Choosing Racial Identity in the United States, 1880-1940*

Ricardo Dahis†, Emily Nix‡ and Nancy Qian§

November 26, 2019

Abstract

This paper documents that many black males experienced a change in racial classification to white in the United States, 1880 -- 1940, while changes in racial classification were negligible for other races. We provide a rich set of descriptive evidence on the lives of black men "passing" for white, such as their patterns of marriage, children, the passing of spouses and children, migration and income.

Keywords: Identity Economics, Cultural Economics, Racial Passing, Economic History, Political Economy. **JEL**: N3, J15

*We thank Ran Abramitzky, Martha Bailey, Leah Platt Boustan, Guillem Cassan, Katherine Eriksson, James Feigenbaum, Joe Ferrie, Claudia Goldin, Larry Katz, Magne Mogstad, Suresh Naidu, Santiago Perez and Marco Tabellini for their insights and numerous useful suggestions; the participants at the U.C. Berkeley Haas Oliver Williamson Seminar, Vancouver School of Economics Economic History Seminar, LSE/NYU Conference on Political Science and Political Economy, VATT Helsinki, Statistics Norway, and Washington Area Economic History Dinner. We are grateful to Suresh Naidu, Gregory Niemesh, Trevon Logan, and John Parman for sharing their data with us. All mistakes are our own. Comments and suggestions are welcome.

†Northwestern University, rdahis@northwestern.edu

‡University of Southern California, enix@usc.edu

§Corresponding author: Northwestern University, NBER, BREAD, CEPR, nancy.qian@kellogg.northwestern.edu

1 Introduction

Akerlof and Kranton (2000) famously argued that "the choice of identity may be the most important 'economic' decision people make. Individuals may – more or less consciously – choose who they want to be" and "previous economic analyses of, for example, poverty, labor supply, and schooling have not considered these possibilities". Since their seminal work, a growing number of papers have provided empirical evidence for this theory along several dimensions of identity, such as religious identity in historical Europe (Botticini and Eckstein, 2012), caste in India (Cassan, 2013) and ethnicity in China (Jia and Persson, 2013).

This paper examines one of the more rigid dimensions of identity -- race -- in the context of the United States during 1880-1940, when race was one of the most important determinants of a person's social and economic opportunities, and where race was considered exogenous and fixed over the lifetime.[1] The social and legal principle of the "one-drop rule" dictated that an individual was "black" if he had one or more black sub-Saharan African ancestors. This was also a context where extensive (often nonconsensual) racial mixing during previous generations caused a substantial share of the black population to have physical traits similar to Europeans.[2] "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).

The magnitude of the phenomenon of passing for white in this historical context has been the subject of heated debates for at least one hundred years. The main difficulty has been the lack of data. Sociologists have attempted to make indirect inferences from aggregate population statistics by calculating the "missing" black (or "extra" white) population

[1] We focus on men because the change in women's names with marriage makes it difficult to link women across censuses (see Section 4).

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

across censuses not accounted for by births, deaths or immigration. These accounting exercises result in a wide range of estimates and face the difficulty that the historical data, particularly vital statistics, are crude and measured with error (e.g. Eckard, 1947; Hart, 1921). Thus, scholars have not been able to achieve any consensus. Some have argued that passing was negligible on average. Others have argued that, for example, the number of black individuals who had passed for white in 1940 was over one-third of the black population at the time. Sollors (1999, p. 281) summarizes the evidence, which are "dramatically heterogeneous and range from hundreds to millions".[3] Recent genetics evidence suggests that the share of individuals who would have been black under the one-drop rule, but self identify as European Americans today are 20% in the United States, 49.6% in Boston and 57% in New York City. The main caveat for these numbers is that only the sample for Boston, which may not be representative of the United States, is randomly selected.[4]

Our study addresses this old question with a new technique from the economic history literature by linking individual census records over time and observing whether an individual who is black in the first census year remains black or becomes white in the next census year. This exercise relaxes the need for accurate vital statistics data.[5] We use a conservative linking algorithm that is highly unlikely to produce false links or over-state passing (i.e., false positive passing). In addition to estimating the number of those who pass, the linked sample allows us to provide descriptive evidence on the motivations and constraints for passing. Finally, to make progress and narrow the range of estimates for the rates of pass-

[3]At the low end, scholars such as Eckard (1947) compare the whole black population in 1930 to those over age ten in 1940 and finds that 26,000 black individuals (approximately 0.25% of the black population) are missing after accounting for mortality and net immigration. In another well-known study, Hart (1921) 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. In the Encyclopedia of African-American Culture and History, Sollors et al. (1995, p. 2108) discusses the wide range of existing estimates. He points out that on the higher end, it has been argued that approximately 30,000 African Americans were passing to white annually, implying that there were five to eight million white individuals who were black under the one-drop rule by the 1940s (which would have been 38% to 62% of the black population of 12.9 million at that time).

Also see, for example, Khanna and Johnson (2010) for a discussion of the range of estimates. For a detailed discussion of the difficulties in accounting exercises, see, for example, Elo and Preston (1997), Eriksson et al. (2018) and Shapiro (1950) and the references therein.

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

[5]For discussions about the difficulty of correctly assessing black mortality in the historical data, see for example, Eriksson et al. (2018), Elo and Preston (1994), and Preston et al. (1996).

ing for the black population, we use several weighted methods to extrapolate the rates of passing in the linked sample to the population.

The main challenge in linking is that historical censuses do not have unique individual identifiers. Thus, the links are made based on names and the few other variables recorded in the census. For our study, the key difficulties are that *i*) many names are common across individuals (even after we impose additional restrictions based on other variables); and *ii*) because there are many more whites than blacks in the population, mistaken links are likely to link a black individual to a white individual and show false passing.

To avoid such false positive results, we use an extremely conservative two-sided linking algorithm, which we will call the 2-Sided Unique Perfect (2SUP) algorithm. We follow standard procedures and link forward in time, looking for a perfect spelling match in names (in addition to restrictions on other variables such as birthplace and age).[6] We then look for a unique link where there is *one and only one* possible match in year $t + 10$. Then, we take the forward-linked individual in year $t + 10$ and link backwards – i.e., search for a perfect spelling match among all the males of all races (i.e., the original black male and all other males) in the previous decade (year t). We only link an individual if there is *one and only one* possible match in year t. This additional restriction mitigates false links because it takes into account that there may be others in the base year who could be equally good or better matches for the linked individual in year $t + 10$. For our study, this mitigates false positive passing from mistaken links because there are more white individuals than black individuals in each census. A mistakenly linked individual who is white in year $t + 10$ will likely link to a white person in year t and be dropped from our sample. We provide a more detailed discussion and examples in the paper.

Within the 2SUP links, we find that over 300,000 black males, or 16.6% of the linked sample, passed for white over the five census intervals during 1880-1940. Amongst the black men who passed for white, approximately 30% reverse-passed back to black in the following census. Since the linked sample is only a part of the total population, the lower bound rate

[6]The way we treat other variables is very similar to the existing methods. See Section 4 for a discussion.

of passing for the population can be obtained by assuming that unlinked individuals never pass for white – i.e., *at least* 1.4% of black males passed for white. However, this lower bound may seem extreme and we will return to discuss alternative methods for extrapolating to the population after discussing the results using the linked sample.

We interpret the observed passing as an *active* choice — i.e., 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, who determines the race in historical censuses.[7] There are two important caveats for our interpretation. The first is the concern that the 2SUP algorithm still creates mistaken links in a way that creates false positive passing – i.e., links a black man to a different man who is white in the subsequent census. To investigate this concern, we provide a large number of sensitivity checks to alternative samples and linking algorithms, as well as several placebo and falsification exercises. For example, we show negligible rates of passing in contexts where the socio-economic incentives to pass are low: from white to black, and across Asian categories of Chinese, Japanese and Korean. The second concern is that changes in racial classification for ostensibly *correctly* linked individuals reflect *inactive* passing rather than an *active* choice: white enumerators may have erroneously coded the race of some black men as white. We believe that this was unlikely because of residential segregation and the scrutiny over race in the historical context.[8] This concern is also difficult to reconcile with the findings of negligible rates of passing from white to black, which should present enumerators with similar difficulties; and across the three Asian races, for whom enumerators would presumably have a similarly, if not more, difficult time distinguishing. See Section 5.4 of the paper for a more detailed discussion.

The second part of our analysis examines when and how individuals passed for white. Amongst other findings, we provide descriptive results consistent with the belief that the lack of social and economic opportunities was an important motivation for an individual to

[7] See the Background and Data sections for a more detailed discussion.

[8] For example, in their study of race counts in the historical census, Strmic-Pawl et al. (2018) state "It is also important to note that from 1880 through 1960, enumerators, trained census officials, collected information from participants and assigned their race... 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".

pass, and that one of the main costs of passing for white was being cut off from his community and family. Individuals were more likely to pass in states where miscegenation was illegal, that were more democratic, where there were better opportunities for education, and if they were unmarried or had fewer children. Those who passed 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. We also provide additional results, such as for individuals who are classified as mulattos or who have distinctively black names. See Section 5.3.

The final part of our study reconsiders the rates of passing for the population. It is interesting to consider the population rates of passing under less extreme assumptions than the lower bound discussed earlier. To do this, we extrapolate from the linked sample using weights that account for the possibility that individuals who are linked may under- or over-represent the population share with similar observable characteristics. We use two types of weights. The first is a standard population weight. The second is a more flexible weight developed by Ager et al. (2019). We find that the implied rates of intercensal passing for the black population are likely to be around 6.8% to 9.9%. In Section 6, we approximate that the stock of black population implied to pass is broadly comparable to the current genetics evidence.

Taken together, the results provide strong evidence for the generalizable insight that identity can be a choice for a sizable number of individuals, even along dimensions as rigidly defined as race in pre-Civil Rights United States. As such, our findings contribute to several branches of the economics literature. First, we add to the literature on the economics of identity, which has traditionally comprised of theoretical studies (e.g., Akerlof and Kranton, 2000; Bénabou and Tirole, 2011).[9] We complement several recent empirical studies that document the correlation between identity and social and economic incentives

[9]We do not present a formal model in the paper because of space constraints, but can provide one upon request. Also see Austen-Smith and Fryer (2005) and Ruebeck et al. (2009) for studies of the contemporary U.S. context, and Bisin et al. (2016) for a study of the context of immigrants in Europe today.

in the context of caste in India (Atkin et al., 2019; Cassan, 2013; Cassan and Vandewalle, 2017); religious identity for Jews in medieval Europe (Botticini and Eckstein, 2012); ethnic identity in contemporary China (Jia and Persson, 2013); racial identity for Native Americans in contemporary United States (Antman et al., 2015); and racial identity in contemporary Brazil (Cornwell et al., 2014). Our findings also complement the recent work of Fouka et al. (2018), which finds that the Great Migration of African Americans increased the strength of identity for white Europeans in the Northeastern United States. The insight that some non-white Americans can choose their racial identity has been highlighted in a few recent studies that document changes in self-identified race in contemporary America (e.g. Austen-Smith and Fryer, 2005; Ruebeck et al., 2009).[10] By demonstrating that some individuals exercised discretion over their racial identity even in the pre-Civil Rights U.S. context, when race was much more rigidly defined, our results emphasize the importance of thinking about race as a choice variable in certain contexts.

Finally, we contribute to the economic history literature by quantifying the magnitude of the "passing" phenomenon. We are most closely related to an earlier unpublished working paper by two of the authors, Nix and Qian (2015), which used one-direction links that potentially suffered from finding false positive passing. The current study makes significant improvements in this respect with the 2SUP link for the reasons discussed earlier. In addition, we use link- and population-based weights to infer a plausible range for the rate of passing for the population. The new paper finds rates of passing that are less than half of those in the earlier paper, both for the linked sample and the population. We also provide a large body of new descriptive evidence on interracial marriages, children and the active choice of passing, which the earlier working paper did not consider.[11] Our work also complements the innovative working paper by Mill and Stein (2016), which focuses on sons who are classified as "mulatto" from families with both black and mulatto children. They find

[10] This topic has received much more attention from sociologists than economists. For example, see Burma (1946); Eckard (1947) for studies of the historical U.S. context, and Liebler et al. (2017); Saperstein (2012).

[11] Also note that Nix and Qian (2015) used the digitized Census data from FamilySearch whereas the current paper uses the digitized Census data from Ancestry. The main benefit of the latter is that we are now able to identify the race of spouses and the number of children for those who are household heads.

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. Our linking method builds on the pioneering works of Ferrie (1996) and Abramitzky et al. (2012).[12] Our findings highlight the potential usefulness of two-direction linking by demonstrating how it can reduce false positive results in some contexts.[13]

This paper is organized as follows. Section 2 discusses the historical background. Section 3 describes the data. Section 4 discusses 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 discusses potential extrapolations from the linked sample to the population. Section 7 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 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.[14]

The black population faced restrictions such as the complete segregation of whites and

[12]For some prominent examples of economic studies which link, see Ferrie (1996), Long and Ferrie (2013a), Abramitzky et al. (2012), Feigenbaum (2014), Abramitzky et al. (2019). The recent paper by Bailey et al. (2017) reviews this literature. Many of the sensitivity checks in this paper are motivated by Bailey et al. (2017). For some relevant examples from outside of economics, see Andoni and Razenshteyn (2015), Enamorado et al. (2018) and Fatemi et al. (2018). The latter papers generally use a probabilistic approach, building on the classic probabilistic framework of Fellegi and Sunter (1969), although they do differ in exactly how they implement the merge procedure.

[13]Most past studies that link use one-direction linking. For a recent example of two-sided linking, see Abramitzky et al. (2019), which demonstrates that the results of from the well-known paper by Abramitzky et al. (2012) are similar when using one- and two-sided linking. See Abramitzky et al. (2019) for a discussion of this and other machine linking methods.

[14]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).

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 (TuskegeeInstitute, 2010).

Blacks earned much less than whites.[15] 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.[16]

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

[15]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).

[16]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.

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.[17] 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.[18] 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).[19] To demonstrate the wide gradient of color for former slaves, emancipated "White and Colored Slaves" were chosen for a propaganda tour of the North in 1863.[20] 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 more educated, had 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 Eu-

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

[18] 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.

[19] 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).

[20] See "White and Colored Slaves" by C. C. Leigh (*Harper's Weekly*, January 30, 1864, p. 71).

ropean Americans have at least 1% African Ancestry (Bryc et al., 2015).[21] 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 et al. (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 pass 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 et al. (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.

Limited sample size or non-random sampling means that the current genetics evidence does not give the actual rate of passing for the United States as a whole. However, they serve as useful benchmarks (sanity checks) for interpreting the implied rates of passing for the population that we present later in the paper.

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;

[21] Note that genetic studies such as the one discussed face numerous caveats from difficulties such as non-random sampling of the population. See Bryc et al. (2015) for a detailed discussion.

> 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.[22]

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".[23] 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 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"

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

[23] Also, see Smith (2006) for a detailed discussion 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.

(Gross, 2009, loc. 1083, 1356).[24] 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..." (Lewis, ed, 2016).[25]

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 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

[24]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".

[25]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 (e.g., incarceration) (Penner and Saperstein, 2008) or attire (Freeman et al., 2011).

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" (Times, ed, 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.[26] 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).[27] Others passed when they were older simply because of the overwhelming discrimination they faced or to provide a better life for their children.[28]

Passing was not always permanent. Sometimes, individuals passed to obtain a job or attend school, and then later reverse pass to black.[29] 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).[30]

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 will investigate this with the data later in the paper.[31]

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

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

[28] See Sharfstein (2011).

[29] See Hobbs (2014).

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

[31] A large body of anecdotal evidence shows that those wishing to pass often completely disassociate 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 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.

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.[32] 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.[33]

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 that of black individuals in years when the two groups are distinct.

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.

See the NBER Working Paper version of this paper for summaries of case studies that illustrate the costs and benefits of passing for white.

[32]The 1850 and 1860 Censuses only reported names of free blacks. Since most of the black population was under slavery, 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.

[33]Father's and mother's names are also available for some years. We currently do not use these variables because of the high number of missing values.

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

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

[36]See Appendix A.1.

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. We will discuss and address this later in the paper.

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. See the next section.

4 2-Sided Unique Perfect (2SUP) Links

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.[38] 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

[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.

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

will be multiple potential links. We also use other "blocking" variables, such as year and place of birth, to restrict the sample.[39] This mitigates, but cannot fully resolve the fundamental problem of multiple potential matches. Thus, for brevity, we will ignore blocking variables for now. Similarly, since our linking algorithm is very similar to existing ones, we will focus the discussion on the key departures from standard linking methods and describe the details more fully in Appendix Section B.

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 et al., 2018; Feigenbaum, 2014; Long and Ferrie, 2013a). To maximize the accuracy of the links, past studies often require perfect spelling matches of first and last names. The first part of our linking algorithm does exactly this.[40] We will call it the Unique Perfect link.[41]

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 Figure 1a, 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

[39] We do not block on variables that can be affected by changing racial classification, such as migration or marriage. See the appendix for more details.

[40] 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.

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

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.[42] To address this, we additionally require the linked 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.[43]

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 Figure 1b. 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 Figure 1a, 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.

2SUP does not entirely remove the possibility of false positives. False positives occur if the following conditions are *all* true about the black individual i observed in base year

[42] 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.

[43] Abramitzky et al. (2019) uses two-directional links to test the robustness of their well-known study, Abramitzky et al. (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).

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; 3) when linking backwards, the linked white individual in year $t+10$ links only to the same 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).[44]

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. For additional evidence that the estimated rates of passing are not driven by false links, we will also report results from many sensitivity checks after presenting the main results.

The main disadvantage of the 2SUP link is that it can potentially generate more Type II errors (missed true links) because it links fewer individuals than other machine-linking algorithms. This increases the challenge for extrapolating statistics from the linked sample to the population, a difficulty that is ubiquitous to all linking studies where the linked sample is not obviously a random sample of the population. We discuss this difficulty and how we address it in more detail in Section 6.

Also note that the linked sample will naturally omit any individual who changed their names when they passed for white. We discuss this more when we examine individuals with distinctively black names.

Tor 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.

[44] Abramitzky et al. (2019) finds that there are approximately 10-15% of names that suffer from transcription error in the historical censuses.

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 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.

Panel C shows that 2SUP links 8.6% of the population of black men under age 55 in base years (1880-1930). Thus, 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% ($0.086 \times .166$). We discuss alternative methods of extrapolating to the population later in the paper, after we present the results with the linked sample.

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.

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).[45] For comparison purposes, the results in this section for other racial categories will also focus on these two linked intervals. We restrict our attention to racial categories with total population sizes of at least fifty thousand individuals in 1920 and 1930.[46]

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. The negligible rates of racial classification change to Native American and Asians could reflect the fact that these groups also faced discrimination, such that black individuals had much less incentives to become Asian, or the possibility that not many black men had appearances which would allow them to pass for these groups.

In row B, we repeat the exercise for white males (i.e., identified as white and under age 55 in the base year). We use 2% of the white male population from each base year for computational feasibility.[47] Table 2 row B shows that only 0.7% of the 2SUP links pass from white to black, which is consistent with the fact that white individuals had little political, social or economic incentives to pass for 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. 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

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

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

[47] We use a smaller sample than for linking black males because the white population is much larger. Note that as with all of the linking exercises, the restriction only applies to the base year sample: the forward link examines the full population of all males in the subsequent census, and the backward link examines the full population of all males in the initial census.

names than black men.

Rows C and D present the rates of racial classification change for Chinese and Japanese men. The links use 100% of the Chinese and Japanese male population in each base year, 1920 and 1930, because of their small size. We find that within the 2SUP sample, 90.5% of Chinese remain Chinese, 8.2% become white, 0.5% become black, and 0.4% become Japanese. Similarly, 91.6% of Japanese remain Japanese, 7% become white, 0.2% become black, 0.4% become Native American ("Indians" in the historical censuses), and 0.2% become Chinese. These patterns are important for two reasons. First, they are consistent with the fact that Asians had little political, social or economic incentives to pass for black, and arguably less incentive to pass for white than black individuals.[48] Second, the negligible rates of passing from Japanese to Chinese, and from Chinese to Japanese go against the concerns of enumerator error. We discuss this more in the next section.

Note that 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.

Row E examines the rates of passing for men classified as Native Americans. As with Chinese and Japanese populations, we use the 100% sample of males classified as Native Americans in each base year because of their small population size. We find that 69% remained Native Americans, 27.8% passed for white, 2.2% became black, and 0.1% became Chinese or Japanese. The high rate of passing to white is consistent with historical evidence that there was much racial mixing prior to our study period and that the fluid definitions used for Native Americans may have made it easier to pass for white.[49]

[48] As with interpreting the black-to-asian passing, differences in the degree of racial mixing in previous generations may also contribute to the rates of racial classification change experienced by Asians.

[49] Sandefur and McKinnell (1986) discuss the history of fluidity in Native American racial identity and the high degree of interracial mixing between Native Americans and whites in the historical United States. They contrast this fluidity with the rigid definitions and rules associated with black versus white individuals. For example, the definition for Native American is rather ambivalent and quite different from the definition for "black". In some censuses, enumerators were explicitly instructed that individuals with mixed white and Indian ancestry are "white" if their community accepts them as white. See Appendix Section A.1 for enumerator instructions. Another example of the ambivalence during this period in distinguishing between Native American and white is the 1924 Racial Integrity Act of Virginia. The act attempted to distinguish white individuals from all non-white individuals, but was notable for including what came to be known as the "Pocahontas Exception" allowing individuals with less than 1/16 Native American Ancestry to be categorized as white, despite the prevailing

5.1.3 Passing Across East Asian Categories

If the observed rates of passing are driven by enumerator error of the type discussed earlier, 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.[50] 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 the accuracy of the 2SUP links and our interpretation that passing was motivated by socio-economic incentives instead of enumerator error.

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 he would need to pass for black. The second is if he is a black man who passed for white and then chooses to "reverse pass" to be black again. The historical evidence provides many examples for the latter. To investigate the percentage of individuals who pass, but then revert to being black in the following census year, we use 2SUP to 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 2SUP 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

standard of one-drop rules for all other racial categories (Zimmer, 2014).

[50] See the Appendix for enumerator instructions for classifying Asians.

in Panel A. Panel D column (2) shows that 30% of those who passed to become white later reverse passed to black.[51]

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 Sensitivity Analysis

Table 3 presents the results from several sensitivity checks motivated by the literature on linking. Because of the computational time required for these, we use three states with a large number of black males: Alabama, Georgia and Louisiana. First, we implement the 2SUP links as in the main analysis for these three states and find that we are able to link 7.9% of black males in these states with the baseline algorithm (column 1). For the linked sample, 14.9% 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. 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 3 rows D and E show that the percentage of the population linked is slightly higher (8.1% vs 7.2%) and the rate of passing in the linked sample is slightly lower (14.6% vs 16.1%) for individuals whose

[51] 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 $t+20$).

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.[52] Table 3 rows F and G show that the rates of passing are similar or higher (15% and 22.4%) with these alternative ways of blocking on age.

Next, we require the links to match on the parents' birth states when these data are available. This is similar to the spirit of using parents' birth states to validate links that do not block on these variables. The estimated rate of passing in Row H is similar to the baseline.

In Row I, we follow the method from Abramitzky et al. (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 higher link rate of 10.3%. Reassuringly, the rate of passing in this sample, 14.7%, is comparable to the baseline, 14.9%.

Finally, we replicate the robustness linking exercise from Long and Ferrie (2013b).[53] Since this algorithm does not link backwards but is otherwise quite similar to our linking algorithm, we find that the results are similar to our forward linking exercise, producing a rate of passing of 24.9% (see Row J).

[52]Note that the latter is similar in spirit to the EM linking algorithm from Abramitzky et al. (2018) because the links formed with the plus or minus one year of the predicted age restriction would be a subset of the links formed by the EM algorithm. Unlike the EM algorithm, we only allow links to be formed if there is a perfect spelling match.

[53]The exercise is described on page 8 of the Online Appendix of Long and Ferrie (2013b) and can be summarized as follows. First, we restrict links to individuals with perfect forward linked first and last names. Second, we require that birth states match. Third, we require that parental birth states match, when available (parental birth states are available for the 1880, 1900, 1910, 1920, and 1930 Censuses). Fourth, we require that age be within plus or minus one year of the predicted age. Fifth, we require that all links are unique. We choose this method (amongst other methods in their paper) to replicate because it is the most conservative and meant to minimize false positives.

Falsification: Literacy Rates

We can also investigate the presence of error in the linking algorithm by conducting a falsification test with the following logic. 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 than those who do not pass, we would be concerned that our findings are confounded.[54]

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, a much higher share of the sample become literate over time: 32.9% for those who pass and 31.4% for those who do not pass. These results provide little evidence for the concern that mistaken links cause false positive passing in the linked sample.

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 African and European descent – comprised a separate category from black. The instructions to enumerators were vague and historians have been dubious about the usefulness of this categorization because of the large proportion of racial mixing in the United States.[55] This concern, along with the need for consistency across years, is why we do not separate the categories in the main results. In this section, we follow earlier studies such as Mill and Stein (2016), which focus on mulatto

[54] In principle, this exercise can be repeated with any variable for which we can plausibly predict the direction of change over time and which we do not use as a blocking variable. The data limits our attention to literacy.

[55] See the Appendix Section 4.

individuals and repeat our investigation for a sample of individuals who are classified as mulatto in the base year.

Table 4 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.[56] We find that in 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.4 row C), where we do not distinguish 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.4 row D).[57]

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 lighter skin and thus could more easily pass for white. At the same time, the

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

[57]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 time frame (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).

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.[58] In other words, the advantage for those categorized as mulatto for passing for white was limited.

5.3.2 Distinctively Black and White Names

Motivated by Cook et al. (2014), we investigate whether the rates of passing are lower for individuals with distinctively black names and who choose to keep the same name in the next census, for whom passing is presumably more difficult. Similarly, we investigate whether the rates of passing for individuals with distinctively white names is higher. We take historically black names from Cook et al. (2014).[59] Table 4 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. However, note that 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 in our sample.

To identify distinctively white names, we adapt Cook et al.'s (2013) method for white males.[60] 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.[61]

[58] "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). 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).

[59] 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 et al. (2014) for more details.

[60] 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.

[61] Note that we are able to link a higher percentage of individuals with distinctively black names, 13.6% (not

5.3.3 Marriage and Passing

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 context.[62] 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.[63] The sample size is slightly reduced relative to the main 2SUP sample since we lose observations with missing values in those two variables.

Table 5 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

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.

[62] For a recent study of the effects of miscegenation laws on interracial births, see Briseno (2013).

[63] 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.

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 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 and Passing

Table 6 investigates the relationship between passing and the number and race of children in the household. It is analogous to Table 5. 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 6 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, 64.9% 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.

Please keep two caveats in mind when interpreting the results in this section. First, as before, we do not distinguish between a child with no African ancestry who is 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. Second, we cannot observe older children who have moved out of the household.

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 from a relatively "blacker" neighborhood to a relatively "whiter" neighborhood.[64] Second, our study coincides with a period of tremendously high internal migration (e.g., Collins and Wanamaker, 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.

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.[65] Table 7 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.[66]

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

[64] See studies such as Logan and Parman (2017), Ananat (2011), Boustan (2010) and Cook et al. (2018) for discussions of segregation.

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

[66] 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 1850-1860 for white males found by Ferrie (2003).

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.[67] Consistent with the Great Migration being a period where 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 $t + 10$.[68] 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.[69]

[67] See Carrington et al. (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.

[68] 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 et al. (2008). Thus, we choose to use counties as the level of comparison.

[69] 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 2a 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.

Figure 2b 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 7 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.[70] 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

[70]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 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 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.[71] 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 8 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 8 control for age category dummy variables, base year and

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

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 8 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. Column (2) uses the Democratic vote share for a given county as a proxy for discrimination in that county.[72] To summarize the meaningful variation, we compute the first principal component for all 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).[73] 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.[74] All of these schooling variables are such that a higher value

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

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

[74] 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

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 and Wanamaker (2017) and control for state fixed effects.[75]

Table 8 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. Dou-

variable that we construct ourselves, the number of black universities in a given state and year.

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

bling 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.

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.[76]

Well-being After Passing for White Table 9 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 earns 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.[77] 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).[78]

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

[76] 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.

[77] 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.

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

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 not be interpreted as a descriptive decomposition exercise. The uninteracted dummy variable for 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 9 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.[79]

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

[79] See the Background section.

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.

6 Extrapolating to the Population

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.[80]

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. In this case, one would conclude that 16.6% of the population of black males under age 55 in the base years 1880-1930 passed for white. See Table 10 column (1). Alternatively, one can take the stance that such differences are meaningful despite the small magnitudes (e.g., they may be correlated with unobserved characteristics which drive the decision to pass for white) and attempt to take them into account in the extrapolation. To be conservative and thorough, this section pursues the second line of inquiry. We employ two extrapolation strategies. Taken together, the results provide useful clues on the range of passing in the black population.

[80] See Appendix Table A.2.

6.1 Weighted Extrapolation

In light of the concerns discussed above, we use population weights to produce less extreme rates of passing for the population. The weights take into account two notions: the share of individuals in the population with the same set of characteristics as a given individual and the share in the linked sample with the same set of characteristics as a given individual. We implement population weights by aggregating the individual observations in the linked sample into cells according to observable characteristics, and then multiply the average rate of passing in each cell by the population weight.[81] 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.[82] We use all permutations of these variables to group individuals into mutually exclusive and collectively exhaustive cells. Table 10 column (2) shows that the implied rate of passing for the population with population weights is 16%.

Alternatively, we can use the weights developed by Ager et al. (2019), which re-weights linked individual observations with the probability that individuals with similar characteristics are linked.[83] Conceptually, the two weights are similar in correcting for the concern that the type of individuals with low link rates are under-represented in the rate of passing estimated in the 2SUP sample. Both weighting schemes assume that individuals with simi-

[81] 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.

[82] 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.

[83] Following Ager et al. (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.

lar observable characteristics will pass for white at the same rate. Table 10 column (3) show that the population rate of passing with such weights is 16.8%.

The extrapolated rates of passing for the population are very similar between the two different types of weights, 16% versus 16.8%, and with the rate of passing in the 2SUP sample, 16.6%. The main caveat for the weighted extrapolations is they omit individuals in characteristic cells where no one is linked from the calculation. Such individuals account for 38% of the population. To address this, we add these individuals back into the weighted extrapolations. This adjustment will mechanically lower the population rate of passing since it adds a large number of individuals assumed to not pass. Table 10 column (4) shows that when we assume that the population weighted rate of passing of 16% only applies to 62% of the population and that 38% of the population do not pass, the implied rate of passing for the population is 9.9% ($0.16 \times 0.62 = 0.099$). We can conduct a similar exercise with Ager et al. (2019) weights.[84] We find that the weighted pass rate decreases to 6.8%.

The estimates in this section suggest that the intercensal rate of passing for black males under age 55 in the base year are likely to be around 6.8% to 9.9%. Recall that the rate of passing suggested by the genetic evidence discussed in Section 2.3 is 20% for a non-random sample of the U.S. population, 57.5% for a non-random sample of the U.S. Northeastern population, and 49.6% of a random sample of the Boston population. To compare our calculated flow of passing to the estimates of the stock of passing from the genetics evidence, we approximate the implied stock of passing from the intercensal flows of 6.8% to 9.9%. Assuming constant rates of fertility and mortality over time, a reverse-pass rate of 30% and a life expectancy of fifty years, our estimates imply that 23.5% to 38.3% of the 1940 black population had passed for white. These crude approximations are comparable to the modern genetics evidence.

The results in this section demonstrate that *prima facie* reasonable assumptions can generate a rate of passing for the population that roughly corresponds to the rate suggested by the modern genetic evidence.

[84] To compute the Probit, we assume that in cells where no one is linked, one individual is linked and he does not pass. Then, we calculate the weights and apply the weights to the rates of passing in the adjusted sample.

6.2 Race Transition Matrix

Table 11 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.[85]

7 Conclusion

The extent of passing from "black" to "white" in pre-Civil rights United States has been a subject of heated debates amongst scholars and the public for the past one hundred years. As we discussed in the Introduction, the difficulties in the historical data have caused the estimates and speculations to vary by orders of magnitude, from 0.26% to 33% or higher. We use relatively new linking methods, which rely less on the problematic historical data than traditional population accounting methods. In addition, we use two-direction linking to minimize the possibility of mistaken links that can produce false positive passing.

Our results show that a significant number of black males passed for white during 1880-1940: 16.6% in the linked sample, and around 6.8% to 9.9% of the black male population were likely to have passed for white. Moreover, the patterns of passing are consistent with the historical evidence. Passing for white was positively associated with higher income. In order to pass for white, individuals needed to relocate to white 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.

We recognize that the magnitude of the change in racial identification is specific to the context of our study. The number of individuals who pass will critically depend on the genetic make up of the population, the racial definitions and the incentives of the context. Nevertheless, the finding that identity can be a choice for many individuals, even along a

[85] The transition matrix using the adjusted Ager (2019) weights are very similar and available upon request.

dimension as rigidly defined as race in the United States under Jim Crow, is a generalizable and potentially powerful insight.

The results of this paper raise many questions for future research. Our findings 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 by the current economics literature. Understanding the extent and magnitude of endogenous racial classification is important for many inquiries. For example, Duncan and Trejo (2011b) and Duncan and Trejo (2011a) point out that understanding selective ethnic attrition is critical for studying immigrant assimilation.[86] Thus, an obvious next step is to better understand the determinants and consequences of racial classification change, as well as to examine the degree of selective ethnic/racial choice in different contexts.

This is important for the historical context as well as the modern U.S. context. Sociologists such as Harris and Sim (2002) have pointed to this, stating that "We know that one's racial classification can vary across contexts and observers, but we know little about the magnitude and patterns of racial fluidity in the United States, and even less about the circumstances that facilitate these shifts". Several recent studies have used modern U.S. data to make progress on this agenda. For example, Antman and Duncan (2015) document that self-identification of race in the United States is associated with incentives from affirmative action. Fryer and Torelli (2010) use the NLSY to document that high achieving black high school students are more likely to "act white". Saperstein and Penner (2012), also using the NLSY, documents that there are significant numbers of changes in racial classification and argue that it is related to socio-economic success, "white people appear to be more successful in part because successful people become white, through either self-identification, external classification, or both". It is interesting to note that race is mostly self-identified in

[86]These two studies use the 2000 U.S. Census to document, for example, that wealthier men with Spanish surnames are less likely to identify as Mexican; and use CPS data to document, for example, that high school dropout rates of third-generation Mexican youth are higher for those who self-identify as Mexican.

the modern U.S. context and the incentives to be part of a given race may differ from the historical context. These changes provide valuable opportunities for delving deeper into the interplay between racial identity and external forces. For example, it would be very interesting to compare racial classification change when race became self reported in the Census after 1960 to earlier periods when the more recent data become available for linking.[87]

It would also be interesting to look beyond the U.S. context to other settings where ethnic and racial variables are strongly influential for important outcomes such as growth, institutions, public goods, and conflict.[88] For example, Montgomery (2011) argues that the growth in the white population in Puerto Rico is explained by changes in racial classification. In many contexts, the data document a strong positive association between ethnic fractionalization and measures of economic and political performance such as civil conflict (Alesina and Ferrara, 2004). The interpretation of this relationship could vary dramatically depending on whether ethnicity is an exogenous variable that is fixed over a person's lifetime or if it is a choice variable that can change at the discretion of the individual.

Finally, it is interesting to point out that our results highlight a break in cultural transmission (i.e., a parent who passed for white is likely to avoid transmitting information to his children that would reveal their African ancestry). The mechanisms for cultural change is an interesting topic for contemplation and research given that the existing literatures from political economy and cultural economics have focused on the opposite force of intergenerational diffusion and the persistence of cultural norms and beliefs (e.g., Algan and Cahuc, 2010; Bisin and Verdier, 2000a; Fernandez and Fogli, 2009).[89] That racial identity change is a time or perhaps cause of a break in cultural transmission is worthy of additional rumination.

[87] Perez and Hirschman (2009) find little changes in racial composition in the Censuses after 1960. They do not link.

[88] For some examples of this vast literature, see Easterly and Levine (1997), Bates (2000), Miguel and Gugerty (2005) and Caselli and Coleman (2013). The studies we discussed in the Introduction have already made some progress on this in the contexts of China, India, and medieval Europe.

[89] Fernandez (2010) overviews the literature. Also, for example, see the well-known theoretical studies of Bisin and Verdier (2000a), Bisin and Verdier (2000b), Bisin and Verdier (2001), and empirical evidence from Algan and Cahuc (2010), Cipriani et al. (2007), Fernandez et al. (2004), Fernandez and Fogli (2009), and Abramitzky et al. (2014).

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, August 2012, *102* (5), 1832–56.

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

_ , _ , _ , James J Feigenbaum, and Santiago Pérez, "Automated linking of historical data," 2019.

_ , Roy Mill, and Santiago Pérez, "Linking Individuals Across Historical Sources: A Fully Automated Approach," 2018.

Ager, Philipp, Leah Boustan, and Katherine Eriksson, "The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War," 2019.

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," *The Quarterly Journal of Economics*, August 2000, *115* (3), 715–753.

Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," February 2004, (10313).

Algan, Yann and Pierre Cahuc, "Inherited Trust and Growth," *American Economic Review*, December 2010, *100* (5), 2060–92.

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

Andoni, Alexandr and Ilya Razenshteyn, "Optimal Data-Dependent Hashing for Approximate Near Neighbors," in "Proceedings of the forty-seventh annual ACM symposium on Theory of computing" ACM 2015, pp. 793–801.

Antman, Francisca and Brian Duncan, "Incentives to Identify: Racial Identity in the Age of Affirmative Action," *Review of Economics and Statistics*, 2015, *97* (3), 710–713.

_ , _ , and Barham, "American Indian Casinos and the Rise in Native American Self-Identification," Technical Report, Citeseer 2015.

Atkin, David, Eve Colson-Sihra, and Moses Shayo, "How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption," 2019.

Austen-Smith, David and Roland G Fryer, "An Economic Analysis of Acting White," *The Quarterly Journal of Economics*, 2005, *120* (2), 551–583.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth," 2017.

Bates, Robert H., "Ethnicity and Development in Africa: A Reappraisal," *American Economic Review*, May 2000, *90* (2), 131–134.

Bénabou, Roland and Jean Tirole, "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Bisin, Alberto and Thierry Verdier, "A Model of Cultural Transmission, Voting and Political Ideology," *European Journal of Political Economy*, March 2000, *16* (1), 5–29.

_ **and** _ , "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *The Quarterly Journal of Economics*, August 2000, *115* (3), 955–988.

_ **and** _ , "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, April 2001, *97* (2), 298–319.

_ **, Eleonora Patacchini, Thierry Verdier, and Yves Zenou**, "Bend it like Beckham: Ethnic Identity and Integration," *European Economic Review*, 2016, *90*, 146–164.

Bodenhorn, Howard, "The Mulatto Advantage: The Biological Consequences of Complexion in Rural Antebellum Virginia," *Journal of Interdisciplinary History*, 2014/10/04 2002, *33* (1), 21–46.

Botticini, M. and Z. Eckstein, *The Chosen Few: How Education Shaped Jewish History* Princeton Economic History of the Western World, Princeton University Press, 2012.

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

_ , "Was Postwar Suburbanization White Flight? Evidence from the Black Migration," *The Quarterly Journal of Economics*, 2010, *125* (1), 417–443.

Briseno, Alonso Gortari, "On the Persistent Effects of Miscegenation," 2013.

Bryc, Katarzyna, Eric Y Durand, J Michael Macpherson, David Reich, and Joanna L Mountain, "The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States," *The American Journal of Human Genetics*, 2015, *96* (1), 37–53.

Burma, John H., "The Measurement of Negro "Passing"," *American Journal of Sociology*, 1946, *52* (1), 18–22.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 2008, *123* (1), 177–218.

_ **and Alan B Krueger**, "School Quality and Black-White Relative Earnings: A Direct Assessment," *The Quarterly Journal of Economics*, February 1992, *107* (1), 151–200.

Carrington, William J, Enrica Detragiache, and Tara Vishwanath, "Migration with Endogenous Moving Costs," *American Economic Review*, 1996, *86* (4), 909–30.

Carruthers, Celeste K. and Marianne H. Wanamaker, "Closing the Gap? The Effect of Private Philanthropy on the Provision of African-American Schooling in the U.S. South," *Journal of Public Economics*, 2013, *101* (C), 53–67.

Carruthers, Celeste K and Marianne H Wanamaker, "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap," *Journal of Labor Economics*, 2017, *35* (3), 655–696.

Caselli, Francesco and Wilbur J. Coleman, "On the Theory of Ethnic Conflict," *Journal of the European Economic Association*, 01 2013, *11*, 161–192.

Cassan, Guilhem, "Identity Based Policies and Identity Manipulation: Evidence from Colonial Punjab," October 2013, (1306).

_ **and Lore Vandewalle**, "Identities and Public Policies: Unintended Effects of Political Reservations for Women in India," 2017.

Cipriani, Marco, Paola Giuliano, and Olivier Jeanne, "Like Mother Like Son? Experimental Evidence on the Transmission of Values from Parents to Children," IZA Discussion Papers 2768, Institute for the Study of Labor (IZA) April 2007.

Collins, William J and Marianne H Wanamaker, "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants," *The Journal of Economic History*, 2015, *75* (4), 947–992.

_ **and Robert A Margo**, "Race and Home Ownership from the End of the Civil War to the Present," *American Economic Review*, 2011, *101* (3), 355–59.

Cook, Lisa D, Trevon D Logan, and John M Parman, "Distinctively Black Names in the American Past," *Explorations in Economic History*, 2014, *53*, 64–82.

_ **,** _ **, and** _ , "Racial Segregation and Southern Lynching," *Social Science History*, 2018, *42* (4), 635–675.

Cornwell, Christopher, Jason M. Rivera, and Ian M. Schmutte, "Wage Discrimination when Identity is Subjective: Evidence from Changes in Employer-Reported Race," Working Paper, University of Georgia 2014.

Davis, F James, *Who Is Black?: One Nation's Definition*, Penn State Press, 2010.

Dawkins, M.A., *Clearly Invisible: Racial Passing and the Color of Cultural Identity*, Baylor University Press, 2012.

Duncan, Brian and Stephen J Trejo, "Intermarriage and the Intergenerational Transmission of Ethnic Identity and Human Capital for Mexican Americans," *Journal of Labor Economics*, 2011, *29* (2), 195–227.

_ **and** _ , "Tracking Intergenerational Progress for Immigrant Groups: The Problem of Ethnic Attrition," *American Economic Review*, 2011, *101* (3), 603–08.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *The Quarterly Journal of Economics*, November 1997, *112* (4), 1203–50.

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

Elo, Irma T and Samuel H Preston, "Estimating African-American mortality from inaccurate data," *Demography*, 1994, *31* (3), 427–458.

_ **and** _ , "Racial and Ethnic Differences in Mortality at Older Ages," *Racial and Ethnic Differences in the Health of Older Americans*, 1997, pp. 10–42.

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai, "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records," 2018.

Eriksson, Katherine, Gregory T Niemesh, and Melissa Thomasson, "Revising Infant Mortality Rates for the Early Twentieth Century United States," *Demography*, 2018, pp. 1–24.

Fatemi, Bahare, Seyed Mehran Kazemi, and David Poole, "Record Linkage to Match Customer Names: A Probabilistic Approach," *arXiv preprint arXiv:1806.10928*, 2018.

Feigenbaum, James, "Automated Census Record Linking," Working Paper, Harvard University 2014.

Fellegi, Ivan P and Alan B Sunter, "A Theory for Record Linkage," *Journal of the American Statistical Association*, 1969, *64* (328), 1183–1210.

Fernandez, Raquel, "Does Culture Matter?," NBER Working Papers 16277, National Bureau of Economic Research August 2010.

_ **, Alessandra Fogli, and Claudia Olivetti**, "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *The Quarterly Journal of Economics*, November 2004, *119* (4), 1249–1299.

_ **and** _ , "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, January 2009, *1* (1), 146–77.

Ferrie, Joseph, *Internal Migration* 2003.

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

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini, "From Immigrants to Americans: Race and Assimilation during the Great Migration," *Harvard Business School BGIE Unit Working Paper*, 2018, (19-018).

Freeman, Jonathan B., Andrew M. Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady, "Looking the Part: Social Status Cues Shape Race Perception," *PLOS ONE*, 09 2011, *6* (9), 1–10.

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

Fryer, Roland G. and Paul Torelli, "An Empirical Analysis of 'Acting White'," *Journal of Public Economics*, 2010, *94* (5-6), 380–396.

Gordon, Linda, *The Great Arizona Orphan Abduction*, Harvard University Press, 1999.

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

Hammer, Michael F, Veronica F Chamberlain, Veronica F Kearney, Daryn Stover, Gina Zhang, Tatiana Karafet, Bruce Walsh, and Alan J Redd, "Population Structure of Y Chromosome SNP Haplogroups in the United States and Forensic Implications for Constructing Y Chromosome STR Databases," *Forensic Science International*, 2006, *164* (1), 45–55.

Harris, David R and Jeremiah Joseph Sim, "Who Is Multiracial? Assessing the Complexity of Lived Race," *American Sociological Review*, 2002, pp. 614–627.

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

Hobbs, Allyson, *A Chosen Exile: A History of Racial Passing in American Life*, Harvard University Press, 2014.

Hochschild, Jennifer L and Brenna Marea Powell, "Racial Reorganization and the United States Census 1850–1930," *Studies in American Political Development*, 2008, *22* (01), 59–96.

Jia, Ruixue and Torsten Persson, "Ethnicity in Children and Mixed Marriages: Theory and Evidence from China," UC San Diego Working Paper, University of California at San Diego 2013.

Keyssar, Alexander, *The Right to Vote: The Contested History of Democracy in the United States*, Basic Books, 2000.

Khanna, Nikki and Cathryn Johnson, "Passing As Black: Racial Identity Work Among Biracial Americans," *Social Psychology Quarterly*, 2010, *73* (4), 380–397.

Lait, AJ and Brian Randell, "An Assessment of Name Matching Algorithms," *Technical Report Series-University of Newcastle Upon Tyne Computing Science*, 1996.

Lewis, Carol, ed., "Color Lines: Racial Passing in America," 2016.

Liebler, Carolyn A, Sonya R Porter, Leticia E Fernandez, James M Noon, and Sharon R Ennis, "America's Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census," *Demography*, 2017, (1), 259 0 284.

Logan, Trevon D and John M Parman, "The National Rise in Residential Segregation," *The Journal of Economic History*, 2017, *77* (1), 127–170.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *The American Economic Review*, 2013, *103* (4), 1109–1137.

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

Margo, Robert A, *Race and Schooling in the South, 1880-1950: An Economic History*, University of Chicago Press, 1990.

Meigs, James B., Richard W. Grant, Rebecca Piccolo, Lenny López, Jose C. Florez, Bianca Porneala, Lisa Marceau, and John B. McKinlay, "Association of African Genetic Ancestry with Fasting Glucose and HbA1c Levels in Non-Diabetic Individuals: The Boston Area Community Health (BACH) Prediabetes Study," *Diabetologia*, Sep 2014, *57* (9), 1850–1858.

Miguel, Edward and Mary Kay Gugerty, "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, December 2005, *89* (11-12), 2325–2368.

Mill, Roy and Luke Stein, "Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America," Technical Report, Arizona State University Working Paper 2016.

Modalsli, Jorgen, "Multigenerational Persistence: Evidence from 146 Years of Administrative Data," 2016.

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

Montgomery, James D, "The Population Dynamics of Black-White-Mulatto Racial Systems," *American Journal of Sociology*, 2011, *117* (1), 46–89.

Nix, Emily and Nancy Qian, "The Fluidity of Race: "Passing" in the United States, 1880-1940," Working Paper 20828, National Bureau of Economic Research January 2015.

Packard, J.M., *American Nightmare: The History of Jim Crow*, St. Martin's Press, 2003.

Penner, Andrew M. and Aliya Saperstein, "How Social Status Shapes Race," *Proceedings of the National Academy of Sciences*, 2008, *105* (50), 19628–19630.

Perez, Anthony Daniel and Charles Hirschman, "The Changing Racial and Ethnic Composition of the US Population: Emerging American Identities," *Population and Development Review*, 2009, *35* (1), 1–51.

Preston, Samuel H, Irma T Elo, Ira Rosenwaike, and Mark Hill, "African-American mortality at older ages: Results of a matching study," *Demography*, 1996, *33* (2), 193–209.

Rockquemore, Kerry Ann and David L Brunsma, *Beyond Black: Biracial Identity in America*, Rowman & Littlefield Publishers, 2007.

Ruebeck, Christopher S., Susan L. Averett, and Howard N. Bodenhorn, "Acting White or Acting Black: Mixed-Race Adolescents' Identity and Behavior," *The B.E. Journal of Economic Analysis and Policy*, March 2009, *9* (1), 1–44.

Sandefur, Gary D and Trudy McKinnell, "American Indian Intermarriage," *Social Science Research*, 1986, *15* (4), 347–371.

Saperstein, Aliya, "Capturing Complexity in the United States: Which Aspects of Race Matter and When?," *Ethnic and Racial Studies*, 2012, *35* (8), 1484–1502.

_ **and Andrew M Penner**, "Racial Fluidity and Inequality in the United States," *American Journal of Sociology*, 2012, *118* (3), 676–727.

Shapiro, S., "Development of Birth Registration and Birth Statistics in the United States," *Population Studies*, 1950, *4* (1), 86–111.

Sharfstein, Daniel, *The Invisible Line: A Secret History of Race in America*, Penguin Group United States, 2011.

Sloan, Nate, "Stanford Historian Re-Examines Practice of Racial 'Passing'," *Stanford Report*, December 2013.

Smedley, Audrey, *Race in North America: Origin and Evolution of a Worldview*, Westview, 1993.

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

Snae, Chakkrit, "A Comparison and Analysis of Name Matching Algorithms.," *Enformatika*, 2007, *19.*

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

Sollors, Werner, Jack Salzman, David Lionel Smith, and Cornel West, "Passing," in "The Encyclopedia of African-American Culture and History," Macmillan, 1995.

Strmic-Pawl, Hephzibah V., Brandon A. Jackson, and Steve Garner, "Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality," *Sociology of Race and Ethnicity*, 2018, *4* (1), 1–13.

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

Times, Chicago Sun, ed., "Five Black Chicagoans Who Passed for White," 2016.

TuskegeeInstitute, "Lynchings: By State and Race, 1882-1968," 2010.

Vile, John R., *Encyclopedia of Constitutional Amendments, Proposed Amendments, and Amending Issues, 1789-2002*, ABC-CLIO, 2003.

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

Woodward, C. Van, *The Strange Career of Jim Crow*, Oxford University Press, 2002.

Zimmer, Carl, "White? Black? A Murky Distinction Grows Still Murkier," *New York Times*, 2014.

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: Sensitivity Checks

		% Link	% Pass for White
		(1)	(2)
		2SUP	
A.	Baseline	7.9%	14.9%
B.	Include if age > 55 in year t	7.5%	15.0%
C.	100% sample	7.9%	14.7%
D.	Heaped Age (age is a multiple of 5 in base year)	7.2%	16.1%
E.	Omit Heaped Age	8.1%	14.6%
F.	+/- 5 year intervals	7.4%	15.0%
G.	+/- 1 year intervals	5.1%	22.4%
H.	Match on mothers' and fathers' birth states	6.2%	14.7%
I.	Unique names (by birth state, +/-5 age interval)	10.3%	14.7%
		One-Direction Link	
J.	Ferrie and Long (2013)*	12.3%	24.9%

Notes: 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 using 2SUP. Exceptions are stated in the row titles. The sample is restricted to individuals that reside in Alabama, Georgia and Louisiana in the base year. Parental birth states are available for 1880 and 1900-1930. *See text for description.

Table 4: Mulattos and Individuals with Distinctively Black or White Names

	(1) % Black	(2) % Mulatto	(3) % 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 5: Marital Status, the Race of the Spouse and Passing

			$Black_t$		$White_t$	
	t	t+10	Obs. (1)	Pass Rate (2)	Obs. (3)	Pass Rate (4)
A.	Single	Single	73814	15.1%	267111	0.7%
B.		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 martial 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 6: The Racial Composition of Children and Passing

			$Black_t$		$White_t$	
	t	t+10	Obs. (1)	Pass Rate (2)	Obs. (3)	Pass Rate (4)
A.	No kids	No Kids	9532	10.1%	26780	0.5%
B.		>=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 7: The Geography of Passing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel I.							
	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 8: The Correlation between Passing and Base Year Characteristics

	Dependent Variable: White in year t+10					
	(1)	(2)	(3)	(4)	(5)	(6)
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 in 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 9: The Correlation between Passing and Income

	Dependent Variables			
	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

Table 10: The Implied Rates of Passing for White in the Black Male Population ($\leqslant$55 Years of Age)

Implied Rates of Passing for the Black Male Population				
			Assume that individuals in cells with zero links do not pass	
Assume 2SUP linked sample is representative	Population weight	Ager et al. (2019) weights	Population weight	Ager et al. (2019) weights
(1)	(2)	(3)	(4)	(5)
16.6%	16.04%	16.8%	9.9%	6.8%

Notes: Column (1) presents the rates of passing for the 10% sample of males identified as black and under age 55 in the base year in the 1880-1930 censuses linked using 2SUP. Column (2) extrapolates using population weights. Column (3) extrapolates using weights developed by Ager et al. (2019). See text for a discussion of columns (4)-(5).

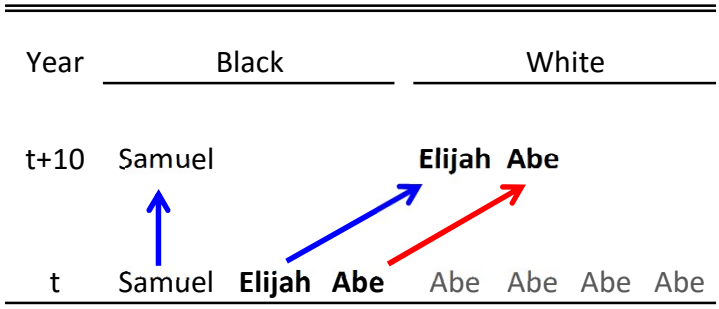
Table 11: Race Transition Matrix – Implied Rates of Passing for the Population, All Races, Population Weights

	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 1: One-sided and 2-sided Unique (Perfect) Links

(a) Unique Perfect Link



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

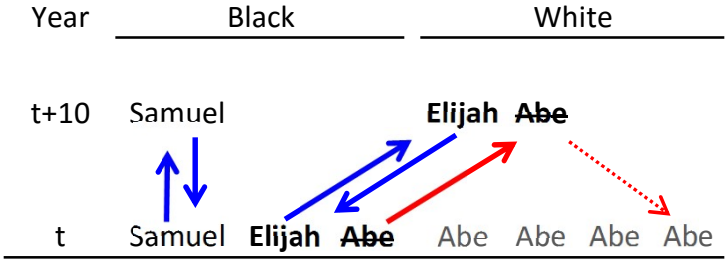
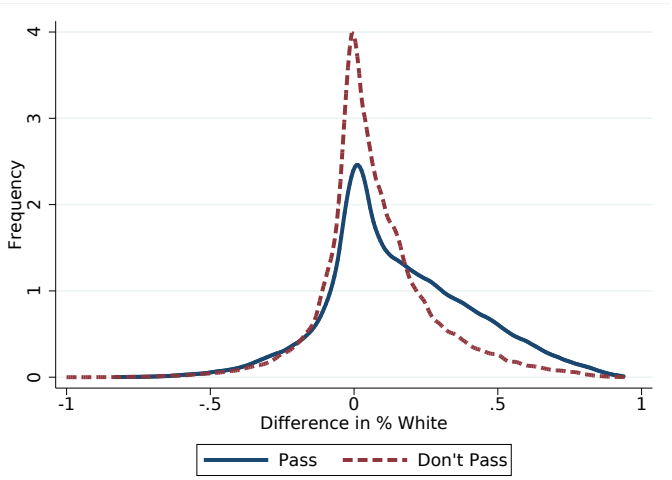
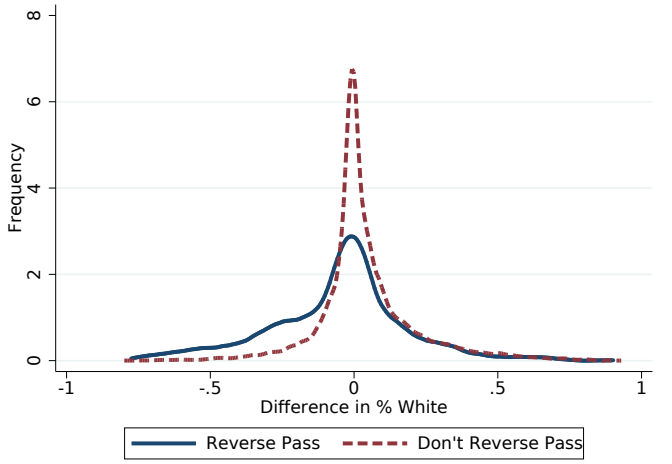


Figure 2: Racial Composition of the County of Residence

(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 U.S. Historical Censuses

A.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 c1asses, 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

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."

A.1.1 Mulatto

In 1880, 1910 and 1920, mulattos – i.e., individuals of African and European descent – comprised a separate category from black. The instructions to enumerators were vague, as described in the previous section. Moreover, interpretation of the instructions for distinguishing between black and mulatto vary across enumerators. "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). 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).

B 2SUP Algorithm

The 2SUP algorithm is illustrated in Appendix Figure A.2. It builds on algorithms used by past studies such as Abramitzky et al. (2014) and Long and Ferrie (2013a). The main departure is in requiring a two-direction unique link, rather than a one-direction unique link. See Section 4 in the paper for a discussion of the advantages and disadvantages of this innovation.

B.1 Constructing the Using Sample

We begin with 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.[90] Then, we create phonetically spelled names for all remaining individuals using Phonex. Phonex, described in Lait and Randell (1996), is a combination of the Soundex and Metaphone methods.[91] Snae (2007)

[90] We follow Feigenbaum (2014) in removing middle names/initials.

[91] 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.

shows that Phonex achieves a higher percentage of the true matches relative to all other methods excluding hybrid methods, and that the Phonex method not only outperforms all other methods in terms of percent of true matches, but it also does so without sacrificing accuracy.[92] We take each black male in year t and search through all males in year $t+10$ to identify all Phonex matches. Phonetic matches are generous in that they are likely to contain the true match absent serious spelling mistakes. The average black male has 6,822 potential Phonex links.

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 $age_t + 7 \leqslant age_t + 10 \geqslant age_t + 13$. The second restriction is to only keep links that have matching birth places (birth states and birth countries for foreign-born individuals).[93] 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.

B.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).[94] 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

[92] 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.

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

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

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.

B.3 Passing for White by Age and Year

Appendix Table A.4 shows the rates of passing for each census interval. The rates of passing are larger, 22.4% versus 15-18%, 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.[95]

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.

[95] 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.

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: 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.3: Literacy

		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%

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

Table A.4: 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.

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: 2SUP Linking Algorithm