CHOOSING RACIAL IDENTITY IN THE UNITED STATES, 1880-1940

Ricardo Dahis
Emily Nix
Nancy Qian

ABSTRACT

This paper documents that a large number of African American men experienced a change in racial identity to white during 1880 to 1940, while analogous changes were negligible for other races. We provide descriptive evidence that is consistent with the conventional wisdom that "passing" for white was a response to severe discrimination, and came at great personal cost. The findings suggest that contrary to traditional economic thinking, racial identity is neither entirely exogenous nor fixed over the lifetime, and responds to incentives.

Ricardo Dahis
Northwestern University
2211 Campus Drive
Office 3386
Evanston, IL 60208
rdahis@u.northwestern.edu

Emily Nix
University College London
Room 223-A
Drayton House
London
United Kingdom
and USC Marshall FBE
e.nix@ucl.ac.uk

Nancy Qian
MEDS
Kellogg SOM
Northwestern University
2001 Sheridan Rd.
Evanston, Il 60208
and NBER
nancy.qian@kellogg.northwestern.edu

1 Introduction

Race is an important determinant of opportunities and outcomes. Most studies in economics treat race as entirely exogenous and fixed over an individual's lifetime. In contrast, other disciplines of social science provide a large body of evidence that some individuals have discretion over their racial identity.[1] Akerlof and Kranton (2000) famously points out that "the choice of identity may be the most important 'economic' decision people make... economic analyses of, for example, poverty, labor supply, and schooling have not considered these possibilities".

The primary goal of our paper is to make progress on this agenda by providing novel evidence on the fluidity of racial identity in a historical context: pre-Civil Rights United States. The findings of our study will emphasize the importance of allowing for endogeneity in conceptualizing racial identity.

In our context, discrimination against African Americans was severe and the "one-drop rule" dictated that an individual was "black" if he had one or more black African ancestors. At the same time, extensive (often nonconsensual) racial mixing during previous generations caused many African Americans to have the physical traits of Europeans. Historical accounts describe some of these individuals leaving behind their black identities to live as white, often at great personal and emotional cost. "Passing [choosing to change one's racial identity from black to white] came into existence during slavery and increased in frequency during the Jim Crow period. Individual African Americans chose to pass to escape discrimination and increase employment opportunities. The costs of passing, however, were high including emotional stress from cutting ties to one's family, condemnation from some segments of the black community, and the constant fear of being 'discovered' by whites" (Rockquemore and Brunsma, 2007, Chap. 1). However, the total magnitude of "passing" for white is unknown and has been a subject of heated debate for over one hundred years.

---

[1]For example, later in the introduction, we discuss the classic sociology studies by Eckard (1947) and Hart (1921) on racial classification change during the early 1900s. More recent studies in sociology, such as Alba, Lindeman, and Insolera (2016) and Saperstein and Penner (2012), examine changes in self-reported racial identity in modern U.S. data.

African American Studies scholar, Werner Sollors, summarizes the existing evidence and describes it as "dramatically heterogeneous and range from hundreds to millions" Sollors (1999, p. 281).

The main difficulty has been the lack of reliable data. Sociologists have attempted to infer from aggregate population statistics the "missing" black (or "extra" white) population across censuses (not accounted for by births, deaths or net migration). These accounting exercises face the difficulty that the historical data, particularly vital statistics and especially for African Americans, are crude and measured with error.[2] For example, at the lower end, Eckard (1947) estimates that 0.25% of black males over age ten in the 1930 census are missing in the 1940 census after accounting for mortality and net migration.[3] On the higher end, Johnson (1925, Figure 37) uses census data to infer that 335,000 (2.5%) black men and women passed to white each year, implying an intercensal rate of around 21%. Recent genetic evidence suggests very high rates of passing, but are difficult to interpret because they are not representative of the population and do not reveal at what point in history passing occurred.[4]

Our study addresses this important question with a relatively new technique from the economic history literature, which links individual census records for men over time. This allows us to observe whether an individual changes race over time without using the relatively more problematic vital statistics and migration data.[5] The linked sample also allows us to provide descriptive evidence on the motivations and constraints for passing.

Using a two-sided linking algorithm that mitigates errors, we find that over 300,000 black males passed for white during 1880 to 1940 (i.e., 16.6% per census interval on average). Amongst the black men who passed for white, approximately 30% "reverse-passed" to black

---

[2]For a detailed discussion of the difficulties in accounting exercises, see, for example, Elo and Preston (1997) and the references within. See, for example, Khanna and Johnson (2010) for a discussion of the range of estimates.

[3]Hart (1921), in another well-known study, conducts a population accounting exercise for white Americans using the 1910 and 1920 Censuses. The "extra" white individuals in 1920 cause him to deduce that a quarter of a million black Americans (2.5% of the black population) changed their race to white.

[4]See the Background Section for a more detailed discussion.

[5]Similar to most previous studies that link individuals across censuses, we focus on men because women are difficult to successfully link across censuses due to the frequency of last name changes with marriage.

in the following census. Since the linked sample is only a part of the total population, the lower bound rate of passing for the population can be obtained by assuming that unlinked individuals never pass for white: our estimates imply that, on average, at least 1.4% of the all black men passed for white each census interval. Taken literally, this is 5.6 times higher than the 0.25% rate from Eckard (1947), which has served as the benchmark lower bound in the literature.[6]

We interpret passing as an *active* choice. It was illegal for black individuals to live as white ones. An observed change in racial classification in the census required a change in lifestyle and situation so that a person would be accepted as white by those he encountered. This includes the census enumerator, who determines the race of respondents in historical censuses. There are two important caveats for our preferred interpretation. The first is the concern that black men are linked to white men who are not their future selves by mistake. We show this is unlikely with a large number of sensitivity checks, alternative linking algorithms and falsification exercises.[7] The second concern is that white enumerators may have erroneously coded the race of correctly linked black men as white without any intention to pass from the black census respondent. We believe that this was unlikely because of residential segregation and the scrutiny over race in the historical context.[8] This concern is also inconsistent with the findings of negligible rates of passing from white to black, which should present enumerators with similar difficulties; and across the Asian races of Chinese, Japanese and Korean, for whom enumerators would presumably have a similarly, if not more, difficult time distinguishing.

The second part of our analysis examines when and how individuals passed for white. This was more likely to occur in states where miscegenation was illegal, that were more democratic, where there were better opportunities for education, and if the individual was

---

[6]The rate of passing for the population implied by our estimates is higher if the extrapolation weights by observables. See Appendix Section F.

[7]See Section C.3 and Table A.2 for sensitivity checks. See Section C.4 and Table A.4 for replication of our results using alternative linking algorithms. See Section 5.2 and Table A.3 for the falsification exercises.

[8]For example, in their study of race counts in the historical census, Strmic-Pawl, Jackson, and Garner (2018) state ".. In this era, the Census Bureau had a near obsession with maintaining the lines among races, specifically with political, economic, and social concerns about safeguarding Whiteness and maintaining the racial hierarchy".

unmarried or had fewer children. Those who passed for white experienced an increase in income, even after controlling for individual characteristics; were more likely to geographically relocate (to communities with a higher share of white residents, and often, out of the South); and their family members would need to pass together or be left behind. These descriptive results are consistent with the historical evidence that discrimination was an important driver in the choice to pass for white, middle class African Americans were more likely to pass, and that one of the main costs was being cut off from one's community and family. We also provide additional results, such as for individuals who are classified as "mulatto" or who have distinctively black names.

In summary, the results show that the lower bound of the number of black men who passed for white is much higher than previously thought. Moreover, the evidence is consistent with passing having been a response to severe discrimination and came at high personal costs. The estimates are specific to our context. However, the insight that identity can be a choice, even along dimensions as rigidly defined as race in pre-Civil Rights United States, is generalizable.

Our work is closely connected to the small but growing number of empirical studies that document the correlation between identity and social and economic incentives in the context of caste in India (Atkin, Colson-Sihra, and Shayo, 2019; Cassan, Keniston, and Kleineberg, 2020; Cassan, 2015; Cassan and Vandewalle, 2017); religious identity for Jews in medieval Europe (Botticini and Eckstein, 2012); ethnic identity in contemporary China (Jia and Persson, 2019); racial identity for Native Americans in contemporary United States (Antman, Duncan, and Barham, 2015); and racial identity in contemporary Brazil (Cornwell, Rivera, and Schmutte, 2017). Our historical context complements the recent work of Fouka, Mazumder, and Tabellini (2018), which finds that the Great Migration of African Americans increased the strength of identity for white Europeans in the Northeastern United States.[9]

---

[9]This literature has traditionally comprised of theoretical studies (e.g., Akerlof and Kranton, 2000; Bénabou and Tirole, 2011).We do not present a formal model in the paper because of space constraints, but our results are highly compatible with the model by Bénabou and Tirole (2011) . Also see Austen-Smith and Fryer (2005) and Ruebeck, Averett, and Bodenhorn (2009) for studies of the contemporary U.S. context, Bodenhorn and Ruebeck (2003) for a discussion of the historical U.S. context, and Bisin, Patacchini, Verdier, and Zenou (2016) for a study of the context of immigrants in Europe today.

This study is related to two strands of the economic history literature. First, we add to works on racial identity in the historical United States by quantifying the magnitude of the "passing" phenomenon and providing a large body of evidence on the correlates of passing.[10] Our findings complement the innovative study by Mill and Stein (2016), which find that 10% to 13% of mulatto sons in the 1910 census pass for white in 1940 (within the linked sample), and that passing is associated with higher income.[11] Second, we add to recent studies that conduct automated linking.[12] The two-sided linking algorithm improves on the one-sided linking algorithms used in existing studies by reducing the possibility of false positives.[13] The recent review of machine-linking methods by Abramitzky, Boustan, Eriksson, Feigenbaum, and Pérez (2019), which is concurrent to our study, recommends two-sided linking as the most rigorous method. To the best of our knowledge, this and our study are the first to emphasize this methodological advantage.[14]

This paper is organized as follows. Section 2 discusses the historical background. Section 3 describes the data. Section 4 summarizes the linking algorithm. Section 5 presents the estimates of the rates of passing as well as the descriptive patterns of passing within the linked sample. Section 6 concludes.

2 Background

2.1 Post-Reconstruction and Jim Crow

The years 1880-1940 coincided with the end of Reconstruction and the start of the Jim Crow laws that preceded the Civil Rights Act of 1964. This was a period when formal and informal

---

[10]See, for example, Ruebeck, Averett, and Bodenhorn (2009) for a discussion about race in the historical U.S. context.

[11]This study focuses on sons who are classified as "mulatto" from families with both black and mulatto children.

[12]For example, see the pioneering studies of Ferrie (1996) and Abramitzky, Boustan, and Eriksson (2012).

[13]An earlier working paper by two of the authors, Nix and Qian (2015), used one-direction links that potentially suffered from finding false positive passing. The current study makes significant improvements in this respect with two-sided linking, which is discussed in more detail in the paper.

[14]The main analysis will demonstrate that the rates of passing that we estimate are much lower than those from one-sided linking. In the Appendix, we also document that the rates of passing we estimate are slightly lower than or similar to alternative two-sided linking algorithms.

discrimination towards the black population severely limited their political, economic and social opportunities relative to the white population. Southern states passed laws intended to disenfranchise the black population (Woodward, 2002, p. 83). These changes significantly reduced the number of black voters.[15]

The black population faced restrictions such as the complete segregation of whites and non-whites in all facilities (e.g., restaurants, schools, water fountains, buses), where the facilities provided to non-whites were usually lower quality than those provided to whites. Many regions practiced strict neighborhood segregation, where public services such as sewers and electricity ended at the boundaries of the white neighborhoods. In other places, particularly urban areas, there could be segregation within buildings (e.g., across floors) (Packard, 2003, p. 102-103). Miscegenation – i.e., interracial marriages – and sometimes even non-marital sexual relationships were also made illegal (Packard, 2003, p. 99). Discrimination was also "enforced" informally by organizations such as the Ku Klux Klan. Non-whites seen as violating white supremacy were often harassed, and sometimes murdered. Between 1882 and 1968, approximately 3,446 African Americans were lynched (Tuskegee Institute, 2010).

Blacks earned much less than whites.[16] Black men and women were shut out of most non-menial jobs (Sharfstein, 2011, p. 255). Sundstrom (1994) shows that the large differences in black and white occupational choices were driven in part by social norms that rejected black workers as supervisors over white workers.[17]

---

[15]For example, in Mississippi less than 9,000 out of 147,000 voting age blacks were registered to vote. In Louisiana, the number of black registered voters decreased from approximately 130,000 in 1896 to 1,342 by 1904. In Georgia, only four percent of all black males were registered to vote (Keyssar, 2000, loc. 2695).

[16]Margo (1990) discusses the striking stability of the black-to-white earnings ratio from 1900 to 1940 and the potential causes of these gaps, with African American men earning between 45%-48% the income of white American men over this entire period. Also see, for example, Carruthers and Wanamaker (2017), Collins and Margo (2011) and Card and Krueger (1992). Wright (2013) discusses how the Civil Rights movement undid many of the barriers to black economic success that existed during our time period and led to large economic gains for black men and women. Returns to skin color are analyzed and discussed in more depth in Bodenhorn (2015).

[17]There was also significant variation in the formal laws which affected the rights and opportunities facing blacks within states, as well as in the informal enforcement of state or federal laws. For example, Carruthers and Wanamaker (2013) document substantial variation in the relative quality of schooling for black students across counties. Keyssar (2000, loc. 3052) notes that the economic qualifications for voting varied across municipalities in New York.

Severe racial discrimination was not isolated to the South. For example, the Ku Klux Klan was based in Indiana during the early 20th Century and had large memberships in Maine and Oregon (Packard, 2003, p. 127). California, which had introduced laws to restrict property ownership of Asians during the 19th Century, extended them to include other non-white races such as black (Packard, 2003, p. 100). When Woodrow Wilson became president, he segregated the District of Columbia's federal agencies, which had been integrated for the previous fifty years (Packard, 2003, p. 123). Many schools in Illinois, Ohio, Pennsylvania and New Jersey were completely segregated, even though it was *de jure* illegal. Between 1913 and 1948, 30 out of the then 48 states enforced anti-miscegenation (mixed-race marriage) laws (Vile, 2003).

2.2 Racial Mixing before 1880

According to the *Trans-Atlantic Slave Trade Database*, a total of 305,326 African Slaves were ever brought to North America. Almost 70% were adult men.[18] By the eve of the Civil War in 1860, there were a total of 4,427,294 individuals classified as black, over 3.9 million of whom were slaves.[19] To understand the magnitude of passing in our study, it is important to note the large number of light skinned people of African extraction by 1880. "By the time that slavery ended, a majority of American Negroes bore in their genetic make-up some degree of white, which is to say European, ancestry" (Packard, 2003, p. 95).[20] To demonstrate the wide gradient of color for former slaves, emancipated "White and Colored Slaves" were used for a propaganda tour of the North in 1863.[21] Past studies have argued that those who had Caucasian features may have had stronger economic incentives to pass for white because they had the most to lose from Jim Crow laws since they were likely to have been

---

[18]See https://www.slavevoyages.org/assessment/estimates.

[19]Children inherited the status of the mother under slavery; the child of a slave woman is always born a slave. Thus a high degree of mixing between white men and black slave women could have contributed to the large increase in the slave population.

[20]Rockquemore and Brunsma (2007) notes that "the vast majority of interracial sex consisted of exploitative unions between white male slave owners and their black female slaves" (Rockquemore and Brunsma, 2007, Chap. 1).

[21]See Appendix Figures A.1-A.2c, which are taken from the article entitled, "White and Colored Slaves" by C. C. Leigh (*Harper's Weekly*, January 30, 1864, p. 71).

more educated, have higher skill jobs, and own property (Bodenhorn, 2002).

2.3 Genetic Evidence of Race Today

The best available genetics evidence shows substantial racial mixing in previous generations. Individuals today who identify as African American are 24% European and 73.2% African on average. Moreover, a significant proportion of individuals who self-report as European Americans have African ancestry: 3.5% have at least 1% (at least one ancestor in the past eleven generations). In Louisiana and South Carolina, 12% of self-identified European Americans have at least 1% African Ancestry (Bryc, Durand, Macpherson, Reich, and Mountain, 2015).[22] Since European-Americans are approximately 72.4% and African-Americans are approximately 12.6% of the U.S. population today, taking literally the possibility that 3.5% of the European-American population are black under the "one-drop rule" implies that approximately 20% of black Americans passed for white. Hammer, Chamberlain, Kearney, Stover, Zhang, Karafet, Walsh, and Redd (2006) conducts a similar exercise with an independent sample. This study does not report national average statistics, and instead compares genetic compositions across regions. They find that for white Americans, the lowest amount of African ancestry is in the Southwest (0.8%) and the highest in the Northeast (10%). Doing a similar calculation as before, these genetic results translate to rates of passing of 57.4% in the Northeast and 4.6% in the Southwest, which could reflect a higher rate of passing in the Northeast or that those who passed migrated to the North. Neither of the genetic studies discussed here use random samples and they may therefore not be representative of the population they study (the United States, the Southwest, or the Northeast). Meigs, Grant, Piccolo, López, Florez, Porneala, Marceau, and McKinlay (2014) obtained a random sample of the population in Boston. They find that 8.63% of the ancestry of European Americans is African, which translates to a rate of passing of 49.6% of the black population.

---

[22]Note that genetic studies such as the one discussed face numerous caveats from difficulties such as non-random sampling of the population. See Bryc, Durand, Macpherson, Reich, and Mountain (2015) for a detailed discussion.

2.4 Defining Race

Racial "science" and eugenics, with beliefs that race captures biological and inherent physical and moral traits, were popular during the period that we study. Much of this was based on Carl Lineaus's 1735 publication, *Systema Naturae*, which classified the races as the following:

> *Africanus*: black, phlegmatic, relaxed; hair black, frizzled; skin silky; nose flat; lips tumid; women without shame, they lactate profusely; crafty, indolent, negligent; anoints himself with grease; governed by caprice.
>
> *Europeaeus*: white, sanguine, muscular; hair long, flowing; eyes blue; gentle, acute, inventive; covers himself with close vestments; governed by laws
>
> (Smedley, 1993, p. 164).

These explicitly racist beliefs led whites to believe that if they had been exposed to blacks, they would be able to infer a person's degree of "blackness" from his appearance and demeanor.[23] The legal definition of black, which was based on the fraction of one's blood that was black, varied across states and over time. During the Jim Crow era, most states used the "one drop rule", which meant that a person is black if she has only one drop of African blood (Packard, 2003, p. 98).

In practice, for many individuals of mixed extract, race was determined by association because this "degree-of-blood rule did not in fact make it impossible for people to cross racial lines" (Gross, 2009, loc. 4123). In his well-known study, Davis (2010, p. 14) points out that "The concept of 'passing' rests on the one-drop rule and on folk beliefs about race and miscegenation, not on biological or historical fact".[24] In describing the successful suit for white identity by a mixed race woman named Alexina Morrison, Gross (2009, p. 55) points out that ".. race was not obvious. Nor did the rule about 'negro' identity... decide the question. More persuasive to the [white] witnesses and jurors at the trial were stories about

---

[23]See Gross (2009, Ch. 7).

[24]Also, see Smith (2006) for a detailed discussion of the difficulties and methods that white individuals developed to distinguish between black and white, given the difficulty of distinguishing based on sight alone when there were many mixed-race individuals.

the hidden marks of race as interpreted by experts, and stories about Alexina's behaviour dancing at white balls, her mingling with white families, her love affairs with white men... separation became the key to whiteness. People who had associated with whites must be whites themselves, just as people who had associated with blacks had to be black... In other words, race by association ... trumped any other sort of physical or documentary evidence" (Gross, 2009, loc. 1083, 1356).[25] As historian Carol Wilson noted when discussing her book on the successful lawsuit of Sally Miller, an enslaved woman who claimed a white identity (and freedom), "Southern whites want desperately to believe that they can tell the difference between white people and black people. And so the fact that white people accept her as a white person, they consider that factual evidence. Well, she must be white, because we think she's white..." Wilson (2016).[26]

2.5 Passing for White

In the context of our study, passing required a person to have physical features that are commonly shared by whites, to behave and dress like a white person, and to associate with white people. Passing most often required a person to move to a white community where the individual was not previously known by others as a black person since "...Caucasian appearance was irrelevant if public knowledge existed of one's black ancestry" (Packard, 2003, p. 96). The exceptionally high rates of internal U.S. migration presumably made it easier for mixed race individuals to move and adopt a white identity.

Our study takes place when the incentives to pass were arguably at their highest since the end of slavery. Jim Crow severely eroded the economic opportunities and civil liberties of anyone identified as black, even as the number of educated and skilled African Americans

---

[25]There are several examples of racial classification by association from lawsuits. See the review by legal historian Ariela Gross (Gross, 2009). In each successful case, the person suing to be legally identified as white would demonstrate that she or he has been accepted by white friends and attended all white functions (e.g., assemblies, balls). The women also sometimes agreed to a physical inspection of her whiteness and provided testimony to her virtuous behavior, which was assumed to be impossible if she was of African extraction. In each case, the judge appealed to the jury to use their "common sense".

[26]In the modern U.S. context, sociologists have documented that the perception of whether an individual is black is positively correlated with socio-economic status expressed by activities such as incarceration (Penner and Saperstein, 2008) or attire (Freeman, Penner, Saperstein, Scheutz, and Ambady, 2011).

grew rapidly in the post-Civil War era. "The harder whites made it for blacks to earn a living, educate their children, and just make it through a single day without threat or insult, the greater the incentives grew for light-skinned blacks to leave their communities and establish themselves as white... the drumbeat for racial purity, the insistence that any African ancestry – a single drop of blood – tainted a person's very existence, accelerated the migration to new identities and lives" (Sharfstein, 2011, p. 235-236). William Pickens of the NAACP stated in 1927, "if passing for white will get a fellow better accommodations on the train, better seats in theater... and may even save his life from a mob, only idiots would fail to seize the advantage of passing, at least occasionally if not permanently" (Janssen, 2016).

Passing was known to have occurred for individuals of all ages. Children sometimes passed from black to white because their parents passed or because parents sent light skinned children to live with white families to allow the children to pass.[27] Some passed as young adults to attend school, obtain a job, or to marry a white person (or a black person who had passed for white).[28] Others passed when they were older simply because of the overwhelming discrimination they faced or to provide a better life for their children.[29]

Passing was not always permanent. Sometimes, individuals passed to obtain a job or attend school, and then later reverse passed to black.[30] Other times, circumstances would force one who had passed as white to reverse pass back to being black. An example is the family of Stephen Wall, who "For the next ten years the family moved neighborhoods repeatedly from white to black to white again" (Sharfstein, 2011, p. 270).[31]

Given that one had to move away from his black community and live with whites to pass, one of the greatest costs associated with passing for white was the near permanent separation from a person's community and family. Spouses and children who could not pass for white would be left behind. We investigate this with the data later in the paper.[32]

---

[27]See Williams (1996) and Dawkins (2012).

[28]See Sharfstein (2011) and Williams (1996).

[29]See Sharfstein (2011).

[30]See Hobbs (2014).

[31]Also see Gordon (1999) and Williams (1996).

[32]A large body of anecdotal evidence shows that those wishing to pass often completely disassociated themselves with their past lives. For example, historian Allyson Hobbs recalls the experience of her relative who passed for white after high school. Her grandmother said to the relative, "you're going to graduate, you're

See the Online Appendix Section A for summaries of case studies that illustrate the costs and benefits of passing for white.

3   Historical Censuses

We use individual-level data from the U.S. historical censuses for the years 1880 - 1940. These were digitized and made available to researchers by Ancestry through the NBER.[33] For each individual, we observe variables such as the first name, last name, age, county of residence, state of residence, state or country of birth, race, gender, relationship to the household head and marital status. Father's and mother's birth states and countries are available for the years 1880 - 1930.

Our study focuses on males because of the difficulties in linking women, who usually change their names after marriage. We restrict our attention to those under age 55 in the base year because higher mortality rates for older ages reduce the number of links.[34] The main exercise divides individuals into two racial categories: white and black. Racial classification was determined by the enumerator in the historical censuses. The categories change over time. To be consistent, "black" in our study includes mulatto individuals, which are separate categories for some years.[35] After we present the main results, we will also compare the rates of passing for mulatto individuals to those of black individuals in years when the two groups are distinct.

---

going to leave Chicago, you're going to go to California, and you're going to become a white woman. And this is the best thing for you". The young girl protested, she didn't want to leave her friends, her family, the only life she'd ever known. And her grandmother said, "no, this is the best thing for you. You'll have the best life chances if you do this" (Sloan, 2013). In his biography, Williams (1996) recounts how his mixed race father passed for white by moving from Indiana to Washington D.C., and married a white woman. In his recount of the experience of the Wall family, Sharfstein (2011) discussed how the children who moved away from their home in Washington D.C. passed for white, and the one son who remained behind and his daughter were classified as black.

[33]The 1850 and 1860 Censuses only reported names of free blacks. Since most of the black population was enslaved, this means that these earlier data contain names for only a small subset of the population in which we are interested. For the 1870 Census, only the 1% sample is currently digitized. The data from 1890 were lost to a fire. Note that Nix and Qian (2015) uses similar data provided by FamilySearch. The current study uses Ancestry because of the availability of information on occupations, the race of spouses and the number and race of children.

[34]We later demonstrate that this restriction does not affect our findings.

[35]See Appendix Section B.1 for details on racial classifications in the Census.

Enumerator instructions were vague.[36] It is generally believed that enumerators inferred the race of respondents based on physical appearance, behavior and association. The requirements for a person of African extraction to be classified as white in the census were presumably similar to the requirements for the legal cases discussed by Gross (2009) in Section 2.

Passing in our context refers to a change in census identification from black (including mulatto individuals) to white from one census to the next. The individual historical census data were not used for other purposes such as employment or taxes.[37] Thus, there was no reason for an individual to pass for white for the census *per se*. Rather, consistent with the historical accounts in the previous section, we assume that the choice to pass required a change in lifestyle and situation so that a person would be accepted as white by those he encountered, including the census enumerator.

Thus, our *prima facie* interpretation is that passing for white in the historical Censuses is an active choice. However, it is also possible that enumerator error results in passive passing in the data by miscoding the race of a light-skinned individual who had no intention of passing. For example, the enumerator often obtained information for the household from one or two individuals. Given the legal and social environment (e.g., residential segregation), enumerators may have assumed that all residents of the household (not in hierarchical relationships) are either white or non-white. However, since mixed-race cohabitation is illegal for the most part, this phenomenon is unlikely to lead to the census data recording as white a black individual who did not mean to pass for white. Also note that most enumerator errors would simply lead to the individual being dropped from the linked sample. We discuss enumerator error in more detail in the next two sections of the paper.

---

[36]See Appendix B.1.

[37]The main purpose of the U.S. census is to determine the number of representatives per state in the house and the number of electoral votes. It is also used to compute aggregate statistics. By law, information that can be used to identify individuals is not released until 72 years later.

4 2-Sided Unique Perfect Links

We identify links as individuals who have a unique link going forwards in time (i.e., a black man i in year t matches one and only one man of any race in year $t + 10$), for which the linked individual in year $t + 10$ also has a unique link going backwards in time (i.e., the linked individual in year $t + 10$ matches one and only one man in year t *who is the initial black man* i). Links must have perfect spelling matches in first and last names, as well as match on other information (e.g., approximate age, birth state, etc.). For brevity, we call this the 2-Sided Unique Perfect (2SUP) link.[38]

The main benefit of the two-sided linking procedure relative to the one-sided procedures used in past studies is that it mitigates the possibility that mistaken links will produce false passing for white. One-direction links identify the individual in year $t + 10$ who is the best match for the black man in year t, but do not require that the black man in year t is the best match for the individual in year $t + 10$. Mistaken links are likely to cause us to over state the number of those who pass from black to white because there are more white men than black men in the population. The problem is mitigated with the 2-sided link because we only identify a link as one that is the best match in both directions.

Our linking algorithm will overstate the true rate of passing only if all of the following conditions are satisfied for the black individual i in year t : 1) there are no black individuals in year $t + 10$ who can be linked to him (e.g., the black individual died in between censuses, or there was an enumerator error); 2) there is one and only one white individual in year $t + 10$ that can be linked to him; and 3) when linking backwards, the linked white individual in year $t + 10$ links only to the initial black individual i in year t (e.g., the true link for the white individual is not in the base year census due to transcription error or because he moved to the U.S. in between the two censuses).[39]

---

[38]For computational feasibility, we use a randomly selected 10% sample of the black male population in each base year t. The restriction only applies to the base year sample. We always link to the full population of all males in the subsequent census, and again to the full population of all males in the initial census when linking backwards. Thus, the restriction should not affect the accuracy of the links.

[39]Our results are similar if we restrict the sample to black men born in the United States in the base year. These are available upon request.

Two common concerns for linking in our context are enumeration error and intentional name changes. Enumerator error can only bias the rate of passing upwards if all three conditions above are true. We will present a large body of evidence in the paper and the Appendix against this possibility. The linked sample will naturally omit any individual who changed his name when he passed for white. We discuss this more when we examine individuals with distinctively black names.

The key advantage of the 2SUP link is that it should contain very few Type I errors (false links). We investigate this issue by manually checking a large sample of 2SUP links. We find little evidence of erroneous links. In addition, we conduct several falsification exercises and a large number of sensitivity checks and show that the rates of passing are similar or higher if we use alternative algorithms from other studies. These are shown later in the paper and in Appendix C.

5 Main Results Using the Linked Sample

5.1 The Rate of Passing in the Linked Sample

5.1.1 Blacks Passing for White

Table 1 Panel A column (1) shows that 16.6% of black males with 2SUP links passed for white during the five census intervals for which we have data. Panel B column (1) shows that this is 30,239 individuals in the 10% sample, and therefore approximately 302,390 individuals in the full population. Our results mean that at least 302,390 black men under age 55 passed for white in the following census during the period 1880-1940. Note that there are five intervals during this period: 1880-1900, 1900-1910, 1910-1920, 1920-1930, 1930-1940. In the paper, we will often refer to this sample as men from the base years of 1880-1930.

Since we only observe race during the census year, we will not be able to account for instances of passing and reverse passing that occur within the same census intervals – i.e., if someone passed for white, but then returned to being black before the next census, they would not be counted as passing in our estimates.

All linking procedures face a similar conceptual difficulty in extrapolating statistics from the linked sample to the population because of the concern that the linked sample may not be representative. Table 1 Panel C column (1) shows that 8.6% of black males in the base year are linked using 2SUP. When we compare the characteristics of individuals in the 2SUP sample and the population, we find that the means are very similar in magnitude, but the large sample size means that the differences, though small, are statistically significant.[40]

This leaves two possible and very different directions for extrapolating population statistics. One is to say that the differences in sample means are not economically meaningful and directly extrapolate the 16.6% rate of passing from the 2SUP sample to the population. Alternatively, one can take the stance that such differences are meaningful despite the small magnitudes. To be conservative and thorough, we consider the second line of inquiry. If we make the extreme assumption that unlinked individuals never pass, the population rate of passing for black men under age 55 would be 1.4% (0.086 × .166). Appendix Section F discusses less extreme extrapolations using weights based on observable characteristics.

For comparison, column (2) provides analogous statistics using the traditional one-direction link, also with unique perfect spelling matches. The substantially higher rates of passing, 26.8%, and higher rates of being linked, 14.1% instead of 8.6%, are consistent with our prior that 2SUP will mitigate false positives relative to one-direction links but link fewer individuals. Henceforth, we will only present estimates using the 2SUP links.

Appendix Table A.6 presents the results disaggregated by census intervals.

5.1.2 Race Transition Matrix

Table 2 presents a race transition matrix for Blacks, Whites, Native Americans, Chinese, Japanese and Koreans. The latter three categories are separately reported only in 1920-1940 (i.e., the 1920-1930 and 1930-1940 linked intervals).[41] For comparison purposes, the results in this section for other racial categories will also focus on these two linked intervals.[42]

---

[40]See Appendix Table A.7.

[41]See Appendix Table A.1 for a comprehensive list of racial categories in each census.

[42]Our results are very similar if we use all available years for each category. They are available upon request.

Panel I reports the rates of links and changes in racial classification within the linked samples for each racial category with at least fifty thousand observations. Row A presents the rates of passing for black to other races for this sample. It shows that 84.8% of black males remain black, 15% pass for white, 0.1% become Native American, and 0.1% are classified as Chinese in the following census.

In row B, we repeat the exercise for white males (i.e., identified as white and under age 55 in the base year).[43] Table 2 row B shows that only 0.7% of the 2SUP links pass from white to black. 99.1% remained white. The rates of passing to the other categories are similarly negligible.

These results, which are aligned with the socio-economic incentives to pass for another race, support our interpretation that a change in racial classification in the census reflects an active choice to pass rather than an inactive process driven by enumerator errors because any difficulty that enumerators had in distinguishing between white and black men should be reflected in both the rates of passing from white-to-black, as well as that of black-to-white.

Rows C and D present the rates of racial classification change for Chinese and Japanese men.[44] We find that 8.2% of Chinese men and 7% of Japanese men become white, and a negligible share change to non-white races other than their own. These patterns are consistent with the two groups of Asian men having faced more discrimination than white men, and similar levels of discrimination to each other.

Row E examines the rates of passing for men classified as Native Americans.[45] We find that 27.8% passed for white and 2.2% became black. This is consistent with historical evidence that there was much racial mixing prior to our study period and that the official definition for Native Americans was more ambiguous than for other groups. In some censuses,

---

[43]We use 2% of the white male population from each base year for computational feasibility. We use a smaller sample than for linking black males because the white population is much larger. Note that the rate of links for white men is substantially higher than for black men, 20% instead of 9.3%. This is consistent with the fact that fewer white men had common names than black men.

[44]The links use 100% of the Chinese and Japanese male population in each base year, 1920 and 1930, because of their small size. The rates of links for Chinese and Japanese men are lower than for black men, 7.3% and 5.4% versus 9.3%. This is consistent with the higher share of common names for Asians in the historical censuses.

[45]We use the 100% sample of males classified as Native Americans in each base year because of their small population size.

enumerators were instructed to classify individuals with mixed white and Indian ancestry as "white" if their community accepts them as white. For other races, the definition was based solely on ancestry.[46]

The main takeaway from these results is that black men are much more likely to pass for white than men of other races who are categorized based on ancestry. This is consistent with the fact that black men arguably faced the most severe discrimination. Equally important is the finding that the rates of white passing to black (or any other race), which includes the "reverse-passing" discussed later in the paper, are negligible. This is consistent with the fact that white men had little incentive to live as another race.[47]

5.1.3 Passing Across East Asian Categories

If the observed rates of passing are driven by enumerator confusion due to similar physical appearances of individuals who belong to different legal racial classifications, we should also observe high rates of passing across the three East Asian groups, which have relatively similar physical appearances. At the same time, these three groups faced broadly similar levels of discrimination and no barriers to inter-racial marriage, such that there were limited incentives to actively pass across groups – i.e., a Chinese individual had little incentive to become legally Japanese or Korean. As such, passing across Asian categories provides a useful placebo.[48] Table 2 Panel II presents the results. In addition to Chinese and Japanese men, we include those identified as Korean, which is a relatively smaller population and thus left out of Panel I. We observe negligible rates of passing across the categories of Chinese, Japanese and Korean – from 0.1% to 0.4%. These findings support our claim that 2SUP links are not subject to frequent enumerator error and our interpretation that passing was motivated by socio-economic incentives.

---

[46]See Appendix Section B.1 for enumerator instructions. See Sandefur and McKinnell (1986) for a discussion of the history of Native American racial identity and interracial mixing with whites.

[47]One may be concerned that the the commonality of names between whites and other races varies across the latter, which could affect the probability that enumerators will miscode a race as white. We address this in Appendix Section D.

[48]See Appendix Section B.1 for enumerator instructions for classifying Asians.

5.1.4 Reverse Passing

Historical accounts note occasions when a person classified as "white" will choose to change their race to "black". The first is if he marries a black woman who cannot pass for white. Given the illegality of miscegenation, this means that the white man would need to pass for black. The second is if he is a black man who passed for white, who subsequently chooses to reverse pass to be black again (e.g., to return to his family if they cannot pass for white). The historical evidence provides many examples for the latter, which we focus on in this section. To investigate the percentage of individuals who pass, but then revert to being black in the following census year, we link individuals across two consecutive census intervals – i.e., link individual i in year t to himself in year $t+10$, and link him in $t+10$ to himself in $t+20$. Table 1 Panel D column (1) presents the rates of passing for the sample linked over two consecutive censuses for comparison. 17% of the sample passed for white. This is comparable to the 16.6% that passed for white in the sample that is linked over one census interval in Panel A. Panel D column (2) shows that 30% of those who passed to become white later reverse passed to black.[49]

These results show that race was fluid and many individuals crossed back and forth across identities. The presence of reverse-passing from "white" to black also implies that the rate of passing from individuals who are born white to black was even lower than the 0.7% estimate from the previous section.

5.2 Falsification Exercises

One way to investigate the concerns of enumerator error or false links is by conducting falsification tests. The first uses literacy. Individuals can only become more literate over time. Changing from literate to illiterate must therefore reflect a mistaken link (or enumerator error). Thus, if we find that the latter change occurs more in the sample of those who pass

---

[49]Recall that interpreting these rates of reverse-passing requires a similar caveat to interpreting the rates of passing shown earlier. Since we only observe individuals in census years, we will undercount reverse passing if the person reverts back to his white identity by the following census (e.g., the individual is black in year t, white in year $t+10$, black in year $t+11$ to $t+19$, and white again in year $t+20$). Also note that reverse-passing is included in the count of whites passing for black in Table 2.

than those who do not pass, we would be concerned that our findings are confounded.

Appendix Table A.3 column (1) shows that amongst those who passed from black to white, 3% changed from literate to illiterate. This is comparable to the 5.1% for those who did not pass in column (2). In contrast, the share of the linked sample to become literate over time is similar or slightly higher for those who pass for white: 32.9% versus 31.4% for those who do not pass. These results do not support the concern that mistaken links cause false positive passing in the linked sample.

A second falsification check uses the ages of children. For each linked individual, we are able to observe all children living in his household and their ages in years t and $t + 10$. We calculate the average reported age of children who are zero to five years old in year t and the average reported age of children who are ten to fifteen years old in year $t + 10$.[50] If the difference between the average ages of the children in years t and $t + 10$ is further from 10 for the children of individuals who passed for white than those who did not pass for white, we would be concerned about the presence of more bad links in the population of those who pass. In row C of Appendix Table A.3, we find that the difference in reported and observed ages of children are similar for the children of those who pass and those who do not pass (9.49 versus 9.52). Note that if we compare the average ages of all individuals who are identified as children of the household head without any restriction on their age, the differences in average age between the two censuses are 5.72 versus 5.14 (not presented in the table). The differences are less than ten because older children move out of the household and younger ones are born. Nevertheless, the similarity between the two groups suggests that the similarity in the restricted sample is not an artifact of the restricted age range.

To show that our results are unlikely to be an artifact of mistaken links, we also conduct a large number of sensitivity exercises and replicate the results using alternative linking

---

[50]This exercise is restricted to linked household heads (for whom we can identify children using the variable for the relationship to the household head) in the base years of 1900-1930. To avoid compositional effects due to fertility and older children leaving the household, we restrict our attention to children who are ages 0 to 5 in the base year. Note that this exercise does not require that age is reported accurately, but only that measurement error (e.g., age heaping) occurs similarly between households where the household head passed for white versus households where he did not pass for white. We focus on household heads with no more than ten children within the specified age range to reduce measurement error.

algorithms. See Appendix Section C.3.

5.3 Descriptive Patterns of Passing for White

Motivated by historical accounts and the existing studies, this section uses the 2SUP sample to examine descriptive patterns of passing, its association with other behaviors such as marriage and migration, and whether passing is correlated with social and economic opportunities.

5.3.1 The Mulatto Category

In 1880, 1910 and 1920, "mulattos" – i.e., individuals of both African and European descent – comprised a separate category from black.[51] The instructions to enumerators were vague and historians have been dubious about the usefulness of this categorization because of the limited number of black Americans without any European extract.[52] This and the need for consistency across years motivates the grouping together of "mulattos" and "blacks" for the main analysis. In this section, we follow earlier studies such as Mill and Stein (2016) and repeat our investigation for a sample of individuals who are classified as mulatto in the base year.

Table 3 Panel A uses a sample of individuals who are categorized as mulatto in 1910 and links them to the 1920 census, when mulatto remains a separate category; and a sample of individuals categorized as mulatto in 1920 and links them to the 1930 census, when mulatto is no longer a category.[53] We find that from 1910 to 1920, only 25.9% remain mulatto, which is consistent with the notion that the mulatto category was very hard for enumerators to systematically and consistently define. 54.2% of individuals become black, while 19.5% become white. The latter is comparable to the 16.6% rate of passing we find in our main 2SUP sample for the same census interval (see Table A.6 row C), where we do not distinguish

---

[51]See Appendix Section B.1 for a detailed discussion on the definition for mulattos each census year.

[52]"After 1920, the U.S. Census Bureau dropped the 'mulatto' category, the government having concluded that at least three-quarters of all American Negroes bore white genes and thus officially specifying people as mulattoes no longer made much sense" (Packard, 2003, p. 98).

[53]Note that in 1910, 25.05% of the "black" (i.e., black + mulatto) population is mulatto, and in 1920, 10.36% is mulatto.

between individuals categorized as mulatto from those categorized as black. In the 1930 census, when mulatto is no longer a category, 85.1% of those in the mulatto category in 1920 become black, whereas 14.3% become white. Again, the rate of passing to white is similar to the rate of 15% from the main 2SUP sample for the same census interval (see Appendix Table A.6 row D).[54]

Panel B presents the rates of passing for those who were categorized as black instead of mulatto in the years when both categories existed. Column (2) shows that between 1910 and 1920, 74% remained black, 10.3% became mulatto and 15.7% passed for white. Between 1920 and 1930, when mulatto was eliminated as a category, 84.8% became black and 15.2% passed for white.

A comparison of Panels A and B yields several insights. First, those who are classified as mulatto in Panel A are less likely to become black, more likely to remain mulatto, and slightly more likely to pass for white than those who are classified as black (in 1910-1920) in Panel B. This supports the belief that individuals in this category were on average more likely to have European physical appearances and thus could more easily pass for white. At the same time, the results show that there is very little persistence in being categorized as mulatto (Panel A column 2) and that the difference in the rates of passing for white is not very large (Panels A and B column 3). This is consistent with the belief that the large share of mixed race individuals made it difficult for census enumerators to accurately and consistently categorize individuals as mulatto or black.[55]

5.3.2 Distinctively Black and White Names

Motivated by Cook, Logan, and Parman (2014), we investigate whether the rates of passing are lower for individuals with distinctively black names and who choose to keep the same

---

[54]Note that Mill and Stein (2016) finds that 10% to 13% of sons who are classified as mulatto in families with black and mulatto children in 1920 become white in 1940. Our estimates may differ slightly from theirs because of the difference in sample (we examine all mulatto males), the difference in time frame (we examine one census interval), or the difference in linking algorithm (they use a one-direction linking algorithm).

[55]The Census stated that "the principal reason for giving up the attempt to separate blacks and mulattoes was the fact that results of the attempt in past censuses had been very imperfect" and "not even approximately accurate" (Hochschild and Powell, 2008, p. 79).

name in the next census, for whom passing is presumably more difficult. Similarly, we investigate whether the rates of passing for black individuals with distinctively white names is higher. We take historically black names from Cook, Logan, and Parman (2014).[56] Table 3 Panel C shows that, as expected, the rate of passing for those with distinctively black names is much lower, 7.4%, as opposed to 16.6% in the full population. These results are likely to understate the rates of passing amongst all those born with a distinctively black name since some may change their names when they pass for white, in which case they will not be linked.

To identify distinctively white names, we adapt Cook, Logan, and Parman's (2013) method for white males.[57] As expected, the rate of passing is higher, at 24.3%. This is consistent with the notion that it is easier to pass for white with a white name, and having such a name may be positively associated with characteristics that enable one to pass for white.[58]

5.3.3 Spouses

Miscegenation (mixed-race marriage) was illegal in many states for much of our context, and existing studies such as Fryer (2007) observe that approximately 0.5% to a little over 1% of all black male marriages are to white women during the Jim Crow era. As in this earlier study, we find that 1.3% of all marriages for black men are to white women. This statistic excludes black men who pass for white when they marry white women, which would presumably be most black men in mixed race marriages since miscegenation was illegal in most of our

---

[56]The names are are Abe, Abraham, Alonzo, Ambrose, Booker, Elijah, Freeman, Isaac, Isaiah, Israel, King, Master, Moses, Pearlie, Percy, Perlie, Purlie, Presley, Presly, Prince, Titus. These names are identified using random samples of black men from the census years 1900 and 1920. See Cook, Logan, and Parman (2014) for more details.

[57]There are two steps: 1) select names that have a count above the median of the distribution of names within a race and have a within-name race frequency larger than its frequency within the same race; 2) select the top twenty most frequent names within each race in this restricted set. Distinctively white names are Albert, Arthur, Carl, Charles, Clarence, David, Edward, Frank, Fred, George, Harry, Jacob, John, Joseph, Louis, Paul, Peter, Thomas, Walter, William.

[58]Note that we are able to link a higher percentage of individuals with distinctively black names, 13.6% (not in tables), than the full population, 8.6%. This is most likely because multiple matches are dropped by our algorithm and there are fewer such cases with distinctively black names. For individuals with distinctively white names, we find a lower rate of links, 7.4% (not in tables), which is consistent with the fact that there are more white people in the population, such that having a distinctively white name means more common names and multiple matches, and therefore a higher likelihood of being dropped from the linked sample.

context. We investigate this phenomenon by examining the pattern of marriage and passing. For this exercise, we use information on marital status and the relationship to the head of the household.[59] The sample size is slightly reduced relative to the main linked sample since we lose observations with missing values in those two variables.

Table 4 columns (1)-(2) examine individuals that were black in the base year and divide the sample according to whether they were single, married to a black spouse, or married to a white spouse in the base year. Then, for each subsample, we further distinguish whether they are single, married to a black spouse, or married to a white spouse in the subsequent census. Column (2) shows that for those who were single in both years, 15.1% passed for white. For those who became married to a black woman, almost no one passed regardless of their status in the base year (0.2% if single or married to a black woman in the base year, 0% if married to a white woman in the base year). In contrast, most of those who became married to a white person passed, regardless of their status in the base year (98.8% if single in the base year, 98.6% if married to a black woman in the base year, 91.1% if married to a white woman in the base year). Note that we cannot distinguish between marrying a white woman and marrying a black woman who has also passed for white. Thus, becoming married to a white woman in the subsequent census year could mean that his original wife passed for white or that he remarried a white woman.

The comparison of rows D, E and F are particularly striking. Black men who are married to a black woman in the base year and pass for white either leave their wives (row D) or become married to a white woman (row F).

In columns (3)-(4), we repeat the exercise for individuals who are classified as white in the base year. The results illustrate a consistent pattern. Rows D - F show that very few white men are married to black women in the base year. When one becomes married to a black woman, he almost always passes for black (rows B, E and H) and none of those who become married to a white woman pass for black (rows C, F and I). Recall that we cannot distinguish between a white man changing his racial classification to black from reverse-passing by a

---

[59]For example, if the person is a "spouse" and "black", and the household head is "married", we then code the household head as being married to a black spouse.

24

black man who has previously passed for white.

These results are consistent with the difficulty of mixed race marriages. And because a black (white) man cohabiting with a woman who is perceived by others as white (black) often faced formal and informal sanctions, these results are also consistent with our interpretation that the estimated pass rates are mostly driven by active decisions to pass rather than enumerator errors.

5.3.4 Children

Table 5 investigates the relationship between passing and the number and race of children in the household. It is analogous to Table 4. The relationship to the household head variable allows us to identify individuals who are children of the household head. Thus, we are able to observe the number of children for each household head and the race of each child. We divide household heads according to whether they had no children, at least one child who was categorized as black, and at least one child who was categorized as white in the base year. The last two categories are not mutually exclusive. We then subdivide each group according to the observed number and race of children in the following census year.

Table 5 Columns (1) and (2) show the number of observations and rates of passing for white amongst black household heads. Rows C, F and I show that regardless of the number or race of children observed in the base year, nearly all individuals who had at least one white child in the following census passed for white (97.8% to 100%). In contrast, Rows B, E, and H show that regardless of the number or race of child in the base year, those who had at least one black child in the following census passed for white at negligible rates (0 to 0.3%).

For those who had no children or at least one black child in the base year census and no children in the subsequent census, the rates of passing range between 10.1% to 15.8% (Rows A and D). These findings are consistent with the fact that children needed to pass for white with parents who passed, or be left behind. For those with a white child in the base year census and no children in the following census, 63.9% passed for white (row G). However,

we note that there are very few individuals with at least one white child in the base year (rows G-I), which is also consistent with the fact that one needed to have the same race as his children or separate from them.

The descriptive statistics in rows D, E and F for individuals who had at least one black child in the base year are particularly striking. It shows that for black male household heads with at least one black child, passing for white results in either leaving your children (row D) or also having your children pass for white (row F).

In column (3) and (4), we document analogous patterns for white household heads. The patterns are consistent.

Two caveats should be kept in mind when interpreting the results in this section. First, as before, we do not distinguish between a child who has always been classified as white from a child who has passed from black to white. For example, some children identified as white in rows G-I of columns (1)-(2) could be black children who have passed, ostensibly in conjunction with their parents passing to white. Similarly, some of the white male household heads observed to change racial classification to black in columns (3) and (4) may be black males who had passed to white, and then chose to reverse-pass. White children of men who have passed for white may be former black children who passed together with their father, or children from new marriages since the previous census interval. Second, we cannot observe older children who have moved out of the household. None of these caveats undermine the main takeaway that men who pass for white cannot live with black children.

5.3.5 Migration

There are two reasons to be interested in migration. First, since there were significant formal and informal sanctions against passing for white, someone who passed would generally need to geographically relocate to a place where no one knows him. Given residential segregation, he is likely to move to a relatively "whiter" neighborhood.[60] Second, our study coincides with a period of tremendously high internal migration (e.g., Collins and Wana-

---

[60]See studies such as Ananat (2011) for a discussion of residential segregation.

maker, 2015) and it is naturally interesting to examine whether the migration patterns for individuals who pass for white differ from those who do not pass for white.[61]

Using the 2SUP linked individuals, we identify those who moved counties within a state, those who moved states, and those who moved out of the South.[62] Table 6 column (1) shows that 49.7% of those who passed for white moved counties within the same state. Columns (2) and (4) show that 38.7% moved states, amongst which 11.9% left the South. Column (3) shows that adding the rates of moving in columns (1) and (2) implies that 88.4% of individuals who passed moved counties within or across states.

Columns (5)-(8) show that the rates of migration for those who did not pass for white are much lower. Column (5) shows that amongst those who did not pass, 22.7% moved counties within a state. Columns (6) and (8) show that 16.5% moved states, amongst which 7.6% left the South. Column (7) shows that a total of 39.2% of those who did not pass moved counties within or across states. The rates of moving for those who remained black are approximately half of the rate of migration for those who passed for white. The large differences are consistent with the necessity of moving in order to pass for white.[63]

Next, we examine the patterns over time. We note that the twenty-year interval in row (B) will naturally experience a higher rate of mobility since the longer interval provides more time for individuals to move. Thus, we focus on the ten-year intervals in rows (C)-(F). Columns (3) and (7) show that the rates of moving are comparable over time for both those who pass and those who do not, with perhaps a slight uptick during 1920-30, which coincides with the Great Migration.[64] Consistent with the Great Migration being a period where

---

[61]Our study also coincides with high (mostly European) international migration into the United States. However, almost all black men in our sample are born in the United States due to the Act Prohibiting Importation of Slaves of 1807 and neglible voluntary migration of black men from other countries during this period. We investigate the relationship between international migration and passing for white in Section 5.3.7.

[62]The South comprises of Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, Tennessee and Virginia.

[63]Note that our estimated cross-county migration rate of 39.2% for individuals who do not pass is comparable to the 47% rate of migration across counties during 1850 to 1860 for white males found by Ferrie, Sutch, Carter, and Bard (2003).

[64]See Carrington, Detragiache, and Vishwanath (1996), Collins and Wanamaker (2015) and Boustan (2009) for examples of studies on the Great Migration. Collins and Wanamaker (2015) links African American males aged 0 to 40 and living in the South in the 1910 Census in the 1% public-use micro-data sample to their future selves in 1930. In total, they link 5,465 men and find that 20.2% have moved out of the South by 1930.

many blacks left the South, Columns (4) and (8) show that there is an uptick in moving out of the South during 1920-30. Interestingly, when we compare the uptick to other years, we find that the relative increase is moderate for those who pass (13.3% versus 10.5 –12.4%), while it is much larger for those who do not pass (11.9% versus 4.3 – 7.6%). This is consistent with the notion that it might have been easier for those who pass to live in the North. Because of this, such individuals were always incentivized to move out of the South. In contrast, those who remained black were less likely to move North outside of the Great Migration.

Finally, we investigate whether those who pass are moving to communities with a higher proportion of white residents. Unfortunately, the census data can only be disaggregated to the county level. Thus, we compare the percentage of the county population that is white in the county of residence in year t and the county of residence in year $t + 10$.[65] We calculate the fraction of individuals (males) that report as white in each county. To see if those who pass for white move to "whiter" counties, we calculate the difference in the percentage white of the county of residence during the current census year (when the individual has passed for white) and the county of residence during the last census year (when the individual reported as black). The historical evidence suggests that we should see an increase in the share of white residents in counties for individuals who pass relative to those who do not pass.[66]

Figure 1a plots the probability density function (PDF) for those who pass for white and those who do not, where the x-axis is the change in the percentage of the county of residence that is white. The PDF for those who pass (illustrated by the thick solid blue line) is to the right of the PDF for those who do not pass (illustrated by the dashed red line). This means that individuals who pass for white are more likely to move to "whiter" counties than those who remain black.

---

[65]The historical census also reports enumeration districts. However, district boundaries change across censuses, while county boundaries are relatively stable. Moreover, we would be concerned that enumeration district boundaries were changing in response to changes in the racial composition, as discussed in Card, Mas, and Rothstein (2008). Thus, we choose to use counties as the level of comparison.

[66]Note that because there are many communities within a county and most counties have mixed populations, we would not expect those who pass for white to move to 100% white counties even if segregation is fully enforced at the community level.

Figure 1b plots the analogous PDF for those who pass and remain white versus those who reverse-pass to black. The figure shows that the relocation pattern of those who reverse pass to being black is a mirror image of the pattern for those who passed for white: reverse passers (illustrated by the thick solid blue line) move to communities with a lower percentage of whites than those who remain white (illustrated by the dashed red line).

These patterns are consistent with the historical evidence that passing required relocation to a white community (and similarly, reverting to one's black identity requires relocating to a less white community).

5.3.6 Regional Differences

Next, we investigate the patterns of passing across regions, which could differ because of the variation in the degree of *de jure* and *de facto* discrimination against blacks across regions. Table 6 Panel II shows the rates of passing for the North (states that were part of the Union during the Civil War), the South (states that were part of the Confederacy during the Civil War), states that allowed slavery at the onset of the Civil War in 1860, and states where 98% of the black male population lived during 1880-1940.[67] Row G examines individuals according to their state of birth. Row H divides the individuals according to where they lived in the base year of the linked interval (year t). This groups together individuals who were born in a given state and still live there, as well as those who moved to the state prior to the base year. Row I divides the sample according to where a linked individual lived in the subsequent census ("year t + 10", or in the case of 1880-1920, "year t + 20").

Row G column (5) shows that individuals born in the Northern states were much more likely to pass, with an average pass rate of 23.8%. In contrast, 15.5% and 16% of those born in the South and former slave states pass for white. Row H shows that the rates of passing

---

[67]Union states include Massachusetts, Connecticut, California, Illinois, Indiana, Iowa, Maine, Michigan, Minnesota, New Hampshire, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Vermont and Wisconsin. Confederate states include Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, Tennessee and Virginia. Slave states include the Confederate states and Delaware, Kentucky, Maryland and Missouri. 98% of the black male population lived in Georgia, Mississippi, Alabama, South, Carolina, North Carolina, Louisiana, Texas, Virginia, Tennessee, Arkansas, Florida, Washington, Kentucky, Maryland, Pennsylvania, New York, Missouri, Ohio, Illinois, New Jersey, Oklahoma, Indiana, West Virginia, Michigan and Kansas.

were higher for individuals who lived in Northern states the census year before changing race, where on average 21.2% of the linked sample passes for white. In contrast, 15.4% and 15.8% individuals who lived in the South and former slave states pass for white. Row I shows that this regional difference is less salient for the state of residence after passing.

We can also examine the rates of passing across more versus less segregated counties. For this exercise, we use the disaggregated segregation measure from Logan and Parman (2017) for the years 1880 and 1940, when the segregation measure is available.[68] This measure provides a nuanced and comprehensive measure of segregation using information on racial similarity of next door neighbors. We estimate two bivariate regressions, where we regress the segregation index in 1880 (1940) on the rates of passing in 1880-1900 (1930-1940). The coefficients are 0.33 and 0.23, respectively. Both are statistically significant at the 1% level. These imply that passing for white was positively associated with the degree of residential racial segregation. These results are not presented in tables.

5.3.7 Descriptive Regressions

Base Year Characteristics   In this section, we investigate whether the factors that influence passing from historical accounts are important on average by examining the correlates of passing for white. Table 7 presents several individual-level regressions. The outcome variable is a dummy variable that equals one if a linked individual changes racial classification from black in year t to white in $t + 10$. Explanatory variables are measured at the base year t. All of the regressions in Table 7 control for age category dummy variables, base year and region fixed effects. The standard errors for the regressions in this table are clustered at the level of variation of the explanatory variable (see the bottom of the Table).

First, we consider the desire to escape severe discrimination as a possible correlate of passing with several proxies for discrimination. Table 7 column (1) examines a dummy variable indicating that mixed marriages are legal in a given state and year. Because miscegenation does not vary within states in a given year, we control for state fixed effects.

---

[68] We use the index of dissimilarity by county, and restrict our examination to counties with at least fifty black men.

Column (2) uses the Democratic vote share for a given county as a proxy for discrimination in that county.[69] To summarize the meaningful variation, we compute the first principal component for all of the elections for U.S. president and the U.S. House of Representatives that have taken place during the census base year and the preceding nine years. This regression controls for county fixed effects. The estimate in column (1) shows that living in states where miscegenation is legal is associated with 31.9 percentage-points less passing for white. In column (2), we find that the Democratic vote share is positively associated with passing for white. However, since the magnitude of a principal component is difficult to interpret, we also present the standardized coefficient in italics. It shows that a one standard deviation increase in the Democratic vote share is associated with a 0.046 standard deviation increase in passing for white. The estimates are statistically significant at the 1% and 10% levels. They are consistent with the notion that an individual is more likely to pass for white in places with more discrimination.

Next, we investigate the possibility that educated or high-skilled mixed race individuals were more incentivized to pass as is suggested by Bodenhorn (2002) and Mill and Stein (2016).[70] We use several proxies. The first proxy is a principal component that captures educational opportunities. The component is constructed from four variables: black-to-white teacher salary, white-to-black pupil-teacher ratio, black-to-white term lengths and the number of black universities.[71] All of these schooling variables are such that a higher value reflects better educational opportunities for the black population. The second proxy is the white-to-black occupational income score ratio. A higher value is associated with fewer economic opportunities for blacks, or that the composition of black workers is low skilled relative to white workers. In the same regression, we include a dummy for whether the individual is literate and his individual occupational income score. We follow Carruthers

---

[69]See https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8611.

[70]Mill and Stein (2016) documents that amongst mulatto males in 1910, more educated individuals were more likely to pass for white in 1940.

[71]We use three measures of the quality of secondary schools for blacks relative to whites taken from Carruthers and Wanamaker (2017): black-to-white teacher salary, white-to-black pupil-teacher ratio and black-to-white term lengths. These variables are available at the county and year level for southern states. We add to this a variable that we construct ourselves, the number of black universities in a given state and year.

and Wanamaker (2017) and control for state fixed effects.[72]

Table 7 column (3) shows that conditional on whether a black man is literate and how much he earned in the base year, he is more likely to pass for white if he is from a county with better educational opportunities for the black population. At the same time, conditional on the educational and income opportunities of his county, a man with a higher occupational income score is more likely to pass for white. These results are consistent with the notion that educated and higher earning mixed-race individuals may have had more to gain from passing for white; or alternatively, they had more to lose from the introduction of Jim Crow.

In column (4), we examine some possible constraints for passing. One narrative which is common in most accounts of passing for white relates to the personal cost from being cut off from one's family and community. All traces of African ancestry must be left behind when an individual passes. He must necessarily relocate. His family must either move with him and pass for white or be left behind (recall our earlier results on the patterns of passing, marriage and children). The estimates in column (4) are consistent with this conventional wisdom. We find that being married and having more children are negatively associated with passing for white (as is having a distinctively black name). The regression controls for county fixed effects and clusters the standard errors at the county-year level.

In column (5), we examine the possibility that the demographic composition of one's community may be associated with the decision to pass. We find that the black population share is uncorrelated with passing for white. However, the share of immigrant population and whether he lives in an urban area are both positively associated with passing. Doubling the share of immigrants is associated with a 34.5 percentage-point higher probability of passing for white. This is consistent with the notion that places with many immigrants or that were urbanizing were places where it might be easier to pass for white (e.g., it is easier for a new arrival to blend in) and also places with more opportunities and innovation. In this case, innovation could also refer to social innovation, where people are willing to be more daring and explore potentially risky or personally costly ideas.

---

[72]They argue that county fixed effects over-control. Note that our results are very similar if we alternatively control for county fixed effects. They are available upon request.

Column (6) shows that the correlation with immigration share is driven by immigrants from Northern Europe, which we define as European countries that do not border the Mediterranean. The latter result may simply be an artifact of the dominance of Northern Europeans in the immigrant population during this period.[73]

Well-being After Passing for White   Table 8 examines the correlation between passing and characteristics measured in year $t + 10$. We focus on outcomes that are measured at the individual level. Column (1) shows that an individual who has passed for white earned a higher income, which is consistent with conventional wisdom that individuals passed for better economic opportunities due at least in part to the high degree of income discrimination against blacks.[74] Note that all regressions control for base year county of residence and base year fixed effects. In addition, we always control for several individual base year characteristics: occupational income score (to address the possibility that individuals who pass for white may have higher earnings potential), whether he lives in an urban area, his marital status, whether he is literate, and age category dummy variables (ages 25-34, 35-44, and 45-54, with 15-24 as the omitted category).[75]

Column (2) shows that an individual who passed is 48% more likely to move counties or states than one who did not pass. This is consistent with our earlier findings on migration.

Column (3) investigates whether moving was associated with a higher income for those who pass. It is similar to the specification in column (1), with several additional right-hand-side variables: a dummy variable for having moved counties within a state, a dummy variable for having moved states, a dummy variable for moving from a rural to an urban area, and each of these variables interacted with a dummy for whether the individual passed for white. Note that because moving is partly an outcome of passing, these estimates should be interpreted as a descriptive decomposition exercise. The uninteracted dummy variable for

---

[73]In examining the country of origin for foreign-born individuals in the census, we find that with the exception of Italy, the countries with the largest numbers of individuals are all Northern European.

[74]The large positive association between earnings and passing for white is consistent with Mill and Stein (2016), which documents a similar pattern for mulatto males who pass to become white between 1910 and 1940.

[75]The results are qualitatively similar if we examine the logarithm of the occupational income score as the dependent variable. They are available upon request.

passing is positive, but smaller than in column (1). This means that individuals who pass for white, but do not move, still experience an increase in income. The interactions with moving state, county and rural-to-urban are large and positive. This means that the gains from passing were much larger for those who passed and moved than those who passed but did not move.

Taken together, the correlational evidence in Table 8 columns (1)-(3) shows that those who passed for white experienced material improvements in well-being after passing, and much of this improvement was accompanied by geographic relocation.

Column (4) investigates the heterogeneous effects of passing for white for those who were classified as mulatto. We do this by adding the dummy variable indicating whether the linked individual is classified as mulatto in the base year, and its interaction with whether the individual passed for white by the following census. The uninteracted coefficient is 0.454 and statistically significant at the 1% level, which means that those classified as mulattos who did not pass for white earned higher income than those classified as black who did not pass for white. This is consistent with the view that lighter skinned individuals earned more, as well as the view that behavior and success can influence the perception of race – i.e., a better dressed or educated individual is more likely to be categorized as mulatto by the enumerator.[76]

The interaction effect is -1.301 and statistically significant at the 1% level. This means that mulatto individuals experienced lower income gains when they passed for white than black individuals who passed for white. This could be because they started from a higher level of income in the base year and thus had less to gain from passing for white. Nevertheless, the sum of the coefficient for passing, 3.178, the interaction effect, -1.301, and the uninteracted coefficient for mulatto, 0.454, is positive. This means that passing for white was still associated with large income gains for those who were classified as mulatto.

Note that the regression results are very similar if we apply the weights discussed in Appendix Section F. These result are available upon request.

---

[76]See the Background section.

6 Conclusion

The extent of passing from "black" to "white" in pre-Civil rights United States has been a subject of heated debate amongst scholars and the public for the past one hundred years. Our findings significantly increase the lower bound of the rates of passing relative to population accounting exercises conducted by sociologists. Moreover, the patterns of passing are consistent with historical evidence that the choice to change racial identity was a response to economic and political incentives, and came at great personal costs. Passing for white was positively associated with higher income. In order to pass for white, individuals needed to relocate to whiter communities. Because miscegenation was illegal, passing required the spouse to pass or be left behind. Similarly, children needed to pass with their parents or be left behind.

The results suggest that racial identity is partly a choice that is endogenous to many of the variables that the economics literature has traditionally examined as outcomes of race. As such, the relationship between race/ethnicity and economic and political outcomes is likely to be much more complex than typically conceived. Understanding the extent and magnitude of endogenous racial classification is important for many inquiries. This is true for both the historical context and the modern one. For example, in a recent review article, Charles and Guryan (2011) states: "An important task for future empirical work on discrimination will be to more thoroughly explore changing racial classification and to examine the theoretical issue of what determines racial identification".

References

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102(5), 1832–56.

——— (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122(3), 467–506.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2019):

"Automated Linking of Historical Data," Working Paper 25825, National Bureau of Economic Research.

ABRAMITZKY, R., R. MILL, AND S. PÉREZ (2020): "Linking Individuals Across Historical Sources: A Fully Automated Approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2), 94–111.

AGER, P., L. BOUSTAN, AND K. ERIKSSON (2019): "The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War," Working Paper 25700, National Bureau of Economic Research.

AKERLOF, G. A., AND R. E. KRANTON (2000): "Economics and Identity," *The Quarterly Journal of Economics*, 115(3), 715–753.

ALBA, R. D., S. LINDEMAN, AND N. E. INSOLERA (2016): "Is Race Really So Fluid? Revisiting Saperstein and Penner's Empirical Claims," *American Journal of Sociology*, 122(1), 247–262.

ANANAT, E. O. (2011): "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality," *American Economic Journal: Applied Economics*, 3(2), 34–66.

ANTMAN, F., B. DUNCAN, AND BARHAM (2015): "American Indian Casinos and the Rise in Native American Self-Identification," Working paper, Citeseer.

ATKIN, D., E. COLSON-SIHRA, AND M. SHAYO (2019): "How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption," Working Paper 25693, National Bureau of Economic Research.

AUSTEN-SMITH, D., AND R. G. FRYER (2005): "An Economic Analysis of Acting White," *The Quarterly Journal of Economics*, 120(2), 551–583.

BAILEY, M., C. COLE, M. HENDERSON, AND C. MASSEY (2017): "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth," Working Paper 24019, National Bureau of Economic Research.

BÉNABOU, R., AND J. TIROLE (2011): "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 126(2), 805–855.

BISIN, A., E. PATACCHINI, T. VERDIER, AND Y. ZENOU (2016): "Bend it like Beckham: Ethnic Identity and Integration," *European Economic Review*, 90, 146–164.

BODENHORN, H. (2002): "The Mulatto Advantage: The Biological Consequences of Complexion in Rural Antebellum Virginia," *Journal of Interdisciplinary History*, 33(1), 21–46.

——— (2015): *The color factor: The economics of African-American well-being in the nineteenth-century South*. Oxford University Press, USA.

BODENHORN, H., AND C. S. RUEBECK (2003): "The Economics of Identity and the Endogeneity of Race," Working Paper 9962, National Bureau of Economic Research.

BOTTICINI, M., AND Z. ECKSTEIN (2012): *The Chosen Few: How Education Shaped Jewish History*, Princeton Economic History of the Western World. Princeton University Press.

BOUSTAN, L. P. (2009): "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940-970," *The Journal of Economic History*, 69(03), 755–782.

BRYC, K., E. Y. DURAND, J. M. MACPHERSON, D. REICH, AND J. L. MOUNTAIN (2015): "The Genetic Ancestry of African Americans, Latinos, and European Americans Across the United States," *The American Journal of Human Genetics*, 96(1), 37–53.

CARD, D., AND A. B. KRUEGER (1992): "School Quality and Black-White Relative Earnings: A Direct Assessment," *The Quarterly Journal of Economics*, 107(1), 151–200.

CARD, D., A. MAS, AND J. ROTHSTEIN (2008): "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 123(1), 177–218.

CARRINGTON, W. J., E. DETRAGIACHE, AND T. VISHWANATH (1996): "Migration with Endogenous Moving Costs," *American Economic Review*, 86(4), 909–30.

CARRUTHERS, C. K., AND M. H. WANAMAKER (2013): "Closing the Gap? The Effect of Private Philanthropy on the Provision of African-American Schooling in the U.S. South," *Journal of Public Economics*, 101(C), 53–67.

CARRUTHERS, C. K., AND M. H. WANAMAKER (2017): "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap," *Journal of Labor Economics*, 35(3), 655–696.

CASSAN, G. (2015): "Identity Based Policies and Identity Manipulation: Evidence from Colonial Punjab," *American Economic Journal: Economic Policy*, 7(4), 103–31.

CASSAN, G., D. KENISTON, AND T. KLEINEBERG (2020): "A Division of Laborers: Identity and Efficiency in India," Working Paper.

CASSAN, G., AND L. VANDEWALLE (2017): "Identities and Public Policies: Unintended Effects of Political Reservations for Women in India," IHEID Working Papers, Graduate Institute of International and Development Studies Working Paper.

CHARLES, K. K., AND J. GURYAN (2011): "Studying Discrimination: Fundamental Challenges and Recent Progress," *Annual Review of Economics*, 3(1), 479–511.

COLLINS, W. J., AND R. A. MARGO (2011): "Race and Home Ownership from the End of the Civil War to the Present," *American Economic Review*, 101(3), 355–59.

COLLINS, W. J., AND M. H. WANAMAKER (2015): "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants," *The Journal of Economic History*, 75(4), 947–992.

COOK, L. D., T. D. LOGAN, AND J. M. PARMAN (2014): "Distinctively Black Names in the American Past," *Explorations in Economic History*, 53, 64–82.

CORNWELL, C., J. RIVERA, AND I. M. SCHMUTTE (2017): "Wage Discrimination When Identity Is Subjective Evidence from Changes in Employer-Reported Race," *Journal of Human Resources*, 52(3), 719–755.

DAVIS, F. J. (2010): *Who Is Black?: One Nation's Definition*. Penn State Press.

DAWKINS, M. (2012): *Clearly Invisible: Racial Passing and the Color of Cultural Identity*. Baylor University Press.

ECKARD, E. W. (1947): "How Many Negroes "Pass"?," *American Journal of Sociology*, 52(6), 498–500.

ELO, I. T., AND S. H. PRESTON (1997): "Racial and Ethnic Differences in Mortality at Older Ages," *Racial and Ethnic Differences in the Health of Older Americans*, pp. 10–42.

FEIGENBAUM, J. (2016): "Automated Census Record Linking: A Machine Learning Approach," Working paper, Harvard University.

FERRIE, J., R. SUTCH, S. CARTER, AND R. BARD (2003): *Historical Statistics of the United States: Millennial Edition*. New York: Cambridge University Press.

FERRIE, J. P. (1996): "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 29(4), 141–156.

FOUKA, V., S. MAZUMDER, AND M. TABELLINI (2018): "From Immigrants to Americans: Race and Assimilation during the Great Migration," *Harvard Business School BGIE Unit Working Paper*, (19-018).

FREEMAN, J. B., A. M. PENNER, A. SAPERSTEIN, M. SCHEUTZ, AND N. AMBADY (2011): "Looking the Part: Social Status Cues Shape Race Perception," *PLOS ONE*, 6(9), 1–10.

FRYER, R. G. (2007): "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century," *Journal of Economic Perspectives*, 21(2), 71–90.

GORDON, L. (1999): *The Great Arizona Orphan Abduction*. Harvard University Press.

GROSS, A. (2009): *What Blood Won't Tell: A History of Race on Trial in America*. Harvard University Press.

HAMMER, M. F., V. F. CHAMBERLAIN, V. F. KEARNEY, D. STOVER, G. ZHANG, T. KARAFET, B. WALSH, AND A. J. REDD (2006): "Population Structure of Y Chromosome SNP Haplogroups in the United States and Forensic Implications for Constructing Y Chromosome STR Databases," *Forensic Science International*, 164(1), 45–55.

HART, H. (1921): *Selective Migration as a Factor in Child Welfare in the United States: With Special Reference to Iowa*, Iowa University Studies, 1st Series.

HOBBS, A. (2014): *A Chosen Exile: A History of Racial Passing in American Life*. Harvard University Press.

HOCHSCHILD, J. L., AND B. M. POWELL (2008): "Racial Reorganization and the United States Census 1850–1930," *Studies in American Political Development*, 22(01), 59–96.

JANSSEN, K. (2016): "Five Black Chicagoans Who Passed for White," *Chicago Sun-Times*.

JIA, R., AND T. PERSSON (2019): "Individual vs. Social Motives in Identity Choice: Theory and Evidence from China," Working Paper 26008, National Bureau of Economic Research.

JOHNSON, C. (1925): *Opportunity*.

KEYSSAR, A. (2000): *The Right to Vote: The Contested History of Democracy in the United States*. Basic Books.

KHANNA, N., AND C. JOHNSON (2010): "Passing As Black: Racial Identity Work Among Biracial Americans," *Social Psychology Quarterly*, 73(4), 380–397.

LAIT, A., AND B. RANDELL (1996): "An Assessment of Name Matching Algorithms," *Technical Report Series-University of Newcastle Upon Tyne Computing Science*.

LOGAN, T. D., AND J. M. PARMAN (2017): "The National Rise in Residential Segregation," *The Journal of Economic History*, 77(1), 127–170.

LONG, J., AND J. FERRIE (2013a): "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *The American Economic Review*, 103(4), 1109–1137.

——— (2013b): "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Reply," *The American Economic Review*, 103(5), 2041–2049.

MANCINI, O. (2002): "Vassar's First Black Graduate: She Passed for White.," *Journal of Blacks in Higher Education*.

MARGO, R. A. (1990): *Race and Schooling in the South, 1880-1950: An Economic History*. University of Chicago Press.

MEIGS, J. B., R. W. GRANT, R. PICCOLO, L. LÓPEZ, J. C. FLOREZ, B. PORNEALA, L. MARCEAU, AND J. B. MCKINLAY (2014): "Association of African Genetic Ancestry with Fasting Glucose and HbA1c Levels in Non-Diabetic Individuals: The Boston Area Community Health (BACH) Prediabetes Study," *Diabetologia*, 57(9), 1850–1858.

MILL, R., AND L. STEIN (2016): "Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America," Working paper, Arizona State University.

MODALSLI, J. (2016): "Multigenerational Persistence: Evidence from 146 Years of Administrative Data," Statistics Norway Working Paper 850.

——— (2017): "Intergenerational Mobility in Norway, 1865–2011," *The Scandinavian Journal of Economics*, 119(1), 34–71.

NIX, E., AND N. QIAN (2015): "The Fluidity of Race: "Passing" in the United States, 1880-1940," Working Paper 20828, National Bureau of Economic Research.

PACKARD, J. (2003): *American Nightmare: The History of Jim Crow*. St. Martin's Press.

PENNER, A. M., AND A. SAPERSTEIN (2008): "How Social Status Shapes Race," *Proceedings of the National Academy of Sciences*, 105(50), 19628–19630.

PERKINS, L. M. (1998): "The Racial Integration of the Seven Sister Colleges," *The Journal of Blacks in Higher Education*, (19), 104–108.

ROCKQUEMORE, K. A., AND D. L. BRUNSMA (2007): *Beyond Black: Biracial Identity in America*. Rowman & Littlefield Publishers.

RUEBECK, C. S., S. L. AVERETT, AND H. N. BODENHORN (2009): "Acting White or Acting Black: Mixed-Race Adolescents' Identity and Behavior," *The B.E. Journal of Economic Analysis and Policy*, 9(1), 1–44.

SANDEFUR, G. D., AND T. MCKINNELL (1986): "American Indian Intermarriage," *Social Science Research*, 15(4), 347–371.

SAPERSTEIN, A., AND A. M. PENNER (2012): "Racial Fluidity and Inequality in the United States," *American Journal of Sociology*, 118(3), 676–727.

SHARFSTEIN, D. (2011): *The Invisible Line: A Secret History of Race in America*. Penguin Group United States.

SLOAN, N. (2013): "Stanford Historian Re-Examines Practice of Racial 'Passing'," *Stanford Report*.

SMEDLEY, A. (1993): *Race in North America: Origin and Evolution of a Worldview*. Westview Press.

SMITH, M. M. (2006): *How Race Is Made: Slavery, Segregation, and the Senses*. University of North Carolina Press.

SNAE, C. (2007): "A Comparison and Analysis of Name Matching Algorithms.," *Enformatika*, 19.

SOLLORS, W. (1999): *Neither Black Nor White Yet Both: Thematic Explorations of Interracial Literature*. Harvard University Press.

STRMIC-PAWL, H. V., B. A. JACKSON, AND S. GARNER (2018): "Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality," *Sociology of Race and Ethnicity*, 4(1), 1–13.

SUNDSTROM, W. A. (1994): "The Color Line: Racial Norms and Discrimination in Urban Labor Markets, 1910–1950," *The Journal of Economic History*, 54(02), 382–396.

THOMAS, R. (1995): "Thyra Johnston, 91, Symbol Of Racial Distinctions, Dies," *The New York Times*.

TUSKEGEE INSTITUTE (2010): "Lynchings: By State and Race, 1882-1968," .

VILE, J. R. (2003): *Encyclopedia of Constitutional Amendments, Proposed Amendments, and Amending Issues, 1789-2002*. ABC-CLIO.

WILLIAMS, G. (1996): *Life on the Color Line: The True Story of a White Boy Who Discovered He Was Black*, National bestseller. Penguin Books.

WILSON, C. (2016): "Color Lines: Racial Passing in America," *Back Story*.

WOODWARD, C. V. (2002): *The Strange Career of Jim Crow*. Oxford University Press.

WRIGHT, G. (2013): *Sharing the Prize*. Harvard University Press.

Table 1: Main Results – Black Passing for White

	2SUP	UP
	(1)	(2)
Panel A: Rate of Passing in the Linked Sample		
	16.6%	26.8%
Panel B: Number Passing in the Linked Sample		
	30239	79937
Panel C: Percentage of the Population in the Linked Sample		
	8.6%	14.1%
Panel D. Reverse-Pass to Black (2SUP)		
% Pass (Black$_t$, White$_{t+10}$): 17%	% Reverse Pass (Black$_t$, White$_{t+10}$, Black$_{t+20}$): 30%	

Notes: In Panels A-C, the observations are the 10% sample of males identified as black and under age 55 in the base year in the 1880-1930 Censuses linked in two consecutive censuses. In Panel D, the observations are linked individuals amongst the 10% sample of males identified as black under age 55 in the 1880-1920 Censuses linked in three consecutive censuses.

Table 2: Race Transition Matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
				I. Passing Across All Racial Categories				
						Race in t+10		
			Link %	Black	White	Chinese	Japanese	Native American
A.	Black		9.3%	84.8%	15.0%	0.1%	<0.01%	0.1%
B.	White		20.0%	0.7%	99.1%	<0.01%	<0.01%	<0.01%
C.	Chinese		7.3%	0.5%	8.2%	90.5%	0.4%	0.1%
D.	Japanese		5.4%	0.2%	7.0%	0.2%	91.6%	0.4%
E.	Native American		14.6%	2.2%	27.8%	0.1%	0.1%	69.0%

II. Passing Across East Asian Categories

	Japanese or Korean to Chinese	Chinese or Korean to Japanese	Chinese or Japanese to Korean
F.	0.3%	0.4%	0.1%

Notes: Observations in row A are the 10% sample of males identified as black and under age 55 in the base year in the 1920-1930 Censuses linked using 2SUP. Observations in row B are the 2% sample of males identified as white and under age 55 in the base year in the 1920-1930 Censuses linked using 2SUP. Observations in rows C-E are the 100% population of males identified as Chinese, Japanese, or Native American (respectively) under the age of 55 from the 1920-1930 Censuses linked using 2SUP. Observations in Row F are the 100% population of males identified as Chinese, Japanese or Korean under the age of 55 from the 1920-1930 Censuses linked using 2SUP.

Table 3: Mulattos and Individuals with Distinctively Black or White Names

	(1)	(2)	(3)
	% Black	% Mulatto	% White
Panel A: Mulatto			
Mulatto in 1910, Race in 1920	54.2%	25.9%	19.5%
Mulatto in 1920, Race in 1930	85.1%	NA	14.3%
Panel B: Black and not Mulatto			
Black in 1910, Race in 1920	74.0%	10.3%	15.7%
Black in 1920, Race in 1930	84.8%	NA	15.2%
Panel C. All Black (including Mulattos), Distinctively Black or White Names, % Pass to White			
Baseline (All names)	16.6%		
Distinctively black names	7.4%		
Distinctively white names	24.3%		

Notes: Observations in Panels A and B are the 10% sample of males identified as either mulatto or black and under age 55 in the base year in the 1910-1920 censuses (unless stated otherwise) linked in two consecutive censuses using 2SUP. In Panels A and B, columns (1)-(3) do not add up to 100% due to the presence of other racial categories. In Panel B, distinctively black names are taken from Cook et. al. (2014): Abe, Abraham, Alonzo, Ambrose, Booker, Elijah, Freeman, Isaac, Isaiah, Israel, King, Master, Moses, Pearlie, Percy, Perlie, Purlie, Presley, Presly, Prince, Titus. Distinctively white names are chosen by the authors using an analogous method (see text): Albert, Arthur, Carl, Charles, Clarence, David, Edward, Frank, Fred, George, Harry, Jacob, John, Joseph, Louis, Paul, Peter, Thomas, Walter, William.

Table 4: Marital Status, the Race of the Spouse and Passing

			Black$_t$		White$_t$	
			Obs.	Pass Rate	Obs.	Pass Rate
	t	t+10	(1)	(2)	(3)	(4)
A.		Single	73814	15.1%	267111	0.7%
B.	Single	Married to Black	25551	0.2%	1151	98.8%
C.		Married to White	6419	98.8%	105539	0.0%
D.	Married to Black	Single	2306	35.7%	3	66.7%
E.		Married to Black	37086	0.2%	15	86.7%
F.		Married to White	6523	98.6%	16	0.0%
G.	Married to White	Single	15	66.7%	3391	4.8%
H.		Married to Black	97	0.0%	1098	99.4%
I.		Married to White	213	91.1%	175593	0.0%

Notes: Observations in columns (1) and (2) are the 10% sample of males identified as black and under age 55 in the base year in the 1880-1930 censuses linked using 2SUP, who have non-missing values for marital status. Observations in columns (3) and (4) are the 2% sample of males identified as white and under age 55 in the base year in the 1880-1930 censuses linked using 2SUP. Additional restrictions regarding marital status and the race of the spouse are stated in the row headings.

Table 5: The Racial Composition of Children and Passing

			Black$_t$		White$_t$	
			Obs.	Pass Rate	Obs.	Pass Rate
	t	t+10	(1)	(2)	(3)	(4)
A.		No Kids	9532	10.1%	26780	0.5%
B.	No Kids	>=1 Black Kid	5434	0.3%	190	94.7%
C.		>= 1 White Kid	2299	98.5%	19093	0.0%
D.	>=1 Black Kid	No Kids	6785	15.8%	12	16.7%
E.		>=1 Black Kid	24513	0.3%	12	91.7%
F.		>= 1 White Kid	3882	97.8%	24	0.0%
G.	>= 1 White Kid	No Kids	36	63.9%	19197	2.3%
H.		>=1 Black Kid	104	0.0%	639	95.0%
I.		>= 1 White Kid	150	100.0%	129027	0.0%

Notes: Observations in columns (1) and (2) are the 10% sample of males identified as black and under age 55 in the base year in the 1880-1930 censuses linked using 2SUP. Observations in columns (3) and (4) are the 2% sample of males identified as white and under age 55 in the base year in the 1880-1930 censuses linked using 2SUP. Additional restrictions regarding the number and race of children are stated in the row headings.

Table 6: The Geography of Passing

Panel I.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Individuals who Pass				Individuals who do not Pass			
	Move County	Move State	Move County or State	Leave South (Confederate States)	Move County	Move State	Move County or State	Leave South (Confederate States)
A. All Years	49.7%	38.7%	88.4%	11.9%	22.7%	16.5%	39.2%	7.6%
B. 1880-1900	51.3%	40.8%	92.1%	11.9%	32.4%	18.9%	51.3%	5.0%
C. 1900-1910	55.0%	35.3%	90.3%	10.5%	26.7%	13.4%	40.1%	4.3%
D. 1910-1920	49.6%	36.1%	85.7%	12.4%	23.9%	16.4%	40.3%	7.6%
E. 1920-1930	50.0%	41.7%	91.7%	13.3%	21.4%	20.5%	41.9%	11.9%
F. 1930-1940	44.7%	39.9%	84.6%	11.4%	17.6%	14.3%	31.9%	7.1%

Panel II. Individuals who Pass, All Years

	All States	North (Union States)	South (Confederate States)	All Slave States in 1860	States Comprising 98% of the Black Population
G. By State of Birth	16.6%	23.8%	15.5%	16.0%	16.5%
H. By State of Residence in Year t	16.6%	21.2%	15.4%	15.8%	16.6%
I. By State of Residence in Year t+10*	16.6%	17.0%	15.8%	16.1%	16.3%

Notes: Observations are the 10% sample of males identified as black and under age 55 in the 1880-1930 Censuses linked in two consecutive censuses using 2SUP. Panel I: Columns (1) and (5) report moving counties within a state; columns (2) and (6) report moving states; columns (3) and (7) are the sums of previous two columns; columns (4) and (8) report those who leave the South: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, Tennessee and Virginia. In Panel II: column (5) includes Massachusetts, Connecticut, California, Illinois, Indiana, Iowa, Maine, Michigan, Minnesota, New Hampshire, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Vermont and Wisconsin. Column (6) includes the same states as Panel I columns (4) and (8). Column (7) includes the states in column (6) and Delaware, Kentucky, Maryland and Missouri. Column (8) includes Georgia, Mississippi, Alabama, South, Carolina, North Carolina, Louisiana, Texas, Virginia, Tennessee, Arkansas, Florida, Washington, Kentucky, Maryland, Pennsylvania, New York, Missouri, Ohio, Illinois, New Jersey, Oklahoma, Indiana, West Virginia, Michigan and Kansas. *State of residence is residence in t+20 for individuals linked over the 1880-1900 interval.

Table 7: The Correlation between Passing and Base Year Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
			Dependent Variable: White in year t+10			
Dep. Var. Mean	**0.170**	**0.171**	**0.145**	**0.158**	**0.170**	**0.170**
Miscegenation Legal	-0.319***					
	(0.106)					
Democratic PCA		0.00659*				
		(0.00350)				
Standardized Coef.		*0.0462*				
Black Educational Opportunity PCA			0.0219***			
			(0.00516)			
Standardized Coef.			*0.0403*			
Literacy (Read and Write)			-0.00361			
			(0.00566)			
WTB Occupational Income Score			-0.00605			
			(0.00805)			
Standardized Coef.			*-0.00589*			
Occupational Score (000's)			0.000946***			
			(0.000350)			
Standardized Coef.			*0.0219*			
Married				-0.0444***		
				(0.00763)		
Number of Head's Children in Household				-0.00639***		
				(0.000727)		
Distinctively Black Name (Cook et al. 2014)				-0.0805***		
				(0.00850)		
County Black Population Share					-0.00863	-0.00643
					(0.0239)	(0.0238)
County All Immigrant Share					0.345**	
					(0.140)	
County Mediterranean Immigrant Share						-0.575
						(0.509)
County Northern European Immigrant Share						0.454***
						(0.163)
County Other Immigrant Share						0.533
						(0.395)
Urban					0.0143***	0.0143***
					(0.00521)	(0.00520)
Observations	77,402	56,465	19,983	44,299	77,402	77,402
R-squared	0.011	0.056	0.007	0.069	0.048	0.048
Region FE	State	County	State	County	County	County
Standard Errors Cluster	State-Year	County-Year	County-Year	Newey-West	County-Year	County-Year

Notes: Observations are the 10% sample of males identified as black and age 15-54 in the base year of the 1910-1930 censuses linked in two consecutive censuses using 2SUP. In column (2), Democratic PCA is the first principal component all elections for U.S. president, the U.S. Congress and the U.S. Senate that has taken place in the past ten years. Column (3) is restricted to Southern states. The black education opportunity PCA is the first principal component of 4 variables: the # of black universities in a state and year; black-to-white teacher salary, white-to-black pupil-teacher ratio and black-to-white term lengths in a county and year. In column (6), Northern European countries comprise of all European countries that do not border the Mediterranean. All regressions control for age category dummy variables (15-24, 25-34, 35-44, 45-54), year fixed effects and region fixed effects as stated in the table. *** p<0.01, ** p<0.05, * p<0.1

Table 8: The Correlation between Passing and Income

	Occ. Score$_{t+10}$	Moved County or State$_{t+10}$	Occ. Score$_{t+10}$	Occ. Score$_{t+10}$
	(1)	(2)	(3)	(4)
Dep. Var. Mean (Std. Dev.)	**16.4 (9.76)**	**0.48 (0.5)**	**16.4 (9.76)**	**16.4 (9.76)**
White$_{t+10}$	3.017***	0.476***	1.828***	3.178***
	(0.120)	(0.00354)	(0.306)	(0.127)
Mulatto$_t$ x White$_{t+10}$				-1.301***
				(0.376)
Mulatto$_t$				0.454***
				(0.130)
Occupational Score$_t$	0.122***	0.000221	0.119***	0.123***
	(0.00727)	(0.000200)	(0.00724)	(0.00726)
Urban$_t$	1.408***	0.0433***	2.800***	1.400***
	(0.126)	(0.00561)	(0.128)	(0.125)
Married$_t$	0.125	-0.0661***	0.190**	0.127
	(0.0932)	(0.00426)	(0.0926)	(0.0932)
Literacy$_t$	0.606***	-0.0419***	0.512***	0.597***
	(0.0840)	(0.00445)	(0.0828)	(0.0841)
Distinctively Black Name$_t$	-0.0379	0.00247	-0.0827	-0.0265
	(0.210)	(0.0114)	(0.206)	(0.210)
Passed$_{t+10}$ x Moved County$_{t+10}$			0.926***	
			(0.359)	
Passed$_{t+10}$ x Moved State$_{t+10}$			1.016***	
			(0.361)	
Passed$_{t+10}$ x Rural to Urban$_{t+10}$			2.768***	
			(0.350)	
Observations	77,440	77,440	77,440	77,440
R-squared	0.092	0.208	0.114	0.092

Notes: Observations are the 10% sample of males identified as black and age 15-54 in the base year in the 1910-1930 censuses linked in two consecutive censuses using 2SUP. All regressions control for age category dummy variables (15-24, 25-34, 35-44, 45-54), year and base-year county of residence fixed effects, as well as age-group dummy variables. Column (3) additionally controls for the uninteracted dummy variables for moving counties, moving states and moving from rural to urban. Robust Newey-West standard errors are presented in the parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Racial Composition of the County of Residence

(a) Passing to white vs. remaining black
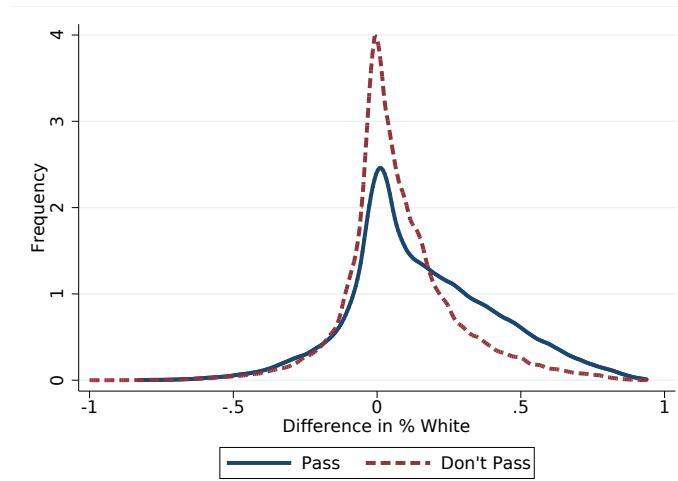
(b) Reverse-passing to black vs. remaining passed for white
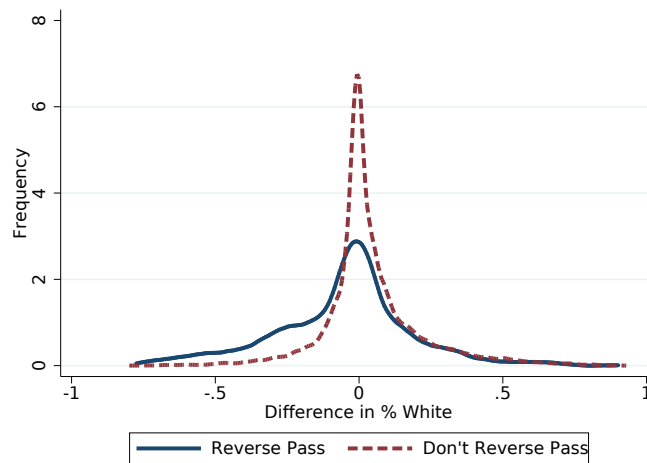
Notes: The y-axis is the PDF. The x-axis is the % white in the county of residence in the current year minus the % white in the county of residence in the previous census year.

Online Appendix

A Case Studies of Racial Passing: Anita Hemmings, Harry Murphy, and the Johnston Family

There are many cases of racial passing that are discussed in detail by historians and biographers. To help illustrate the environment and fix ideas, we provide three cases here: Anita Hemmings, Harry Murphy, and Dr. Johnston and his family.

Anita Hemmings, shown in Appendix Figure A.3 panel A, had parents who were both identified as "black". She attended Vassar College as a "white" student, was discovered to be black in 1897, but still graduated. She later married a black man. Both passed for white and raised their children as white. Her daughter, Ellen Love, also attended Vassar as a white woman (Mancini, 2002).

Harry Murphy is an example of an individual who started passing for white due to enumerator error, then decided to actively pass, and then reverse-passed to black. His photo is in Appendix Figure A.3 Panel B. When he entered the Navy, an official checked the box for "white" for his race. This allowed him to participate in the Navy's V-12 program for training officers and take classes at Ole Miss in 1945 as a white student, where no one questioned his identity. When the V-12 program ended, he transferred to Morehouse College in his hometown of Atlanta, where he returned to living as a black man. He later moved to New York City, where he may have lived as a white man again (Hobbs, 2014).

Dr. Johnston and his wife Thyra Johnston were both born black. Dr. Johnston attended the University of Chicago Medical School as one of two black students. Upon graduation, Dr. Johnston failed to secure a radiology position in the few hospitals that would hire black interns. However, when his race was not mentioned, he was able to obtain a position in a Maine hospital. He and his family moved there and passed for white for a number of years. In 1940, the Navy began to recruit Dr. Johnston but then ended the process upon hearing reports that Dr. Johnston might be black, which led Dr. Johnston to reveal to his children their background (Thomas, 1995). The photo of Dr. Johnston and his family is shown in

Appendix Figure A.3.

B   U.S. Historical Censuses

B.1   Race Categories

Racial categories used in each Census (1880-1940) are reported in Appendix Table A.1. "Black" and "Negro" are synonymous and interchangeably used by the Censuses. For the years that "mulatto" is reported as a separate category, our study defines "black" to be any individual in either the black or mulatto census categories. Below, we report the enumerator instructions for the relevant racial categories for this paper - black, white, and mulatto - for each decade.

- **1880:** "It must not be assumed that, where nothing is written in this column, "white" is to be understood. The column is always to be filled. Be particularly careful in reporting the class mulatto. The word is here generic, and includes quadroons, octoroons, and all persons having any perceptible trace of African blood. Important scientific results depend upon the correct determination of this class in schedules 1 and 5."

- **1900:** No specific instructions given, apart from black being defined as "negro or of negro descent" when listing the categories. Enumerator instructions state "Under these words write "White" "Black" (negro or of negro descent), "Indian" "Chinese" or "Japanese," as the case may be."

- **1910:** "Write "w" for white; "B" for black; "Mu" for mulatto; "Ch" for Chinese; "JP" for Japanese; "In" for Indian. For all persons not falling within one of these classes, write "Ot" (for other), and write on the left-hand margin of the schedule the race of the person so indicated. For census purposes, the term 'black' (B) includes all persons who are evidently full-blooded negroes, while the term 'mulatto' (Mu) includes all other persons having some proportion or perceptible trace of negro blood."

- **1920:** "Write "w" for white; "B" for black; "Mu" for mulatto; "In" for Indian; "Ch" for

2

Chinese; "Jp" for Japanese; "Fil" for Filipino; "Hin" for Hindu; "Kor" for Korean. For all persons not falling within one of these classes, write "Ot" (for other), and write on the left-hand margin of the schedule the race of the person so indicated. For census purposes the term 'black' (B) includes all Negroes of full blood, while the term 'mulatto' (Mu) includes all Negroes having some proportion of white blood."

- **1930:** Write "W" for white; "Neg" for Negro; "Mex" for Mexican; "In" for Indian; "Ch" for Chinese; "Jp" for Japanese; "Fil" for Filipino; "Hin" for Hindu; and "Kor" for Korean. For a person of any other race, write the race in full." Additional instructions are given for "Negro" and "Indian": "A person of mixed white and Negro blood should be returned as a Negro, no matter how small the percentage of Negro blood. Both black and mulatto persons are to be returned as Negroes, without distinction. A person of mixed Indian and Negro blood should be returned a Negro, unless the Indian blood predominates and the status as an Indian is generally accepted in the community. A person of mixed white and Indian blood should be returned as Indian, except where the percentage of Indian blood is very small, or where he is regarded as a white person by those in the community where he lives."

- **1940:** "A person of mixed white and Negro blood should be returned as a Negro, no matter how small the percentage of Negro blood. Both black and mulatto persons are to be returned as Negroes, without distinction. A person of mixed Indian and Negro blood should be returned as a Negro, unless the Indian blood very definitely predominates and he is universally accepted in the community as an Indian. A person of mixed white and Indian blood should be returned as Indian, if enrolled on an Indian Agency or Reservation roll; or if not so enrolled, if the proportion of Indian blood is one-fourth or more, or if the person is regarded as an Indian in the community where he lives."

C Linking

C.1 2-Sided Linking for Reducing False Positives

This section motivates the two-sided linking algorithm (as opposed to the one-sided linking typically used in past studies). The main concern for linking is Type I error (which in our context means observing a race change by error). This difficulty is particularly important for our study because there are many more white individuals than black individuals in the population. Thus, forward linking alone is likely to incorrectly link a black individual to a white individual and cause us to overstate the number of individuals who pass.[77] Additionally requiring the linked individual in year $t + 10$ to have a unique perfect spelling link in year t addresses this problem. The main conceptual advantage of 2SUP is that, by additionally linking backwards, it accounts for the possibility that there may exist other individuals in year t who are equally good or better links for the individual identified as a match by the one-direction link.

To identify a change in race, we need to trace individuals over time. The main difficulty in linking individuals over time is that there are no unique individual identifiers in the historical censuses to form the link.[78] The most important variables for distinguishing individuals are a person's first and last names. However, names are usually insufficient for constructing unique identifiers because most names are common to more than one individual and there will be multiple potential links. We also use other "blocking" variables, such as year of birth and place of birth, to restrict the sample.[79] This mitigates, but cannot fully resolve the fundamental problem of multiple potential matches.

A standard practice in the economic history literature is to drop all observations for which there are multiple potential links and keep only individuals from year t for whom a link can be formed with one and only one individual in year $t + 10$ (e.g., Abramitzky, Mill,

---

[77]Note that in principle, we can also generate incorrect links between black individuals in year t to other black individuals in year $t + 10$. But this will not lead us to overstate the number of passers.

[78]Social Security numbers were introduced in the United States in 1935 with the New Deal.

[79]We do not block on variables that can be affected by changing racial classification, such as migration or marriage.

and Pérez, 2020; Feigenbaum, 2016; Long and Ferrie, 2013a). To maximize the accuracy of the links, past studies often require perfect spelling matches of first and last names. When we link forwards we do exactly this.[80] We will call it the Unique Perfect link.[81] We link forwards from 1880 to 1900, from 1900 to 1910, and so on, up until 1940.

However, one-direction Unique Perfect links do not rule out the possibility of false positive findings in our context – i.e., we identify someone as passing for white because we incorrectly link a black individual to a white individual with the same name. Consider the simple example illustrated in Appendix Figure A.4a, where in year t, there are three hypothetical black males, Samuel, Elijah and Abe, and there are no white Samuels or Elijahs and many white Abes. In year $t + 10$, there is one and only one black Samuel, and one and only one white Elijah and Abe. The standard one-direction Unique Perfect linking algorithm will link the black Samuel$_t$ to the black Samuel$_{t+10}$, the black Elijah$_t$ to the white Elijah$_{t+10}$ and the black Abe$_t$ to the white Abe$_{t+10}$. We will observe that one stayed black (Samuel$_t$) and two passed for white (Elijah$_t$ and Abe$_t$).

Consider the possibility that while Elijah truly passed in our example, Abe did not. The one and only one Abe in year $t + 10$ is actually the future self of one of the many white Abes from year t (and the black Abe could have disappeared for reasons such as mortality, enumerator error, or transcription error). One-direction linking cannot take into account the possibility that there may be other individuals from year t who may be better matches with the linked individual in year $t + 10$.

This difficulty is particularly important for our study because there are many more white individuals than black individuals in the population. Thus, the problem we just described is likely to incorrectly link a black individual to a white individual and cause us to overstate the number of individuals who pass.[82] To address this, we additionally require the linked

---

[80]Because we require perfect first and last name links, we will be unable to link an individual who changes his name in order to pass. This will cause us to understate the true rate of passing, even when we compute pass rates with weights. Note that the same logic applies to other blocking variables. For example, if individuals who pass intentionally change their birth state, they will not be in the linked sample. This exclusion will lower the rate of passing in the linked sample.

[81]To see the rates of passing in samples that link forward and allow for slight misspelling of the name, see Nix and Qian (2015).

[82]Note that in principle, we can also generate incorrect links between black individuals in year t to other black

individual in year $t + 10$ to have a unique perfect spelling link in year t. For brevity, we call this the 2-Sided Unique Perfect (2SUP) link.[83]

Note that 2SUP is not equivalent to a one-sided link where the sample in the initial year has already dropped all duplicates within a specified age band. 2SUP is more conservative because it requires that the match be unique in both directions.

The main advantage of 2SUP is that, by additionally linking backwards, it accounts for the possibility that there may exist other individuals in year t who are equally good or better links for the individual identified as a match by the one-direction link. To illustrate this, examine Appendix Figure A.4b. It shows that 2SUP will be similar to the one-direction linking algorithm and identify Samuel and Elijah as links. It also shows that Abe_{t+10} can be linked to many individuals in year t and so the algorithm will drop Abe from the sample. Relative to the one-direction link illustrated in Appendix Figure A.4a, imposing the additional backwards link has dropped the false link and reduced the number of individuals observed as passing in this example from two to one. Since there are more whites than blacks in the population, the additional individuals dropped are more likely to be white. Thus, the restriction imposed by the two-direction linking will, on average, reduce false positive passing in the linked sample. Figure A.4b also shows that simple enumeration errors are unlikely to result in false positive passing. If Abe is black in year t and never intends to pass for white in $t + 10$, but is incorrectly enumerated as white, the white Abe in $t + 10$ will not have a unique link when linking backwards, and Abe will drop out of the sample.

For example, Abramitzky, Boustan, Eriksson, Feigenbaum, and Pérez (2019) finds that there are approximately 10-15% of names that suffer from transcription error in the historical censuses. Such errors will generally lead to the individual being dropped from the linked sample. *Enumeration error can only lead to false positive passing if in addition to mistakingly reporting Abe as white in $t + 10$, there are no white (or other black) Abes in year t.*

---

individuals in year $t + 10$. But this will not lead us to overstate the number of passers.

[83]Abramitzky, Boustan, Eriksson, Feigenbaum, and Pérez (2019) uses two-directional links to test the robustness of their well-known study, Abramitzky, Boustan, and Eriksson (2012). There are similarities between our linking algorithm and the full population linking with restrictions to unique matches used to match historical Norwegian data in Modalsli (2016) and Modalsli (2017).

For computational feasibility, we use a randomly selected 10% sample of the black male population in each base year t. The restriction only applies to the base year sample. We always link to the full population of all males in the subsequent census, and again to the full population of all males in the initial census when linking backwards. Thus, the restriction should not affect the accuracy of the links.

C.2   2SUP Algorithm

This section describes the 2SUP algorithm in detail. See Appendix Figure A.5 for an illustration.

C.2.1   Constructing the Using Sample

We begin with a 10% random sample of the black male population in year t and all males in the following census in year $t + 10$. For each sample, we drop observations with missing values for first or last names or age. We do not use information on the middle name.[84] Then, we create phonetically spelled names for all remaining individuals using Phonex.[85] We take each black male in year t and search through all males in year $t + 10$ to identify all Phonex matches. Phonex matches are generous in that they are the most likely to contain the true match absent serious spelling mistakes (Lait and Randell, 1996; Snae, 2007). The average black male has 6,822 potential Phonex links. This step reduces the computation intensity of the later step of looking for perfect spelling matches by limiting the number of individuals in the other census that we search through.

The next step is to restrict this sample. The first restriction we impose is to only keep the links where the reported age of the individual in year $t + 10$ is within a six-year interval of the predicted age of the individual from year t, e.g., keep if $age_t + 7 \leqslant age_t + 10 \geqslant$

---

[84] We follow Feigenbaum (2016) in removing middle names/initials.

[85] Phonex first retains the first letter of the name. Next, all vowels are dropped and the consonants are replaced with a number, with consonants that sound the same (for example, c and k) assigned the same number. Lastly, all but the first three numbers are dropped, so that all translations consist of a letter followed by three numbers.

Another popular method in the literature is NYSIIS, which has been used by studies such as Ferrie (1996). The main difference between Soundex and NYSIIS is that the latter preserves the position of vowels in its translation of names - all vowels are replaced with the letter A - and more than 4 characters are retained. Snae (2007) shows that compared to NYSIIS, Phonex yields twice the number of true matches without sacrificing accuracy.

$age_t + 13$. The second restriction is to only keep links that have matching birth places (birth states and birth countries for foreign-born individuals).[86] In other words, we block on age and the state of birth. At this point, there are twenty-four potential links for the average black male in year t. We call this sample the "Using Sample" and everything we discuss henceforth uses the Using Sample.

C.2.2 2SUP

To maximize accuracy in the links that are formed, we restrict links to those with perfect spelling matches, which we measure using the Jaro-Winkler score (i.e., we require a score of two), in the Using Sample.[87] We only keep individuals in year t that have one and only one perfect spelling match in year $t + 10$. All others are dropped. This forms the Unique Perfect Sample. For the one-direction link, we closely follow existing studies, with the key difference being that we do not require the race to be the same to form a link.

We depart from the literature by requiring the linked individual in year $t + 10$ to have a unique perfect spelling link in year t. In practice, we repeat the linking algorithm, starting with the linked individuals from year $t + 10$ and for each, we search through the full population of males in year t for possible links. We only keep individuals that have one and only one perfect spelling match in year t. Thus, 2SUP links are individuals in year t that have a unique forward link in year $t + 10$, and also a unique backwards link in year t.

C.3 Sensitivity Analysis

Appendix Table A.2 presents the results from several sensitivity checks motivated by the literature on linking, and in particular, Bailey, Cole, Henderson, and Massey (2017). To reduce computational intensity, we restrict our attention to black men born in three states with a large black population: Alabama, Georgia and Louisiana. First, we implement the baseline 2SUP algorithm as in the main analysis and find that we are able to link 8.1% of

---

[86]We manually correct misspelled birth states and countries so that there are no losses in observations due to mistaken spellings.

[87]Note that we have also examined less restrictive spelling matches, where we allow for some misspellings. The results are similar.

black males born in these states (column 1). For the linked sample, 13.5% pass for white (column 2).

Next, we investigate what happens if we include all black males in the base year instead of restricting the sample to those under 55 years of age (as we will continue to do for the remaining exercises). We find that both the percentage of the population linked and the rate of passing in the linked sample are very similar to the baseline. Another check is to use the 100% sample instead of the 10% sample. This unsurprisingly produces similar results, given that the 10% sample is randomly selected from the full population.

A common problem in linking comes from age heaping in the historical data, where many more individuals report ages that are products of five (relative to other ages). To investigate whether our estimates for passing are biased by age heaping, we divide the data into individuals whose age is a product of five and everyone else. Table A.2 rows D and E show that the percentage of the population linked is slightly higher (8.4% vs 7.5%) and the rate of passing in the linked sample is slightly lower (13.1% vs 14.4%) for individuals whose ages are not products of five. But the difference is small.

The main linked sample allows links to be formed within individuals who are within within plus or minus three years of the predicted age. Here, we alternatively expand the age interval to be within plus or minus five years of the predicted age, or shrink the interval to be within plus or minus one year of the predicted age. Table A.2 rows F and G show that the rates of passing are similar or higher (13.6% and 15%) with these alternative ways of blocking on age.

Next, we require the links to match on parents' birth states for years where the latter data are available.[88] The estimated rate of passing in Row H is similar to the baseline. One caveat for this robustness check is that most individuals are born in the same states as their parents, which limits the additional information of matching on parents' birth states. To address this, we impose the added restriction that either the mother's or the father's birth state must be

---

[88]Parental birth state is reported in the censuses of 1880 and 1900-1930. Bailey, Cole, Henderson, and Massey (2017) suggests using parents' birth states to validate links (that do not already block on these variables). This is analogous to what we do.

different from the linked individual's birth state. Row I shows that the rate of passing is 13.6%, which is similar to the baseline. Interestingly, note that the rate of links declines from 8% in Row H to 1% in Row I, which implies that 12.5% ($0.01/0.08 = 0.125$) of linked individuals are born in different birth states from at least one parent. This is comparable to our finding that 13.4% to 18.9% of black men who do not pass migrate across states prior to the Great Migration (see Table 6 Panel I column 6).[89]

Finally, in Row J, we follow the method from Abramitzky, Boustan, and Eriksson (2014) and use a sample restricted to black males with unique name-birth state-predicted age interval combinations. The rate of links should be higher for this subsample because there should be fewer multiple matches, which are dropped by the 2SUP algorithm. Indeed, we find a slightly higher link rate of 8.2%. Reassuringly, the rate of passing in this sample, 13.4%, is comparable to the baseline, 13.5%.

C.4 Alternative Linking Methods

This section provides a brief discussion of alternative linking algorithms presented in Appendix Table A.4. The rate of passing for white within the linked sample generated by 2SUP is smaller than those generated by all of the alternative algorithms.

ABE Exact, ABE NYSII and ABE JW In earlier works, the linking was only conducted in one direction, linking an individual forward in time from year t to t + 10. In a recent working paper, Abramitzky, Boustan, Eriksson, Feigenbaum, and Pérez (2019) additionally recommends two-sided linking and provides a Stata command for producing the results. We follow the nomenclature from this paper to refer to the respective algorithms (e.g., ABE Exact, ABE NYSII, ABE JW, and EM). The replications implement their code for data cleaning and linking, which is publicly available at https://ranabr.people.stanford.edu/matching-codes (March 2020 version).

Each algorithm identifies a set of links going forward in time. The procedure is then

---

[89]Note that for individuals to have a different birth state from a parent, that parent must move from his or her birth state prior to giving birth to the child. These migrating parents will be a subset of total migration.

repeated linking backwards in time. Only individuals with unique links forwards and backwards in time are ultimately identified as "linked". Conceptually, the two sided linking is similar to 2SUP, which is consistent with the fact that we find broadly similar rates of linking and rates of passing across methods in Table A.4. However, 2SUP produces the most conservative (i.e., lowest) rates of passing.

The only difference between ABE Exact and ABE NYSII is that one uses the usual spelling while the other uses the NYSII standardized names as linking variables. In the base year, only individuals with unique first name-last name-age-birthplace combinations are kept in the sample. Then, the algorithm looks at individuals in the next census year and identifies perfect matches on first and last names, birthplace and age. An individual is linked if there is one and only one link. If there are multiple links, the individual from the base year is dropped from the sample. If there are no links, the age window expands by one year in each direct, i.e., ± 1 years instead of a perfect match. A link is formed if there is one and only one link. If there are multiple links, the individual from the base year is dropped from the sample. If there are still no links, then the procedure is repeated by expanding the age to within two years in each direction, i.e., ± 2 years instead of ± 1 years or a perfect match. ABE Exact and ABE NYSIIS produces link rates of 13.4% and 16.4%. The rates of passing in the linked samples are 19.4% and 21.2%.

2SUP differs in requiring that the match be unique within ± 3 years (without expanding the age window in stages). This will cause 2SUP to have a lower rate of links (8.2% versus 13.4% and 16.4%) because there are cases where an individual has multiple links within the ± 3 years (which will cause 2SUP to drop the individual), but have exactly one match within a narrower age window (which will cause the link to be formed by ABE Exact and ABE NYSII).

ABE JW first conducts a crude linking exercise to restrict the sample. Individuals in the base year are linked to those in the next census interval with the same birth place, the same first letter for the first and last name, and where the reported age in the next census is within ± 5 years of the predicted age. A Jaro-Winkler (JW) score, which measures the

string distance between two names, is calculated for the individual in the base year and all possible matches in the following census of this restricted sample. For each potential link, the JW score ranges from 0 (a perfect spelling match) to 1. The ABE JW algorithm restricts the sample to potential links where the first and last names each have JW scores of 0.1 or less. A link is identified if the individual in the base year has one and only one potential link in the following census year. If there are multiple links, then the potential link with the closest age is identified as a link if there is only one potential link with that closest age and the next closest age distance is at least two years further away.[90] ABE JW links 6.8% of the sample. The rate of passing in the linked sample is 13.9%.

This algorithm will produce a lower rate of links than 2SUP because there are individuals with a perfect spelling match (i.e., JW score 0 for first name and for last name) and age match ± 3 years (where 2SUP will identify as a link), but who also have two or more potential links that have the same closest age and JW scores of less than 0.1. Conceptually, ABE JW is a more conservative linking algorithm than 2SUP because it drops "perfect" links if there are other possible individuals that could be noisily linked in terms of names. Table A.4 shows that the link rate for ABE JW is 6.8%, lower than the 8.2% for 2SUP. The rate of passing for white is slightly higher, 13.9% in the ABE JW sample versus 13.4% in the 2SUP sample.

EM   Expectation Maximization (EM) was initially developed by computer scientists and applied to linking historical data in Abramitzky, Mill, and Pérez (2020) and further discussed in Abramitzky, Boustan, Eriksson, Feigenbaum, and Pérez (2019).[91]

The algorithm begins by conducting the same crude linking exercise as ABE JW to restrict the sample, and also constructs JW scores for the names of the restricted samples the same way. The EM algorithm combines the two measures of distances - distance between names and age - to calculate a probability score that any potential link is the true link. A link is formed when the potential link with the highest probability score has a score that is

---

[90]The authors discuss that this is more cautious than the slight variation of taking the link with the closest age provided it is unique.

[91]We use the Stata and R-code shared by the authors at https://ranabr.people.stanford.edu/matching-codes from March 2020.

both sufficiently high and also sufficiently distant from the next best link. The thresholds for the probability score that the link with the highest score must exceed, and which the second highest must fall below are arbitrary. The percent of the population that is linked will decline as the first threshold increases and as the second threshold decreases. If no such link satisfies both criteria, then the individual is dropped from the sample.

The key advantage of EM over other algorithms is that it transforms the discretized information from the quality of the matches of names and age into a continuous variable, and it takes into account the possibility that enumerator error may cause the true link to not have the highest probability score. The limitation of EM is that the choices of the thresholds which are key to determining the links are arbitrary. For our replication exercise, we use thresholds of 90 and 85, which is similar to Abramitzky, Mill, and Pérez (2020). EM links 6.3% of the sample. The rate of passing in the linked sample is 14.8%.

Ferrie and Long   This is the only algorithm in Table A.4 that is not two-sided. We replicate it given its seminal importance in the literature of machine linking historical data. The well-known paper, Long and Ferrie (2013b), conducts many linking algorithms. We implement the one the authors describe as the most conservative in their appendix.[92] It is very similar to the one-direction Unique Perfect links presented in Table 1 of the paper. Links require perfect spelling matches of first and last names, birth states and parental births states (when available in the 1880, 1900, 1910, 1920, and 1930 Censuses). In addition, it requires a ± 1 year age match. Finally, it requires that all links are unique. A link is only formed if there is one and only one individual in year $t + 10$ which is a potential link for the individual in the base year t. This method links 9.9% of the sample. The rate of passing in the linked sample is 21.4%.[93]

Agreement Across Algorithms   Appendix Table A.5 presents the similarities of link probabilities across algorithms. Row A shows that the percent of individuals in the population

---

[92]We base our code on the discussion on page 8 of the Online Appendix of Long and Ferrie (2013b).

[93]One machine-linking method we do not replicate is Feigenbaum (2016), which implements a training algorithm that matches on race.

that are similarly linked or unlinked using 2SUP versus alternative algorithms range from 86.4% to 91.4%. Appendix Table A.5 Row B shows that amongst individuals in year t that are linked by both 2SUP and the alternative algorithm, the share which is linked to the same individual in year $t + 10$ ranges from 93.1% to 99.2%.

D Name Similarity

Here we consider the possibility that black and white men are more likely to share similar names than with other races. Specifically, the concern is that Chinese and Japanese men have more differentiated names from white men, such that enumerators are less likely to mistake them for white (as compared to mistaking a black man for a white man). To address this, we compare black and Asian name similarity with white men. First, we calculate the number of white men that have the same first and last name as a black man in the same state of residence. We look at the state of residence because this is arguably the population that is the most salient in the mind of the enumerator (e.g., a white enumerator in New York may not care or know about the distribution of names in other states). Given the Census instructions and the low literacy rates at the time, the first time an enumerator learns the name of the individual is most likely in hearing it (rather than seeing it written down). Thus, we use the Phonex soundings of the name for this exercise. When we do this, we find that for the average black man, there are approximately 111 white men in the same state of residence (averaged across all census years) with the same phonetic name. When we repeat the exercise for other races, we find the following analogous numbers: 174.7 for white, 114 for Chinese, 35 for Japanese and 59 for Native Americans. These numbers show substantial variation in name similarity with whites across races, with names for Native American and Japanese men being less common with whites. However, the names of Chinese and black men were similar in their commonality with white men.

E   Passing for White by Age and Year

Appendix Table A.6 shows the rates of passing for each census interval. The rates of passing are larger, 22.4% versus 15-18.1%, for the 1880-1900 interval. This is most likely because the accumulation of the number of individuals who passed for white over twenty years is larger than over ten years. The fact that it is not exactly twice the rates of passing for a ten-year interval is likely to be partly due to reverse passing.[94]

If we compare the four ten-year intervals in rows (B) to (E), we see that there is a slight decline in the rates of passing from 18.1% to 15%. This could be due to changes in the incentives or constraints for passing. It is also consistent with the notion that many individuals who wished to pass when Jim Crow began chose to do so in the earlier census intervals and remained white. Thus, they would not need to pass again and the rates of passing would naturally stabilize at slightly lower rates.

Next, we examine passing by age. The results show comparable rates and no distinct pattern of passing across ages groups other than that adults are slightly more likely to pass than children. This is consistent with the historical accounts that the decision to pass was often made as a family and at different points in life. Note that the rate of links declines for the older group (row K, age 45-54). This is consistent with the rise of mortality rates at higher ages.

F   Weighted Extrapolations

The population weights take into account two factors: the share of individuals in the population with the same set of characteristics and the share in the linked sample with the same set of characteristics. We aggregate the individual observations in the linked sample into cells according to observable characteristics, and then multiply the average rate of passing

---

[94]Linking over twenty years also results in a lower rate of links because the number of people who cannot be linked due to mortality will be higher if we look over a twenty-year period than a ten-year period.

in each cell by the population weight.[95] The weight is increasing in the share of the population with similar characteristics, and decreasing in the share of the linked sample with similar characteristics. The sample is aggregated along the following dimensions: age, literacy, whether an individual lives in an urban area, marital status, whether he has a distinctively black name, relationship to the household head, birth place, state of residence, household size, and the census year.[96] We use all permutations of these variables to group individuals into mutually exclusive and collectively exhaustive cells. The implied rate of passing for the population with population weights is 16%.

Alternatively, we can use probit weights in the spirit of Ager, Boustan, and Eriksson (2019), which re-weight linked individuals with the probability that those with similar characteristics are linked.[97] Conceptually, the two weights are similar in correcting for the possibility that linked individuals are not representative of the population. Both weighting schemes assume that individuals with similar observable characteristics will pass for white at the same rate. The population rate of passing with probit weights is 16.8%, which is very similar to the population-weighted rate of passing (16%), as well as the rate of passing in the 2SUP sample (16.6%).

One caveat for the weighted extrapolations is the interpretation for individuals with characteristic combinations that are not shared with any linked individuals. To see this, consider the population-weighted rates of passing, for which this problem is the most transparent. In this case, all individuals with characteristic combinations where no one with the

---

[95]Our approach is equivalent to multiplying each observation by a post-stratification weight (which is the standard method used for adjusting survey data to be representative) given by $\frac{PopShare_i | X_i}{2SUPSampleShare_i | X_i}$, and then taking the average pass rate across post-stratification weighted individuals.

[96]Birth place refers to birth states for U.S.-born individuals and birth countries for foreign-born individuals. Note that we do not use variables such as the number of children because we only observe the number of children for the household head, when, in reality, many adults who are not household heads may also have children. To maximize the amount of information used, we create dummy variables for all values of the variables listed above. We interpret missing values as simply another value that the given variable can take. Thus, observations that report missing values for these variables are not omitted.

[97]Following Ager, Boustan, and Eriksson (2019), we construct a weight for each observation: $\frac{1 - P_i(M_i = 1 | X_i)}{P_i(M_i = 1 | X_i)} \times \frac{q}{1-q}$. The propensity of being matched $P_i(M_i = 1 | X_i)$ is calculated using a probit of the probability of being linked conditional on the covariates X_i, and q is the proportion of records linked. We use the same observable characteristics for calculating this weight as for the population weights. Note that unlike for population weights, we use linear and quadratic measures of household size for computational feasibility.

exact same combination is linked will be dropped from the calculation because there are no linked individuals with which to calculate the cell-level rate of passing. Such individuals account for 38% of the population. A straightforward way to address this is to add these individuals back into the weighted extrapolations with the assumption that none of them passed for white. This adjustment will mechanically lower the population rate of passing to 9.9% ($0.16 \times (1 - .38) = 0.099$). The weighted estimates using probit weights mechanically include individuals who do not share the exact characteristic combination with anyone who is linked. However, to address the same conceptual concern, we make the following adjustment. We compute the probit, assuming that for characteristic combinations with no one linked, one individual is linked and he does not pass. Then, we calculate the weights and apply them to estimate the rates of passing in the adjusted sample. The adjusted weighted rate of passing is 6.8%.

Appendix Table A.8 presents the adjusted population-weighted race-transition matrix. The results are consistent with the incentives to change race and with the patterns seen in the race-transition matrix (Table 2) using only the linked sample. Recall that the transition matrix uses data for 1920 and later. Thus, the rates of passing from black to white will slightly differ from the full sample results discussed earlier.[98]

Interpreting these extrapolations assumes that selection (i.e., who decides to pass for white) is based on observables. Thus, they should be interpreted as merely illustrative.

---

[98]The transition matrix using the adjusted probit weights are very similar and available upon request.

Table A.1: Racial Categories in the U.S. Censuses, 1880-1940

	1880	1900	1910	1920	1930	1940
White	x	x	x	x	x	x
Black	x	x	x	x	x	x
Mulatto	x		x	x		
Indian	x	x	x	x	x	x
Chinese	x	x	x	x	x	x
Japanese		x	x	x	x	x
Korean				x	x	x
Filipino				x	x	x
Mexican					x	
Hindu				x	x	x
Other			x	x	x	x

Table A.2: Sensitivity Checks

		% Link	% Pass for White
		(1)	(2)
		2SUP	
A.	Baseline 10%	8.1%	13.5%
B.	100% Sample	8.2%	13.4%
C.	Include if age > 55 in year t	7.8%	13.5%
D.	Heaped Age (age is a multiple of 5 in base year)	7.5%	14.4%
E.	Omit Heaped Age	8.4%	13.1%
F.	+/- 5 year intervals	8.0%	13.6%
G.	+/- 1 year intervals	7.8%	15.0%
H.	Match on mothers' and fathers' birth states	8.0%	13.3%
I.	Match on mothers' and fathers' birth states, and birth state different from father's or mother's.	1.0%	13.6%
J.	Unique names (by birth state, +/-5 age interval)	8.2%	13.4%

Notes: Observations are the 100% population of males identified as black and under age 55 in the base year in the 1880-1930 Censuses linked in two consecutive censuses using 2SUP. Exceptions are stated in the row titles. The sample is restricted to individuals born in Alabama, Georgia and Louisiana in the base year. Parental birth states are available for 1880 and 1900-1930.

Table A.3: Falsification Exercises

		Individuals Who Pass	Individuals Who Do Not Pass
		(1)	(2)
A.	Literate to Illiterate	3.0%	5.1%
B.	Illiterate to Literate	32.9%	31.4%
C.	Difference in Avg. Children's Age	9.49	9.52

Notes: Observations are the 10% sample of males identified as black in the 1880-1930 Censuses linked with 2SUP. Row C restricts to the 1900-1930 Censuses linked with 2SUP, to children aged 0-5 in the base year, and to households with 10 or fewer children.

Table A.4: Replications Using Alternative Linking Methods

		% Link	% Pass for White
		(1)	(2)
A.	Dahis, Nix and Qian (DNQ) - 2SUP	8.2%	13.4%
B.	Abramitsky, Boustan and Eriksson - Exact	13.4%	19.4%
C.	Abramitsky, Boustan and Eriksson - NYSIIS	16.4%	21.2%
D.	Abramitsky, Boustan and Eriksson - JW	6.8%	13.9%
E.	Expectation-Maximization	6.3%	14.8%
F.	Long and Ferrie (2013)*	9.9%	21.4%

Notes: See Appendix Section D for detailed descriptions of each methodology. All replication exercises are restricted to men born in the following robustness states: Alabama, Georgia, and Louisiana. For these replication exercise we match 100% of the population in the base decades to 100% of the population in the next decades and then report the link and pass rates for black men.

Table A.5: Agreement in Links Across Linking Algorithms

		Alternative Algorithms		
	Sample	ABE-Exact	ABE-NYSII	ABE-JW	EM
		(1)	(2)	(3)	(4)
A.	Linked or Not Linked with 2SUP Full Population in Year *t* and Alt. Algorithm	91.4%	86.4%	91.1%	90.4%
B.	Linked to the Same Individual Linked by both 2SUP and in Next Census Alt. Algorithm	97.9%	93.1%	99.2%	95.0%

Notes: Observations are the 100% population of males identified as black and under age 55 in the base y the 1880-1930 Censuses linked in two consecutive censuses using 2SUP and the alternative algorithms s in the column headings. The sample is restricted to individuals who are born in Alabama, Georgia and Louisiana in the base year.

Table A.6: Passing by Age and Census Interval

	% Link	% Pass
	(1)	(2)
A. 1880-1900	6.4%	22.4%
B. 1900-1910	8.4%	18.1%
C. 1910-1920	9.0%	16.6%
D. 1920-1930	9.4%	15.0%
E. 1930-1940	9.6%	15.0%
F. Age < 5	10.9%	15.4%
G. Age 5-14	9.2%	14.3%
H. Age 15-24	8.2%	17.4%
I. Age 25-34	8.5%	18.2%
J. Age 35-44	8.4%	18.1%
K. Age 45-54	7.3%	18.8%

Notes: Observations are the 10% sample of males identified as black in the 1880-1930 Censuses linked using 2SUP.

Table A.7: Balance Statistics for 2SUP Sample

| | Using Sample | | | 2SUP | | | Not 2SUP | | | 2SUP- Not 2SUP | |
	Mean (1)	SD (2)	Obs. (3)	Mean (4)	SD (5)	Obs. (6)	Mean (7)	SD (8)	Obs. (9)	Dif in Means (10)	p-val (11)
Age	20.89	14.21	1,625,681	20.34	14.44	181,915	20.96	14.18	1,443,766	-0.62	<0.001
Literacy	0.54	0.50	1,423,591	0.54	0.50	163,685	0.54	0.50	1,259,906	<0.01	<0.001
Occupational Score	8.90	9.61	1,129,609	8.42	9.68	131,845	8.96	9.60	997,764	-0.54	<0.001
Head of Household	0.32	0.47	1,625,681	0.33	0.47	181,915	0.32	0.47	1,443,766	0.01	<0.001
Single	0.64	0.48	1,619,050	0.64	0.48	181,320	0.64	0.48	1,437,730	<0.01	<0.001
Married	0.33	0.47	1,619,050	0.34	0.47	181,320	0.33	0.47	1,437,730	0.01	<0.001
Number of Children	1.94	2.30	516,374	2.19	2.42	59,208	1.91	2.28	457,166	0.28	<0.001
North (Union States)	0.12	0.33	1,625,681	0.14	0.34	181,915	0.12	0.32	1,443,766	0.02	<0.001
Distinctively Black Name	0.01	0.12	1,625,681	0.02	0.15	181,915	0.01	0.11	1,443,766	0.01	<0.001
Lives in Birth State	0.70	0.46	1,625,681	0.74	0.44	181,915	0.70	0.46	1,443,766	0.05	<0.001

Notes: Observations are the 10% sample of males identified as black and under age 55 in the base year in the 1880-1940 censuses. Additional restrictions are stated in the column headings.

Table A.8: Race Transition Matrix – Implied Rates of Passing for the Population, All Races, Population Weights

	Implied Rates of Passing for the Population					
	(1)	(2)	(3)	(4)	(5)	(6)
		Race in t+10				
	Race in t (below)	Black	White	Chinese	Japanese	Native American
A.	Black	91.4%	8.5%	0.1%	<0.01%	<0.01%
B.	White	0.6%	99.3%	<0.01%	<0.01%	<0.01%
C.	Chinese	0.1%	1.6%	98.2%	0.1%	<0.01%
D.	Japanese	<0.01%	1.8%	0.0%	97.9%	0.1%
E.	Native American	1.1%	16.9%	0.1%	<0.01%	81.5%

Notes : Results from this table are obtained by extrapolating the estimates from Table 2 using population weights to the full population. See text.

Figure A.1: Emancipated Slaves in 1863

Source: "White and Colored Slaves" by C. C. Leigh (*Harper's Weekly*, January 30, 1864, p. 71).

Figure A.2: Emancipated Slaves in 1863

(c) Black Children Turned Away from White Schools

(b) Rebecca

(a) Isaac and Rosa

Source: "White and Colored Slaves" by C. C. Leigh (*Harper's Weekly*, January 30, 1864, p. 71).

Figure A.3: Illustrative Examples of Individuals who Passed from Black to White

(a) Anita Hemmings

(b) Harry S Murphy

(c) Dr. Johnston and Family

Notes: A: Anita Hemmings passed in order to attend Vassar which she graduated from in 1897 (Perkins, 1998). B: Harry S. Murphy attended the University of Mississippi from 1945-1946 as a white man. He later returned to life as a black man. When Ole Miss violently resisted James Meredith's integration of the campus in 1962, Murphy stated "they're fighting a battle they don't know they lost years ago" (Hobbs, 2014). C: Dr. Johnston and his family passed for white in the 1920s and 1930s to practice medicine (Hobbs, 2014).

Figure A.4: One-sided and 2-sided Unique (Perfect) Links

(a) Unique Perfect Link

Year	Black			White			
t+10	Samuel			**Elijah**	**Abe**		
t	Samuel	**Elijah**	**Abe**	Abe	Abe	Abe	Abe

(b) 2-Sided Unique Perfect (2SUP) Link

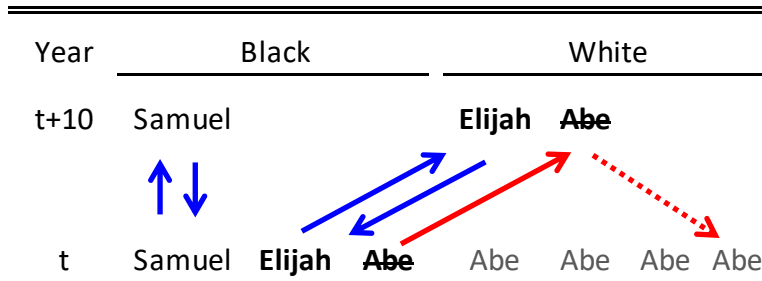| Year	Black			White			
t+10	Samuel			**Elijah**	~~**Abe**~~		
t	Samuel	**Elijah**	~~**Abe**~~	Abe	Abe	Abe	Abe
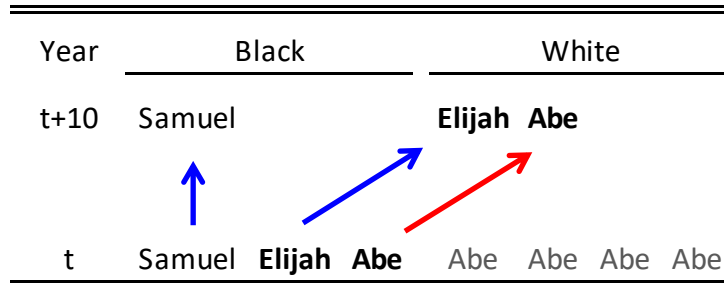
Figure A.5: 2SUP Linking Algorithm

Black Male Population in *t*   All Male Population in *t+10*

Keep: first and last name non-
missing, age non-missing.

Base Population in *t*   Target Population in *t+10*

Link on Phonex names

Avg. # potential links: 6,822. Population in sample: 98%.

Keep if pot. matches are within +/- 3 years and have same birth place

"Using Sample"
Avg. potential links: 24. Population in sample: 80%.

Keep if perfect spelling
match (JW2)

Keep if 1 potential match

Unique Perfect Sample
Avg. potential links: 1. Population in sample: 14.1%.

Link backwards to *t.* Repeat
procedure

2-Sided Unique Perfect (2SUP) Sample
Avg. potential links: 1 FB. Population in sample: 8.6%.