td>
</tr>
<tr>
<td></td>
<td>-4.25***</td>
<td>-1.59***</td>
<td>-7.86***</td>
<td>-1.86***</td>
<td>-0.74***</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.05)</td>
<td>(0.45)</td>
<td>(0.21)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Controls (age, age²)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Observations</td>
<td>163,549</td>
<td>163,549</td>
<td>163,549</td>
<td>163,549</td>
<td>163,549</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>9,078</td>
<td>9,078</td>
<td>9,078</td>
<td>9,078</td>
<td>9,078</td>
</tr>
</tbody>
</table>

Notes: This table shows the Free-Enslaved gap in education across different measures: Literacy, years of education, and the probability of holding a high school, college, or graduate degree. As literacy is reported only until 1930, we use educational attainment beyond second grade as a proxy in 1940. All estimates are for Black prime-age men in 1940. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

I.2 The Free-Enslaved Gap in Intergenerational Mobility

To determine if the persisting Free-Enslaved gap is purely the result of large initial differences or if there are also differences in intergenerational mobility, we estimate the intergenerational mobility of person $i$ between 1870 and 1940 as the following:

$$ r_{i,1940} = \alpha + \beta_1 s_i + \beta_2 r_{i,1870} + \beta_3 s_i \times r_{i,1870} + \epsilon_i, $$

(14)

where $r_{i,1940}$ is the percentile rank of $i$’s 1940 income or wealth in the Black distribution and $r_{i,1870}$ is the percentile rank of the income or wealth occupied by the prime-age male ancestor of $i$ in 1870. Upward mobility is defined as the increase in percentile ranks from one generation to the next. To reduce the impact of measurement error, we use the average across 1870 and 1880 for ancestors if possible (Ward, 2021).

Figure I.21 shows that upward mobility is significantly lower for descendants of the Enslaved. Across all parts of the 1870–1880 income distribution and the 1870 wealth distribution, descendants of the Enslaved reach lower expected income and wealth levels in 1940.
FIGURE I.21: Gaps in Intergenerational Mobility (1870-1940)

(A) Income-Rank Mobility

(B) Wealth-Rank Mobility

Notes: This figure shows the estimated intergenerational mobility of free Black (in black), formerly enslaved (in red), and white prime-age (20-54) men. The left panel shows income-rank mobility conditional on the average rank of an ancestor’s occupational income score (OCCSCORE) in 1870 and 1880 (Ward, 2021). The right panel shows wealth-rank mobility conditional on the rank of an ancestor’s wealth (personal property and real property) in 1870. 1940 total income is predicted (see Appendix A.5). 1940 house values are measured conditional on ownership. The lines reflect the coefficients estimated in Equation 14, including robust confidence bands at the 95 percent level. Bubbles reflect binned averages. Percentile ranks reflect rank that each income occupies in the year-specific national income distribution of Black prime-age men. See Data Appendix A for details on the sample and data.

REFERENCES

### I.3 The Free-Enslaved Gap Based on the Distribution of Last Names

**Table I.9: Free-Enslaved Gap Based on the Distribution of Last Names (1940)**

<table>
<thead>
<tr>
<th></th>
<th>Education (Years)</th>
<th>Wage Income (USD)</th>
<th>Homeownership (%)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean: 5.71</td>
<td>Mean: 598.74</td>
<td>Mean: 21.89</td>
<td>Mean: 1,599.75</td>
</tr>
<tr>
<td><strong>P(Ancestor Enslaved)</strong></td>
<td>-0.47***</td>
<td>-0.54***</td>
<td>-13.73***</td>
<td>-29.89***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(5.26)</td>
<td>(7.17)</td>
<td>(0.21)</td>
</tr>
<tr>
<td></td>
<td>-2.43***</td>
<td>-2.43***</td>
<td>-630.53***</td>
<td>-970.17***</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(277.72)</td>
<td>(506.45)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Name-measure</th>
<th>Exact NYSIIS</th>
<th>Exact NYSIIS</th>
<th>Exact NYSIIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls (age, age$^2$)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R$^2$</td>
<td>0.03</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>2,859,747</td>
<td>2,821,235</td>
<td>2,842,572</td>
<td>601,789</td>
</tr>
</tbody>
</table>

**Notes:** This table repeats Table 1 showing the gap in years of education, total income, homeownership, and house value among prime-age (20-54) male descendants of free vs. enslaved Black Americans in 1940. The sample includes the entire universe of prime-age Black men, not just those linkable. The coefficients can be interpreted as a 100 percentage point increase in the likelihood of descending from the Enslaved based on their (exact) last name. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Robust standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

### I.4 The Free-Enslaved Gap for Free Without Physical and Human Capital

**Table I.10: Free-Enslaved Gap (1940) for Free Without Physical and Human Capital in 1860**

<table>
<thead>
<tr>
<th></th>
<th>Education (Years)</th>
<th>Wage Income (USD)</th>
<th>Homeownership (%)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean: 5.83</td>
<td>Mean: 381.64</td>
<td>Mean: 29.08</td>
<td>Mean: 1,380.43</td>
</tr>
<tr>
<td><strong>Ancestor Enslaved</strong></td>
<td>-1.00***</td>
<td>-0.12</td>
<td>26.85</td>
<td>-1.42</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(21.13)</td>
<td>(21.44)</td>
<td>(2.00)</td>
</tr>
<tr>
<td></td>
<td>-343.74***</td>
<td>440.28***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1870 State of Birth-FE</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Controls (age, age$^2$)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R$^2$</td>
<td>0.03</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Observations</td>
<td>71,574</td>
<td>71,574</td>
<td>67,672</td>
<td>67,672</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the gap in years of education, total income, homeownership rate, and house value among prime-age (20-54) male descendants of a subset of the Free vs. enslaved Black Americans in 1940. Among the Free, we only include those whose ancestors had no measurable physical capital (real and personal property) or human capital (literacy) in 1860. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Columns 1, 3, 5, and 7 repeat Table 1 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. Appendix Figure I.20 shows the evolution of the conditional Free-Enslaved gap over time. See Data Appendix A for details. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 

89
### I.5 The Free-Enslaved Gap between and within Ancestor’s Birthplace

#### Table I.11: Free-Enslaved Gap (1940) between and within Ancestor’s Birthplace

<table>
<thead>
<tr>
<th>Ancestor Enslaved</th>
<th>Education (Years)</th>
<th>Wage Income (USD)</th>
<th>Home Ownership (%)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean: 5.91</td>
<td>Mean: 388.01</td>
<td>Mean: 29.48</td>
<td>Mean: 1,412.17</td>
<td></td>
</tr>
<tr>
<td>-1.49***</td>
<td>-0.41***</td>
<td>-137.00***</td>
<td>-120.22**</td>
<td></td>
</tr>
<tr>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(8.51)</td>
<td>(9.84)</td>
<td></td>
</tr>
<tr>
<td>-6.76***</td>
<td>-1.61</td>
<td>-574.06***</td>
<td>8.40</td>
<td></td>
</tr>
<tr>
<td>(0.86)</td>
<td>(1.04)</td>
<td>(90.08)</td>
<td>(115.61)</td>
<td></td>
</tr>
</tbody>
</table>

| 1870 State of Birth-FE | N | Y | N | Y | N | Y | N | Y |
| Adjusted R² | 0.04 | 0.08 | 0.04 | 0.07 | 0.01 | 0.03 | 0.01 | 0.03 |
| Observations | 75,583 | 75,583 | 71,474 | 71,474 | 76,048 | 76,048 | 21,873 | 21,873 |
| Ancestor Free | 4,617 | 4,617 | 4,371 | 4,371 | 4,640 | 4,640 | 1,624 | 1,624 |

**Notes:** This table shows the gap in years of education, total income, homeownership rate, and house value among prime-age (20-54) male descendants of free vs. enslaved Black Americans in 1940. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Columns 1, 3, 5, and 7 repeat Table 1 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and the Enslaved. Appendix Figure I.20 shows the evolution of the conditional Free-Enslaved gap over time. See Data Appendix A for details. Standard errors are clustered at the family level and are shown in parentheses. *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).

#### Table I.12: Free-Enslaved Gap (2000) at Census Block Level between and within Ancestor’s Birthplace

<table>
<thead>
<tr>
<th>Ancestor Enslaved</th>
<th>HS Degree (%)</th>
<th>College Degree (%)</th>
<th>Income (USD)</th>
<th>House Value (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean: 69.20</td>
<td>Mean: 12.32</td>
<td>Mean: 30,143.90</td>
<td>Mean: 88,830.12</td>
<td></td>
</tr>
<tr>
<td>-2.57***</td>
<td>-0.89</td>
<td>-2.07***</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td>(0.74)</td>
<td>(0.82)</td>
<td>(0.78)</td>
<td>(0.78)</td>
<td></td>
</tr>
<tr>
<td>-5,032.50***</td>
<td>-1,014.92</td>
<td>-1,391.02***</td>
<td>-780.04</td>
<td></td>
</tr>
<tr>
<td>(921.89)</td>
<td>(1,005.32)</td>
<td>(3,498.95)</td>
<td>(3,829.19)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Tract×Race×Sex</th>
<th>Tract×Race×Sex</th>
<th>Tract×Race</th>
<th>Tract×Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>1870 State of Birth-FE</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Controls (age, age²)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>11,931</td>
<td>11,931</td>
<td>11,931</td>
<td>11,931</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>863</td>
<td>863</td>
<td>863</td>
<td>863</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the Free-Enslaved gap in the fraction of people who hold a high school degree, the fraction of people who hold a college degree, the median income earned, and the median house value in 2000. Columns 1, 3, 5, and 7 repeat Table 2 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership and therefore exclude zeros. Each person is assigned the respective value of the census block in which they lived at the time of death. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).
Figure I.22: Free-Enslaved Gap in 1940 Years of Education by 1870 Ancestor Birthplace

Notes: This figure shows the gaps between descendants of Free and Enslaved in 1940 years of education by 1870 ancestor state of birth. The comparison is made between prime-age (20-54 years) male descendants in each census decade. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included, minimizing bias due to the fact that the Free by definition have a link to 1850 or 1860. Both panels control for age and include 95 percent confidence bands that are clustered at the family level.

I.6 The Free-Enslaved Gap in Mortality

Table I.13: Mortality Gaps between Descendants of the Free and the Enslaved (2000)

<table>
<thead>
<tr>
<th>Ancestor Enslaved</th>
<th>Life Expectancy (Years)</th>
<th>Fatal Police Encounters (per 100K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean: 74.94</td>
<td>Mean: 21.31</td>
</tr>
<tr>
<td></td>
<td>-1.08***</td>
<td>3.40***</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>Level</td>
<td>Tract</td>
<td>ZIP</td>
</tr>
<tr>
<td>Controls (age, age^2)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Observations</td>
<td>26,181</td>
<td>25,246</td>
</tr>
<tr>
<td>Ancestor Free</td>
<td>1,686</td>
<td>1,423</td>
</tr>
</tbody>
</table>

Notes: This table shows the Free-Enslaved gap in life expectancy and fatal police encounters per 100,000 residents. Each person is assigned the respective value of the census tract or five-digit ZIP code in which they lived at the time of death. Sample means are computed for the combined sample of the Free and the Enslaved. See Data Appendix A for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
I.7 The Intergenerational Effect of Exposure to State-Specific Factors

FIGURE I.23: Effect of 1870 Ancestor’s State of Birth on 1940 Outcomes

(A) Years of Education

(B) Literacy Rate

(C) Homeownership Rate

(D) Wage Income

(E) Total Predicted Income

Notes: This figure shows the 1870 ancestor state of birth fixed effect estimates on years of education, literacy rate, homeownership rate, wage income, and predicted total income in 1940. A state’s fixed effect is the deviation from the population-weighted average across all states after controlling for a quadratic function of age. The sample includes only Black prime-age men whose ancestors can be located in 1870.
**Figure I.24: Effect of Ancestor’s State of Birth of Enslavement on 2000 Outcomes**

- **(A) Income (2000 USD)**
- **(B) High School Degree (%)**
- **(C) College Degree (%)**

**Notes:** This figure shows the 1870 ancestor state of birth fixed effect estimates on income, the likelihood of holding a high school degree or college degree, and house values in 2000 for descendants of the Enslaved. A state’s fixed effect is the deviation from the population-weighted average across all states (baseline mean) after controlling for a quadratic function of age. The sample includes men and women whose ancestors can be located in 1870. All county fixed effects based on ten observations or less are discarded.
**Figure I.25: Effect of 1870 Ancestor’s State of Residence on 1940 Outcomes**

(A) Years of Education

(B) Literacy Rate

(C) Wage Income

(D) Homeownership Rate

(E) Total Predicted Income

Notes: This figure shows the 1870 ancestor state of residence fixed effect estimates on years of education, literacy rate, homeownership rate, wage income, and predicted total income in 1940. A state’s fixed effect is the deviation from the population-weighted average across all states after controlling for a quadratic function of age. The sample includes only Black prime-age men whose ancestors can be located in 1870.
I.8 The Persistence of Place Effects

**Figure I.26: Persistence of a County’s Capacity to Generate Upward Mobility**

![Graph showing the correlation between own childhood county effect on income rank in 2000s and 1870-ancestor county effect on years of education in 1940.](image)

Correlation = 0.503

**Notes:** This figure repeats Figure 14 at the county level. We use empirical Bayesian shrinkage as described in Appendix G. See Data Appendix A for details on the sample and data.

**Figure I.27: Persistence of a County’s Capacity to Generate Upward Mobility**

![Graph showing the correlation between own childhood county effect on income rank in 2000s and 1870-ancestor county effect on years of education in 1940.](image)

Correlation = 0.204

**Notes:** This figure repeats Figure 14 using estimates of a county’s effect on intergenerational mobility conditional on a child growing up with parents at the 75th (rather than the 25th) percentile of the national household income distribution. See Data Appendix A for details on the sample and data.
I.9 Additional Results from the Border Discontinuity Design

**Figure I.28: Pooled Regression Discontinuity (RD) Estimate**

![Graph showing RD estimate](image)

Discontinuity: -0.28*** (0.09)

Notes: This figure shows the RD estimate in 1940 years of education for Black families who were freed across state borders with different Jim Crow intensity in 1865. The left half of the figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level.
**Figure I.29: No Border Discontinuities in Additional Location Characteristics**

(A) Migration Cost to the North (1860)
- Discontinuity: 0.11 (0.36)

(B) Population Density (1860)
- Discontinuity: \(-4.76^* (2.88)\)

(C) Average Income
- Discontinuity: -0.67 (0.51)

(D) Age of enslaved people
- Discontinuity: 0.03 (0.58)

Notes: This figures shows the RD estimate in additional counties’ characteristics in 1860 across state borders with different Jim Crow intensity in 1865. Migration costs reflect estimates by Donaldson and Hornbeck (2016). Average income is calculated based on occupational income scores. The sample is restricted to policy borders. The left half of each figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix A for details on the sample and data.
Figure I.30: Different Bandwidths for the Pooled Regression Discontinuity (RD) Estimate

Notes: This figure shows the RD estimate in 1940 years of education for Black families who were freed across state borders with different Jim Crow intensity in 1865. The analysis is limited to “policy borders” where Jim Crow intensity differed more than across the median border. Panels (A) to (D) show 100, 150, 200, and 250 kilometer bandwidths respectively. The left half of each figure represents more oppressive states (in red); the right half represents less oppressive states (in black). Each dot represents the average across a decile of the population close to the border. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix A for details on the sample and data.

J. Model Appendix

Using the model in Section 4.3, the average difference between descendants of the Enslaved and the Free in 1940 is a consistent estimator of

\[
E[y_{i,1} \mid s_i = 1] - E[y_{i,1} \mid s_i = 0] = (\lambda + \rho^0_{i(0)}) (E[\alpha_{i,0} \mid s_i = 1] - E[\alpha_{i,0} \mid s_i = 0]) + E[\rho_{i(0)} (\gamma^0_{i(0)} - \delta) + \gamma^1_{i(0)} - \mid s_i = 1] - E[\rho_{i(0)} \gamma^0_{i(0)} + \gamma^1_{i(1)} - \mid s_i = 0].
\]

The difference between the two groups thus reflects 1) any potential differences in ability between the two groups, 2) different exposure to locations over time (as a result of slavery and of potential selection), and 3) the inherent disadvantage of descending from an enslaved person.
conditional on environment and ability. We think of the “effect of descending from enslaved people” as the expected difference between the two groups holding constant any confounding factors (i.e., ability). That is, we define the conditional treatment effect as

\[
CTE(a_{i,0}) \equiv \mathbb{E}[y_{i,1} | s_i = 1, \alpha_{i,0}] - \mathbb{E}[y_{i,1} | s_i = 0, \alpha_{i,0}] =
\]

\[
= \mathbb{E}\{\rho_{I(i,0)}\gamma_{I(i,0)}^0 + \gamma_{I(i,1)}^1 - \rho_{I(i,0)}\delta | s_i = 1, \alpha_{i,0} \} - \mathbb{E}\{\rho_{I(i,0)}\gamma_{I(i,0)}^0 + \gamma_{I(i,1)}^1 | s_i = 0, \alpha_{i,0} \},
\]

(16)

and we define the average treatment effect as

\[
ATE \equiv \int CTE(a_{i,0})dF(\alpha_{i,0}).
\]

(17)

From here, the Free-Enslaved gap in (15) is equal to

\[
\mathbb{E}[y_{i,1} | s_i = 1] - \mathbb{E}[y_{i,1} | s_i = 0] = ATE - B,
\]

(18)

where the (negative of) the selection bias \( B \) is defined in Section 4.3. That is, the gaps between the two groups reflect the “treatment effect” of descending from an Enslaved ancestor but also potential differences in ability between the two groups that have persisted over time.

### J.1 Selection into Being Free

Naturally, if being free before the Civil War was a matter of pure chance, the differences between the Free and the Enslaved have a causal interpretation.

**Assumption A1.** Free-Enslaved status was independent of ability. That is,

\[
\alpha_{i,0} \perp \perp s_i.
\]

That is, under Assumption A1, the bias in (18) is zero such that (18) shows that the unconditional gap is equal to the average treatment effect of being enslaved until the Civil War (1861–1865). A priori, Assumption A1 is strong. However, the plausibility of the assumption depends crucially on the conditions under which freedom was attained.

There were five main channels into freedom between the Revolutionary War (1775–1783) and the abolition of slavery in 1865: 1) by emancipation through abolition of slavery in the North in the late 18th and early 19th century, 2) by manumission through one’s master, 3) by manumission through self-purchase, 4) by manumission through purchase by a third party, or 5) by running away. A person born to a free mother inherited their mother’s freedom. In rare occasions, enslaved people were unintentionally freed by accompanying their masters on a trip to a free state. Setting foot on free soil freed enslaved people by law and some sued to enforce their rights (see, e.g., Rose, 2009).

In 1860, around half of the free population was born in the North, which we argue is a reasonable approximation of the share of the free families who were freed through general eman-
cipation in the North. Within the remaining half, it is hard to estimate the share of people who were freed “legally” and those who ran away. While the 1850 and 1860 censuses suggest the number of runaway slaves is less than 2,000 (out of a total population of around 4 million), the true number is likely much higher (Franklin and Schweninger, 2000).

Dittmar and Naidu (2012) use runaway slave advertisements placed in Southern newspapers between 1840 and 1860 and suggest that such advertisements were placed for around 8,000 runaway slaves throughout those two decades. However, the authors also point out that “it is clear that among the many absconders only a small fraction remained at large for a lengthy period.” The odds of a successful escape were especially small in the Lower South. This is corroborated by the fact that in a Pennsylvania census of Free Black Americans, only 2 out of 314 people who were not born free indicated that they attained freedom through escape. It is therefore safe to conclude that the vast majority of those who became free in the South did so through manumission (as opposed to escape).

Since slavery had been de facto abolished in the North by 1850 (see Table A.2), the enslaved people there were freed non-selectively. That is, as long as one is willing to assume that those enslaved in the North were not inherently different from those enslaved in the South, those in the North were freed entirely independent of any observed or unobserved characteristics. In the South, the degree of selection into manumission varied largely across time and locations. Around the 1780s, the early years after the Revolutionary War, there was a stream of manumissions motivated by morality or religion. In later antebellum years, manumission turned into an instrument to uphold slavery (Berlin, 1974). It did not, in most cases, arise from anti-slavery sentiments. On the contrary, many owners manumitted their slaves as a reward for loyalty and by doing so “reinforced rather than challenged the values, assumptions, and discipline of slavery” (Wolf, 2006, p. 44).

One could imagine that the practice of manumission induced a degree of selection into being free. Indeed, some quantitative evidence on the presence of selection into manumission exists. Cole (2005) finds that in Louisiana, manumitted people were 62.5 percent female (43.6 percent in the enslaved population) and much more likely to be “mulatto” (38.5 percent) than the slave population (5.8 percent). This is consistent with the observation that manumission in the Lower South was reserved for “illicit offspring, special favorites, or least productive slaves” (Berlin, 1974). Bodenhorn (2011), too, finds evidence of preferential manumission for people of mixed race in Virginia. Similarly, Berlin (1974) argues that skilled slaves had a larger chance of accumulating enough wealth to be manumitted through self-purchase. Little is known about selection into being manumitted through purchase by other people (usually other free Black people). Runaways, however, “as a group, had always been more skilled, sophisticated, and aggressive than the mass of slaves” (Berlin, 1974, p. 160). Table J.14 summarizes the discussion.

To assess the potential bias that may result from selection into freedom on observables, we re-estimate the Free-Enslaved gap using only the Free who had not accumulated any measurable physical or human capital by the end of slavery. Even for this subsample, we find large

---

### Table J.14: Relative prevalence of and selectivity in different roads to freedom

<table>
<thead>
<tr>
<th></th>
<th>%</th>
<th>Degree of selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emancipation in North</td>
<td>≈ 50</td>
<td>None</td>
</tr>
<tr>
<td>Manumission by master</td>
<td>30-40</td>
<td>Varied across time and locations</td>
</tr>
<tr>
<td>Manumission by self-purchase</td>
<td>5-10</td>
<td>Potentially high</td>
</tr>
<tr>
<td>Manumission by a third buyer</td>
<td>5-10</td>
<td>Unknown</td>
</tr>
<tr>
<td>Escape</td>
<td>&lt; 5</td>
<td>Potentially high</td>
</tr>
</tbody>
</table>

**Notes:** This table indicates a rough breakdown of the relative probability of attaining freedom in various ways. The percentage emancipated in the North is estimated by the fraction of free Black people born in the North in the 1860 census. The fraction that escaped is a conservative upper bound given the observations mentioned in the text. The remaining probability is attributed to manumissions. The distribution within manumissions is derived from (Bodenhorn, 2011): 10-20 percent through self-purchase, 10-20 percent through a third buyer, and the remaining 60-80 percent by the master.

Free-Enslaved gaps in 1940 (see Appendix Table I.10). This result provides additional evidence against there being substantial bias in the Free-Enslaved gap through selection into freedom.

### J.2 Bounding the Selection Bias

The assumption of independence of Free-Enslaved status and ability (Assumption A1) may be strong. In this section, we show that under weaker assumptions, we can provide plausible bounds for the average treatment effect defined in (17). Assumptions B1 and B2 are the key assumptions.

**Assumption B1.** The initial (pre-1865) distribution across states of the Free and the Enslaved was not a result of any potential selection into being free. That is,

\[
E \left[ \rho_{l(i,0)} \left( \gamma^0_{l(i,0)} - \delta s_i \right) \mid s_i = s \right] = \int E \left[ \rho_{l(i,0)} \left( \gamma^0_{l(i,0)} - \delta s_i \right) \mid s_i = s, \alpha_i,0 \right] dF(\alpha_i,0) \quad \forall s \in \{0, 1\}.
\]

Assumption B1 requires that, in the presence of any selection into being free, selection did not affect the overall geographic distribution of either the Free or Enslaved. Note that Assumption A1 is sufficient but not necessary for Assumption B1 to be satisfied. To make the assumption as plausible as possible, we let \(l(i,0)\) be the state of birth in 1870 (as opposed to the state of residence). For the formerly enslaved population, this thus refers to the state where they were enslaved at their time of birth. In Section 5.3, we argue in detail that selection into enslavement location was likely limited, mainly because the enslaved population did not enjoy freedom of movement (this condition is sufficient but not necessary for Assumption B1 to hold). The free population in 1870 consisted both of people who were born free and those who were born enslaved and freed at some point in their life before 1860. For the latter group, \(l(i,0)\) also refers to the location of enslavement.

The remaining threat to Assumption B1 is that those who were born free may have had parents who selected themselves into certain locations in the first half of the 19th century. Around half of the freeborn descended from people living in the North who were (non-selectively) freed
by general emancipation in the Northern states. One may be concerned, however, that selection arose through the free population in the South. Their geographic distribution could have differed from the counterfactual geographic distribution of the Free under no selection into being free. We argue this is unlikely to have made a meaningful difference. Migration was limited among the Free, and therefore their state of residence was strongly influenced by their family’s location during enslavement. For some of the Free, this was simply due to a lack of migration opportunities (Morgan, 2012, p. 490). However, for many others, there was also a strong unwillingness to migrate. A contemporaneous observer in Charleston, South Carolina wrote that “the Free People [of color] will never emigrate; they have so little to hope for, and so much to dread, from any change of place, that they will adhere to the spot of their nativity, under the pressure of any inconvenience.”

Furthermore, even though political and civil liberties for Black people were greater in the North, economic opportunities tended to be considerably larger in the South, which “suggests why so few southern freemen sought to leave the slave states for ‘freedom’ in the North or, for that matter, why so few northern blacks went South in search of better jobs. While southern free Negroes could expand their political and social liberties greatly by migrating to a free state, they could do so only at the expense of their economic opportunities (Berlin, 1976, p. 313).” All in all, this meant that in the antebellum period there was no clear “destination” in which free people could select themselves. Therefore, we argue that it is implausible that selection into being free affected the geographic distribution of the Free by much.

Assumption B2. Selection into freedom in the South was, if anything, stronger than overall selection into freedom. That is,

$$\sum_{l \in L} \left( \lambda + \rho_l(i,0) \right) \left( E \left[ \alpha_{i,0} \mid s_i = 0, l(i,0) = l \right] - E \left[ \alpha_{i,0} \mid s_i = 1, l(i,0) = l \right] \right) w(l) \geq$$

$$E \left[ \left( \lambda + \rho_l(i,0) \right) \alpha_{i,0} \mid s_i = 0 \right] - E \left[ \left( \lambda + \rho_l(i,0) \right) \alpha_{i,0} \mid s_i = 1 \right],$$

where

$$w(l) = \frac{Pr \left( s_i = 1 \mid l(i,0) = l \right) Pr \left( s_i = 0 \mid l(i,0) = l \right) Pr \left( l(i,0) = l \right)}{\sum_{l \in L} Pr \left( s_i = 1 \mid l(i,0) = l \right) Pr \left( s_i = 0 \mid l(i,0) = l \right) Pr \left( l(i,0) = l \right)} \tag{19}$$

is the least squares regression weight for location l.

Note that $w(l)$ is positive only for locations where both free and enslaved people lived before Emancipation. It is thus zero for locations in the North and is largest in the Upper South where there were substantial populations of both enslaved and free people. Since the Free in the North mostly achieved freedom through general emancipation, they would likely have been less selected than the Free in the South. Also, if there was any “negative selection” into being enslaved up to 1865, this would likely have been stronger in the Upper South (where more people were free) than in the Lower South. This is corroborated by the fact that free Black Americans in the Upper South tended to have a far lower socioeconomic status than those in the Lower South (see Appendix Figure J.31). Last, if anything, one would expect post-slavery relative mobility to

---

have been lower in the South (i.e., a higher $\rho_l$). All observations point to Assumption B2 being satisfied.

To bound the bias under Assumptions B1 and B2, we need to impose two more assumptions that we argue are almost trivial.

**Figure J.31: Free Black Americans’ Average Wealth in 1860**

Notes: This figure shows the average total wealth (real and personal property) among free Black Americans in the 1860 census by county. The maps are truncated to omit the western half of the country, which at the time was only sparsely populated.

**Assumption B3.** In the absence of selection into being free before the Civil War, descendants of the free population would have been able to locate themselves at least as favorably as descendants of the Enslaved conditional on initial location. That is,

$$
\sum_{l \in L} \int \mathbb{E}[\gamma^1_{l(i,1)} | s_i = 0, l(i,0) = l, a_{i,0}] dF(a_{i,0}) w(l) \geq 
\sum_{l \in L} \int \mathbb{E}[\gamma^1_{l(i,1)} | s_i = 1, l(i,0) = l, a_{i,0}] dF(a_{i,0}) w(l).
$$

Assumption B3 rules out the possibility that being enslaved had a positive causal effect on the 1940 location given ancestor location. The Free tended to have more resources and networks available to migrate to places of opportunity. Indeed, Appendix Figure C.6 shows that the Free were considerably more likely to move from the South to the North between 1870 and 1940 (even when controlling for ancestor county).

**Assumption B4.** Had the Free been endowed with the same geographical distribution as the Enslaved,
the differential selection into their location in 1940 would have, if anything, been stronger. That is,

\[
\sum_{l \in \mathcal{L}} \left( \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 0, l(i,0) = l] - \int \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 0, l(i,0) = l, a_{i,0}] dF(a_{i,0}) \right) w(l) - \\
\sum_{l \in \mathcal{L}} \left( \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 1, l(i,0) = l] - \int \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 1, l(i,0) = l, a_{i,0}] dF(a_{i,0}) \right) w(l) \geq \\
\left( \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 0] - \int \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 0, a_{i,0}] dF(a_{i,0}) \right) - \\
\left( \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 1] - \int \mathbb{E}[\alpha_{l(i,1)} \mid s_i = 1, a_{i,0}] dF(a_{i,0}) \right)
\]

Loosely speaking, Assumption B4 requires that the differential selection into location would have been larger had the geographical distribution of the Free and the Enslaved been the same, entirely in the South, and predominantly in the Lower South.

We argue that Assumption B4 is very likely to hold. The potential for selection into location is larger whenever the current location is less beneficial. For instance, for the Free who were concentrated in the North and Upper South, there was relatively little room for selective migration since the Great Migration showed (by revealed preference) that they were already in the most favorable locations for Black people at the time. On the other hand, for the enslaved population in the South (especially the Lower South), the potential for an effect of selective migration was large since those locations tended to provide the least opportunities for progress. Had they been concentrated in the Upper South, this potential would have likely been smaller, if anything. Therefore, under this counterfactual geographical distribution, it seems plausible that selective migration would have been more important for the Free, while it would have been less important for the enslaved population. Both tend to the direction of satisfying Assumption B4.

**Proposition 1.** Under Assumptions B1, B2, B3, and B4, the Free-Enslaved gap in 1940 conditional on ancestor location defines a bound on the sum of the (negative of the) selection bias and the direct effect of descending from enslaved ancestors in (18) from below. Formally,

\[
\sum_{l \in \mathcal{L}} (\mathbb{E}[y_{i,1} \mid s_i = 0, l(i,0) = l] - \mathbb{E}[y_{i,1} \mid s_i = 1, l(i,0) = l]) w(l) \geq B + \rho_{l(i,0)} \delta.
\]

**Proof:** Using Assumption B2, the Free-Enslaved gap conditional on location at \( t = 0 \) can be bound by

\[
\sum_{l \in \mathcal{L}} (\mathbb{E}[y_{i,1} \mid s_i = 0, l(i,0) = l] - \mathbb{E}[y_{i,1} \mid s_i = 1, l(i,0) = l]) w(l) \geq \\
\mathbb{E} \left[ \left( \lambda + \rho_{l(i,0)} \right) a_{i,0} \mid s_i = 0 \right] - \mathbb{E} \left[ \left( \lambda + \rho_{l(i,0)} \right) a_{i,0} \mid s_i = 1 \right] + \\
\sum_{l \in \mathcal{L}} (\mathbb{E}[\gamma_{l(i,1)} \mid s_i = 0, l(i,0) = l] - \mathbb{E}[\gamma_{l(i,1)} \mid s_i = 1, l(i,0) = l]) w(l) + \rho_{l(i,0)} \delta.
\]
Assumptions B3 and B4 ensure we can bound this further by

\[
\sum_{l \in \mathcal{L}} (\mathbb{E}[y_{i,1} \mid s_i = 0, l(i,0) = l] - \mathbb{E}[y_{i,1} \mid s_i = 1, l(i,0) = l]) w(l) \geq \\
\mathbb{E}\left[\left(\lambda + \rho_l(i,0)\right) \alpha_{i,0} \mid s_i = 0\right] - \mathbb{E}\left[\left(\lambda + \rho_l(i,0)\right) \alpha_{i,0} \mid s_i = 1\right] + \\
\left(\mathbb{E}[\gamma^1_{l(i,1)} \mid s_i = 0] - \int \mathbb{E}[\gamma^1_{l(i,1)} \mid s_i = 0, a_{i,0}] dF(a_{i,0})\right) - \\
\left(\mathbb{E}[\gamma^1_{l(i,1)} \mid s_i = 1] - \int \mathbb{E}[\gamma^1_{l(i,1)} \mid s_i = 1, a_{i,0}] dF(a_{i,0})\right) + \rho_l(i,0) \delta = B + \rho_l(i,0) \delta, \tag{23}
\]

where the last equality follows from (18) and Assumption B1. \(\square\)

The left-hand side in (21) is exactly the estimand in an ordinary least squares regression of the outcome \(y_{i,1}\) on ancestor slave status \(s_i\) and ancestor location \(l(i,0)\) (Angrist and Pischke, 2008).

**References**


