

The Black-White Lifetime Earnings Gap

Ezra Karger* and Anthony Wray†

October 22, 2024

Abstract

The average white male born in 1900 earned 2.6 times more labor income over their lifetime than the average Black male. This gap is nearly twice as large as the more commonly studied cross-sectional Black-white earnings gap because 48% of Black males born in 1900 died before the age of 30 as compared to just 26% of white males. We calibrate a model of optimal consumption in a world with mortality risk to data describing the life-cycle earnings and survival probabilities of Black and white males born between 1900 and 1970. We find that convergence in Black and white mortality rates led to a 50% reduction in Black-white welfare gaps between the 1900 and 1920 birth cohorts, even as cross-sectional Black-white income gaps for those cohorts remained relatively constant. However, the Black-white welfare gap stagnated for the 1920 to 1970 birth cohorts as gaps in Black-white life expectancy and income remained stable and large.

*Karger: Federal Reserve Bank of Chicago (ezra.karger@chi.frb.org).

†Wray: Department of Economics, University of Southern Denmark (wray@sam.sdu.dk).

We thank Derek Neal, Azeem Shaikh, and participants of the Explorations in Economic History Special Issue Workshop on Wealth and Income Inequality around the World at the Federal Reserve Bank of Chicago for helpful comments. The views we express in this paper are our own and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

1 Introduction

Economists rely on repeated cross-sections of income and wealth to track inequality over time (Piketty 2003; Piketty and Saez 2003; Anand and Segal 2008; Derenoncourt et al. 2024), but such measures provide an incomplete description of inequality when there are differences in life expectancy between groups and over time. If unequal access to healthcare causes a randomly selected 20% of one group—but not a second group—to die before adulthood, and sharp improvements in healthcare access suddenly eliminate the gap in mortality, then this increase in life expectancy has large effects on the relative welfare of the two groups. However, standard measures of cross-sectional income inequality would not change due to this increase in healthcare access.¹

Economists recognize that mortality is a form of sample selection when measuring welfare. For example, Kanbur and Mukherjee (2007) construct measures of poverty that account for the “glaring paradox in all commonly used measures of poverty. The death of a poor person, because of poverty, reduces poverty according to these measures.”² While many researchers claim to address this concern by proposing measures of inequality based on a birth cohort’s adult outcomes, they often instead calculate inequality in the cross-section and fail to account for sample attrition due to mortality. For instance, in a pioneering study of cross-sectional Black-white gaps in educational attainment and earnings for adults using census microdata, Smith and Welch (1989, p. 33) claim that “[a]mong men born in this century, there has been a substantial narrowing of racial difference in years of school completed,” but the authors focus only on those Black and white men alive in later censuses as adults, ignoring the large number of men, often with low levels of education, who died before they could be enumerated in decennial censuses.

In this paper, we explore the importance of mortality in the measurement of Black-white in-

¹This example ignores any general equilibrium effects for the sake of exposition.

²Currie (2011, p. 11) makes a related argument: “A possible drawback to using data on births [to measure welfare effects of pollution abatement] is that pollution could affect the probability of a conception or of a live birth. If we suppose that pollution abatement would lead to fewer fetal deaths, and more births, and that the marginal fetus lost due to pollution is more vulnerable and less healthy than others, then focusing on births will tend to understate the beneficial effects of abatement by increasing the number of less healthy infants whose birth weight is recorded.”

come inequality. This is a widely measured quantity that has changed dramatically over the past 100 years. For example, Bayer and Charles (2018) show that the Black-white gap in median income shrank from roughly 1 log point in 1940 to 0.66 log points in 1950 before stagnating from 1950 to 2014.³ And earlier work by Card and Krueger (1993, p. 85) states that “The narrowing of the Black-white earnings gap between 1960 and the mid-1970’s represents one of the most significant episodes of relative progress for African Americans in U.S. history.” Cross-sectional earnings gaps represent differences in the relative price of available white and Black labor in a given year, and that is itself a potential outcome of interest in discussions of discrimination and inequality in the labor market. But economists and policymakers often point to these trends as measures of the relative lifetime outcomes of Black and white men.

A more complete way to think about inequality between birth cohorts is to measure the lifetime outcomes for these groups from birth. From this starting point, mortality is important. For example, 48% of Black males born in 1900 died before the age of 30 compared to only 26% of white males (Figure 1). But cohorts born between 1900 and 1920 experienced a reduction in mortality in both absolute and relative terms and these shares declined to 22% and 15%, respectively, by 1920. Much of this improvement in health was concentrated early in life as the racial gap in infant mortality rates closed by nearly 80 percent over the two decades (Figure 2). Measures of income inequality that do not adjust for these changes in racial health disparities may not correctly capture key trends in inequality.

We use census microdata and mortality data from vital statistics to provide the first long-run estimates of lifetime earnings by race in the US that cover the first half of the 20th century. After we calculate lifetime earnings by race, we construct a ‘full income’ measure of welfare based on income, life expectancy, and a simple structural model of optimal consumption. In the model, newborns of different types face known exogenous sequences of mortality risk and income each year. We use this model to discuss trends in Black-white inequality in the 20th century.

³See Figure 9A in Bayer and Charles (2018).

We find that the Black-white lifetime earnings gap for the 1900 birth cohort, the earliest in our sample, was almost twice as large as the cross-sectional earnings gap for working age men in 1940. The average white male born in 1900 earned 2.6 times more labor income over their lifetime than the average Black male (Figure 5). We attribute this difference to the higher mortality risk faced by Black men and show that the lifetime earnings gap closed considerably between 1900 and 1920, primarily due to improvements in Black life expectancy. Interestingly, this improvement occurred despite cross-sectional income gaps remaining relatively constant during this period (Figure 4). Lifetime earnings gaps remained fairly large and constant after the 1920 birth cohort, reflecting stagnation in both the Black-white mortality difference and cross-sectional earnings differences.

We conclude by calibrating our model of optimal consumption to Census data, showing that the closing of the lifetime earnings gap translated into welfare improvements for Black men. We show that the declining mortality risk for Black men relative to white men between 1900 and 1920 halved the Black-white welfare gap. Despite this initial progress, convergence then slowed, leaving a significant Black-white welfare gap today. This stagnation in the narrowing of the welfare gap is driven by persistent Black-white gaps in both earnings and mortality rates. These findings highlight the limitations of using cross-sectional earnings alone to measure Black-white inequality in the US.

Our paper makes contributions to three literature. First, our study contributes to literature on Black-white inequality in the US during the 20th century. Related literature measures the racial gap in wealth (Derenoncourt et al. 2024) and intergenerational mobility (Jácome et al. 2021), while Althoff and Reichardt (2024) attribute racial disparities to the legacy of slavery and restrictive Jim Crow institutions. The theoretical component of our paper relates to recent work by Brouillette et al. (2021), who calculate Black-white welfare gaps over time, with a focus on a more recent period (1984–2019), and find substantial convergence in Black and white welfare using a model that accounts for life expectancy, consumption, leisure, and inequality. We contribute to this literature by highlighting the role of mortality risk in driving Black-white inequality in lifetime earnings. We

also extend welfare comparisons back to the first-half of the 20th century.

Second, we contribute to work estimating lifetime earnings in different global and historical contexts. Researchers use administrative datasets to calculate lifetime earnings in Germany, Norway, and the United States (Bönke et al. 2015; Tamborini et al. 2015; Bhuller et al. 2017; Guvenen et al. 2022), but prior work only separates out lifetime earnings estimates by race using synthetic cohorts from more recent time periods.⁴ Our paper presents the first estimates of lifetime earnings by race across the 20th century.

Third, our work relates to a set of papers that explore the welfare effects of increasing life expectancy by considering the economic value of health improvements. Murphy and Topel (2006) use a structural model of health and labor decisions to estimate that increases in life expectancy since 1900 were worth trillions of dollars to society. In related work, Murphy and Topel (2005) show that increases in life expectancy since 1968 were worth more for Black men than they were for white men, while Becker et al. (2005) and Gallardo-Albarrán (2019) measure cross-country inequality by combining GDP and life expectancy to construct a measure of “full income.” Becker et al. (2005) show that cross-country inequality shrank significantly between 1960 and 2000 when they measure inequality using this combined ‘full income’ metric. Gallardo-Albarrán (2019) uses a similar method to show that GDP-based measures of welfare understate cross-country improvements in living standards in the early 1900s. In contrast, we measure the economic value of health improvements at the individual level.

2 Historical Background on the Black-white Health Gap

As a stark reflection of the prevailing racial disparities in health, Black males had significantly shorter lifespans than white males in the early 1900s. Figure 3 illustrates that white males born in 1900 had an average life expectancy at birth of 47 years while the life expectancy of Black males from the same birth cohort averaged only 33 years.⁵ Life expectancy increased dramatically across

⁴For example, Table 4-2 of Jena et al. (2010) examines lifetime race-based earnings gaps from 1970–2000.

⁵The life expectancy numbers come directly from the CDC (National Center for Health Statistics 2015) and rely on a variety of statistical procedures (Arias 2015).

birth cohorts between 1900 and 1940 and the Black-white gap in male life expectancy shrank from 14.1 to 10.6 years. Despite accounting for a sharp 11.8 year decline in life expectancy for both white and Black individuals from the 1918 birth cohort (Noymer and Garenne 2000), the 1918 influenza pandemic contributed to a reduction in racial inequality in influenza and pneumonia mortality with the non-white/white gap narrowing by 75% compared to previous years.⁶ Eiermann et al. (2022) provide evidence that this change in the mortality gap was driven by “immunological imprinting” from the 1890-92 influenza pandemic as urban White populations were more likely to have been exposure to the earlier pandemic, leaving them more vulnerable to the new H1N1 strain that emerged in 1918.

Both the level shift and the narrowing of the racial gap in life expectancy were largely driven by changes in infant mortality. Figure 2 shows a secular decline in Black and white infant mortality rates over the 20th century, punctuated by rapid convergence between 1900 and 1920, with the racial gap narrowing from 120 to 25 deaths per 1,000 live births. A consequence of the racial mortality gap is that while urban infectious disease mortality in the US declined dramatically during the epidemiological transition in the first-half of the 20th century, the decline occurred later in the South where the Black population faced a higher risk of death from infectious diseases (Feigenbaum et al. 2019).

Recent literature shows that discrimination contributed to the persistence of the racial health gap. Eli et al. (2023) argue that racial discrimination and physician bias reduced the size of pensions awarded to Black Civil War veterans compared to White veterans, which explains much of the racial gap in longevity in their sample. At the same time, improvements in nutrition and health care access helped to reduce racial inequalities in the early-20th century. Clay and Schmick (2020) show that the arrival of the boll weevil led to an increase in food production that improved early-

⁶This 11.8 year drop in life expectancy in Figure 1 is based on a cumulation of the age-specific decreases in life expectancy for US-born individuals in 1918. In other words, males born in 1918 did not have a realized lifespan 11.8 years lower than males born in 1917, but if one had given ever male born in 1918 the age-specific mortality rates of cohorts in 1918 as they proceeded through life, the projected life expectancy of males born in 1918 would have been 11.8 years lower than the projected life expectancy of males born a year earlier.

life conditions for Black children through better nutrition and a decline in pellagra. In addition to nutrition, access to hospitals and drugs also mattered for mortality. Private funding that upgraded hospitals and improved physician quality led to a reduction in the Black-White infant mortality gap (Hollingsworth et al. 2024), while the introduction of sulfa drugs during the 1930s reduced Black-White differences in pneumonia mortality, though the latter did not translate into a closing of socioeconomic gaps (Bhalotra and Venkataramani 2011).

Economic factors also played an important role in racial health disparities (Boustan and Margo 2016). The Black-White gap in economic status that existed around the peak of discriminatory “Jim Crow” era in 1900 remained large in the post-WWII era despite the Great Migration from the South and Civil Rights legislation during the intervening years. Collins and Wanamaker (2022) find that Black men were much less likely to experience upward mobility from the bottom of the occupational income distribution the White men, pointing to the role of intergenerational mobility in the persistence of the disparities. On the other hand, World War II had the opposite effect. During the 1940s, higher World War II casualty rates among semi-skilled White soldiers provided employment opportunities to Black southerners who experienced upward mobility from low-skilled to semi-skilled jobs (Ferrara 2022), but improvements in the labor market outcomes of Black men were only temporary in the aftermath of World War I (Ferrara 2023).⁷

3 Data

3.1 Census Microdata

We use 1910 to 1940 US full-count decennial census data, nationally representative samples of the 1950 to 2010 censuses, and annual 2006 to 2014 American Community Survey (ACS) microdata from IPUMS to estimate earnings gaps by race and birth cohort (Ruggles et al. 2024a,b; Karger and Wray 2024).⁸ For all analyses, we subset to Black and white men born in the U.S.

⁷Wartime casualties do not explain the mortality gap. White soldiers faced higher casualty rates in World War II (Ferrara 2022) and Black men voluntarily enlisted at lower rates while there was no racial gap in draft enlistment (Qian and Tabellini 2021).

⁸We do not use the 2001-2005 ACS waves because they do not include institutionalized individuals in their sampling frames. We also do not use the 2015 and 2016 ACS data because the Census Bureau modified the race variable

between 1900 and 1970. We use sample weights throughout so that our estimates are nationally representative of the American-born Black and white male populations.

We use the census microdata to calculate three separate measures of annual earnings: pre-tax labor earnings, post-tax labor earnings, and total pre-tax earnings. We start with labor earnings from 1940-2014 and total earnings from 1950-2014.⁹ In all years, we assume that all males have zero earnings before the age of 16, because these younger men were often not in the sampling frame for the income questions in the decennial censuses.¹⁰ Because no earnings measures are available before 1940, we must first calculate earnings for males before 1940 so that we can estimate lifetime earnings for the 1900–1970 birth cohorts (See Appendix A for more details).

We begin by estimating non-labor earnings for each male in the 1940 census using data from 1950. We regress 1950 non-labor earnings at the individual level on indicators for race, age, occupation, industry, state of birth, and state of residence.¹¹ We then use the regression coefficient estimates from this model to predict non-labor earnings for each male in 1940.¹² At this point, we have measured labor and total income for 1940 onward. To estimate the pre-1940 earnings of men, we use the same imputation procedure that we used to impute 1940 non-labor earnings. We regress labor and total earnings in 1940 for people in our sample on indicators for race, age, occupation,

in those years, complicating comparisons between pre-2015 and 2015-2016 data.

⁹In 1940, only labor earnings are available. Also, in 1940 1.5% of respondents were institutionalized and these respondents were not asked to report labor income. We assume that all of these respondents have zero earnings, which does not affect any of this paper’s conclusions.

¹⁰Males under the age of 16 had very low measurable earnings in the early 1900s. According to Table 2 of the U.S. Census of Manufacturers report on the Earnings of Wage-Earners, less than 1% of total earnings were paid to workers under the age of 16 (United States. Bureau of the Census 1908). While this evidence is from only one (large) sector of the economy, anything close to a 1% change in the total earnings of any group in our data will not affect, in any way, the conclusions of this paper.

¹¹There are roughly 300 distinct occupations and 20 distinct industries in the census microdata. We also separately regress an indicator for whether or not each person worked on those indicators and included the predicted value from that regression in the wage regression.

¹²We do not correct for top-coding since it only affects approximately 0-2% of observations in Census earnings data. To confirm that this decision has no impact on any of our results, we replicate our main results by assuming that the income distribution in each year closely matched the income distribution in 2000, when top-coding had a negligible effect on reported income. For example, in 1950, which has the most top-coding of any census year, roughly 2% of labor income values are top-coded. In 2000, the 98th percentile of the labor income distribution is roughly \$190,000. The average of income values in the top 2% of the 2000 income distribution is roughly \$300,000, so we multiply all top-coded income values in 1950 by $\frac{300000}{190000} = 1.6$. In 1960–1990, the top-coding ‘ratio’ is around 1.15.

industry, state of birth, and state of residence. We use these regression coefficients to predict the amount of earned income for all men in 1910, 1920, and 1930.

We calculate post-tax earnings by merging federal individual personal income tax rates (Tax Foundation 2013) and personal exemption amounts (Tax Policy Center 2015) from the appropriate year onto each individual in the 1940-2014 census and ACS data. We apply these tax rates and personal exemptions to each person's nominal labor earnings to estimate each person's post-tax earnings. Finally, we inflate all earnings variables to 2015 dollars using the CPI-U (IPUMS USA 2024). We rely on federal income tax rates and not state or local income tax rates for two reasons. First, there is no standardized time series of state income tax rates in the early or mid 1950s.¹³ Second, federal income taxes are the largest source of income tax revenue. For example, in 2015, federal, state, and local personal income tax revenue totaled 1.5 trillion dollars, 336 billion dollars, and 32 billion dollars respectively.¹⁴ The disparity between federal and non-federal income tax revenue was even larger in the early 1900s.

We then collapse the census microdata into estimates of the average income for each age-by-year-by-race cell. We linearly interpolate these income averages for each age-by-race cell in years between adjacent census years.¹⁵ For example, we take the average income of white 16 year olds in 1940 and 1950, and linearly interpolate average incomes of white 16 year olds for the years 1941-1949. We also linearly interpolate population counts for each age-by-race cell across missing years. Because pre-1940 earnings information was based on the relationship between 1940 demographic information and earnings, we lastly multiply all pre-1940 earnings cells by the ratio of average earnings per person in that year to average earnings per person in 1940 (United States. Bureau of

¹³Bakija (2019) constructs a federal and state income tax calculator, but the calculator is incomplete and not yet available to researchers.

¹⁴See <https://www.taxpolicycenter.org/statistics/revenue-government-level>.

¹⁵Each age-by-race cell has roughly ten snapshots of census and ACS data that we use to interpolate income values. Linear interpolation can potentially impact estimates of earnings in either direction. On one hand, interpolation may smooth over shocks to earnings that occur in intracensal years and are not captured by our data. On the other hand, interpolation can exacerbate the influence of temporary shocks to earnings if such deviations occur in census years (Dezhbakhsh and Levy 1994, 2022). Our results are virtually identical when we interpolate income data using more flexible cubic splines.

the Census 1975).¹⁶ In a small number of sparsely populated cells describing respondents at young ages, the imputed earnings are very slightly negative, and thus we zero out the earnings.

3.2 Survivorship by Age and Race

We draw on estimates of period survival probabilities through age $t \in \{1, 5, 10, 15, \dots, 100\}$, which we denote s_t^p , for Black and white men in decennial census years k from 1900 to the present. These probabilities are constructed by the Center for Disease Control (CDC) (Table 20 in Arias 2015). The CDC constructs these survival probabilities using census data, medicare records, and birth and death certificates. The survival probabilities s_t^p in year k apply to a hypothetical “period life table cohort” that experiences the age-specific death rates of the actual population in year k . We linearly interpolate the survival probabilities between ages and census years for each race. For each birth cohort, $p_t = 1 - s_t^p/s_{t-1}^p$ represents the age-specific mortality risk (i.e. the probability of dying at age t conditional on surviving to age $t - 1$) experienced by the cohort. Then, the product of these age-specific probabilities through age a , which we denote $s_a = \prod_{j=1}^a (1 - p_j)$, measures the probability of survival through age a for a birth cohort.¹⁷ In Figure 1, we plot these cohort survival probabilities through age 30 by race and birth cohort. We see that the largest Black-white convergence in the probability of survival took place between the 1900 and 1920 birth cohorts. In Section 4.2 we explore the implications of this racial convergence in mortality for lifetime earnings.

3.3 Validation of Survival Probabilities using Census Data

We validate the cohort survivorship rates by comparing birth counts published by the CDC to imputed birth counts implied by the combination of census population counts and the survival probabilities. The CDC data report birth counts separately for Black and white males from 1959-present and for white and non-white males from 1910–1959 (Hamilton et al. 2003).¹⁸ In years prior to 1959, when the CDC did not separate out births of Black children from births of all non-

¹⁶See series D722-727 on page 44 of Chapter D (Labor) for more details about how the series was constructed.

¹⁷To obtain survival probabilities for recent cohorts, we need to project future survival. We assume that the age-specific survival probabilities for the 2011 period life table cohort apply to all future years.

¹⁸We rely on birth counts separated by the race of the child, when available, rather than the race of the mother. The CDC data described here rely heavily on birth and death certificates.

white children, we assume that the number of non-white and non-Black births equals 55,000, which is the number of non-white non-Black births in 1959. We impute birth counts by applying the cohort survival probabilities described in Section 3.2 to population counts in the census and ACS microdata (Ruggles et al. 2024a,b) in the appropriate birth cohort, race, and age group. This produces implied birth counts from each age-by-birth-cohort-by-race cell in the census microdata. In Figure 6, we plot the number of Black (panel A) and white (panel B) males born each year from 1900 to 1970. In each panel we report the raw CDC birth counts and the median number of implied births from cells for ages 5-60 in the census.¹⁹ The birth counts implied by cohort survivorship rates applied to census counts closely match the CDC birth counts, although the CDC birth counts tend to be roughly 5% larger for both Black and white males.

3.4 Under Registration of Births

CDC mortality rates for 1900 and 1910 rely on the 10 death registration states that collected complete vital statistics information at the time, while the 1920 rates use 34 states with high-quality data. Black births were under-reported in the early 1900s because of the CDC’s reliance on death registration states and because of poor birth certificate quality, which in turn biases Black infant mortality rates. Eriksson et al. (2018) show that Black infant mortality rates are biased upward relative to white infant mortality rates because Black newborns are less likely to be enumerated. For example, in 1940, 94% of white births were registered on birth certificates, compared to only 82% of Black births, which is consistent with earlier work showing that Black births were under-counted on birth certificates by roughly 20% in 1880–1940 (Coale and Rives 1973). Eriksson et al. (2018) argue that the Black infant mortality rate for the 1915 birth cohort was 11.1%. Meanwhile, the CDC data (Table 20 in Arias 2015) implies a Black infant mortality rate of roughly 14.9%. While this difference is quantitatively large, our estimates of lifetime earnings gaps are largely driven by death rates before the age of 20. Moreover, this 3.8pp underestimation of the mortality

¹⁹We subset to these age ranges because at ages 0-4 and 61+, census population counts have various rounding, under-counting, and small sample size issues in some years.

rate is not quantitatively important for our welfare estimates when compared to the probability of death before the age of 20, which was roughly 29% for the 1915 birth cohort of Black males.

3.5 Sample Selection

Several factors may generate sample selection that leads to underestimates of the lifetime earnings gap. For example, we do not observe whether there is a socioeconomic gradient to infant mortality. If infant mortality is higher among poor families, then had the children from these families survived to adulthood, they would likely have earned lower incomes. Thus, the observed distribution of incomes would be positively selected from the population at birth, meaning that later-in-life earnings may be overstated. Likewise, the surviving population at older ages is positively selected on health, implying that late-in-life mortality risk is understated. To the extent that the surviving Black population is more positively selected on income and health than the surviving white population, we would underestimate the racial income and welfare gaps and our estimates of the racial lifetime earnings gap would represent a lower bound.

4 Results

4.1 Cross-Sectional Earnings Gaps

In Table 1, we report the average earnings of Black and white 30 year-old men every ten years, beginning in 1930 and ending in 2020.²⁰ Earnings are reported in 2014 dollars and inflated using the CPI-U (IPUMS USA 2024). These data correspond to average age-30 earnings for the 1900 through 1970 birth cohorts. We do not subset to wage-earners, so men with no earnings are included in the average. White earnings increase by roughly 60% over this time period while Black earnings grow by roughly 50%. The white-to-Black ratio in average earnings at age 30 increases from 1930 to 1950 before declining between 1960 and 1980. Post-tax labor earnings (column 2 of each panel) have a slightly smaller white-to-Black earnings ratio for the simple reason that income tax rates in the U.S. are progressive and white males had higher average incomes than Black males. Total earnings, which includes government transfers, investment, and other income, display similar

²⁰We impute earnings for the pre-1940 and post-1914 years using the methods described in the data section.

trends to labor earnings. Table 1 shows that white males born in 1900 who were alive at age 30 earned an average of 1.5 to 1.6 times the amount that Black males born in 1900 earned at age 30.

In Panel A of Figure 4, we plot the ratio of white to Black labor earnings for living men in different age groups to better visualize the evolution of cross-sectional white earnings relative to Black earnings over time.²¹ The figure separates out this ratio for living men in each birth cohort and in three age bins: men in their 30s, 40s, and 50s. To produce each line, we average the ratio of white to Black earnings for each age in that group. Recall that we impute data in non-decennial census years before 2000. Also, for the 1975 birth cohorts and the men in their 50s, the white/Black earnings ratio is projected forward in years that have no available data (2015-2029). The figure shows that the white-to-Black ratio of average earnings is roughly 2 for the 1900 birth cohort in each age bin. The gap steadily drops to around 1.7 in the 1945 birth cohort before slowly increasing back up to 1.8-2.1 depending on the age bin. If we used this type of standard inequality metric as a measure of the evolution of Black-white inequality for children born in different years, we might assume that there was no progress in the expected adult labor market outcomes of white men relative to Black men between 1900 and 1970. But as we discussed above, this is misleading.

Lastly, in Panel B of Figure 4 we show how the ratio of white to Black labor earnings evolves over the life-cycle for four different birth cohorts from 1900 to 1975. For more recent cohorts, we impute data for later ages. The plot is noisy, because the sample sizes for a specific birth cohort-by-age-by-race cell are not particularly large (especially for Black males), but the general trend is that the relative earnings of white relative to Black males increases from 1.5 at young ages to closer to 2 at higher ages.

4.2 Lifetime Earnings Gaps

We estimate Black and white lifetime earnings for each birth cohort using the decennial census and ACS microdata described in the data section, imputing earnings for unavailable years and

²¹The graphs for post-tax labor earnings and total earnings are virtually indistinguishable and show all the same trends.

ages. In addition, to estimate lifetime earnings for more recent birth cohorts, we assume that real earnings will grow for each birth cohort-by-race group at a 1.5% rate from 2014 through 2060. We also assume that no one earns any income after age 89. We calculate two measures of lifetime earnings: undiscounted and discounted. For the discounted lifetime earnings, we discount each annual income datapoint back to the birth year for each cohort (assuming an annual discount rate of 0.96). We then sum the average earnings from age 16-89 for each race in each birth cohort, weighted by the interpolated CDC mortality rates. This is a measure of the ‘expected’ earnings for each birth cohort, if the men in that birth cohort only knew their race, expected annual income at each age conditional on race, and expected annual mortality rates conditional on race, with no other information about ability or environment.

In the previous section, we calculated cross-sectional earning gaps, which are standard and reported widely in the literature on Black-white inequality. We now provide the first estimates of lifetime earnings for birth cohorts as early as 1900 in the U.S. In Table 2, we report the average undiscounted lifetime labor earnings, lifetime post-tax labor earnings, and lifetime total earnings for Black and white birth cohorts born every five years from 1900 to 1970. In Table 3, we report discounted lifetime earnings ($\beta = 0.96$). Lifetime earnings roughly tripled for white men and increased by a factor of six for Black men when we compare the 1900 and 1940 birth cohort. But lifetime earnings then increased only slightly from the 1940 to 1970 birth cohorts. The ‘No Death’ column reports the lifetime earnings each newborn male could expect to receive in the given cohort if they died at age 90 and received the average earnings for the living males in their birth cohort*race cell at each age. As expected, the hypothetical lifetime earnings in a world with no death are roughly 45-50% larger for white males than their actual average lifetime earnings. For Black males, the hypothetical lifetime earnings in a world with no mortality until age 90 are 2.3-2.6 times larger than the actual lifetime earnings.

We also include, in Tables 2 and 3, a column titled “+VSL” that measures the economic value of the difference in life expectancy for each birth cohort of Black and white men. We calculate

this value as follows: we begin with a \$2.2 million estimate of the value of a statistical life for the 1970 birth cohort.²² We follow the U.S. government in assuming that the value of a statistical life year is constant across ages and races (Aldy and Viscusi 2008). Because the lifetime income and life expectancy of the average male born in the U.S. grew tremendously between 1900 and the present, we scale our estimate of the VSL by the changes in lifetime income and the changes in life expectancy each year. To do this, we first assume that the \$1.5 million estimate comes from the 1970 birth cohort. White males born in 1970 had a life expectancy of 68 years. So, the 1970 birth cohort of white men valued each life-year at around \$32,000 2015 dollars, which is 0.074-times their discounted lifetime earnings of \$432,000. So, we use the average lifetime earnings of white men multiplied by 0.074 as a measure of the value of a statistical life year for each birth cohort. This means that the statistical value of a life year increases as the average lifetime income of men grows. Here, we are assuming that a 1% change in lifetime income increases the statistical value of a life-year by 1%. This is not unrealistic, since recent estimates of the income elasticity of the value of a statistical life year range from 0.7 to 1.1 (Viscusi and Masterman 2017). The “+VSL gap” column shows the product of the statistical value of a life year and the Black-white difference in life expectancy for each birth cohort. When we report discounted lifetime earnings in Table 3, we interpret the “+VSL gap” bonus as a future lump sum payment transferred at the life expectancy of Black individuals from a given birth cohort and discount the “+VSL gap” bonus accordingly.

In Tables 4 and 5, we report the ratio of white to Black undiscounted and discounted lifetime earnings, for all three measures of earnings (labor, post-tax labor, and total earnings). We also report in the ‘No Death’ column what the ratio of white to Black lifetime earnings would be in a

²²Ashenfelter and Greenstone (2004) use changes in speed limits to estimate the causal effect of a change in mortality risk on time saved. By putting a value on each hour of travel time saved equal to the average hourly wage, they estimate that the value of a statistical life is \$1.5 million in 1997 dollars, which is \$2.22 million in 2015 dollars. Recent papers by Ioannidis et al. (2017) and (Doucouliagos et al. 2012, p. 13) show that standard hedonic measures of the VSL suffer from “severe publication selection bias.” The authors correct for this publication bias by conducting a meta-analysis of the VSL parameter after restricting to studies with a sufficient amount of power. They estimate a VSL of \$1.47 million in 2000 dollars, which is virtually identical to the estimate of the social value of a statistical life from Ashenfelter and Greenstone (2004).

world where everyone lived through age 89 and died at age 90. White males born in 1900 earned roughly 2.6 times as much over their lifetime as Black males born in 1900 did. This ratio fell to 2.0 for the 1920 birth cohorts and has remained virtually the same through the 1970 birth cohort. The last column in each panel of Tables 4 and 5 presents the ratio of white to Black expected consumption from the model we present in the next section.

It is clear that white men born in the early 1900s had significantly higher lifetime earnings than equivalent cohorts of Black men. But white men could also expect to live much longer than Black men in all birth cohorts. One simple way to combine both of these facts into one welfare metric is to add the value of the additional life-years that white males could expect to receive at birth to their lifetime income. The thought experiment here is that if white and Black males could all expect to live for the same number of years, but still expected to receive the lifetime earnings of their respective race-by-birth cohort, then the white birth cohort should be indifferent between receiving X additional years of expected life or receiving a lump sum payment at birth equal to the discounted value of a life-year times X . In Table 2, we presented this value of the Black-white gap in life expectancy for each birth cohort in the '+VSL' column. In the third column of each panel in Tables 4 and 5, we take the ratio of white lifetime earnings plus the value of the extra years of life white cohorts received to Black lifetime earnings. In Figure 5, we plot these measures of relative white and Black welfare annually – undiscounted in Panel A ($\beta = 1$) and discounted in Panel B ($\beta = 0.96$).

This method for combining lifetime earnings and life expectancy gaps into one welfare measure is useful because it is in dollars, but it is somewhat ad hoc. We must assume a fixed value of each year of life for a given birth cohort, independent of how long they live and independent of their annual income profile. In other words, there are no decreasing returns to life expectancy and there is no interaction between the return to additional income and the return to lower mortality risk. So, below we calibrate a simple model of consumption in a world with mortality risk to the census data.

5 Model

We consider a simple model of optimal consumption in a world with annual mortality risk. Time is discrete with T periods, $1, \dots, T$. Each person born in year 0 plans to consume c_a at each age a if they are alive, which provides them with utility at age a of $u(c_a)$. p_t is the independent and exogenous probability that a person dies at age t , so the probability of surviving through age a is $s_a = \prod_{j=1}^a (1 - p_j)$. Earnings at each age a , $y_a \geq 0$, are exogenous and expected lifetime utility, given a sequence of consumption $\{c_a\}$, is

$$U = \sum_{a=1}^T \beta^a s_a u(c_a)$$

where β is the standard discount rate.

At birth, individuals can sign contracts to borrow or loan money intertemporally at a constant annual interest rate r , so the budget constraint is:

$$\sum_{a=1}^T \frac{s_a c_a}{(1+r)^a} \leq \sum_{a=1}^T \frac{s_a y_a}{(1+r)^a}.$$

In other words, a person can sign a contract to borrow against future expected income to finance consumption, but she can only borrow against income that the lender expects her to receive given the commonly known probability of dying each period and the commonly known income process. This budget constraint assumes a complete annuities market in which each newborn agrees to send their income each year to the bank, and in return, they receive a predetermined amount of money from the bank each year that they are alive. The amount of money the bank sends each person is governed by the consumer's budget constraint, which states that the lifetime net present value of consumption must be no larger than the lifetime net present value of income. This condition is equivalent to a zero profit constraint for the bank that is borrowing and loaning money, so this is a world where many small banks compete in this annuity market with exogenous interest rate r .

At birth, each newborn chooses a potential sequence of consumption $\{c_1 \dots c_T\}$ to maximize expected lifetime utility subject to the budget constraint. Consumption is the total dollar value the newborn will consume in a given period if they are alive, so expected consumption in period t is $s_t c_t$. The Lagrangian for this maximization problem is:

$$\mathcal{L} = \sum_{a=1}^T \beta^a s_a u(c_a) + \lambda \sum_{a=1}^T \left(\frac{s_a (y_a - c_a)}{(1+r)^a} \right)$$

Differentiating with respect to c_t gives us a sequence of first order constraints that yield the standard Euler equation: $\beta(1+r) = \frac{u'(c_t)}{u'(c_{t+1})}$. For simplicity, if we assume that $u(c) = \log(c)$, the Euler equation simplifies to $c_{t+1} = \beta(1+r)c_t = [\beta(1+r)]^t c_1$, which implies that $c_t = [\beta(1+r)]^{t-1} c_1$. The logic follows identically for CRRA utility, but solving for a closed-form value of consumption each period is more difficult. Returning to the budget constraint and plugging in the Euler equation from the previous step, we see that:

$$\sum_{a=1}^T \frac{\beta^{a-1} s_a c_1}{1+r} = \sum_{a=1}^T \frac{s_a y_a}{(1+r)^a}.$$

So,

$$c_1 = \frac{\sum_{a=1}^T \frac{s_a y_a}{(1+r)^{a-1}}}{\sum_{a=1}^T s_a \beta^{a-1}}.$$

This means that for all time periods t ,

$$c_t = \frac{[\beta(1+r)]^{t-1} \sum_{a=1}^T \frac{s_a y_a}{(1+r)^{a-1}}}{\sum_{a=1}^T s_a \beta^{a-1}}.$$

With the above framework, we calculate, at birth, the amount each person expects to consume at each age as a function of known parameters. Thus, we can deterministically calculate the utility

a newborn can expect to receive given (1) constant parameters β and r , and (2) sequences of probabilities of dying at each age $\{p_a\}$ and incomes at each age $\{y_a\}$. Notice that T is the lowest age a such that $p_a = 1$. Because the decennial census sometimes top-codes age at 90, we will assume that everyone dies by the age of 90, and therefore no one consumes anything after age 89.

5.1 Comparative Statics

We can easily see that consumption in all periods is increasing in income, but the effect of mortality risk, p_a , on consumption in each period is more complicated. With many applications of the chain rule, we arrive at this partial derivative of consumption in period t conditional on being alive (c_t) with respect to the probability of dying (p_s) in a given period $s > 1$:

$$\frac{\partial c_t}{\partial p_s} = \frac{[\beta(1+r)]^{t-1}}{(1-p_s)} \left(\frac{\left(\sum_{a=1}^{s-1} \frac{s_a y_a}{(1+r)^{a-1}} \right) \left(\sum_{a=s}^T [s_a \beta^{a-1}] \right) - \left(\sum_{a=s}^T \frac{s_a y_a}{(1+r)^{a-1}} \right) \left(\sum_{a=1}^{s-1} [s_a \beta^{a-1}] \right)}{\left[\sum_{a=1}^T s_a \beta^{a-1} \right]^2} \right)$$

This partial derivative can be either positive or negative depending on the distribution of earnings and mortality over a lifespan. While this might seem unrealistic, recall that c_t does not represent realized consumption at age t . Instead, it represents consumption at age t if the representative agent lives to age t .

Using the partial derivative $\frac{\partial c_t}{\partial p_s}$, we once again use several applications of the chain rule to calculate how utility is affected by an increased probability of dying in a given period:

$$\frac{\partial U}{\partial p_s} = \frac{1}{(1-p_s)} \sum_{t=1}^T \beta^t s_t \left(\frac{\left(\sum_{a=1}^{s-1} \frac{s_a y_a}{(1+r)^{a-1}} \right) \left(\sum_{a=s}^T [s_a \beta^{a-1}] \right) - \left(\sum_{a=s}^T \frac{s_a y_a}{(1+r)^{a-1}} \right) \left(\sum_{a=1}^{s-1} [s_a \beta^{a-1}] \right)}{\left(\sum_{a=1}^T \frac{s_a y_a}{(1+r)^{a-1}} \right) \left(\sum_{a=1}^T s_a \beta^{a-1} \right)} - \log(c_t) \right)$$

Note that this sum is made up of two terms. The first, itself a complicated ratio of terms, is bounded below by -1 and above by 1 , so a sufficient, but not necessary condition for $\frac{\partial U}{\partial p_s} < 0$ is

if, for all t , $\log(c_t) > 1$. This partial derivative can sometimes be positive because if consumption is sufficiently close to zero, utility is negative, and thus an increase in the probability of dying increases expected utility by moving it closer to zero. However, for any plausible sequence of income, including all sequences of income that we see in census microdata for race*birth cohort cells, this partial derivative will be negative and an increase in mortality risk reduces utility.

Lastly, one value of this model is that it allows mortality risk to affect people differently as a function of their expected annual income. Equivalently, income shocks will have different effects on people with higher or lower mortality risk. To measure this, we calculate the cross partial derivative of c_t with respect to p_s and y_s , and we see that:

$$\frac{\partial c_t}{\partial y_s \partial p_s} = \frac{\beta^{t-1} (1+r)^{t-s} s_s}{(1-p_s) \left(\sum_{a=1}^T s_a \beta^{t-1} \right)^2} * \left[- \sum_{j=1}^{s-1} s_j \beta^{j-1} \right]$$

This term is always negative, which means that at higher values of income, an increase in mortality risk in a given period has an increasingly negative effect on the level of expected consumption in all periods.

We can now compare the expected utility of Black and white males over time and within a birth cohort by calculating levels of utility separately for Black and white newborns with known annual incomes (from census microdata) and probabilities of dying (from the CDC). Throughout the section below, we assume a discount rate of $\beta = 1$ or $\beta = 0.96$ and an interest rate of $r = 0.02$.

5.2 Calibration

We calibrate this model separately to any sequence of annual survivorship probabilities and earnings. We take each Black and white birth cohort from 1900 to 1970 and calculate the expected consumption and utility for that cohort. In Tables 4 to 5 and Figure 5, we show the ratio of dollars of lifetime expected consumption for white and Black birth cohorts each year. These closely match the general time trends of earnings-based inequality measures, with the largest convergence in expected lifetime consumption between the 1900 and 1920 birth cohorts. But this measure of

relative consumption calibrated to total earnings (in the last column of Table 4) shows continual gradual progress in Black-white inequality between the 1920 and 1970 birth cohorts.

In Panel A of Figure 7, we plot the level of utility for Black and white males in each birth cohort. Typically, levels of utility are unitless, but in this model they represent the net present value of future expected log consumption. As the figure illustrates, levels of utility have weakly monotonically increased from the 1900 to 1970 birth cohorts for Black and white males. In Panel B, we plot the gap between Black and white utility levels.²³ The largest convergence in Black and white utility occurred between the 1910 and 1920 birth cohorts. Since then, the gap is roughly constant. Black males born in 1970 have earnings and mortality risk that net them welfare equivalent to the welfare received by white males born in 1930.

6 Robustness

Our Black-white lifetime earnings and welfare gap results are robust to alternative specification choices and sample restrictions. We have already shown in Sections 4.2 and 5.2 that our results are unaffected by the choice of discount rate ($\beta = 1$ or $\beta = 0.96$). Here, we modify the model for the imputation of earnings in 1910 to 1940 in three ways: (1) we allow all variables used to predict earnings to vary by race, (2) we use non-wage income from the 1960 census rather than 1950 to impute non-wage income in 1940, and (3) we use the 1940 occupational income estimates from Ward (2023) and Collins and Wanamaker (2022) in place of our imputed earnings for 1910 to 1940.²⁴ We also include income at ages 14-15 in calculating lifetime earnings to account for racial differences in labor force participation at young ages.²⁵ We report these results in Figure 9

²³In these figures, future utility is not discounted ($\beta = 1$). Figure 8 illustrates that results are similar when we apply a discount parameter of $\beta = 0.96$.

²⁴The White-Black lifetime earnings gap for the 1900 birth cohort when using occupational income from Ward (2023) and Collins and Wanamaker (2022) is slightly larger compared to the range of estimates for the imputation of 1910 to 1940 earnings. This difference may reflect the fact that the occupational income estimates from Ward (2023) and Collins and Wanamaker (2022) do not adjust for the life-cycle profile in earnings and impute pre-1950 earnings at all ages using average occupational income. Thus we prefer our imputation method for computing lifetime earnings.

²⁵If child labor is not reported in the census, this adjustment will not fully correct for the fact that Black children were more likely to work than white children to work on the family farm (Logan 2022). However, given that the earnings gap is somewhat smaller when incorporating earnings at ages 14-15 compared to our main specification, it seems unlikely that this difference can explain the larger White-Black gap in earnings for the 1900 to 1920 cohorts.

using total income as the measure of earnings and show that the tendency for the White-Black gap in earnings and welfare to close between 1900 and 1920 continues to hold.

7 Conclusion

48% of Black males born in 1900 died before the age of 30 as compared to only 26% of white males. These mortality rates declined to 22% and 15%, respectively, by the 1920 birth cohort, and this sudden convergence had large effects on relative welfare. Black-white cross-sectional earnings gaps are often used as a measure of inequality, but these gaps ignore the large (albeit incomplete) convergence of Black and white mortality rates from 1900 to the present. We calculate two ‘full income’ measures of Black-white inequality that combine life-cycle earnings and mortality rates to measure welfare.

In the 1940 census, the average working-age white male reported earnings twice as large as the average working-age Black male, but because Black males had significantly higher mortality rates at each age, white males born in 1900 could expect to have 2.6 times the lifetime labor earnings of Black males born in 1900. This ratio fell to 2.0 for the 1920 birth cohort and has remained virtually unchanged from the 1920 to 1970 birth cohorts. A simple structural model of consumption in a world with mortality risk similarly shows that almost all of the convergence in Black-white welfare over the past century came from the rapid but incomplete convergence in Black and white mortality rates between 1900 and 1920 birth cohorts. Black-white welfare gaps have declined only slightly since the 1920 birth cohort.

This paper focuses on the racial gap in mortality risk as the primary channel that drives the racial gap in lifetime earnings. While we assume a common discount rate for households after accounting for mortality risk, other race-based factors such as discrimination in lending and differences in savings rates, earnings risk, occupational hazards, and access to credit may contribute to Black households discounting the future more than White households. We leave it to future research to explore these factors using a similar framework as in this paper.

References

- Aldy, Joseph E. and W. Kip Viscusi (2008) “Adjusting the Value of a Statistical Life for Age and Cohort Effects,” *The Review of Economics and Statistics*, Vol. 90, No. 3, pp. 573–581.
- Althoff, Lukas and Hugo Reichardt (2024) “Jim Crow and Black Economic Progress After Slavery,” Working Paper.
- Anand, Sudhir and Paul Segal (2008) “What Do We Know about Global Income Inequality?” *Journal of Economic Literature*, Vol. 46, No. 1, pp. 57–94.
- Arias, Elizabeth (2015) “United States Life Tables, 2011,” *National Vital Statistics Reports*, Vol. 64, No. 11, pp. 1–63, Hyattsville, MD: National Center for Health Statistics.
- Ashenfelter, Orley and Michael Greenstone (2004) “Using Mandated Speed Limits to Measure the Value of a Statistical Life,” *Journal of Political Economy*, Vol. 112, No. S1.
- Bakija, Jon (2019) “Documentation for a Comprehensive Historical U.S Federal and State Income Tax Calculator Program,” Working Paper.
- Bayer, Patrick and Kerwin Kofi Charles (2018) “Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940,” *The Quarterly Journal of Economics*, Vol. 133, No. 3, pp. 1459–1501.
- Becker, Gary S., Tomas J. Philipson, and Rodrigo R. Soares (2005) “The Quantity and Quality of Life and the Evolution of World Inequality,” *American Economic Review*, Vol. 95, No. 1, pp. 277–291.
- Bhalotra, Sonia and Atheendar Venkataramani (2011) “The Captain of the Men of Death and His Shadow: Long-Run Impacts of Early Life Pneumonia Exposure,” IZA Discussion Paper No. 6041.
- Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes (2017) “Life-Cycle Earnings, Education Premiums, and Internal Rates of Return,” *Journal of Labor Economics*, Vol. 35, No. 4, pp. 993–1030.
- Boustan, Leah and Robert A. Margo (2016) “Racial Differences in Health in the United States: A Long-Run Perspective,” in John Komlos and Inas R. Kelly eds. *The Oxford Handbook of Economics and Human Biology*: Oxford University Press.
- Brouillette, Jean-Felix, Charles I. Jones, and Peter J. Klenow (2021) “Race and Economic Well-Being in the United States,” December, National Bureau of Economic Research Working Paper No. 29539.
- Bönke, Timm, Giacomo Corneo, and Holger Lüthen (2015) “Lifetime Earnings Inequality in Germany,” *Journal of Labor Economics*, Vol. 33, No. 1, pp. 171–208.
- Card, David and Alan B. Krueger (1993) “Trends in Relative Black-White Earnings Revisited,” *The American Economic Review*, Vol. 83, No. 2, pp. 85–91.
- Clay, Karen and Ethan J. Schmick (2020) “The Impact of an Environmental Shock on Black-White Inequality: Evidence from the Boll Weevil,” National Bureau of Economic Research Working Paper No. 27101.
- Coale, Ansley J. and Norfleet W. Rives (1973) “A Statistical Reconstruction of the Black Population of the United States 1880-1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility,” *Population Index*, Vol. 39, No. 1, pp. 3–36.
- Collins, William J. and Marianne H. Wanamaker (2022) “African American Intergenerational Economic Mobility since 1880,” *American Economic Journal: Applied Economics*, Vol. 14, No. 3, pp. 84–117.
- Currie, Janet (2011) “Inequality at Birth: Some Causes and Consequences,” *The American Economic Review*, Vol. 101, No. 3, pp. 1–22, Publisher: American Economic Association.
- Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick (2024) “Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860–2020*,” *The Quarterly Journal of Economics*, Vol. 139, No. 2, pp. 693–750.
- Dezhbakhsh, Hashem and Daniel Levy (1994) “Periodic properties of interpolated time series,” *Economics Letters*, Vol. 44, No. 3, pp. 221–228.

- (2022) “Interpolation and shock persistence of prewar U.S. macroeconomic time series: A reconsideration,” *Economics Letters*, Vol. 213, p. 110386.
- Doucouliaos, Chris, T. D. Stanley, and Margaret Giles (2012) “Are estimates of the value of a statistical life exaggerated?” *Journal of Health Economics*, Vol. 31, No. 1, pp. 197–206.
- Eiermann, Martin, Elizabeth Wrigley-Field, James J. Feigenbaum, Jonas Helgertz, Elaine Hernandez, and Courtney E. Boen (2022) “Racial Disparities in Mortality During the 1918 Influenza Pandemic in United States Cities,” *Demography*, Vol. 59, No. 5, pp. 1953–1979.
- Eli, Shari J., Trevon D. Logan, and Boriana Miloucheva (2023) “The Enduring Effects of Racial Discrimination on Income and Health,” *Journal of Economic Literature*, Vol. 61, No. 3, pp. 924–940.
- Eriksson, Katherine, Gregory T. Niemesh, and Melissa Thomasson (2018) “Revising Infant Mortality Rates for the Early Twentieth Century United States,” *Demography*, Vol. 55, No. 6, pp. 2001–2024.
- Feigenbaum, James J., Christopher Muller, and Elizabeth Wrigley-Field (2019) “Regional and Racial Inequality in Infectious Disease Mortality in U.S. Cities, 1900–1948,” *Demography*, Vol. 56, No. 4, pp. 1371–1388.
- Ferrara, Andreas (2022) “World War II and Black Economic Progress,” *Journal of Labor Economics*, Vol. 40, No. 4, pp. 1053–1091.
- (2023) “Wars and the Labor Market Outcomes of Minorities in the U.S.,” October.
- Gallardo-Albarrán, Daniel (2019) “Missed opportunities? Human welfare in Western Europe and the United States, 1913–1950,” *Explorations in Economic History*, Vol. 72, pp. 57–73.
- Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner (2022) “Lifetime Earnings in the United States over Six Decades,” *American Economic Journal: Applied Economics*, Vol. 14, No. 4, pp. 446–479.
- Hamilton, Brady E., Paul D. Sutton, and Stephanie J. Ventura (2003) “Revised birth and fertility rates for the 1990s and new rates for Hispanic populations, 2000 and 2001: United States,” *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, Vol. 51, No. 12, pp. 1–94.
- Hollingsworth, Alex, Krzysztof Karbownik, Melissa A. Thomasson, and Anthony Wray (2024) “The Gift of a Lifetime: The Hospital, Modern Medicine, and Mortality,” *American Economic Review*, Vol. 114, No. 7, pp. 2201–2238.
- Ioannidis, John P. A., T. D. Stanley, and Hristos Doucouliagos (2017) “The Power of Bias in Economics Research,” *The Economic Journal*, Vol. 127, No. 605, pp. F236–F265.
- IPUMS USA (2024) “Note on Adjusting Dollar Amount Variables for Inflation.”
- Jena, Anupam B., Tomas J. Philipson, and Eric Sun (2010) *Health and wealth disparities in the United States*, Washington, D.C: AEI press.
- Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu (2021) “Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century,” National Bureau of Economic Research Working Paper No. 29289.
- Kanbur, Ravi and Diganta Mukherjee (2007) “Premature Mortality and Poverty Measurement,” *Bulletin of Economic Research*, Vol. 59, No. 4, pp. 339–359.
- Karger, Ezra and Anthony Wray (2024) “Data and Code for: The Black-White Lifetime Earnings Gap,” Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2024-10-22. <https://doi.org/10.3886/E209621V1>.
- Logan, Trevon D. (2022) “American Enslavement and the Recovery of Black Economic History,” *Journal of Economic Perspectives*, Vol. 36, No. 2, pp. 81–98.
- Murphy, Kevin M. and Robert H. Topel (2005) “Black-white differences in the economic value of improving health,” *Perspectives in Biology and Medicine*, Vol. 48, No. 1 Suppl, pp. S176–194.
- (2006) “The Value of Health and Longevity,” *Journal of Political Economy*, Vol. 114, No. 5, pp.

871–904.

National Center for Health Statistics (2015) “NCHS - Death rates and life expectancy at birth.”

Noymer, Andrew and Michel Garenne (2000) “The 1918 Influenza Epidemic’s Effects on Sex Differentials in Mortality in the United States,” *Population and development review*, Vol. 26, No. 3, pp. 565–581.

Piketty, Thomas (2003) “Income Inequality in France, 1901–1998,” *Journal of Political Economy*, Vol. 111, No. 5, pp. 1004–1042.

Piketty, Thomas and Emmanuel Saez (2003) “Income Inequality in the United States, 1913–1998*,” *The Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 1–41.

Qian, Nancy and Marco Tabellini (2021) “Racial Discrimination and the Social Contract: Evidence from U.S. Army Enlistment during WWII,” National Bureau of Economic Research Working Paper No. 29482.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler (2024a) *IPUMS USA: Version 15.0 [dataset]*, Minneapolis, MN: IPUMS.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren (2024b) *IPUMS Ancestry Full Count Data: Version 4.0 [dataset]*, Minneapolis, MN: IPUMS.

Smith, James P. and Finis R. Welch (1989) “Black Economic Progress After Myrdal,” *Journal of Economic Literature*, Vol. 27, No. 2, pp. 519–564.

Tamborini, Christopher R., ChangHwan Kim, and Arthur Sakamoto (2015) “Education and Lifetime Earnings in the United States,” *Demography*, Vol. 52, No. 4, pp. 1383–1407.

Tax Foundation (2013) “Federal Individual Income Tax Rates History, Nominal Dollars, Income Years 1913–2013.”

Tax Policy Center (2015) “U.S. Individual Income Tax: Personal Exemptions and Lowest and Highest Tax Bracket, Tax Rates and Tax Base for Regular Tax, Tax Years 1913–2015.”

United States. Bureau of the Census (1908) *Census of manufactures: 1905. Earnings of wage-earners.*, Bulletin 93, Washington, DC: Government Printing Office.

——— (1975) *Historical statistics of the United States, colonial times to 1970*, No. 93: US Department of Commerce, Bureau of the Census.

Viscusi, W. Kip and Clayton J. Masterman (2017) “Income Elasticities and Global Values of a Statistical Life,” *Journal of Benefit-Cost Analysis*, Vol. 8, No. 2, pp. 226–250.

Ward, Zachary (2023) “Intergenerational Mobility in American History: Accounting for Race and Measurement Error,” *American Economic Review*, Vol. 113, No. 12, pp. 3213–3248.

8 Tables

Table 1. Average Earnings at Age 30

Year	White Earnings			Black Earnings			White/Black Earnings		
	Labor	Post-tax	Total	Labor	Post-tax	Total	Labor	Post-tax	Total
1930	24,186	24,111	31,615	15,707	15,680	19,265	1.54	1.54	1.64
1940	19,627	19,271	26,318	12,282	12,059	15,360	1.60	1.60	1.71
1950	23,862	20,465	29,471	13,430	12,009	15,295	1.78	1.70	1.93
1960	36,993	30,471	42,820	20,916	18,088	22,732	1.77	1.68	1.88
1970	51,640	41,801	57,100	33,743	28,390	35,683	1.53	1.47	1.60
1980	46,917	38,443	52,454	31,770	27,045	34,241	1.48	1.42	1.53
1990	45,179	37,418	49,895	26,154	22,311	27,737	1.73	1.68	1.80
2000	47,061	38,613	51,566	29,963	25,242	32,172	1.57	1.53	1.60
2010	38,740	32,687	42,294	21,751	18,868	24,163	1.78	1.73	1.75
2020	38,535	32,463	41,337	24,061	20,689	25,418	1.60	1.57	1.63

Notes: Earnings data from default IPUMS census and ACS microdata samples (Ruggles et al. 2024a,b). Units are 2014 dollars. Labor earnings are only available in the Census data from 1940-2014 and total earnings only available from 1950-2014. We do not use ACS data from 2015-2016 because the race variable is not easily comparable to 1900-2014 census and ACS data. Data are subset to US-born Black and white men who report being age 30 in the census and ACS. Each cell is the average earnings across all living white or Black 30-year old men, including men who earn zero dollars. Labor earnings rely on the INCWAGE IPUMS variable, Total earnings rely on the INCTOT IPUMS variable. Post-tax earnings apply federal income tax rates to the INCWAGE variable. Earnings values in 2020 are equal to 2014 earnings values. 1930 labor and total income values and 1940 total income values are imputed as we describe in the text of the paper. For example, we regress 1940 labor earnings at the individual level on indicators for race, age, occupation, industry, state of birth, and state of residence. We then use these regression coefficients to predict labor earnings for each male in 1930.

Table 2. Undiscounted Lifetime Earnings

Birth Cohort	Whites							Blacks					
	Labor		Post-Tax		Total		+VSL Gap	Labor		Post-Tax		Total	
	Earnings	No Death	Earnings	No Death	Earnings	No Death		Earnings	No Death	Earnings	No Death	Earnings	No Death
1900	845	1303	769	1169	1259	2290	162	321	756	304	705	408	1159
1905	924	1389	817	1212	1389	2472	226	362	798	335	728	458	1225
1910	1039	1502	891	1277	1578	2721	209	439	872	397	779	557	1360
1915	1214	1657	1013	1374	1853	3012	257	571	980	506	860	726	1552
1920	1426	1850	1164	1504	2166	3293	172	712	1094	620	946	913	1733
1925	1623	2031	1305	1631	2447	3535	317	821	1205	704	1029	1072	1924
1930	1799	2190	1436	1747	2690	3725	303	925	1306	783	1103	1231	2106
1935	1975	2349	1573	1870	2920	3916	260	1035	1424	869	1194	1393	2297
1940	2139	2499	1704	1990	3139	4117	308	1134	1522	947	1269	1531	2443
1945	2235	2577	1785	2058	3276	4243	252	1191	1547	993	1289	1629	2519
1950	2228	2548	1792	2049	3283	4246	224	1182	1499	989	1253	1656	2519
1955	2227	2543	1801	2055	3302	4294	181	1135	1418	955	1192	1636	2482
1960	2257	2567	1830	2079	3370	4393	193	1121	1381	945	1164	1660	2504
1965	2314	2618	1876	2121	3487	4546	201	1166	1416	983	1193	1752	2616
1970	2407	2711	1952	2197	3651	4757	261	1220	1465	1028	1234	1861	2755

Notes: Earnings data from default IPUMS census and ACS microdata samples (Ruggles et al. 2024a,b). Units are thousands of 2014 dollars. Labor earnings are only available in the Census data from 1940-2014 and total earnings only available from 1950-2014. We do not use ACS data from 2015-2016 because the race variable is not easily comparable to 1900-2014 census and ACS data. Data subset to American-born Black and white men who report being age 30 in the census and ACS. Each cell represents the average lifetime earnings for white and Black males born in the given birth cohort. Earnings are calculated by summing average earnings for each birth cohort and race across all ages, weighted by the age-specific CDC mortality rate, and assuming that all men live only until age 89. Labor earnings rely on the INCWAGE IPUMS variable, Total earnings rely on the INCTOT IPUMS variable. Post-tax earnings apply federal income tax rates to the INCWAGE variable. Earnings values after 2014 are equal to 2014 earnings values, assuming an additional 1.5% annual growth rate. 1910, 1920, and 1930 labor and total income values and 1940 total income values are imputed as we describe in the text of the paper. For example, we regress 1940 labor earnings at the individual level on indicators for race, age, occupation, industry, state of birth, and state of residence. We then use these regression coefficients to predict labor earnings for each male in 1930. In non-surveyed years from 1916-2014, earnings are linearly interpolated at the age-by-race-by-birth year level before being summed to construct lifetime earnings measures. Pre-1940 earnings are adjusted to account for annual variation in wages. This has no effect on our results. The ‘No Death’ columns assume that all newborns live until age 89. The ‘+VSL Gap’ column estimates the value of the Black-white gap, calculated as described in the text.

Table 3. Discounted Lifetime Earnings ($\beta = 0.96$)

Birth Cohort	Whites							Blacks					
	Labor		Post-Tax		Total		+VSL Gap	Labor		Post-Tax		Total	
	Earnings	No Death	Earnings	No Death	Earnings	No Death		Earnings	No Death	Earnings	No Death	Earnings	No Death
1900	183	257	174	243	239	353	53	77	157	75	151	88	189
1905	191	263	178	244	250	363	79	83	158	79	150	93	189
1910	204	274	185	246	268	378	59	95	165	89	153	106	197
1915	229	294	201	256	300	405	60	119	181	109	165	131	215
1920	265	325	224	274	343	443	28	146	203	131	180	161	240
1925	300	359	248	295	384	480	53	164	220	144	192	185	265
1930	331	387	269	313	420	511	46	180	235	154	201	208	289
1935	367	420	295	337	459	545	34	201	257	170	218	235	318
1940	402	452	323	363	497	579	40	225	283	190	237	262	346
1945	419	466	338	376	517	595	27	242	295	203	247	282	362
1950	417	460	339	374	514	588	21	241	287	203	242	283	356
1955	415	456	339	373	510	583	16	229	269	194	228	272	339
1960	416	454	341	373	510	581	17	222	257	189	218	266	328
1965	417	453	342	371	512	581	17	225	258	191	219	272	333
1970	423	458	347	375	520	589	23	231	262	196	222	282	342

Notes: See note to Table 2. This table only differs in that the discount rate is set to $\beta = 0.96$.

Table 4. Undiscounted White/Black Lifetime Earnings

Birth Cohort	Labor				Post-Tax				Total			
	Earnings	No Death	+VSL Gap	Cnsm.	Earnings	No Death	+VSL Gap	Cnsm.	Earnings	No Death	+VSL Gap	Cnsm.
1900	2.63	1.72	3.14	2.80	2.53	1.66	3.06	2.72	3.09	1.98	3.48	3.24
1905	2.56	1.74	3.18	2.69	2.44	1.66	3.11	2.60	3.03	2.02	3.53	3.16
1910	2.37	1.72	2.84	2.49	2.24	1.64	2.77	2.38	2.83	2.00	3.21	2.95
1915	2.13	1.69	2.58	2.23	2.00	1.60	2.51	2.11	2.55	1.94	2.91	2.65
1920	2.00	1.69	2.25	2.10	1.88	1.59	2.15	1.98	2.37	1.90	2.56	2.47
1925	1.98	1.69	2.36	2.10	1.85	1.59	2.31	1.97	2.28	1.84	2.58	2.39
1930	1.95	1.68	2.27	2.08	1.83	1.58	2.22	1.96	2.18	1.77	2.43	2.30
1935	1.91	1.65	2.16	2.03	1.81	1.57	2.11	1.93	2.10	1.71	2.28	2.20
1940	1.89	1.64	2.16	1.98	1.80	1.57	2.12	1.89	2.05	1.69	2.25	2.13
1945	1.88	1.67	2.09	1.94	1.80	1.60	2.05	1.86	2.01	1.68	2.17	2.07
1950	1.88	1.70	2.07	1.93	1.81	1.63	2.04	1.86	1.98	1.69	2.12	2.03
1955	1.96	1.79	2.12	2.00	1.89	1.72	2.08	1.92	2.02	1.73	2.13	2.06
1960	2.01	1.86	2.19	2.04	1.94	1.79	2.14	1.96	2.03	1.75	2.15	2.08
1965	1.98	1.85	2.16	2.00	1.91	1.78	2.11	1.93	1.99	1.74	2.10	2.02
1970	1.97	1.85	2.19	1.97	1.90	1.78	2.15	1.90	1.96	1.73	2.10	1.98

Notes: See note to Table 2. This is the ratio of white to Black lifetime undiscounted earnings.

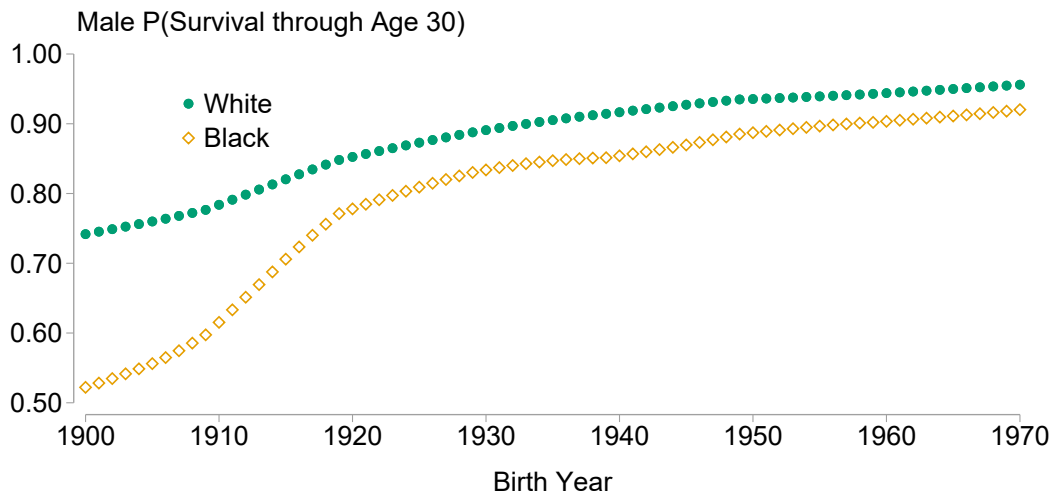
Table 5. Discounted White/Black Lifetime Earnings ($\beta = 0.96$)

Birth Cohort	Labor				Post-Tax				Total			
	Earnings	No Death	+VSL Gap	Cnsm.	Earnings	No Death	+VSL Gap	Cnsm.	Earnings	No Death	+VSL Gap	Cnsm.
1900	2.36	1.64	3.05	2.81	2.32	1.61	3.02	2.73	2.72	1.86	3.32	3.25
1905	2.31	1.66	3.26	2.70	2.25	1.62	3.24	2.60	2.69	1.91	3.54	3.16
1910	2.16	1.66	2.78	2.47	2.08	1.61	2.74	2.36	2.54	1.93	3.09	2.93
1915	1.93	1.62	2.43	2.20	1.84	1.55	2.39	2.08	2.29	1.89	2.74	2.61
1920	1.81	1.61	2.01	2.06	1.71	1.52	1.93	1.93	2.13	1.84	2.30	2.41
1925	1.83	1.63	2.15	2.04	1.72	1.54	2.09	1.92	2.07	1.81	2.36	2.33
1930	1.84	1.65	2.10	2.02	1.74	1.56	2.04	1.91	2.02	1.77	2.24	2.24
1935	1.82	1.63	1.99	1.99	1.73	1.55	1.93	1.88	1.95	1.71	2.10	2.15
1940	1.78	1.60	1.96	1.95	1.70	1.53	1.91	1.86	1.90	1.67	2.05	2.09
1945	1.73	1.58	1.85	1.91	1.66	1.52	1.80	1.83	1.83	1.64	1.93	2.03
1950	1.73	1.60	1.82	1.90	1.67	1.54	1.77	1.83	1.82	1.65	1.89	1.99
1955	1.81	1.69	1.88	1.97	1.75	1.63	1.83	1.89	1.88	1.72	1.94	2.03
1960	1.88	1.77	1.95	2.01	1.81	1.71	1.90	1.93	1.92	1.77	1.98	2.05
1965	1.85	1.76	1.93	1.97	1.79	1.70	1.88	1.90	1.88	1.75	1.95	1.99
1970	1.83	1.75	1.93	1.94	1.77	1.69	1.89	1.87	1.84	1.72	1.92	1.95

Notes: See note to Table 2. This is the ratio of white to Black lifetime earnings with annual discount rate $\beta = 0.96$.

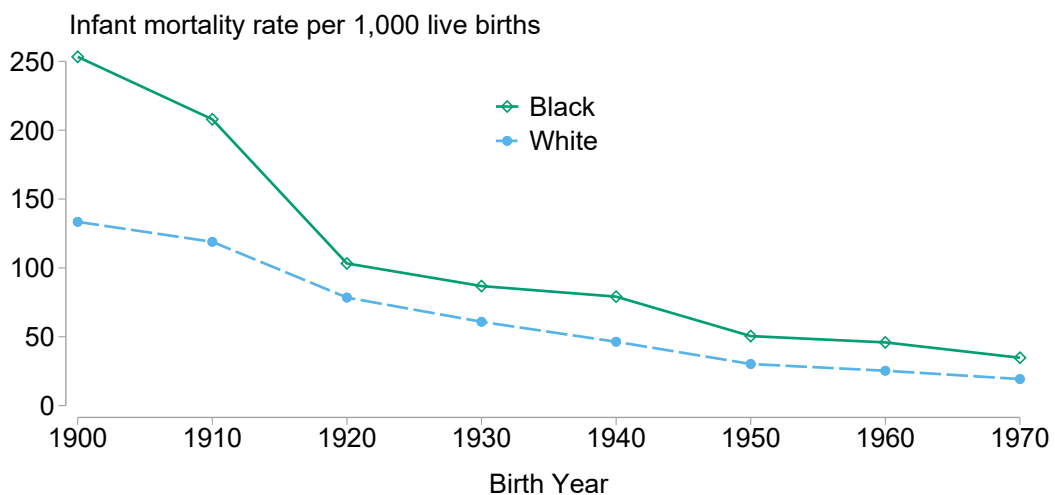
9 Figures

Figure 1. Male Survivorship by Race and Birth Year



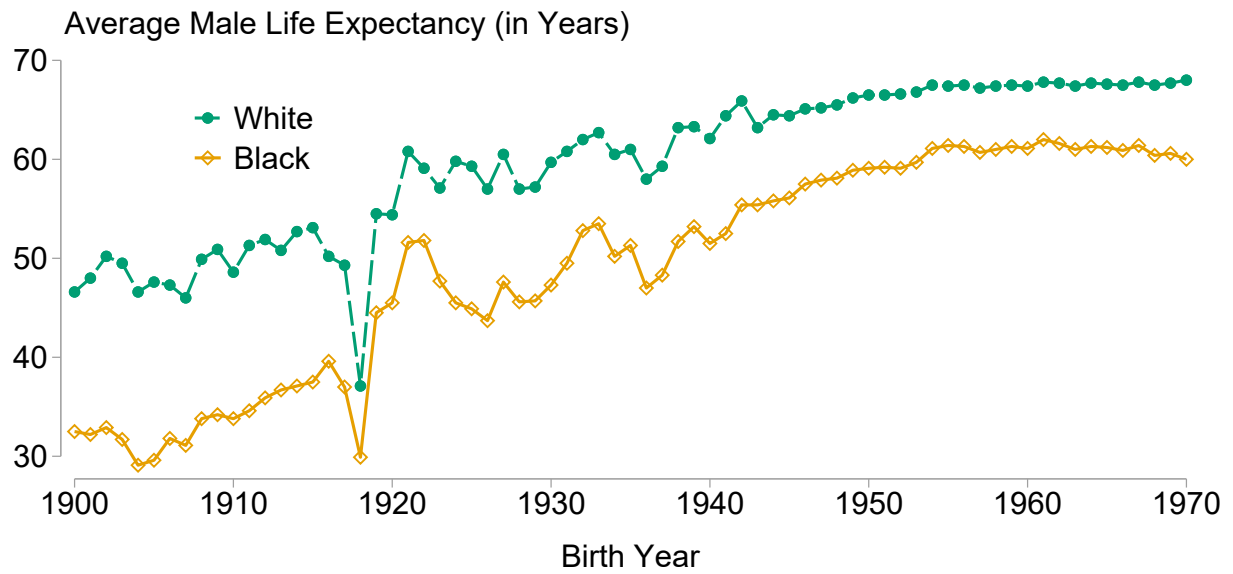
Notes: Probability of survival through age t by race and birth cohort computed from survival probabilities for a hypothetical period life table cohort that experiences the age-specific death rates of the actual population in a given year. See Section 3.2 for more details. Source: Period survival probabilities are taken directly from the CDC data (Arias 2015).

Figure 2. Infant Mortality By Race and Birth Year



Notes: This figure plots Black and white infant mortality rates per 1,000 live births. Birth counts by race are based on published CDC data (Hamilton et al. 2003). The number of infant deaths by race are derived from birth counts and the probability of survival from age 0 to 1 taken directly from CDC data (Arias 2015).

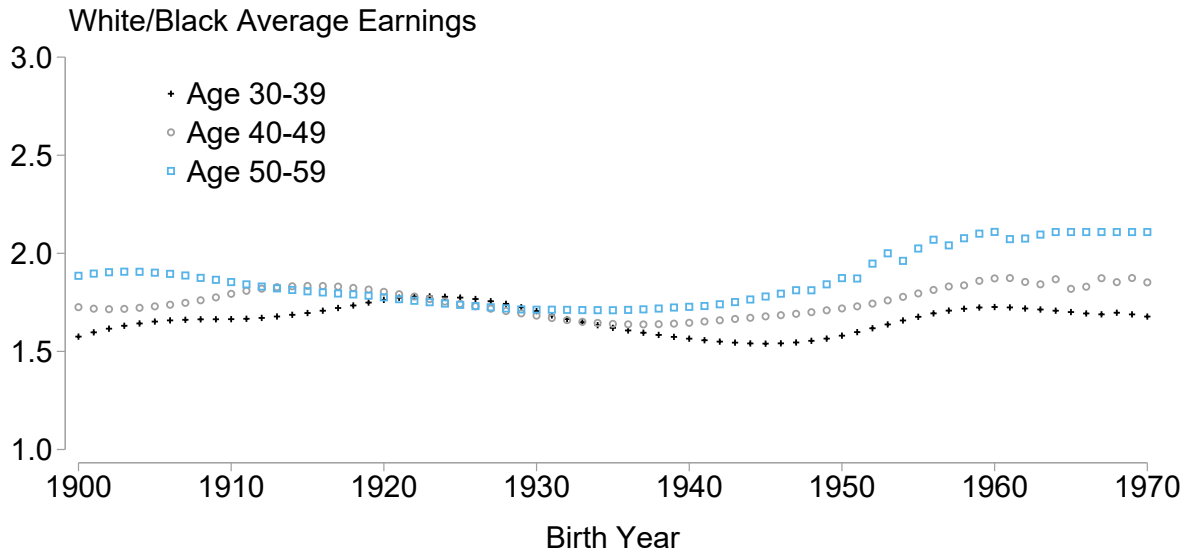
Figure 3. Life Expectancy by Race and Birth Year



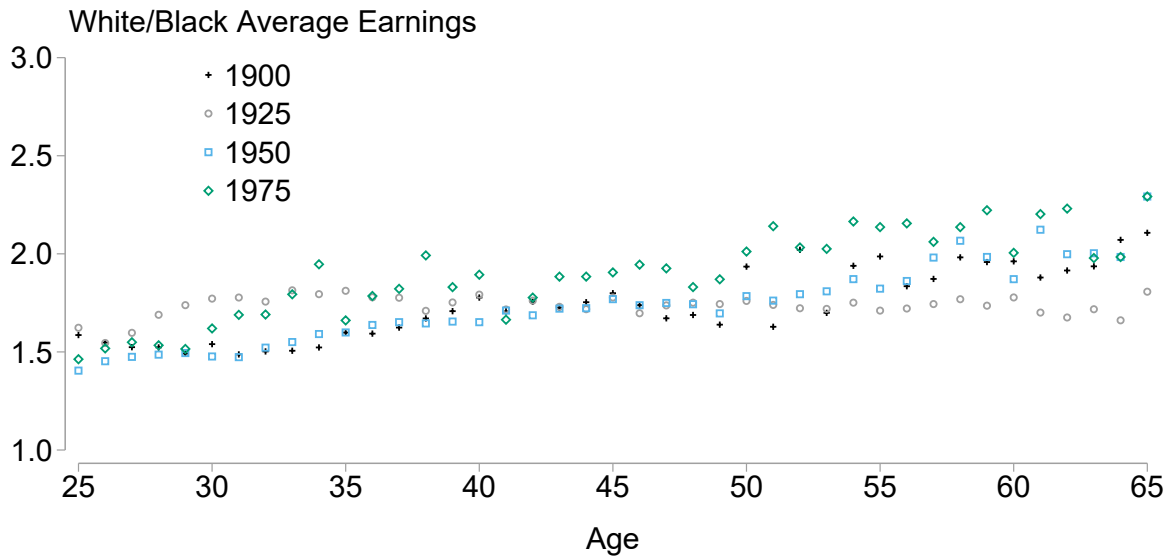
Notes: Life expectancy taken directly from CDC data (National Center for Health Statistics 2015). The life expectancy measures are calculated using each year's mortality rates for each cohort. See Arias (2015) for details.

Figure 4. White-to-Black Ratio of Average Earnings

(a) By Birth Year for 3 Age Groups



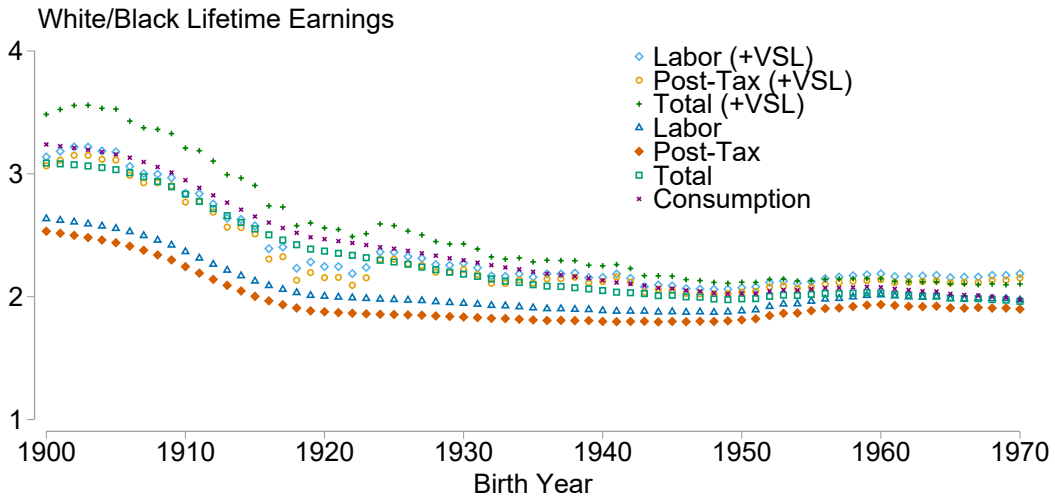
(b) By Age for 4 Birth Years



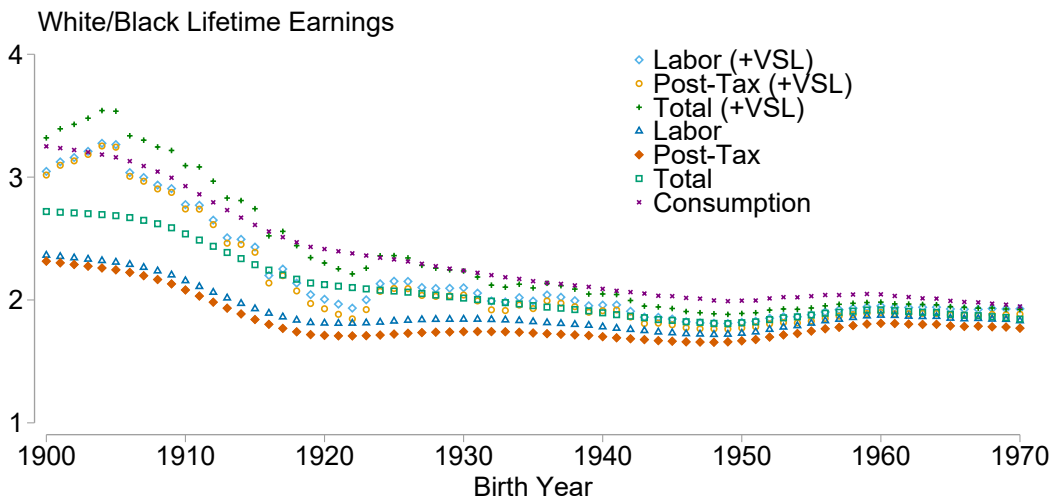
Notes: See note to Table 2. Earnings are linearly interpolated across race-by-age-by-birth year cells in non-census years and then averaged within 10-year age bins for each birth cohort to produce these scatter plots at an annual level. We assume that earnings after 2014 are equal to earnings in 2014 (within race-by-age cells). This is only relevant for the oldest age group in this figure.

Figure 5. White-to-Black Ratio of Lifetime Earnings

(a) Undiscounted ($\beta = 1$)



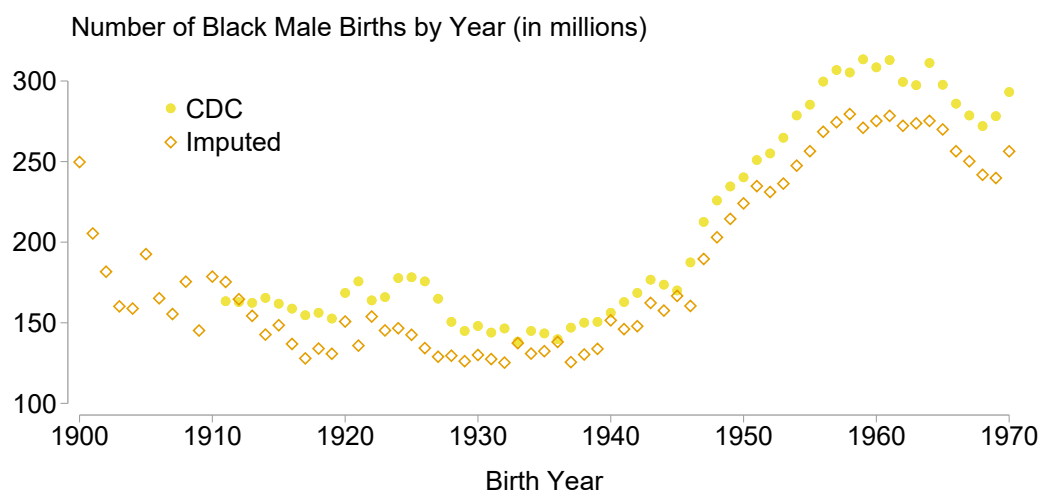
(b) Discounted ($\beta = 0.96$)



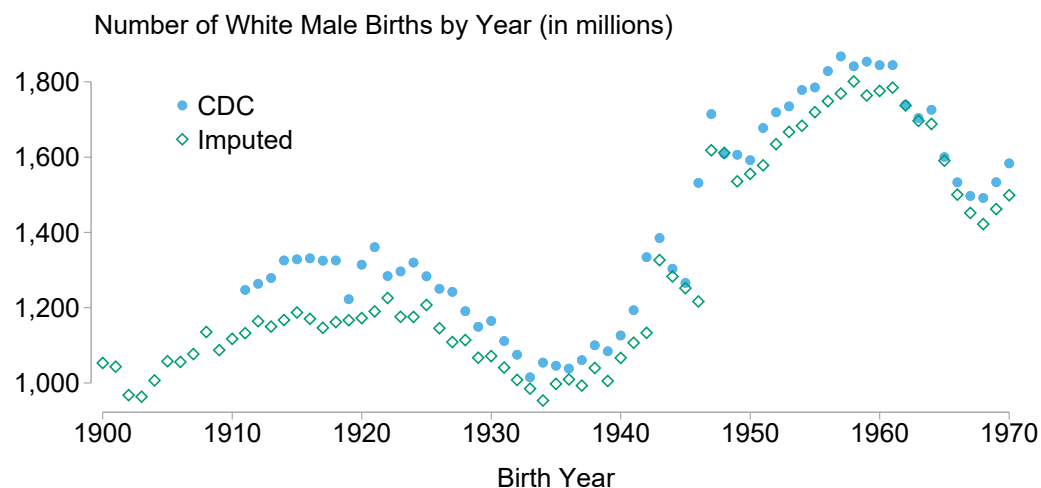
Notes: See note to Table 2. This figure shows the ratio of white to Black lifetime earnings for each birth cohort. Earnings are undiscounted ($\beta = 1$) in Panel A and discounted ($\beta = 0.96$) in Panel B. The +VSL lines correspond to the thought experiment we describe in the text, where we take the ratio of undiscounted white lifetime earnings + the value of the additional life-years each white male can expect to receive to undiscounted Black lifetime earnings. The consumption line corresponds to the lifetime expected consumption of white males divided by the lifetime expected consumption of Black males born in each birth cohort. Here, consumption is calculated assuming $\beta = 1$, but this ratio is quite similar when the consumption line is calculated using $\beta = 0.96$.

Figure 6. Validation of Survivorship Rates: Raw vs. Implied Birth Counts by Race

(a) Black males

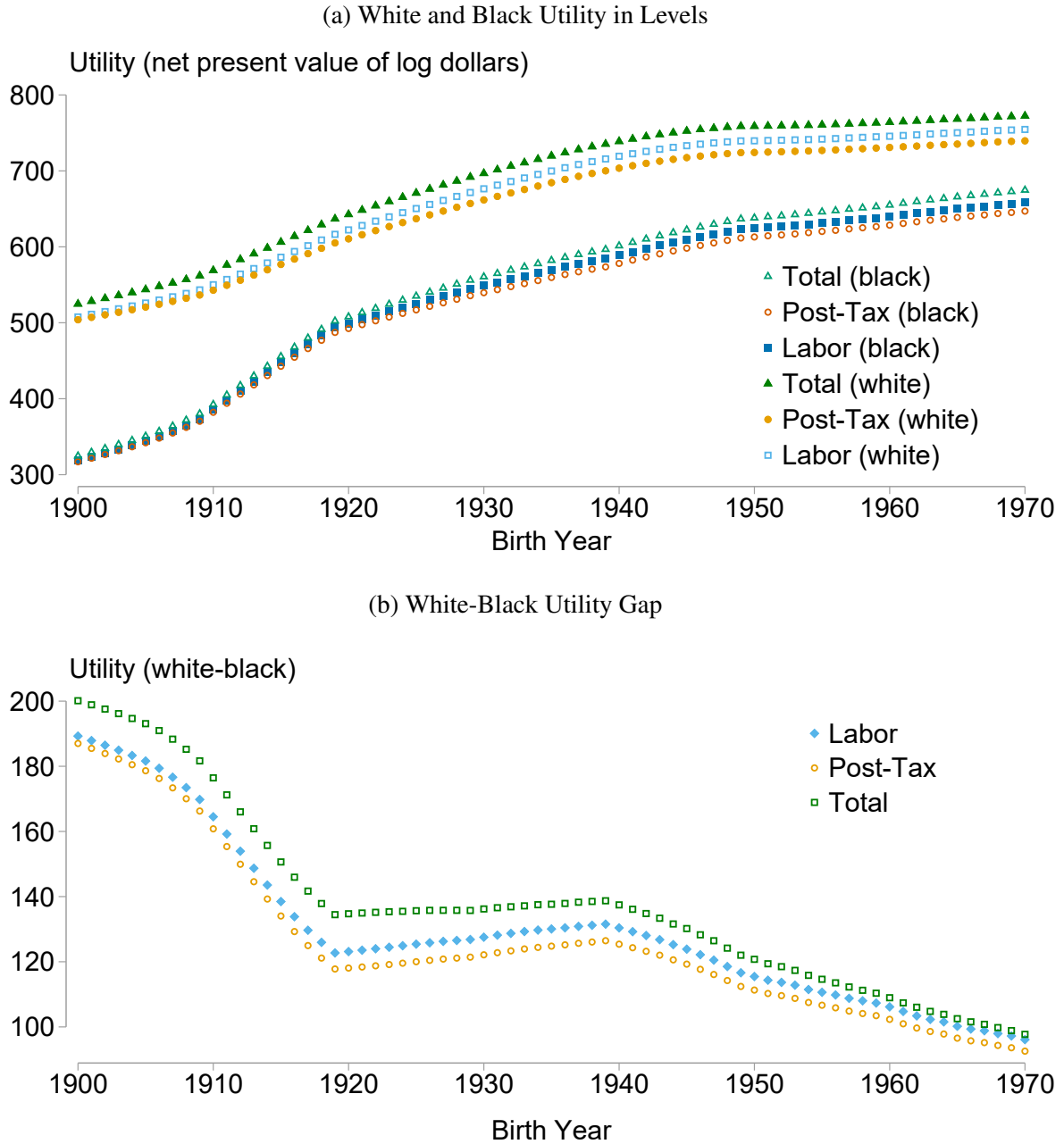


(b) White males



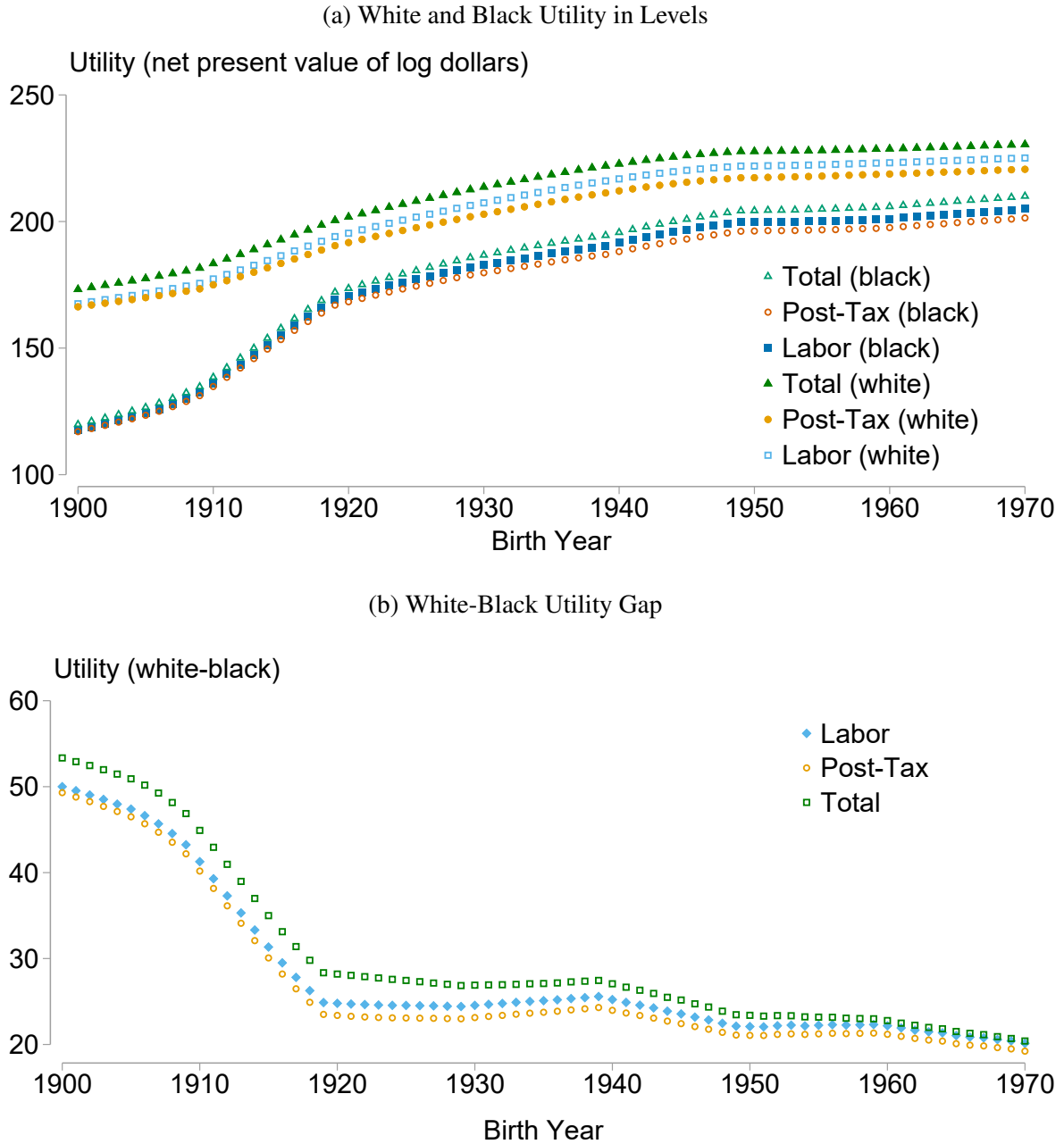
Notes: CDC birth counts for Black (panel A) and white (panel B) males based on published CDC data (Hamilton et al. 2003). Imputed birth counts apply age-specific mortality rates to birth counts from census microdata (Ruggles et al. 2024a,b). The imputed data point for each birth year equals the median imputed birth count from age 5-60 census age groups.

Figure 7. Undiscounted White and Black Utility by Birth Cohort ($\beta = 1$)



Notes: In Panel A, each line is the level of utility for white or Black males, extracted from the structural model we calibrate in the paper. Earnings are not discounted ($\beta = 1$). Panel B is simply the level difference in white and Black utility from Panel A. We include it to emphasize the large convergence in white and Black utility between the 1900 and 1920 birth cohorts, followed by general stagnation.

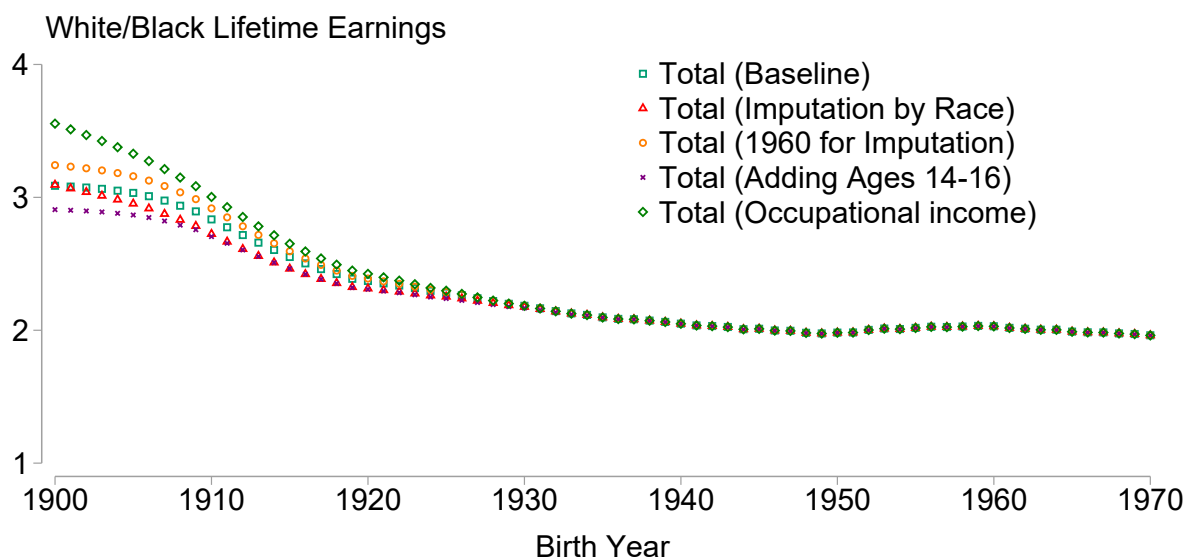
Figure 8. Discounted White and Black Utility by Birth Cohort ($\beta = 0.96$)



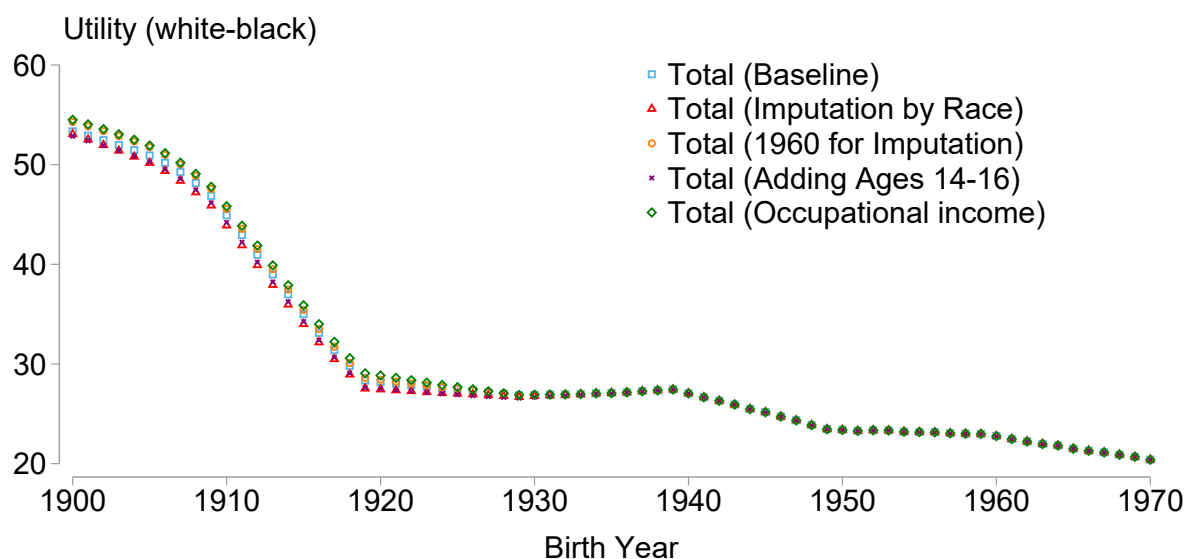
Notes: In Panel A, each line is the level of utility for white or Black males, extracted from the structural model we calibrate in the paper. We assume a discount rate of $\beta = 0.96$. Panel B is simply the level difference in white and Black utility from Panel A. We include it to emphasize the large convergence in white and Black utility between the 1900 and 1920 birth cohorts, followed by general stagnation.

Figure 9. Robustness of White-Black Lifetime Earnings Gap and Utility Gap Estimates

(a) White-Black Lifetime Earnings Gap



(b) White-Black Utility Gap



Notes: Panel A shows the White-Black gap in total lifetime income for each birth cohort across various specification choices. The “Baseline” specification reproduces the estimates shown in Figure 5. The “Imputation by Race” specification modifies the imputation of non-wage earnings in 1940 and both wage income and non-wage earnings in 1910 to 1930 by interacting all model covariates with an indicator for race. The “1960 for Imputation” specification uses the 1960 census to impute non-wage earnings in 1940 rather than the 1950 census. The “Adding Ages 14-15” specification includes earnings at ages 14-15 in the lifetime earnings calculation, rather than starting at age 16. The “Occupational income” specification uses the average occupational income by race and region in 1960 from Collins and Wanamaker (2022) to impute total income in 1910 to 1940. See Appendix A for more details on the imputation procedures.

Appendix: For Online Publication

The Black-White Lifetime Earnings Gap

Ezra Karger and Anthony Wray

A Imputation of Earnings

Pre-processing of data In this appendix section we describe the process used for the imputation of earnings for years where we do not observe earnings data. We use the complete-count 1940 US Census and the 5% sample of the 1950 census from IPUMS, restricting the sample to Black and White males. Prior to imputing earnings, we assign zero wage income to all individuals enumerated as prisoners in the 1940 census, to be consistent with other census years. We also adjust for inflation by converting nominal earnings to constant 1999 dollars using the IPUMS CPI99 variable (<https://usa.ipums.org/usa/cpi99.shtml>). We then adjust the base year to 2014 dollars by dividing by a factor of 0.704.

A predictor for negligible non-wage income We define non-wage income in 1950 as the difference between total income and wage income. We then define *negligible* non-wage income as non-wage income in the range $[-5, 5]$ dollars. Next, we estimate a model for whether an individual has negligible non-wage income in 1950:

$$1(\text{Negligible Non-Wage Income}_i = 1) = \alpha + \gamma_r + \delta_a + \theta_o + \phi_n + \pi_b + \chi_s + \epsilon_i \quad (1)$$

where γ_r are fixed effects for race, δ_a are fixed effects for age, θ_o are fixed effects for occupation (*OCC1950* codes), ϕ_n are fixed effects for industry (*IND1950* codes), π_b are fixed effects for state of birth, and χ_s are fixed effects for state of residence. We use this model to predict whether an individual has negligible non-wage income in 1940. In combination with the fitted values from this model, we construct a categorical variable for the ventiles of predicted non-wage income across all individuals in the 5% sample of the 1950 census and the 100% sample of the 1940 census. We also construct a categorical variable for the ventiles of wage income across the pooled 1940 and 1950 sample and use these two categorical variables as predictors when imputing 1940 non-wage income.

Imputation of non-wage earnings in 1940 We estimate a model for non-wage earnings in 1950:

$$\begin{aligned} \text{Non-Wage Income}_i &= \alpha + \sum_{p=1}^{20} \beta_p 1(\text{Ventile of predicted non-wage income} = p) \\ &+ \sum_{q=1}^{20} \eta_q 1(\text{Ventile of wage income} = q) \\ &+ \gamma_r + \delta_a + \theta_o + \phi_n + \pi_b + \chi_s + \epsilon_i \end{aligned} \quad (2)$$

where the ventiles of predicted non-wage income, ventiles of wage income and fixed effects are defined above. Using this model, we predict non-wage income in 1940 and add to this value the reported wage income in the 1940 census to obtain an estimate of total income in 1940.

Imputation of earnings in 1910 to 1930 Now we impute wage income and total income in the 1910 to 1930 censuses based on the wage income and estimated total income from 1940. We modify the model in Equation 1 to additionally use unemployment status in 1940 as a predictor. We proxy

for unemployment status in 1940 by using an indicator for *negligible* wage income variable in the 1940 census (i.e. reported wage income in the range $[-5, 5]$). We use this model to predict whether an individual is unemployed in 1940. In combination with the fitted values from this model, we construct a categorical variable for the ventiles of predicted unemployment status across all individuals in the 5% sample of the 1950 census and the 100% sample of the 1940 census. We then incorporate the ventiles of predicted unemployment status as a predictor when imputing 1910-1930 wage income in the following equation:

$$\begin{aligned} \text{Type of pre-1940 Income}_i = & \alpha + \sum_{p=1}^{20} \beta_p 1(\text{Ventile of predicted unemployment status} = p) \\ & + \gamma_r + \delta_a + \theta_o + \phi_n + \pi_b + \chi_s + \epsilon_i \end{aligned} \quad (3)$$

where Type of pre-1940 Income is either wage income or total income. Using these models, we predict wage income and total income in 1910, 1920, and 1930.

Robustness In Section 6, we explore the robustness of our main results to variations on the strategy for the imputation of earnings. Here we provide additional details on the adjustments to the imputation procedure. In separate specifications, we do the following:

1. Modify the prediction models (Equations 1 to 3) by incorporating interactions of race with all other fixed effects.
2. Use the 5% sample of the 1960 census rather than the 1% sample of the 1950 census (Ruggles et al. 2024a) when predicting non-wage income in 1940 due to concerns about small cell sizes in the 1950 census (Collins and Wanamaker 2022). This approach also more closely aligns our income estimates with the occupational earnings score of Collins and Wanamaker (2022) that we implement below.
3. Use the 1940 occupational income estimates from Collins and Wanamaker (2022) rather than our imputed earnings for 1910 to 1940. The Collins and Wanamaker (2022) approach imputes 1940 non-wage income using the average non-wage income from the 5% sample of the 1960 census for each race, region, and occupation cell. We add these estimates to the wage income reported in the 1940 census and apply this measure of total income to all years between 1910 and 1940. In this specification, we do not make any changes to income measures drawn from the 1950 census onward. Our implementation of the Collins and Wanamaker (2022) occupational income score uses the data set provided in the replication package of Ward (2023). See Kosack and Ward (2020) and Ward (2020) for more details.
4. Include income earned at ages 14 to 15 in our computation of lifetime income.

Appendix References

- Collins, William J. and Marianne H. Wanamaker (2022) “African American Intergenerational Economic Mobility since 1880,” *American Economic Journal: Applied Economics*, Vol. 14, No. 3, pp. 84–117.
- Kosack, Edward and Zachary Ward (2020) “El Sueño Americano? The Generational Progress of Mexican Americans Prior to World War II,” *The Journal of Economic History*, Vol. 80, No. 4, pp. 961–995.
- Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler (2024) *IPUMS USA: Version 15.0 [dataset]*, Minneapolis, MN: IPUMS.
- Ward, Zachary (2020) “The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration,” *American Economic Journal: Applied Economics*, Vol. 12, No. 4, pp. 73–102.
- (2023) “Intergenerational Mobility in American History: Accounting for Race and Measurement Error,” *American Economic Review*, Vol. 113, No. 12, pp. 3213–3248.