The historical incarceration penalty in the U.S. *

Desmond Ang, Ellora Derenoncourt, Kyle Hancock, and Jing Wu

July 17, 2024

Abstract

The U.S. has both the highest incarceration rate and the largest prison population worldwide. Furthermore, this reliance on incarceration in criminal justice policy is not only a modern phenomenon but dates back to the late 19th century. An active literature seeks to understand the social and economic impacts of this policy in the U.S., including the labor market effects of incarceration, reaching mixed conclusions. In this paper, we trace the labor market trajectories of incarcerated vs. non-incarcerated individuals in the U.S. during the first major increase in incarceration, from 1870-1940. Using data on millions of historical inmates from census and state prison ledgers, we provide new descriptive statistics on the evolution of incarceration in the U.S. We then provide estimates of the incarceration penalty, the difference in labor market outcomes such as occupation, earnings, employment, and labor force participation between incarcerated and non-incarcerated men across this full time period. We find a penalty in occupation-based income scores ranging from 5-15% over the period, with the penalty increasing over time as incarceration rates increased. Using double-linked samples and a linked sample of brothers, we show the penalty remains even after controlling for observed and unobserved measures of family background. Finally, we estimate a "pseudo" event study around year of prison admission for state prisoners linked to census and show that prior labor market trajectory cannot explain the penalty.

Keywords: Incarceration, labor markets, economic history
JEL Classification: N11, N12, J31, K42

*Desmond Ang: Harvard University. Email: desmond_ang@hks.harvard.edu; Ellora Derenoncourt: Princeton University. Email: ellora.derenoncourt@princeton.edu; Kyle Hancock: Princeton University. Email: kh6458@princeton.edu; Jing Wu: Princeton University. Email: jw73@princeton.edu. We thank Ran Abramitzky, Leah Boustan, Alexander Egol, Ilyana Kuziemko, and seminar participants at UC San Diego, Stanford University, the Paris School of Economics, and in the Economics of Racism series for many fruitful discussions and feedback. We also thank all the participants of the NBER Summer Institute Development of the American Economy program meeting and the Stanford Institute for Theoretical Economics.

1 Introduction

The U.S. has both the highest incarceration rate and the largest prison population worldwide. How has this practice affected the evolution of inequality in the United States, both historically and in the present? An active social sciences literature seeks to understand the social and economic impacts of contemporary U.S. criminal justice policy, including the labor market effects of incarceration, yet the conclusions in this literature remain mixed (Pager, 2003; Western and Pettit, 2010; Mueller-Smith, 2015; Doleac and Hansen, 2020; Garin et al., 2023).

We study this question in the context of the historical United States, when incarceration rates experienced a dramatic increase from the late 19th century to the early 20th century. We follow the differential labor market trajectories of incarcerated and non-incarcerated individuals over a 60-year period, providing estimates of the "incarceration penalty," defined as the difference in occupation-based earnings scores, income, and labor force participation between incarcerated and non-incarcerated men from 1880 to 1940.

To do so, we use data on millions of historical inmates identified in full-count censuses and administrative state prison data between 1870-1940. To identify incarcerated individuals in census, we apply approaches developed by Abramitzky, Boustan, Jácome, et al. (2023) and Eriksson (2020) to each census from 1870 to 1940. This method identifies incarcerated individuals using original census enumerator entries in the occupation, relationship-to-household-head, and grouped quarters fields. This generates one of our core samples of incarcerated individuals. We use this sample to generate new descriptive facts about the incarcerated population and incarceration rate over time and to estimate a long-run labor market penalty associated with incarceration by linking incarcerated and non-incarcerated individuals from one census to the next.

Our second core sample of incarcerated individuals comes from historical state inmate ledgers. The ledgers consist of individual records of prisoners admitted into state prison over time. We link these records to census and provide an estimate of the incarceration penalty for this sample.

Our main specification estimates the association between past incarceration and occupation-based earnings scores in a linked sample of men with controls for literacy, age, farm status, birth state, and county of residence. We also include controls for a proxy measure of parental income, the average parental occupation score among individuals of the same cohort and first name. Our estimates of the incarceration penalty range from around 5% in 1880 and 1900 to 10-15% by 1930 and 1940 in census and from 10-20% in the inmate ledger sample between 1910 and 1940. Across the entire period, we find that the incarceration penalty is higher for white individuals than for non-white individuals in the census sample, which includes inmates of jails, but lower than the non-white penalty in our sample of state prisoners.

We examine the penalty associated with incarceration on a variety of other labor-market related

outcomes, including labor force participation, occupational choice, refined income score measures, and annual income (1940 only). We find a negative association between past incarceration and labor force participation. However, our overall penalty estimate is not solely driven by the extensive margin—past incarceration is associated with reduced earnings scores even among those with a gainful occupation. We further find that past incarceration is associated with a 26% fall in annual income in 1940. Our estimates of the penalty are not sensitive to the inclusion of controls for geography or parental income proxies, suggesting that neither geographic mobility nor our coarse measure of family background can explain the reduction in earnings associated with past incarceration.

We then test for selection into incarceration along two dimensions: family background and prior labor market trajectory. To assess the role of negative selection based on family background, we construct doubly-linked samples of individuals who we observe in their childhood census, in addition to the census where we measure incarceration status and the subsequent census where we measure labor market outcomes. Directly controlling for parental occupational earnings scores does not change our estimate of the penalty, helping us rule out selection on this measure of family background as the reason for a negative incarceration penalty.

Further, we construct a double-linked sample of brothers, allowing us to estimate the penalty in a specification with childhood household fixed effects, thus allowing us to account for selection on unobservable family background characteristics. We find that controlling for family fixed effects does reduce the penalty. However as the baseline penalty is larger in the double-linked brothers sample, we recover penalty estimates that are strikingly close to our baseline estimates from the single linked sample when we control for family fixed effects in the brothers sample. Thus, even conditioning on identical childhood household, there remains a significant penalty associated with past incarceration, particularly in the later years of our period of study.

To test for selection into incarceration by prior labor market trajectory, we use a pseudo-event study around year of prison admission. Specifically, we link state prisoners between two census decades to the previous census and estimate the association between occupation-based earnings and incarceration, separately by year until prison admission. We also estimate the association between past prison admission and occupation-based earnings in the next census, separately by year since prison admission. The results show that future incarceration does not have a significant assocation with past earnings and this relationship does not change with year until prison admission. After prison admission, incarceration status is associated with a large immediate drop in earnings scores which becomes smaller over time but plateaus at a lower level of earnings score relative to pre-incarceration estimates. These results are consistent with no selection based on prior occupational and labor market trajectories into incarceration.

We discuss alternative interpretations of the penalty, including the direct human capital effects

of the incarceration experience, changes in the return to human capital over time, and changes in the effect of criminal history on one's labor market prospects. Finally, we compare our estimates of the penalty to modern estimates. Although these latter have a wide range due to different samples and estimation approaches, we find that our historical penalty estimates are in the ballpark of estimates of the penalty for a much later period (1980s and 1990s).

Our work relates to a growing quantitative literature on historical incarceration in the U.S. Recent work by Wu (2024) documents the responsiveness of incarceration rates to economic conditions in Missouri between 1880 and 1930, showing that economic booms decrease White incarceration rates but do not affect Black incarceration rates, which are instead correlated with political factors and racial violence. Abramitzky, Boustan, Jácome, et al. (2023) documents differences in the incarceration rates between immigrants and the U.S.-born from 1850 until the present day, showing that incarceration rates for immigrants were initially higher than the U.S.-born before 1870, but converged with and then fell below U.S.-born white men's incarceration rates during the late 19th and 20th centuries. Eriksson (2020) shows that the construction of schools for Black children in the Jim Crow South lowered incarceration rates of exposed cohorts of Black men. Migrating North during the first Great Migration, by contrast, increased Black men's probability of incarceration relative to brothers who did not migrate (Eriksson, 2019). Relatedly, Muller (2012) finds that Black migration into the North during this period accounts for at least part of the rise in the racial disparity in incarceration rates between 1880 and 1950. We contribute to this existing quantitative literature on historical incarceration by studying the association between past incarceration and labor market outcomes, a phenomenon not studied by the prior literature. We also document novel descriptive facts on incarceration rates by state and demographic group.

A large literature in sociology and a growing literature in economics examines the relationship between criminal justice history and labor market outcomes in the more modern period.[1] Agan et al. (2022) use IRS tax return data linked to criminal court records data from several jurisdictions to study the effects of first charges on labor market outcomes including tax filing and W-2 earnings, finding a near 50% drop in the latter after a first charge. Garin et al. (2023) use judge IV designs as well as sentencing law changes in North Carolina and Ohio to study the effect of incarceration on W-2 earnings and find no effect, with little earnings to begin with in the untreated group (those not sentenced to prison or with exogenously shorter sentences).

We contribute to this literature by studying the relationship between past incarceration and earnings in the historical U.S. Advantages of our context include the ability to leverage complete count census data that allows us to track the outcomes of incarcerated people across the entire country and to examine effects on self-reported occupation and wage income (as opposed to earnings re-

[1] A non-exhaustive list of studies in sociology related to incarceration and economic inequality includes Pager (2003) and Western and Pettit (2010).

ported on tax forms). We are able to study long-run effects of past incarceration and to examine the effects of incarceration on labor market outcomes over a time period of rapid increases in the U.S. incarceration rate and rapid changes in the U.S. economy. We're able to provide insights on the evolution of the gap in earnings between those with past incarceration experience and those without over a 60-year period, from 1880-1940.

The paper is organized as follows. Section 2 provides historical background on the evolution of incarceration in the U.S. and presents new descriptive facts of the geographic and demographic variation in incarceration from 1870-1940. Section 3 describes our procedure for estimating the incarceration penalty and presents and discusses our penalty estimates. Section 5 concludes.

2 Evolution of incarceration in the U.S., 1780-1940

In this section we provide an overview of the history of incarceration in the U.S., starting with the period between the American Revolution and the Civil War. We then provide a quantitative, descriptive history of incarceration after the Civil War using our counts of the incarcerated by demographic group and geography. We begin by describing our method for classifying individuals as incarcerated using information from the restricted full count U.S. census. We then describe the evolution of the incarcerated population both in terms of levels and as a rate of the total population, and we compare our counts to those from historical correctional statistics published by the Bureau of Justice ("BJS").

2.1 Early incarceration in the U.S., 1780-1870

Following the American Revolution, the widespread colonial reliance on public corporal punishment was done away with, and local jails were supplemented and sometimes replaced by penitentiary houses which featured "home-styled" environments with communal living, working, and lax security (McLennan, 1999).

At the turn of the 19th century, prevailing philosophical frameworks broadly identified reform as a function of the carceral system, and fortified jails featuring individual cells began to be constructed. Northern cities led these innovations, and by 1820, a predominant philosophy of incarceration was established at Auburn Penitentiary in New York, where incarcerated people were required to maintain silence, live in solitary confinement, and work during the day in small workshops.

These innovations reflected a pervasive religious sentiment that silent introspection and obedience to the strict rule of wardens would prepare incarcerated individuals for re-entry into society. The so-called "Auburn System" spread widely throughout the North in subsequent decades. Mean-

while, adoption of the System and construction of prisons in the South lagged behind, as southern states relied more heavily on older systems of corporal punishment and local county jails (McKelvey, 1936).

The postbellum period saw dramatic changes in the nature of incarceration that rendered the Auburn System defunct. Burgeoning incarceration rates made the isolation of incarcerated individuals prohibitively expensive. Progressive-era reforms professionalized and consolidated control of jails and police, while the New Penology movement, using the lens of nascent social science to analyze the causes of crime and recidivism, swayed the opinions of legislators against the "despotic model" of prior prisons (McLennan, 1999) and toward more social interaction between those incarcerated. Lastly, technological advancements in mechanization and increasing industrialization offered states a way to recover the rising costs of incarceration and even profit from it by putting incarcerated individuals to work in factory-like environments, and the related system of "convict-leasing" characterized incarceration in many Southern states (Friedman, 1994).

Below, we characterize the growth in incarceration in the postbellum period, presenting incarceration counts and rates overall as well as by demographic group and geography. The descriptive results in this section are based on our counts of the incarcerated population using the method described below. We compare our estimates to official historical corrections statistics from BJS as well as to counts from Abramitzky, Boustan, Jácome, et al. (2023), whose method we build on in order to identify the incarcerated population.

2.2 Identifying the incarcerated population in the full-count U.S. census

In addition to the Bureau of Justice, the U.S. Census Bureau also counted the population residing in correctional institutions, and microdata from the historical census are available from 1850 to 1940. A core part of our methodology depends on identifying incarcerated individuals in any given census year[2] by leveraging rich information from census (Ruggles, Flood, et al., 2021). All residences in Census were categorized into households, group quarters, or unoccupied units. Group quarters are institutional or other group living arrangements besides households and include, but are not limited to, correctional institutions, mental institutions, military barracks, and college dormitories.

Our core approach for identifying incarcerated individuals in the historical U.S. closely follows the approach developed by Eriksson (2020) and elaborated by Abramitzky, Boustan, Jácome, et al. (2023). At the heart of this approach is making use of original census enumerator entries into key fields–such as occupation and relationship to household head–that when combined with group quarters status, reveal the individual to be an inmate of a correctional institution. These original census enumerator entries are available in the restricted version of the IPUMS full count census

[2]Except the 1890 census, for which the microdata were lost in a fire.

data available on the National Bureau of Economics Research ("NBER") server. We elaborate the variables used for identifying the incarcerated and the step-by-step procedure below.

First, we include individuals whose relationship to the household head indicates that they were incarcerated. This includes people who have the following words in their relationship string variable: "Prisoner," "Convict," or "Jail." We also exclude individuals whose relationship string variable includes words that indicate an employment relationship with the institution, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." This is due to the fact that many staff at correctional facilities live within the prison but in separate housing in the historical period we study (Morris and Rothman, 1998). We further exclude individuals whose relationship indicates a familial relationship, such as "Wife," "Daughter," or "Son." Second, We include individuals whose relationship to the household head string includes both "inmate" and words that indicate a correctional institution: "Jail," "Prison," "Penitentiary," "Reformatory," or "Correction." This is to distinguish incarcerated people from inmates in other forms of group quarters such as mental institutions. Third, we turn to the group quarter variable if the information from the relationship to the household head is not sufficient. For individuals whose relationship to the household head is "inmate" (without any additional words) or a missing value, "X," or a "*", we classify them as incarcerated only if their group quarter string variable contains words that indicate correctional facilities, such as "Prison," "Jail," "Penitentiary," "Reformatory," "Correction," "Convict," "Delinquent," and "Penal."

We repeat a similar procedure using the occupation string variable instead of the relationship to household head to identify more incarcerated individuals. The first two steps are the same. For the third step, we use group quarter variables if the occupation string indicates the lack of occupation ("No occupation," "without occupation," "None," or "Nothing") or if the occupation string is missing ("X," "*").

Finally, we classify individuals as incarcerated when it can be inferred by combining more than one variable. For people whose relationship to the household head string is a single word "Inmate," we classify them as incarcerated if their group quarter type indicates correctional institutions. For people whose relationship to the household head string is missing, we classify them as incarcerated if their relationship to the household head type is "Institutional inmates." Finally, for people whose occupation string is a single word "Inmate," we classify them as incarcerated if their group quarter type indicates correctional institutions and their relationship to the household head string doesn't suggest otherwise by including words other than "Inmate."[3] Such adjustments make sure we include as many incarcerated people as possible while still being conservative with regard to

[3]We further make the restriction so that if an individual whose relationship string is missing is designated as incarcerated only under one of the two methods, we only recognize them as incarcerated if their "relate" variable code categorizes them as institutional inmates or if the "relate" variable string corresponds to the term "Inmate."

Type I errors.

Based on data availability in different censuses, we adjust the procedure accordingly. For example, censuses between 1850 and 1870 do not include the relationship to household head string variable, so we only conduct the classification using the occupation string variable and include individuals if their relationship code is "institutional inmates" and their group quarter type corresponds to correctional institutions. The details of the classification process for each census are explained in Appendix A.

We face two types of potential errors in the classification of incarcerated individuals. First, some non-incarcerated people might be classified as incarcerated (Type I error); second, some incarcerated people might be classified as non-incarcerated (Type II error). Because the incarceration rate was well below 1% throughout the period, Type I error imposes a bigger threat to the empirical findings, because even a very small Type I error rate will contaminate the classified incarcerated group by a big margin[4]. Because of the inconsistent definition of group quarter variables over time (Eriksson, 2020) and the common practice of correctional staff living onsite, using group quarter status alone in identifying incarcerated individuals is likely to lead to substantial Type I error, threatening the validity of our empirical design. Thus, we take a conservative approach and only classify individuals as incarcerated if the original string of the relevant variables directly indicated their incarceration status. When one variable is not sufficient, we combine information from multiple variables to verify their incarceration status.

Figure 1 shows the total number of incarcerated people identified with our method in the Census. We are able to replicate the sample used in Abramitzky, Boustan, Jácome, et al. (2023). The level and trend of our series align closely with the official statistics by the Bureau of Justice Statistics, further strengthening our confidence in the methodology.

2.3 Descriptive patterns in U.S. incarceration, 1870-1940

Figure 1 Panel (a) presents our counts of the incarcerated population from 1870 to 1940 (dashed red line) based on the methodology described in Section 2.2. The figure also presents comparison numbers that come from the BJS publication "Historical Corrections Statistics in the United States, 1850-1984" (Cahalan, Parsons, et al., 1986) (black triangles and squares). Comparison numbers for all classes of the incarcerated, including those sentenced and not yet sentenced and across adult and juvenile facilities, are only available for 1880, 1923, and 1940. As we are not able to distinguish between adult and juvenile facilities or the sentenced and unsentenced population, these figures are the closest comparison figures from BJS.

[4]For example, in a world where group A and B accounts for 1% and 99% of the population respectively, misclassifying 1% of group B will lead to 49.7% false positive rate in classified group A (0.99/(1+0.99) = 49.7%).

Our count of incarcerated closely aligns with the BJS count for 1880 (60,303 vs. 69,228) and 1940 (343,218 vs. 317,168). The 1923 BJS count (163,889) is consistent with the trend we estimate between 1920 and 1930.

Our figures for 1910 are likely an undercount, judging by the number of sentenced juvenile and adult inmates, a lower bound for the total incarcerated population in 1910. We identify 71,738 individuals as incarcerated using the method described above, but as the relationship variable is missing in this year, we are not able to capture all inmates of correctional institutions. We also demonstrate that we exactly replicate counts of the incarcerated by Abramitzky, Boustan, Jácome, et al. (2023) when we restrict to their age and nativity sample (see the blue diamonds for the Abramitzky, Boustan, Jácome, et al. (2023) sample and the dashed green line for our replication).

Figure 1 Panel (b) plots our estimated incarceration rates for different age groups. In the solid black line, we plot the incarceration rate among all age groups; the dashed green line shows the incarceration rate for 15-44 year-olds; and the dashed red line shows the incarceration rate for 16-26 year-olds, the age range for the sample for which we estimate the incarceration penalty. Rates are substantially higher among 16-26 year-olds who make up a disproportionate share of the incarcerated.

Patterns by demographic group Figure 2 plots the rates among men by racial group. We focus on men here as they make up the majority of the incarcerated population. In all years, the rates have been higher for non-white throughout the whole period, with a major increase in the incarceration rate for non-white men between 1870 and 1880. The increase from 1920 to 1940 was also higher for non-white men.

Geographic differences in incarceration rates We provide new evidence on spatial patterns in incarceration rates by racial group using our counts of the incarcerated. We focus on state-level rates due to the fact that below the state level, incarceration rates may reflect the presence of a local jail or correctional institution. We examine state-level variation in incarceration rates over time and separately by racial group. Figure 4 presents state-level maps of incarceration rates for white men from 1870 to 1940. Rates are typically low in the Upper Midwest and high in the South East, Northeast, and West. For non-white men, of whom Black men form the vast majority, rates are particularly high in the border states and Midwest from 1900 to 1940. In 1870 and 1880 rates were also high in parts of the South.

3 The incarceration penalty from 1880-1940

In this section, we estimate the incarceration penalty based on two samples of incarcerated individuals from different sources. We first link the identified incarcerated individuals in Censuses as described in 2.2 to the next Census and observe their earnings gap with people who were not incarcerated. We then introduce the administrative records on incarcerated from prison ledger to directly observe individuals who were admitted into prisons and link them to the following Census for their labor market outcomes. Furthermore, we apply alternative linking methods and labor market outcome measures to examine robustness and heterogeneity in the incarceration penalties.

3.1 Estimating the incarceration penalty in Census

Linked samples To trace individuals' labor market outcomes after incarceration, we link the census in which we identify incarceration status to the next census in which we observe individuals' labor market outcomes ten years later, using automated record-linking methods (Abramitzky, Boustan, Pérez, et al., 2020). This name-based method links one individual in one census to the next when there is a unique match based on the phonetic sounds of their names, year of birth, and state of birth, conditional on their race. The method treats phonetically similar names as equivalent (NYSIIS), and allows for name misspellings due to the possible errors in both census enumeration and records digitization. In robustness checks in Section C, we utilize alternative linking methods to identify individuals across censuses.

We make several important restrictions to our analysis sample. First, we only include men in our estimation. Because women frequently changed their last names upon marriage, name-based methods fail to match most married women. Moreover, the incarcerated population was predominantly men. Second, we exclude immigrants from the sample, who exhibited different patterns in their economic mobility and criminal justice involvement in this period (Abramitzky and Boustan, 2022; Abramitzky, Boustan, Jácome, et al., 2023), and instead focus on the U.S.-born. Third, we focus on the cohorts who were between ages 16 and 26 at the first census when we observe their incarceration status, so we can observe their labor market outcomes when they are 26-36 years old. The same 10-year age band in a life cycle allows for direct comparison across decades because every birth year cohort will be included exactly once in the analysis. We chose 16-26 because most incarceration took place in early adulthood, and labor market outcomes during this period tended to stabilize after age 25. This choice enables us to capture most of the incarceration spells in the census and observe representative long-run outcomes in the following census.

Representativeness of the Linked Sample Because information used in census linking, such as names and race, are not randomly distributed in the population, the selection into the linked sample can potentially bias the findings. We examine how representative the linked sample is compared to the general population. We investigate selection into linking along three key socioeconomic variables: age, literacy, and immigrant status. Figure 5 shows the comparison over the period we study, for both the incarcerated group and the non-incarcerated group. The linked group tends to be younger, have more education, and less likely to be born outside the U.S. This pattern is consistent with past work using name-based census linking (Abramitzky, Platt Boustan, et al., 2021; Ager, Boustan, and Eriksson, 2021). However, the extent of selection doesn't change over time, giving us confidence that the change in selection does not pose a threat to the trend in the incarceration penalty we estimate. The summary statistics tables in Appendix C show that this pattern holds across racial groups and over time. In Section 3.4, we further use alternative linking methods to address the concern of the bias introduced by linking censuses.

Main Specification We estimate the following equation to capture the relationship between past incarceration and individuals' occupation scores:

$$OccScore_{it} = \alpha + \beta Incarcerated_{i,t-10} + X'_{it}B + \epsilon_{it} \tag{1}$$

Where i denotes the individual and t denotes census year. $OccScore_{it}$ is the individual's occupational income score observed in Census t, which reflects the median total income of a person in that occupation in 1950 and is a common measure for socioeconomic status in historical Censuses[5] (Ruggles, Flood, et al., 2021). $Incarcerated_{i,t-10} \in [0, 1]$ is the individual's incarceration status in census $t-10$ as identified in Section 2.2. X'_{it} represents a vector of control variables that can affect people's labor market outcomes, including age and a quadratic in age, literacy, residence on a farm, county of residence in census year t, and birth state. We then normalize the incarceration penalty to be in percentage terms relative to the average occupational income score. This allows for consistent comparison of our penalty estimates across decades and groups.

To further control for family background, we include a control the average parental occupational income score associated with individuals' first names, following the methodology in Olivetti and Paserman (2015). The key insight is that the first name provides information about parents' socioeconomic status at the time of the individuals' birth, reflecting preferences stratified across different groups. Although direct information on family background is unavailable in the linked sample between Census t and Census $t-10$, we address this limitation by examining the distri-

[5]For people who were incarcerated in Census t, we assign zero to their occupational income score.

bution of parental occupational scores by first name in Census $t - 20$, when the individuals were in the 6-16 age range. This age range ensures a likelihood of co-residence with parents. We then compute the average parental occupational income score for each distinct first name in the whole population and use them as proxies for family background. This method allows us to control for family background in a larger sample than one that links across three censuses.

3.2 Baseline results

Figure 6 shows the estimated incarceration penalty for each Census year. The incarceration penalty increased after 1900 from around 5% to 10-15% of the average occupational income score. Dropping county fixed effects and parental occupational income scores does not change the incarceration penalty, which suggests that the penalty is not attributable to differences in our first-name-based proxy for family background between the incarcerated people and the rest of the population.

Incarceration Penalty by Race There exist persistent racial disparities in incarceration rates in the U.S. after Reconstruction (Muller, 2018; Poyker, 2019). To examine the differential labor market impacts of incarceration histories by race, we re-estimate the incarceration penalty separately for the white and non-white populations. Figure 7 shows that while the two racial groups exhibit similar trends in incarceration penalty over time, the incarceration penalty is higher for the white population than for the non-white population in every Census, and the difference is statistically significant in most years.

3.3 Other Labor Market Outcomes

The prior analyses of income proxies are limited by the fact that individuals with non-occupational values for the occupation variable are given zero income, meaning those estimates of the incarceration penalty combine effects on labor force participation, effects on earnings, and effects on occupational choice. We next move to separately analyze the incarceration penalty along these dimensions for the years in which we have the requisite data.

Labor force participation Data for the modern concept of labor force participation are only available in the 1910, 1930, and 1940 censuses; however, individuals across all censuses were able to report non-occupational responses to enumerators in lieu of an occupation, and they were not required to be currently employed to report an occupation. With these facts in mind, we construct a labor force participation variable across the entire sample that equals 1 when an individual reports an occupation, and 0 if they reported a non-occupational response (e.g. homemaker/housewife, student, "other non-occupational response").

Figure 8 mirrors Figures 6 and 7, showing results from estimating Equation 1 using our constructed labor force participation variable. Figure C5 plots the estimates using the official labor force variable, with very similar results in the years for which we have data.

Labor force participation penalties are much smaller than the occupation-based income score penalties, hovering around 2 percentage points in 1910 and 1930 before falling to 5 percentage points in 1940. By race, the penalty is larger for the non-white population in 1910, but afterwards is similar across racial groups.

Occupation score conditional on labor force participation Figure 9 plots estimates of the penalty on the occupation score, conditioning on having an occupational response. Consistent with the relatively small estimates for labor force participation, Panel (a) shows that the penalty conditioning on nonzero occupation score is very similar to the overall penalty; however, there is no additional deepening of the penalty in 1940. In fact, the results by race show that the non-white penalty decreases in 1940.

Farm ladder Lastly, we examine the association between incarceration and an individual's position of the "farm ladder." The outcome conditions on reporting any occupation related to farming: it then equals 1 when an individual reports an occupation belonging to the "Farmers" group of occupations (farm owners, tenants, or managers) versus the "Farm Laborers" group of occupation codes. Figure 10 shows that incarceration is associated with a lower position on the farm ladder, and this association becomes larger throughout the 20th century. The penalty is similar by race until 1930, during which the non-white penalty reverses course while the white penalty continues downwards, ultimately resulting in a significant difference in the penalties by 1940.

3.4 Robustness and Heterogeneity in Incarceration Penalty

Incarceration Classification Because we do not observe full records of prison admissions and releases, we infer incarceration status based on information in full-count Censuses. However, censuses only provide a snapshot of every incarcerated individual at the time of census enumeration. If someone's incarceration spell was fully in between two censuses, then we will mislabel them as never-incarcerated.

We argue that this limitation should not unduly affect our estimates of the penalty for two reasons. First, using the snapshot instead of the complete flow of prisons undercounts the size of the incarcerated population. But because only a tiny proportion of the total population was ever incarcerated, such misclassification does not meaningfully bias the estimation. Second, using external full-count prison records in Missouri utilized in Wu (2024), we examine the changes in

incarceration spell length for our study period. As shown in C1, there was no significant change in the mean and median of sentence lengths, suggesting that the level of bias didn't change over time. Therefore, as long as states did not manipulate the prison population because of censuses, both the trend and the magnitude of the estimated incarceration penalty can be properly interpreted.

Alternative Linking Methods One concern with linking individuals across censuses using names is that individuals may be matched to the wrong person with similar attributes. If these linking errors systematically take place, the estimated incarceration penalty could be biased. To address this issue, we utilize two alternative methods.

First, we use Census Tree linking developed in Price et al. (2021) and Buckles et al. (2023). It uses family links provided by users of genealogy websites to train machine-learning methods and create new links between censuses. Second, we use the Multigenerational Longitudinal Panel (MLP) used in Ruggles, Fitch, et al. (2019). It develops a two-step probabilistic model that uses both demographic information as well as cohabitation within the same household across censuses to identify people across censuses. Figure C2 shows the estimated incarceration penalty using the two alternative linking methods with the same specification as in Section 3.1. Other than 1880, the two methods generate results very similar to our main name-based linking methods, further validating the robustness to the linking method.

Alternative Income Measures To supplement our usage of the standard occupation-based income score, we present estimates of the penalty with two alternative measures of income: Lasso-adjusted Industry, Demographic, and Occupation (LIDO) scores and real annual wages for 1940. Figure C3 plots the penalty for the full sample and by race, this time using LIDO scores (Saavedra and Twinam, 2020). LIDO is a similar outcome to the occupation score, as it assigns median earnings from the 1950 census to individuals in earlier censuses, but in addition to occupation, it uses information about an individuals' race, sex, state of residence, age, and industry.

Saavedra and Twinam (2020) provide evidence that this extra information results in a better prediction of income, and that its usage of additional demographic information allows it to uncover within-occupation racial and gender gaps. This is important in providing a robustness test of our overall penalty and racial gap estimates, as it is possible that differential wage premia within occupations due to industry, place, or demographic characteristics could bias the penalty estimates in either direction.

The results are smaller in magnitude but show a similar deepening of the penalty over time, except for the sharp drop between 1930 and 1940. The racial dynamics are similar as well, showing an initially insignificant penalty for the non-white population that grows while still being smaller than the white penalty.

Figure C4 shows the results using annual earnings in the year prior, which includes zeros, as the outcome variable. This variable is only available for 1940. The penalty is larger then the baseline penalty in Figure 6, as the formerly incarcerated are estimated to earn around 26% less than the control group. Panel (b) shows the results by race, with the same qualitative finding as all other estimates: a smaller penalty for the non-white population.

In summary, these alternative measures of income deliver similar results to our baseline estimates of the earnings penalty using the occupation score, giving us confidence that idiosyncrasies in the construction of the standard occupational income score are not driving the results.

Geography of Incarceration Penalty Given the substantial variation in incarceration rates across states in the U.S., we examine geographic heterogeneity in the incarceration penalty. We separately estimate the incarceration penalty for the local population in each state and region for comparison. As shown in Figure 11, there were changes in the relative magnitude of incarceration penalty across regions. Nevertheless, a few patterns arise: First, the Northeast consistently had a high incarceration penalty; Second, the incarceration penalty in the Midwest and the South remained unchanged over the period, for both white and non-white populations. Third, the West saw a large increase in the incarceration penalty in the period.

Figure 12 further shows the changes in incarceration penalty by state. At this more granular level, the regional trends identified in Figure 11 are present, as well as the additional insight that variation in the penalty across states decreased over time. The figure also confirms that, contrary to common belief, the South actually had a lower incarceration penalty compared to states in the Northeast and the West. The coastal states faced increasingly negative relationships between past incarceration and economic outcomes.

How do incarceration rates at the state level relate to state-level estimates of the incarceration penalty? Figure 13 shows scatter plots with lines of best fit for 1910 through 1940, correlating states' incarceration rates with their incarceration penalties. Note that while the slopes change sign sporadically, are all close to zero and are never statistically significant. In a world without a causal effect of incarceration on earnings, we might expect to see a positive relationship between incarceration rates and the penalty. For instance, if selection into incarceration was such that only those in the bottom of the earnings distribution became incarcerated, an increase in the incarceration rate would lead the marginal incarcerated individual to have higher potential returns in the labor market, thereby decreasing the penalty. Another scenario which might produce a positive correlation would be if worsening economic outcomes increased both crime and incarceration as well as decreased the control groups' earnings, mechanically shrinking the penalty. A lack of correlation does not rule out these stories as possible contributors, but does make it seem less likely they can entirely explain the patterns seen here. With this in mind, we next turn to a more thorough

examination of selection in explaining our results.

3.5 Estimating the incarceration penalty in Prison Ledgers

We further employ newly digitized prison ledger data to estimate the incarceration penalty. The dataset covers administrative records on 1.9 million individuals admitted into correctional facilities between 1800 and 1970. It was collected and digitized from inmate ledgers in 22 states and 4 federal prisons. The variables available are different by state, and to link to Censuses using Abramitzky, Platt Boustan, et al., 2021, the minimal information we need is: full name, birth year, and the place of birth. This limits our sample to 686 thousand individuals in Washington, Texas, Pennsylvania, New York, New Mexico, Nevada, Missouri, Iowa, and California.

To trace individuals' labor market outcomes, We link individuals in the inmate ledger to the first Census following their prison admission via Abramitzky, Platt Boustan, et al., 2021 to obtain their income measures after incarceration, similar to Section 3.1. The linkage rates range between 16 percent to 22 percent for different decades, similar to other papers using the same method. To make the results comparable with the findings using Census data, we make similar restrictions to our analysis sample. We only include native-born men who were between age 26-36 at the next Census after prison admission. This means we are examining the outcomes for the same birth cohort as in the Census data, but we were able to include individuals incarcerated between two censuses who would have been omitted from the incarcerated population identified in the full-count Census.

We estimate the following equation to capture the relationship between past incarceration and individuals' future income:

$$Income_{it_0} = \alpha + \beta Incarcerated_{i,t\in[t_0-10,t_0)} + X'_{it_0}B + \epsilon_{it_0} \tag{2}$$

Where i denotes the individual and t_0 denotes the census year. $Income_{it_0}$ is the individual's income in year t_0, measured as the occupational income scored observed in Census t_0, which reflects the median total income of a person in that occupation in 1950. We used an array of different income measures developed in recent economic history literature for the outcome. $Incarcerated_{i,t\in[t_0-10,t_0)}$ is the individual's incarceration history between year t_0-10 and year t_0 and is measured as whether they are linked to any records in the inmate ledger. X'_{it_0} represents a vector of characteristics that can affect labor market outcomes, including age and a quadratic in age, literacy, residence on a farm, and the county residence in Census year t_0. We then normalize the incarceration penalty to be in percentage terms relative to the average occupational income score of individuals who were never incarcerated. We further estimate the equation separately by race for a consistent comparison across decades and groups.

Figure 17 shows the estimated incarceration penalty for each Census year, separately by race. The incarceration penalty increased from around 10% in 1910 to 20% in 1940 compared to the average occupational income score. The upward trend is similar to the results in the full-count Census, two differences in the results arise: firstly, the magnitude is bigger from the inmate ledger sample. Considering the fact that the incarceration spell is closer to the labor market outcome for individuals in the inmate ledger sample on average, this fact is hardly surprising. However, we also find that the incarceration penalty is larger for the nonwhite population and is significantly negative throughout the period, which is different from the Census sample.

4 Examining the role of selection in incarceration penalty

There is a great deal of debate in the literature on modern effects of incarceration on earnings regarding the magnitude, or even existence, of a causal impact. Garin et al. (2023) find using data from North Carolina and Ohio, that despite a dip in earnings stemming from incapacitation during the year of incarceration, incarceration has no effect within five years of the incarceration spell. By contrast, other work documents significant causal effects of incarceration on labor market earnings (see Mueller-Smith (2015)). In this section, we examine the role of the selection into incarceration in explaining the historical incarceration penalty in the U.S. We first build double-linked samples across three censuses to directly control for childhood environment and to use siblings as comparison groups to isolate the causal impact of incarceration on long-run labor market outcomes. We then link the inmate ledger directly to the census prior to incarceration to estimate the income trajectory leading up to incarceration. Both exercises show that the incarceration penalty is unlikely to be explained by the negative selection into incarceration.

4.1 Double-Linked sample to control for selection

Building on the linked sample in Section 3.1, we further use the same name-based methods to link individuals to their childhood Census when they were 6-16 years old. Because of the limitations in linking across censuses, this further limits the sample size, but double-linking offers the advantage of being able to directly observe childhood environment and familial relationships. We focus on 1920-1940 because small samples prohibit the estimation in 1910, and the double-linked sample between 1880-1870-1860 has too small an incarcerated group to conduct meaningful analysis.

Directly Controlling for Family Background We estimate the incarceration penalty and include the direct measure of parental occupation-based income scores in the specification:

$$OccScore_{it} = \alpha + \beta Incarcerated_{i,t-10} + X'_{it}B + \gamma ParentalOccScore_{i,t-20} + \epsilon_{it} \quad (3)$$

Where i denotes individuals and t denotes census year. $ParentalOccScore_{i,t-20}$ denotes the highest occupational income score within the household in Census $t-20$, when the individual was 6-16 years old. The rest of the specification is the same as in Equation 1.

Panel (a) in Figure 14 shows the estimated incarcerated penalty with and without direct controls of parental occupational income score. The incarceration penalty increased over time, and including family background has no impact of the estimated penalty on occupation score. This suggests little role for selection based on parental income into incarceration, at least after controlling for literary, geography, and farm status in adulthood.

Brothers Comparison There are many unobserved characteristics of the family environment in ϵ_{it} that are not captured by the parental occupational income score. If they are correlated with the possibility of incarceration and the labor market outcome in adulthood, our interpretation of the estimated β would be a combination of the selection on unobservables and the causal impact of incarceration. To get as close as possible to a causal interpretation, we further limit the sample to brothers within the same household to control for unobservable childhood characteristics.

We estimate the incarceration penalty and include childhood home fixed effects in the specification:

$$OccScore_{it} = \alpha + \beta Incarcerated_{i,t-10} + X'_{it}B + \gamma ParentalOccScore_{i,t-20} + \lambda_{j(i,t-20)} + \epsilon_{it} \quad (4)$$

Where i denotes individuals, t denotes census year, and $j(i, t-20)$ denotes the unique household identifier that individual i resided in during census year $t-20$. The rest of the specification is the same as in Equation 1. The family fixed effect $\lambda_{j(i,t-20)}$ ensures that the estimated β is derived from the outcomes difference between brothers growing up in the same household who were both 26-36 years old in year t, and some of them were incarcerated in yer $t-10$ while the other ones in the family were not.

Panel (b) in Figure 14 shows the estimated incarcerated penalty using the brothers' sample. The incarceration penalty increased over time, but compared to the main specification where we do not include family fixed effects, the incarceration penalty did shrink. In fact, there was no significantly negative penalty in the labor market for people incarcerated in 1910. These results have two interpretations: first, they may document some degree of negative selection on family unobservables into incarceration. The dynamic stability of our estimates, however, suggests that the relative importance of this selection did not change over time. An alternative interpretation

is that there may be spillovers from one brother's incarceration spell onto other members of the family, including the brother whose earnings serve as a comparison point. In the case of such negative spillovers, the brothers comparison will understate potential causal effects of incarceration on earnings.

Characteristics of the Double-Linked Sample The double-linked sample is not a random subsample of the total population, introducing potential bias in penalty estimation. Figure 15 shows the comparison between double-linked sample and the full sample. Similar to Section 3.1, the double-linked sample tends to be younger, more literate, and less likely to be foreign-born, and the extent of selection is even higher than in the main specification. Appendix C further shows a summary statistics comparison for the double-linked sample.

Method Comparison with Fixed Sample One challenge in comparing the results between different methods is that the estimation sample also changes. By linking an additional census, the sample size is greatly reduced, and the potential selection into the linked sample can confound our interpretation on selection into incarceration. In Figure 14, the magnitude of the estimated incarceration penalty is much bigger than in the main specification as in Figure 7. We address the issue by applying different estimation methods on the same sample of brothers that are all linked across three censuses.

Figure 16 shows the estimated incarceration penalty on the sample size but with different methods as specified in Equation 1, 3, and 4. The magnitude of the estimated incarceration penalty using parental occupational income score proxy remains higher than in Figure 7, indicating that linking another census indeed introduced selection into the sample. However, we yield similar conclusions from comparing different methods on the same sample: directly controlling for parental occupational income score has no impact on the result, while brothers' comparison significantly shrinks the incarceration penalty. This suggests that the selection into incarceration was, if anything, on family unobservables rather than on family observable characteristics. However, the incarceration penalty still increases over time and is significantly negative. Under the interpretation of the brothers comparison as a causal estimate, we find a negative impact of incarceration on future labor market outcomes and that this effect grew over the first half of the 20th century.

4.2 Income Dynamics in the Inmate Ledger Sample

In the full-count Census sample, we address the issue of selection by controlling for family background in multiple ways. However, because Censuses only happen once every ten years, the individuals in our sample were too young in the Census before incarceration for us to directly observe

their own socio-economic characteristics. It is possible that individuals encountered other negative shocks right before incarceration that were not captured in the family information ten years prior. We approach this issue by linking individuals in the inmate ledger dataset to the previous census before prison admission to examine the dynamics of their occupational income scores in the years leading up to incarceration. We estimate the following equation:

$$Income_{it_0} = \alpha + \beta Incarcerated_{i,t_0+t} + X'_{it_0}B + \epsilon_{it_0} \tag{5}$$

Where i denotes the individual and t_0 denotes the census year. $Income_{it_0}$ is the individual's income in year t_0, measured as the occupational income score observed in Census t_0, which reflects the median total income of a person in that occupation in 1950. $Incarcerated_{i,t_0-t}$ indicates whether the individual was admitted into prison in $t_0 - t$ and is measured as whether they are linked to any records in the inmate ledger. X'_{it_0} represents a vector of characteristics that can affect labor market outcomes, including age and a quadratic in age, literacy, residence on a farm, and the county residence in Census year t_0. We limit the sample to people who were either incarcerated in year $t_0 + t$ or were never incarcerated throughout the decade in order to exclude individuals who were to be incarcerated in other years from the control group, and β is the penalty for being incarcerated t years later. We pool all decades together for statistical power but include decade fixed effects in the regression to control for trends over time.

Following a similar logic, we utilize the variation in the time lapsed between prison admission in the inmate ledger and labor market outcomes observed in the Census in order to estimate the incarceration penalties every year after incarceration, under the assumption that there is no difference in overall inmate composition over time within a decade. For every year between prison admission and the following Census, we then estimate the following equation:

$$Income_{it_0} = \alpha + \beta_t Incarcerated_{i,t_0-t} + X'_{it_0}B + \epsilon_{it_0} \tag{6}$$

$Incarcerated_{i,t_0-t}$ indicates whether the individual was admitted into prison in $t_0 - t$, and all other specifications are the same as in Equation 2. We limit the sample to people who were either incarcerated in year $t_0 - t$ or were never incarcerated throughout the decade so as to exclude individuals who were incarcerated in other years from the control group, and β_t is the penalty for having been incarcerated t years ago. We pool all decades together for statistical power but include decade fixed effects in the regression to control for trends over time.

We then plot the estimated penalty against the number of years lapsed until the incarceration spells in 5 in the same graph as the dynamics after incarceration as estimated in 6. Figure 21 plots the trajectory of labor market income in a pseudo-event-study fashion. It shows no evidence of a

downward trend in income in the years leading up to prison admission. After prison admission, there is an immediate large penalty of 40% - 60% in occupational income score resulting from incapacitation, and the penalty slowly decreases over time. However, even ten years after incarceration, there is still a 10% incarceration penalty, which is close to the estimates from the Census sample. The gradual recovery and ultimate stagnation suggest that negative selection alone cannot explain the incarceration penalty.

To ensure that the result is not driven by the difference in the analysis sample, we construct a dataset that links inmate ledger records to both the pre-incarceration and post-incarceration censuses for robustness. We estimate the same specifications as in Equation 5 and Equation 6 on the same double-linked sample. Figure C7 shows that the overall patterns remain the same, even though the estimates are noisier due to a smaller sample size.

Additionally, we estimate the incarceration penalty for other labor market outcomes for the mechanism. We use labor market participation, implied from having a positive occupational income score in the Census, as the main outcome instead in Equation 6. Results in 20 show that the penalty for labor market participation exhibits similar patterns as for occupational income scores. There is an immediate large negative penalty after incarceration followed by a gradual decrease of the penalty, but the labor market participation rates for incarcerated people did not fully recover ten years later. When we limit the sample to individuals with non-zero occupational income scores, there is no recovery in income as shown in 20. In summary, the recovery in income after incarceration is only driven by the extensive margin of labor force participation.

4.3 Interpretations of the incarceration penalty

The above section discussed selection as a potential explanation for the negative association between past incarceration and earnings. Here we discuss a number of other interpretations of the incarceration penalty estimates presented thus far.

There are a number of mechanisms through which incarceration may have a direct effect on occupation and earnings. One is through the effect of incarceration on work experience and being out of the civilian labor force for the duration of one's sentence. A penalty driven by time out of the civilian labor force could increase with increasing returns to experience even in the absence of sentence length increases. Alternatively, sentence lengths could be increasing, thus leading to longer spells of no work experience in the civilian labor force.

During this period, convict labor was pervasive throughout the U.S. For some of the incarcerated population, incarceration may not have represented an interruption in work experience. On the other hand, some of the incarcerated may have been employed in sectors they would be unlikely to work in outside of prison, leading to a potential mismatch between skills acquired as convict

laborers compared to skills demanded in the civilian labor market.

A final factor is how criminal records were treated in the labor market. If employers begin to discriminate against workers with past criminal justice involvement, this could result in an increase in the penalty.

One final potential mechanism behind a rising penalty is increased skewness in the earnings distribution and changes in the structure of the economy. As income inequality increased in the early 20th century, the penalty associated with incarceration, whether from selection or direct causes, could increase over time. In ongoing work, we explore the extent to which changes in the occupation-based earnings distribution can explain changes in the penalty between 1920 and 1940.

4.4 Comparison to modern estimates of the penalty

How do our estimates of the penalty associated with past incarceration compare to modern estimates?

Early literature estimating the effect of incarceration on earnings used panel data with fixed effects (Raphael, 2007; Western, 2002) or court records linked to unemployment insurance earnings data (Grogger, 1995; Waldfogel, 1994; Needles, 1996) to compare earnings conditional on employment between the formerly incarcerated and varying control groups. Most studies found significant negative drops in income associated with incarceration. Estimates of the effect of one year of incarceration ranged from 10-16%, and effects of being ever incarcerated results ranged from 2-28%, with heterogeneity by race and comparison group of the studies. As discussed previously, studies using judge IV designs typically find smaller or null effects on earnings or employment (Garin et al., 2023; Kling, 2006; Loeffler, 2013) with the exception of Mueller-Smith (2015) who finds that individuals convicted of felonies earn 0.34 less log income per year incarcerated, with no significant effects for those convicted of misdemeanors.

The panel data approach, which compares the earnings of the formerly-incarcerated to the never incarcerated, is the most similar to our analysis,[6] and our estimates from the latest period (being incarcerated in 1930) are on the lower-to-middle end of the range of penalties estimated in this literature, which is of interest given the increasing trend identified across our estimates.

One departure of our estimates from modern studies is the heterogeneity by race. Research has found stronger effects of incarceration on earnings for Black individuals (Silver, D'Amato, and Wooldredge, 2024; Joseph, 2003; Lyons and Pettit, 2011; Apel and Powell, 2019), contrary to our findings on the non-white population, which almost universally show smaller or equal effects

[6]It is not an identical control group as it is possible that our control group, those aged 16-26 and not incarcerated in decade $t-1$, become incarcerated and are released sometime before we see them again in decade t, but this is presumably a small proportion of the overall control group, at least for the larger samples presented in the double-linked analysis.

relative to the white population.

Lastly, estimates on employment and labor force participation have a wide variance in the literature. Some studies have found similar large and negative short-run effects as on earnings (Freeman, 1991; Waldfogel, 1994), but others have found more muted (Brown, 2019) or even positive effects in the short run that diminish over time (Lalonde and Cho, 2008; Loeffler, 2013), typically due to participation in jobs-training programs during incarceration. We find null effects on labor force participation until the last two decades of our sample, after which we find evidence of long-lasting impact of incarceration on labor force participation.

5 Conclusion

By 1940, incarceration rates in the U.S. reached about 50% of the level they are today and about a third of the 20th century peak in incarceration rates in the early 2000s. The first wave of "mass incarceration" took place between 1920 and 1940, with the U.S. already surpassing rates in similar economies at the time (Cahalan, Parsons, et al., 1986).

What has been the trajectory in outcomes for formerly incarcerated individuals throughout the historical period in the U.S., and what can we learn about the sources of differences in earnings between the formerly incarcerated and those not incarcerated?

In this paper, we document the penalty in labor market outcomes associated with past incarceration experience over a 60-year period in U.S. history, from 1880 to 1940. We appy two techniques to historical full count census data in the U.S. to construct our estimates of the penalty. First, we build on existing recent literature and make use of the fact that original census enumerator entries into fields such as occupation, relationship-to-household-head, and group quarters reveal individuals' status as inmates of correctional institutions. Second, we use links across censuses generated by automated linking methods to match individuals observed incarcerated or not in a given census to the following census where we can measure their labor market outcomes.

Using this approach, we estimate a penalty in occupation-based income scores associated with past incarceration experience that ranges from 5% to 15%. This penalty is robust to inclusions of proxies for parental income and to direct controls for parental income and family fixed effects in double-linked samples and double-linked samples of brothers. These results suggest that selection into incarceration based on family characteristics is unlikely to explain the penalty we estimate.

In ongoing work, we investigate different interpretations of the penalty, including the direct effect of incarceration on earnings through withdrawal from the labor market as well as the changing treatment of criminal justice history in the labor market.

References

Abramitzky, Ran and Leah Boustan (2022). *Streets of gold: America's untold story of immigrant success*. PublicAffairs.

Abramitzky, Ran, Leah Boustan, Santiago Pérez, and Myera Rashid (2020). "Census Linking Project: Version 2.0 [dataset]". In.

Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, Santiago Pérez, and Juan David Torres (2023). *Law-Abiding Immigrants: The Incarceration Gap Between Immigrants and the US-born, 1850–2020*. Tech. rep. National Bureau of Economic Research.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez (2021). "Automated Linking of Historical Data". In: *Journal of Economic Literature* 59.3, pp. 865–918. DOI: 10.3386/w25825.

Agan, AY, A Garin, D Koustas, A Mas, and CS Yang (2022). "The impact of criminal records on employment, earnings, and tax filing". In: *Statistics on Income (SOI) Working Paper*.

Ager, Philipp, Leah Boustan, and Katherine Eriksson (2021). "The intergenerational effects of a large wealth shock: white southerners after the Civil War". In: *American Economic Review* 111.11, pp. 3767–3794.

Apel, Robert and Kathleen Powell (2019). "Level of criminal justice contact and early adult wage inequality". In: *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5.1, pp. 198–222.

Brown, Christian (2019). "Incarceration and earnings: Distributional and long-term effects". In: *Journal of Labor Research* 40.1, pp. 58–83.

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

Cahalan, Margaret Werner, Lee Anne Parsons, et al. (1986). *Historical corrections statistics in the United States, 1850-1984*. US Department of Justice, Bureau of Justice Statistics Washington, DC.

Doleac, Jennifer L and Benjamin Hansen (2020). "The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden". In: *Journal of Labor Economics* 38.2, pp. 321–374.

Eriksson, Katherine (2019). "Moving North and into jail? The great migration and black incarceration". In: *Journal of Economic Behavior & Organization* 159, pp. 526–538.

— (2020). "Education and incarceration in the Jim Crow South: Evidence from Rosenwald schools". In: *Journal of Human Resources* 55.1, pp. 43–75.

Freeman, Richard B (1991). *Crime and the employment of disadvantaged youths*.

Friedman, Lawrence M (1994). *Crime and punishment in American history*. Basic Books.

Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan K Rose, Yotam Shem-Tov, and Jeffrey Weaver (2023). "The Impact of Incarceration on Employment, Earnings, and Tax Filing". In: *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2023-108.

Grogger, Jeffrey (1995). "The effect of arrests on the employment and earnings of young men". In: *The Quarterly Journal of Economics* 110.1, pp. 51–71.

Joseph, Mark V (2003). "The economic consequences of a criminal background." In.

Kling, Jeffrey R (2006). "Incarceration length, employment, and earnings". In: *American Economic Review* 96.3, pp. 863–876.

Lalonde, Robert J and Rosa M Cho (2008). "The impact of incarceration in state prison on the employment prospects of women". In: *Journal of Quantitative Criminology* 24, pp. 243–265.

Loeffler, Charles E (2013). "Does imprisonment alter the life course? Evidence on crime and employment from a natural experiment". In: *Criminology* 51.1, pp. 137–166.

Lyons, Christopher J. and Becky Pettit (2011). "Compounded Disadvantage: Race, Incarceration, and Wage Growth". In: *Social Problems* 58.2, pp. 257–280.

McKelvey, Blake (1936). "American prisons: A study in American social history prior to 1915". In: *(No Title)*.

McLennan, Rebecca Mary (1999). *Citizens and criminals: The rise of the American carceral state, 1890–1935*. Columbia University.

Morris, Norval and David J Rothman (1998). *The Oxford history of the prison: The practice of punishment in Western society*. Oxford University Press, USA.

Mueller-Smith, Michael (2015). "The criminal and labor market impacts of incarceration". In: *Unpublished working paper* 18.

Muller, Christopher (2012). "Northward migration and the rise of racial disparity in American incarceration, 1880–1950". In: *American Journal of Sociology* 118.2, pp. 281–326.

— (2018). "Freedom and convict leasing in the postbellum south". In: *American Journal of Sociology* 124.2, pp. 367–405.

Needles, Karen E (1996). "Go directly to jail and do not collect? A long-term study of recidivism, employment, and earnings patterns among prison releases". In: *Journal of research in crime and delinquency* 33.4, pp. 471–496.

Olivetti, Claudia and M Daniele Paserman (2015). "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940". In: *American Economic Review* 105.8, pp. 2695–2724.

Pager, Devah (2003). "The mark of a criminal record". In: *American journal of sociology* 108.5, pp. 937–975.

Poyker, Michael (2019). “Economic Consequences of the US Convict Labor System”. In: *Institute for New Economic Thinking Working Paper Series* 91.

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

Raphael, Steven (2007). “Early incarceration spells and the transition to adulthood”. In: *The price of independence: The economics of early adulthood*, pp. 278–305.

Ruggles, Steven, Catherine Fitch, Ron Goeken, J David Hacker, Jonas Helgertz, Evan Roberts, Matt Sobek, Kelly Thompson, John Robert Warren, and Jacob Wellington (2019). “IPUMS Multigenerational Longitudinal Panel.””. In.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, José Pacas, and Matthew Sobek (2021). “IPUMS USA: Version 11.0 [dataset]. Minneapolis, Minnesota: IPUMS”. In.

Saavedra, Martin and Tate Twinam (2020). “A machine learning approach to improving occupational income scores”. In: *Explorations in Economic History* 75, p. 101304.

Silver, Ian A, Christopher D’Amato, and John Wooldredge (2024). “Exacerbating inequality over the life-course: examining race differences in the reciprocal effects between incarceration and income”. In: *Social Forces* 102.3, pp. 839–860.

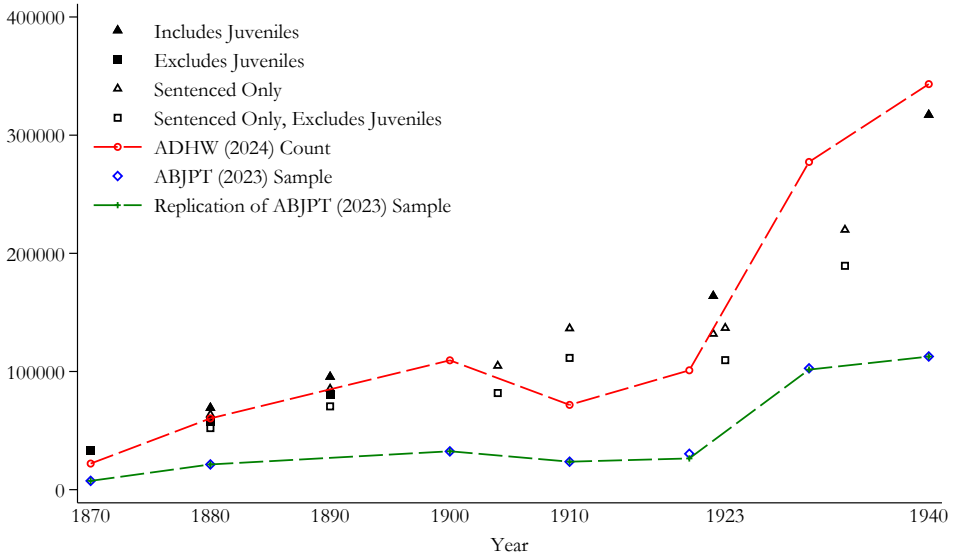
Waldfogel, Joel (1994). “The effect of criminal conviction on income and the trust" reposed in the workmen"”. In: *Journal of Human Resources*, pp. 62–81.

Western, Bruce (2002). “The impact of incarceration on wage mobility and inequality”. In: *American sociological review* 67.4, pp. 526–546.

Western, Bruce and Becky Pettit (2010). “Incarceration & social inequality”. In: *Daedalus* 139.3, pp. 8–19.

Wu, Jing (2024). “Economic Conditions and Incarceration in the United States: Evidence from Missouri, 1850-1930”. In: *Available at SSRN 4569615*.

Figure 1: Incarcerated population in the U.S., 1870-1940



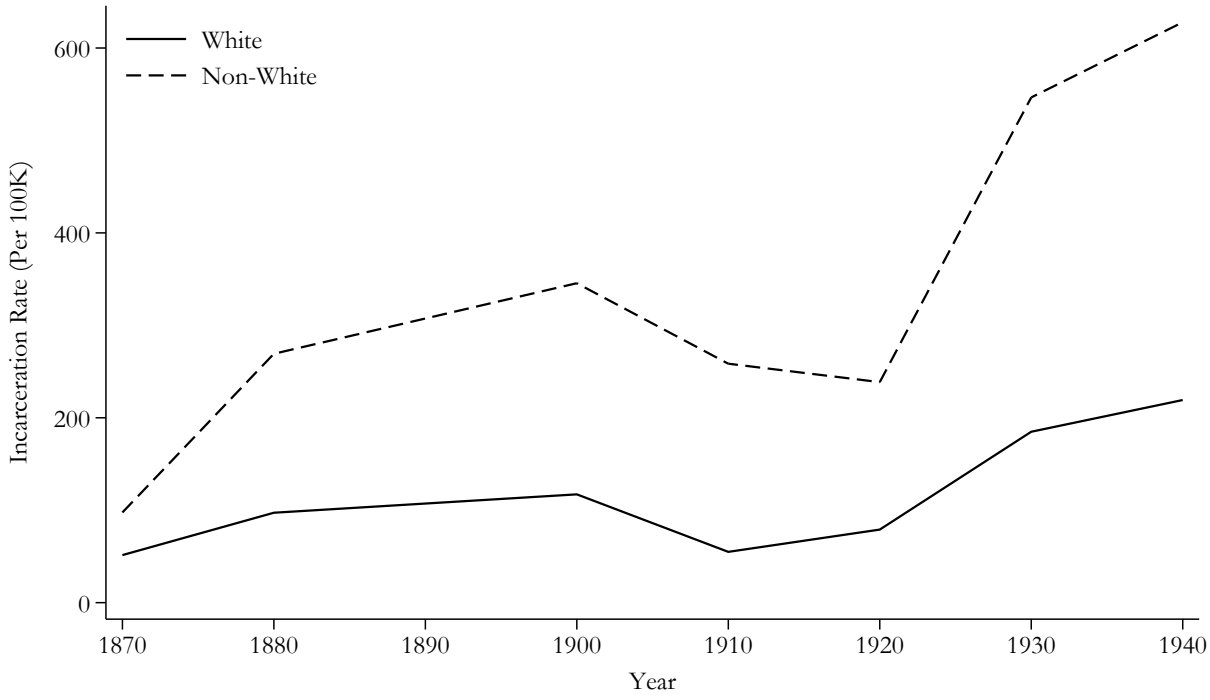
(a) Incarcerated Population Across Sources



(b) Incarceration Rate Estimates

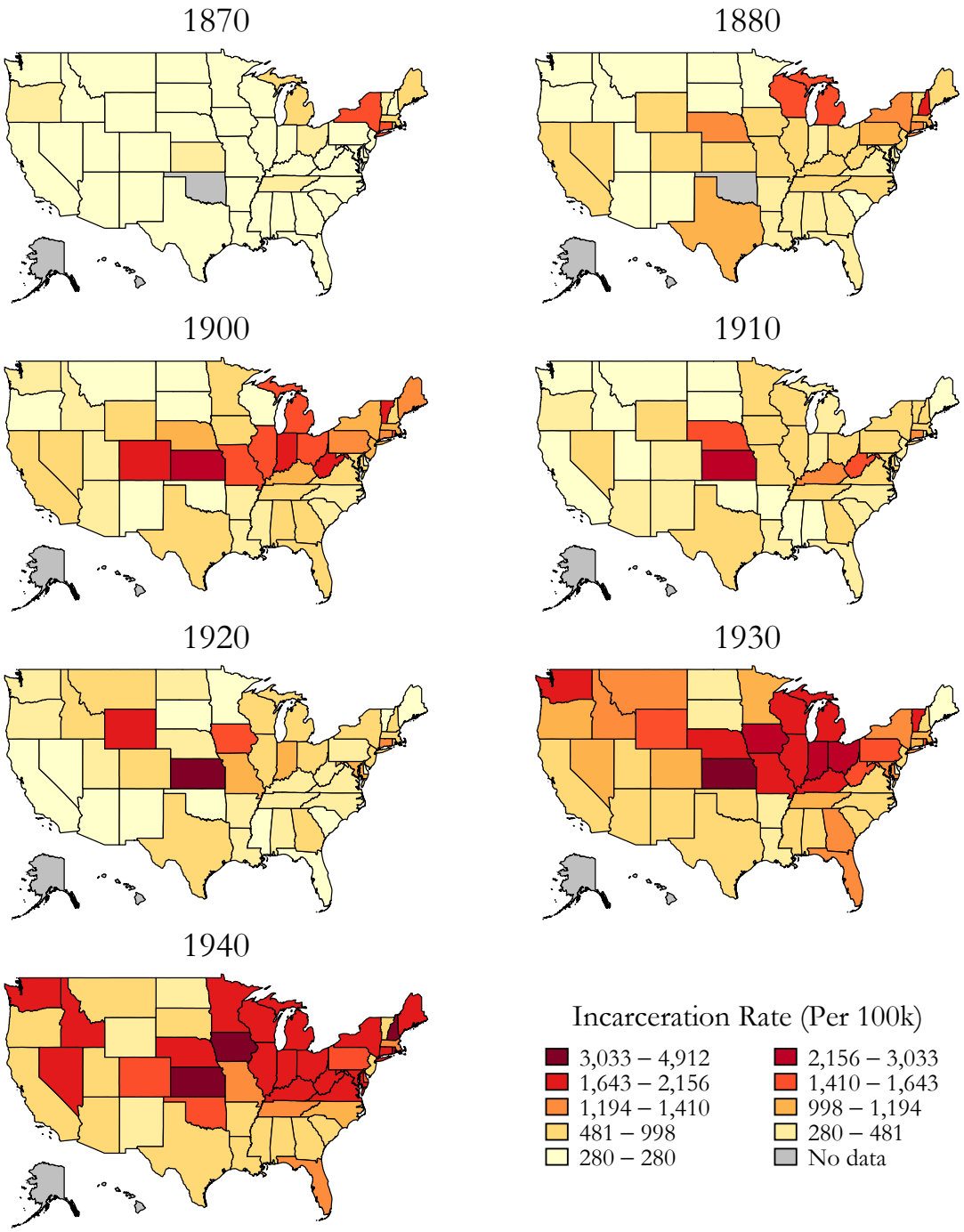
Notes: Panel (a) of this figure plots various estimates of the number of incarcerated individuals over time. Panel (b) plots the authors' estimates of incarceration rates per 100,000. Counts by sentence and juvenile status come from an aggregation of special census reports by the Bureau of Justice Statistics (Cahalan, Parsons, et al., 1986). The DHW (2024) counts are derived from the authors' own calculations, and the remaining counts and rates are from the authors' replication of Abramitzky, Boustan, Jácome, et al. (2023).
Sources: NBER restricted full count U.S. census: 1870-1940, Cahalan, Parsons, et al. (1986) and Abramitzky, Boustan, Jácome, et al. (2023)

Figure 2: Incarceration rates in the U.S., 1870-1940



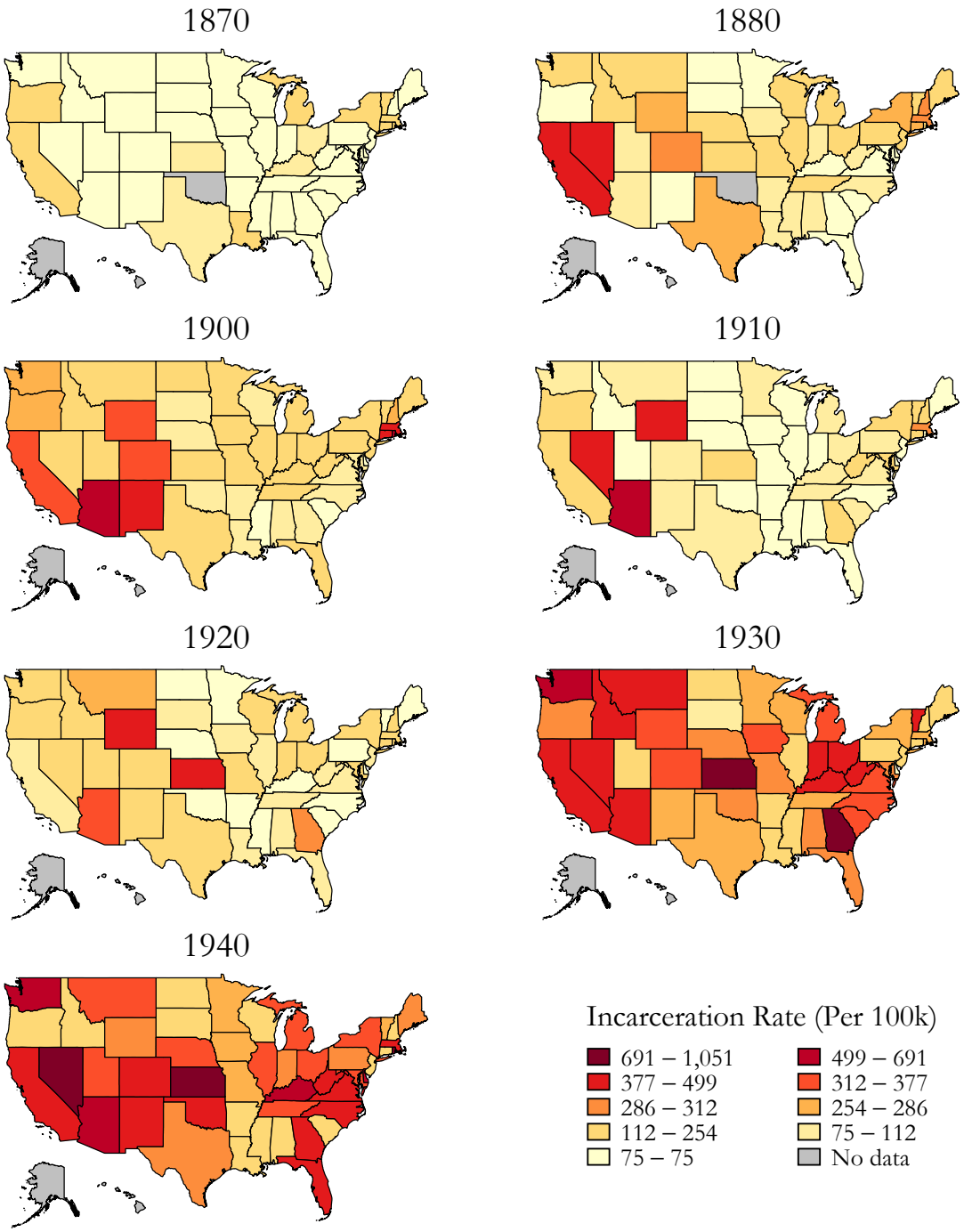
Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time by race using historical full-count census data.
Sources: NBER restricted full count U.S. census: 1870-1940

Figure 3: Non-White Men's Incarceration Rates by State, 1870-1940



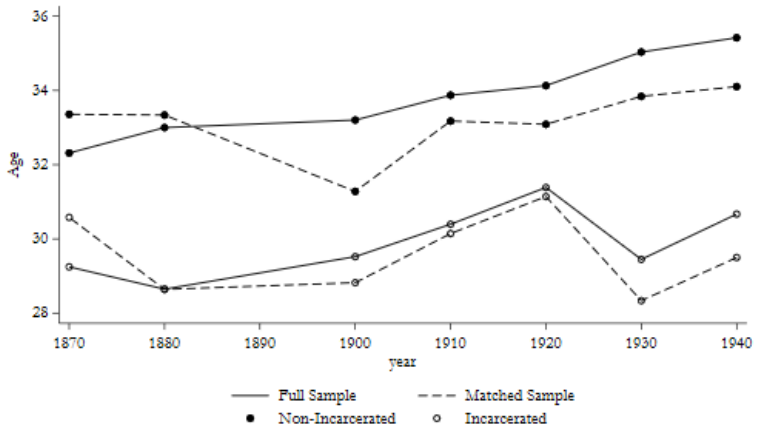
Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time and by state for the male, non-white population using historical full-count census data.
Sources: NBER restricted full count U.S. census: 1870-1940

Figure 4: White Men's Incarceration Rates by State, 1870-1940

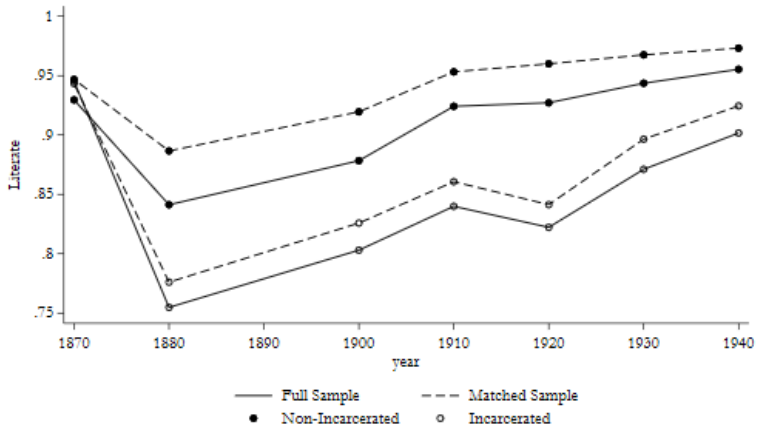


Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time and by state for the male, white population using historical full-count census data.
Sources: NBER restricted full count U.S. census: 1870-1940

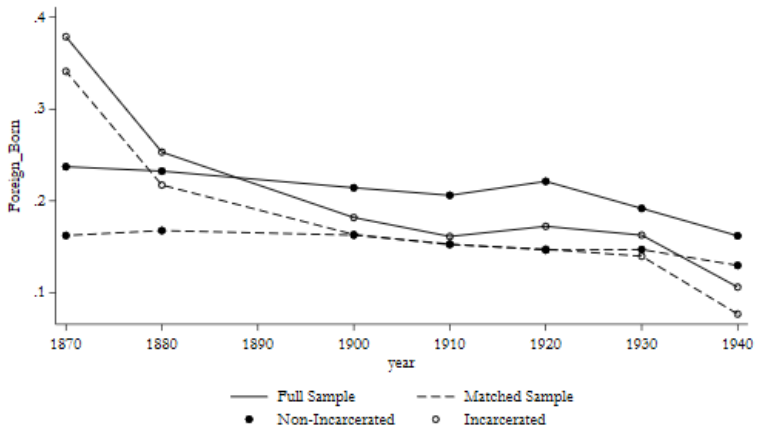
Figure 5: Characteristics of Matched Sample



(a) Average Age



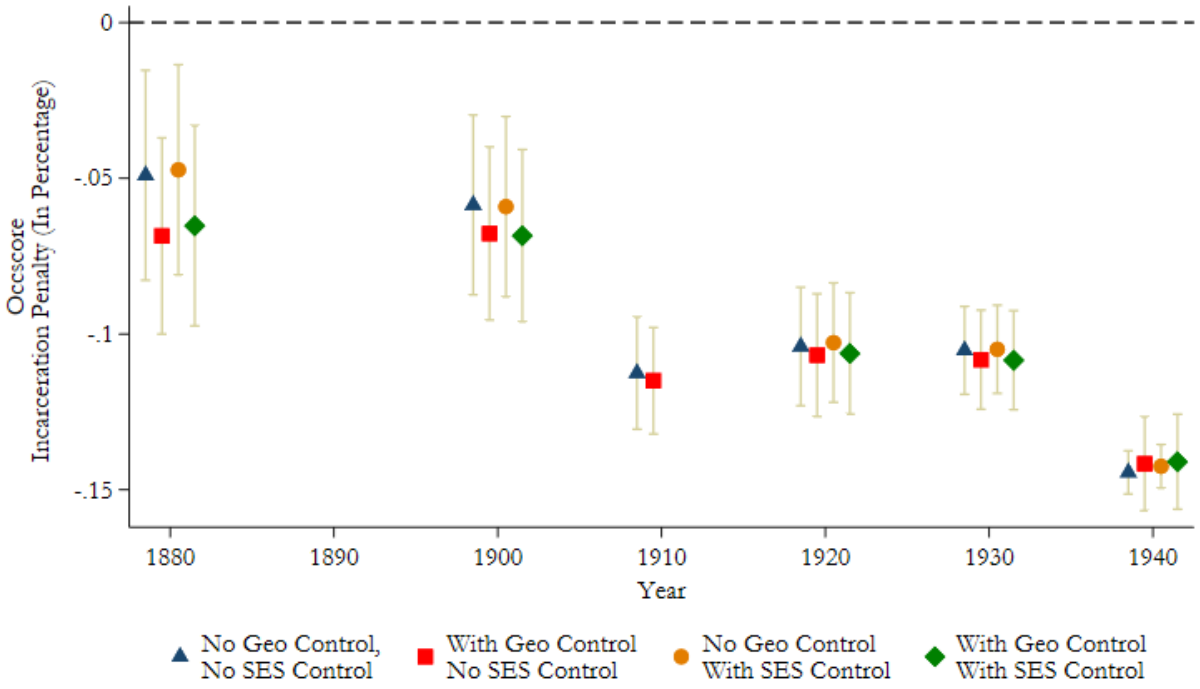
(b) Literacy Rates



(c) Share of Foreign-Born

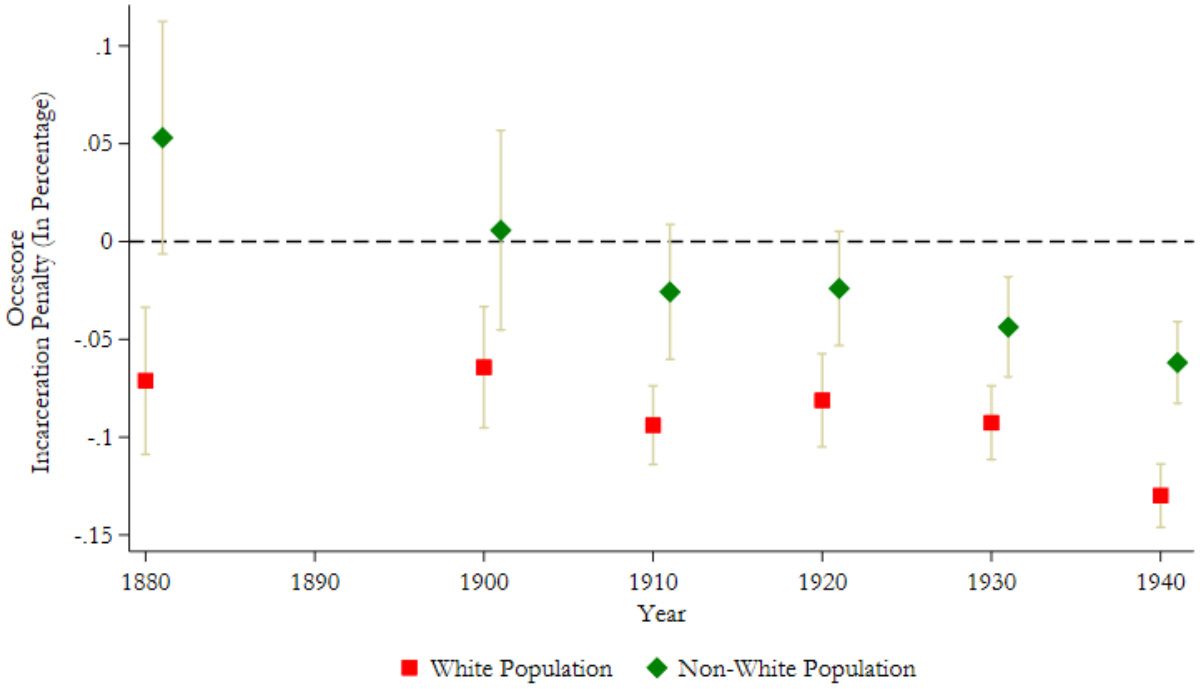
Notes: This figure compares the means of the age, literacy rate, and share of foreign-born for all men in decade $t-1$ (Full Sample) versus the subset of men linked to t using the links provided in Abramitzky, Boustan, Pérez, et al. (2020) (Matched Sample), with additional subgroup analysis conditioning on whether individuals were identified as incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 6: Incarceration Penalty, 1880-1940



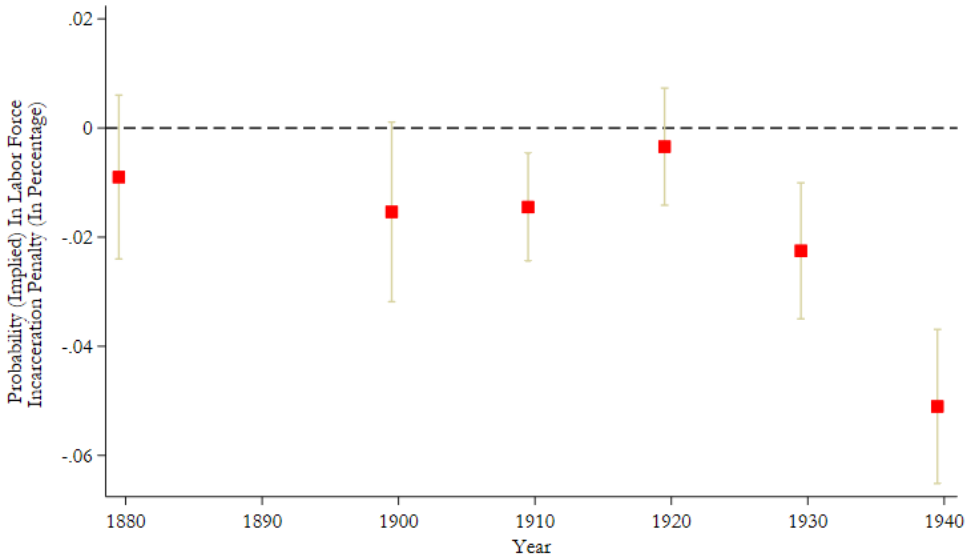
Notes: This figure plots β from Equation 1, which we interpret as the "incarceration penalty," estimated using data from the historical census. The coefficient is normalized by the mean of occscore in decade t among all linked males that were not incarcerated in decade $t-1$, and represents the percent difference in occscore in decade t between those linked males who were aged 16-26 and incarcerated in decade $t-1$ and those who were not incarcerated. "Geo Control" indicates the inclusion of county of residence in decade t fixed effects, while "SES Control" indicates the inclusion of a control for average parental occupational income score based on the name of an individual's parent, as described in the text. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 7: Incarceration Penalty by Race, 1880-1940

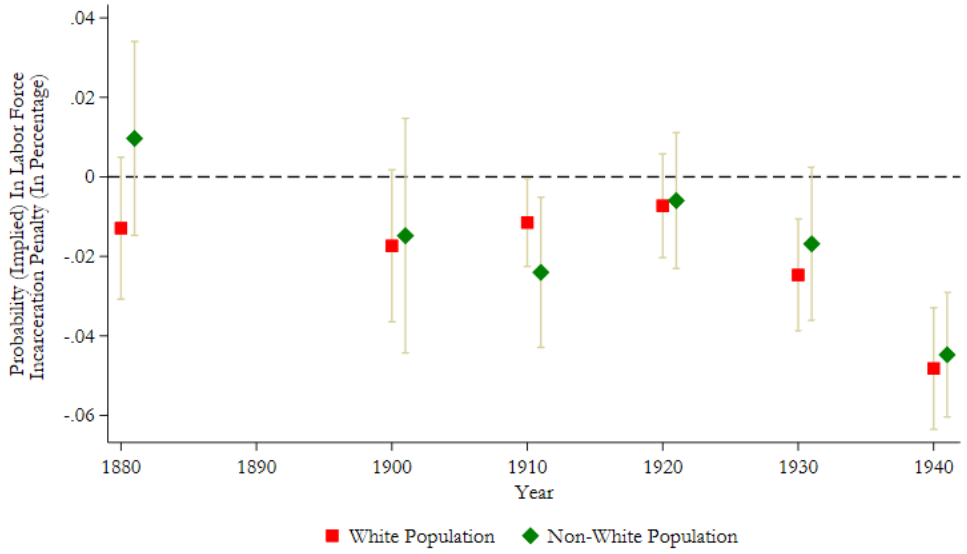


Notes: This figure replicates Figure 6, instead estimating Equation 1 twice, each restricting to a different racial subgroup. Regression estimates reported here include county fixed effects, but do not include average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 8: Implied Labor Force Participation Incarceration Penalty, 1910-1940



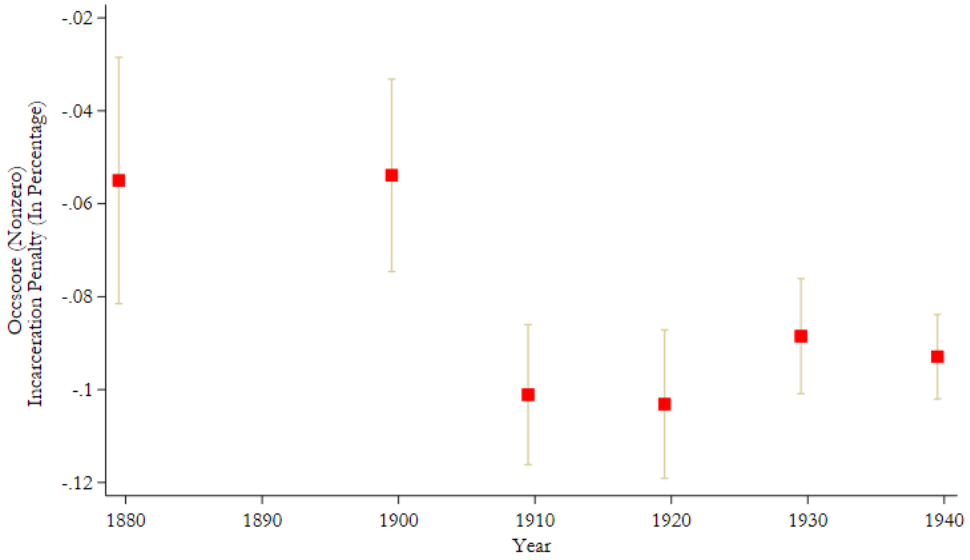
(a) Full Sample



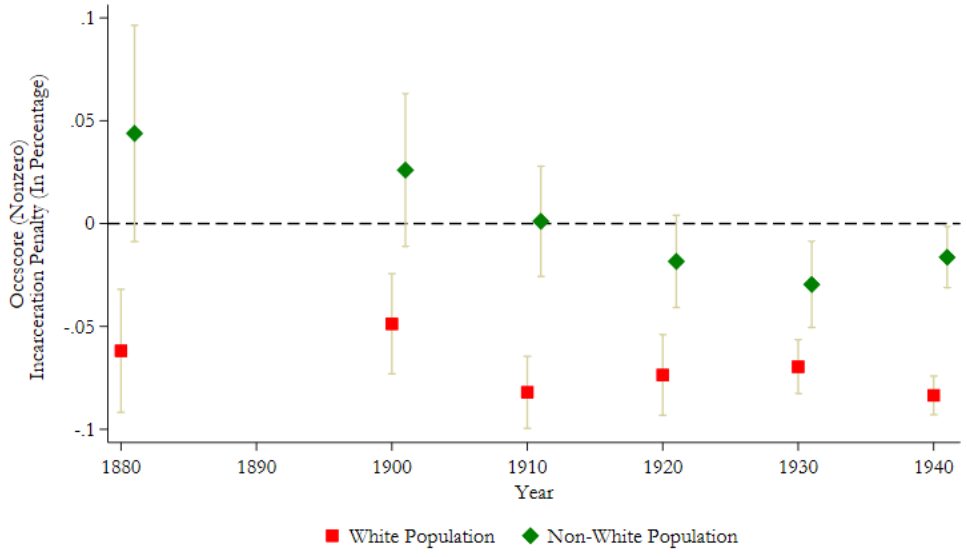
(b) By Race

This figure calculates the incarceration penalty using our constructed labor force participation indicator (whether an individual reported an occupation) as the outcome. See figure notes for Figure 6 for more information on the specification and sample. Panel (a) plots the penalty for the entire population, while Panel (b) shows results separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 9: Nonzero Occscore Incarceration Penalty, 1910-1940



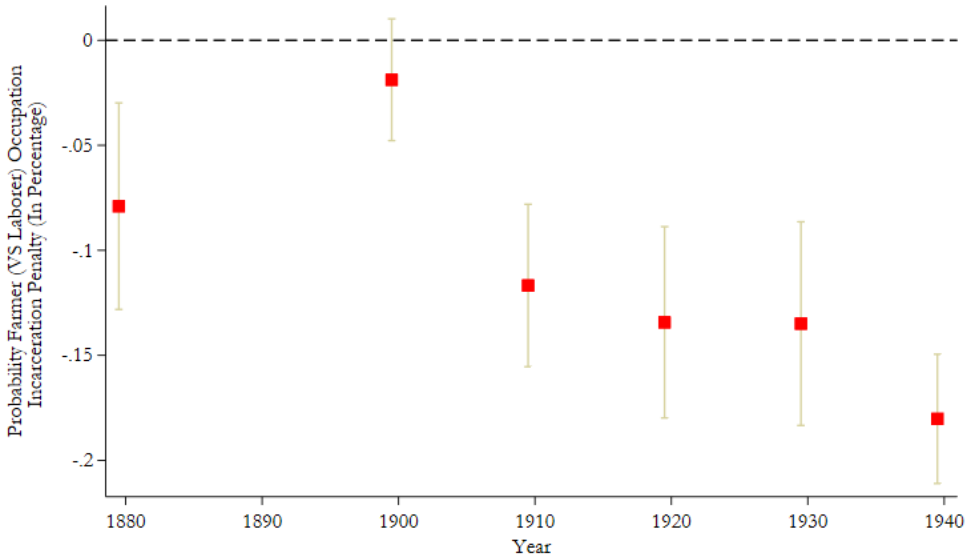
(a) Full Sample



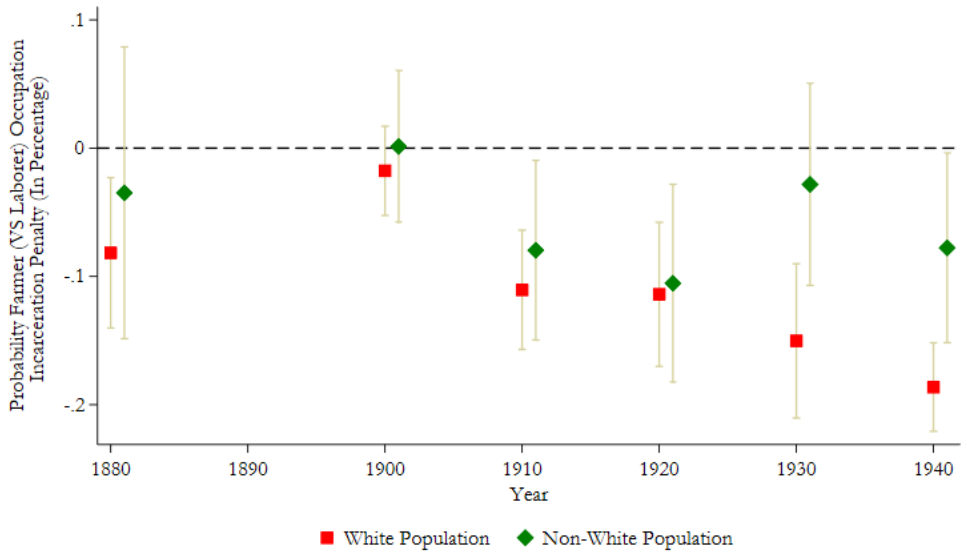
(b) By Race

Notes: This figure calculates the incarceration penalty on occscore, restricting the sample to those that have a nonzero occscore. See figure notes for Figure 6 for more information on the specification and sample. Panel (a) plots the penalty for the entire population, while Panel (b) shows results separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 10: Farm Ladder Incarceration Penalty, 1880-1940



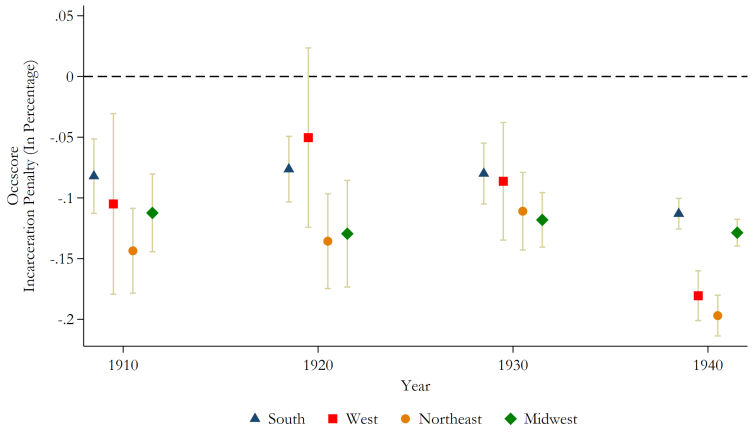
(a) Full Sample



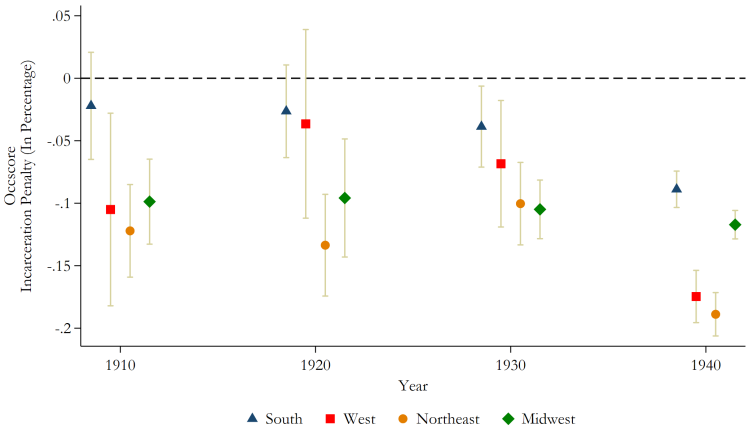
(b) By Race

Notes: This figure calculates the incarceration penalty on an indicator variable for having a "Farmer" occupation as opposed to a "Farm Laborer" occupation, conditional on reporting a farm related occupation. See figure notes for Figure 6 for more information on the specification and sample. Panel (a) plots the penalty for the entire population, while Panel (b) shows results separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

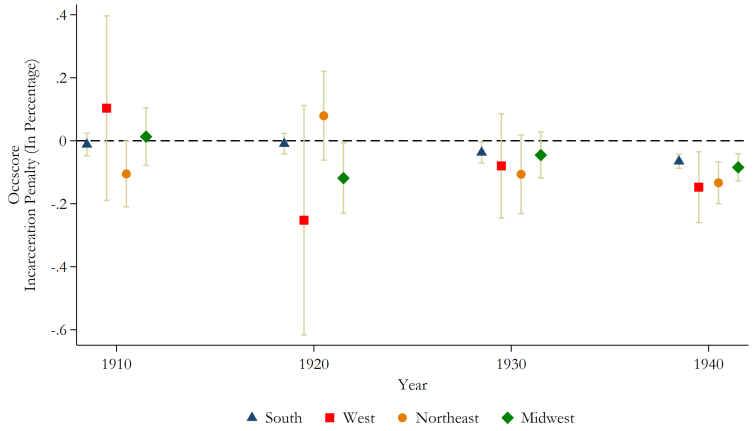
Figure 11: Incarceration Penalty by Region, 1910-1940



(a) All Population



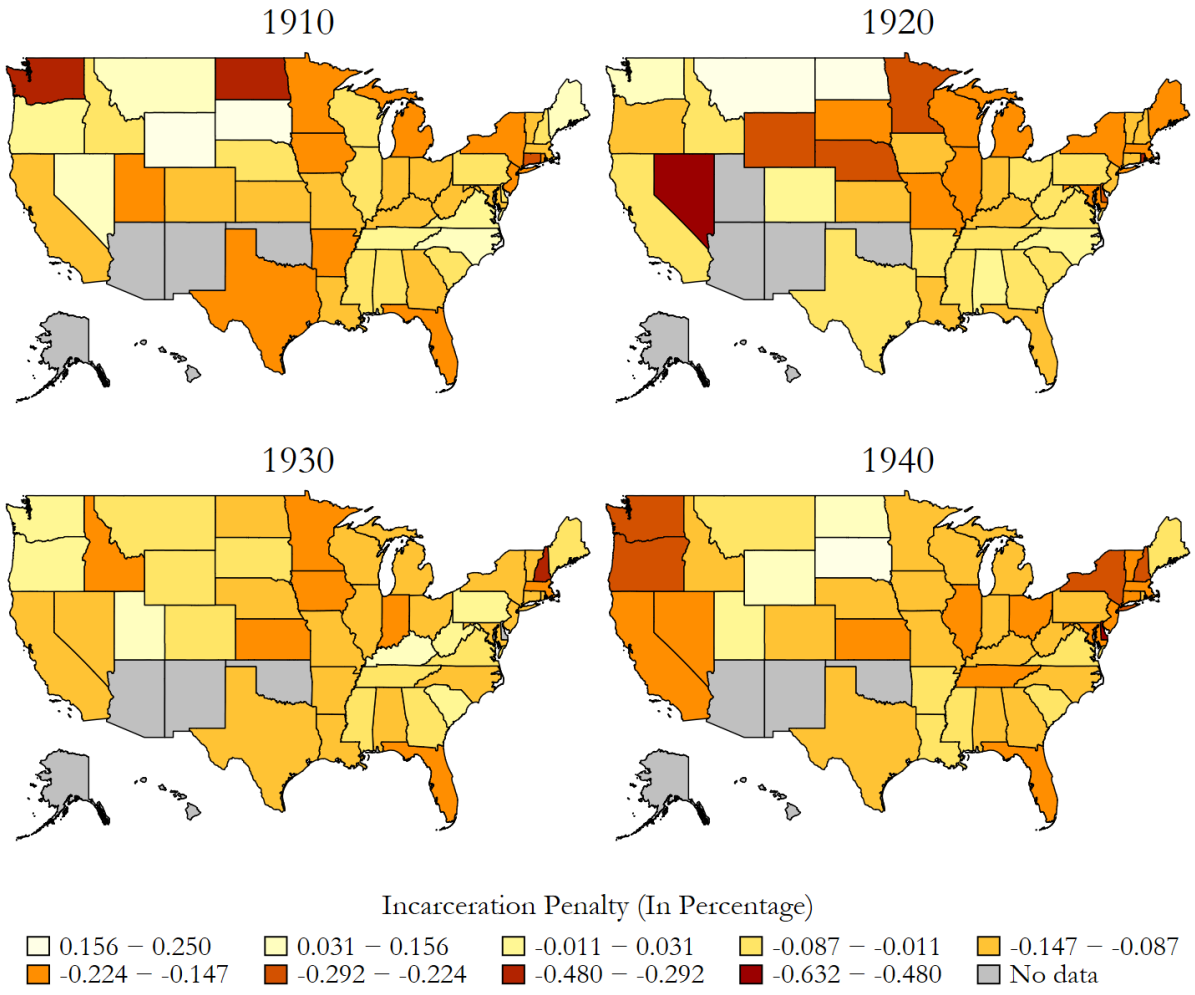
(b) White Population



(c) Non-White Population

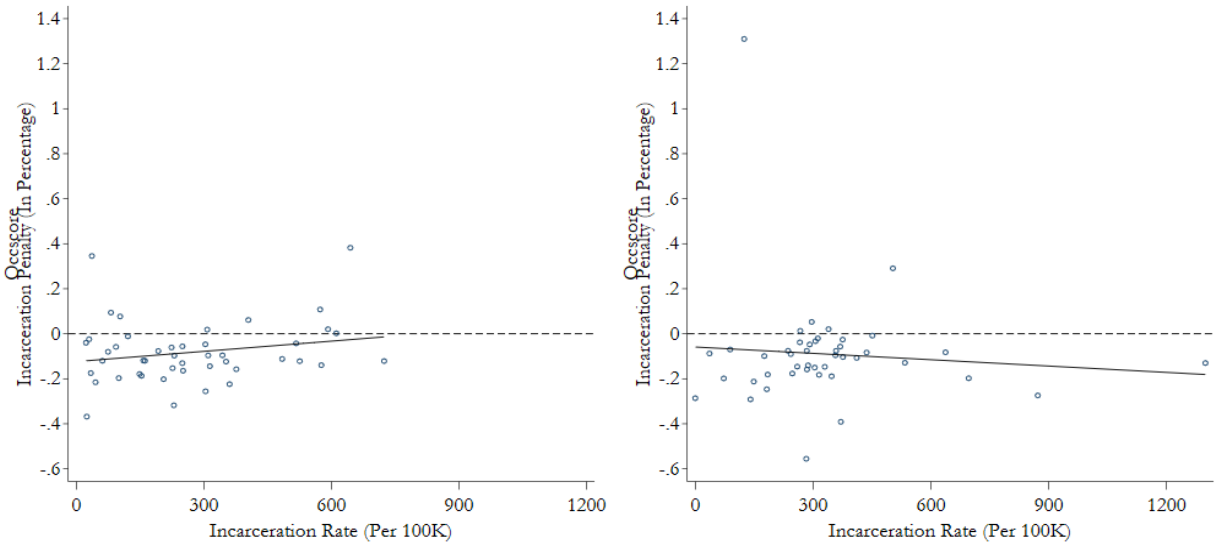
Notes: This figure plots estimates of the incarceration penalty on occscore for each region over time. Panel (a) shows results for the entire population, Panel (b) restricts to the white population in each region, and Panel (c) restricts to the non-white population in each region. There are no county of residence fixed effects, nor average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1900-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 12: Incarceration Penalty by State, 1910-1940



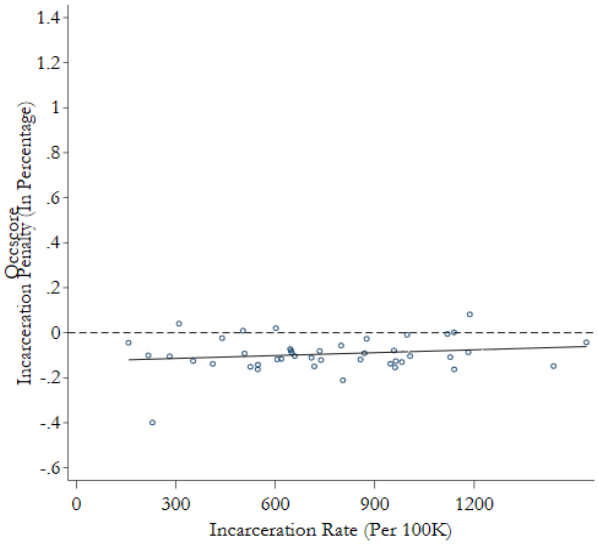
Notes: This figure plots ranges of estimates of the incarceration penalty on occscore for each state over time. There are no county of residence fixed effects, nor average parental occupational income score controls. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1900-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 13: Correlation Between Incarceration Penalty and Rate, 1910-1940

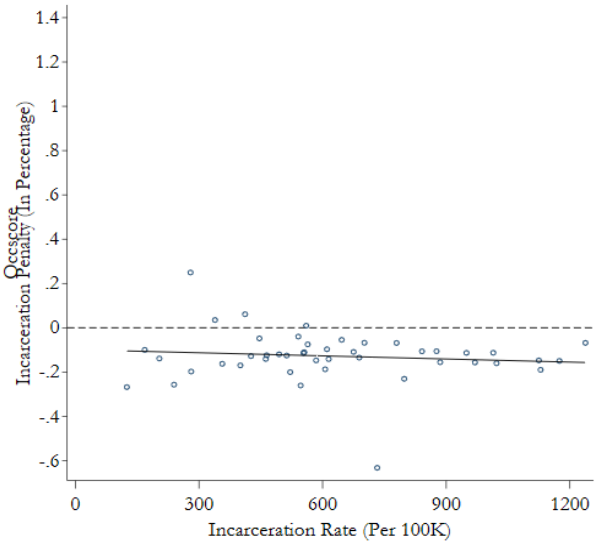


(a) 1910

(b) 1920



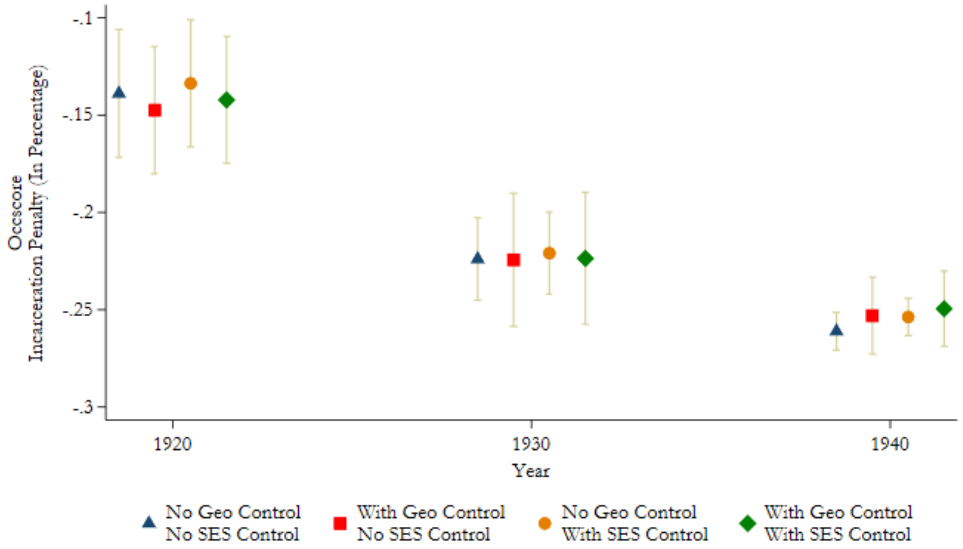
(c) 1930



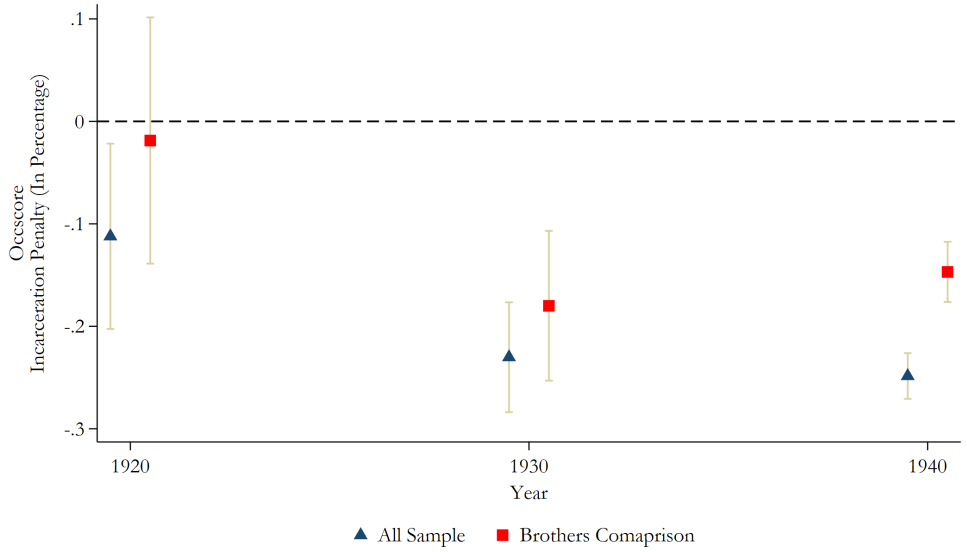
(d) 1940

Notes: This figure presents a scatter plot of state-level penalties (which are mapped out in Figure 12) against state-level incarceration rates per 100,000 individuals. The solid line denotes the slope coefficient of a regression of the penalty against the incarceration rate if a state has an estimated penalty in the given decade.
Sources: NBER restricted full count U.S. census: 1900-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 14: Incarceration Penalty in Double Linked Sample, 1920-1940



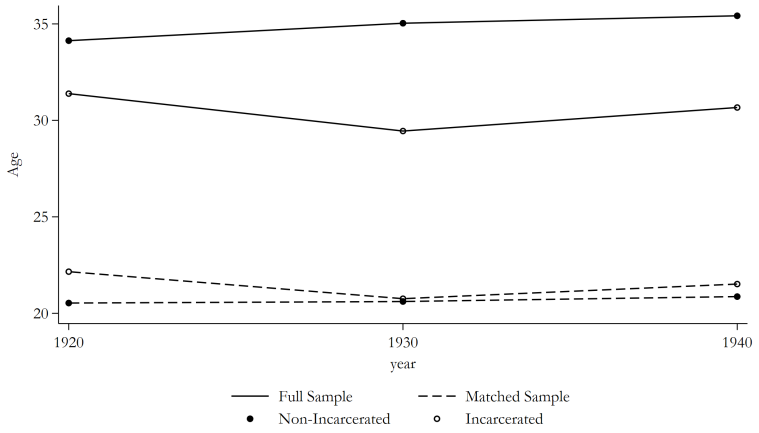
(a) Direct Control of Family Background



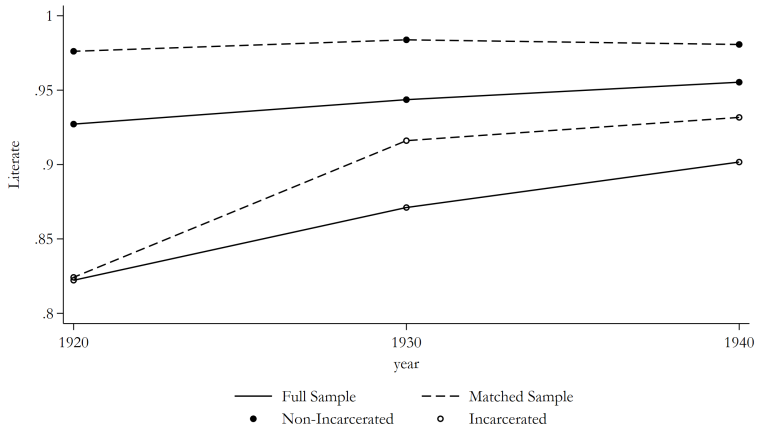
(b) Brothers Comparison

Notes: This figure plots estimates of the incarceration penalty on occscore from Equation 3, which restricts the sample to men aged 16-26 in decade $t-1$ that were able to be linked to decades $t-2$ and t. Panel (a) shows results for four specifications of the estimating equation on the double-linked sample. "Geo Control" indicates the inclusion of county of residence in decade t fixed effects, while "SES Control" indicates the inclusion of individual level controls for the largest occupational income score of the individual's parents, identified in their household in decade $t-2$. Panel (b) plots the "No Geo Control, No SES Control" coefficient from Panel (a) alongside estimates from Equation 4, which restricts the sample to brothers growing up in the same household who were both 26-36 years old in decade $t-1$ and includes family fixed effects. All regressions in both panels include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1910-1940, Abramitzky, Boustan, Pérez, et al. (2020)

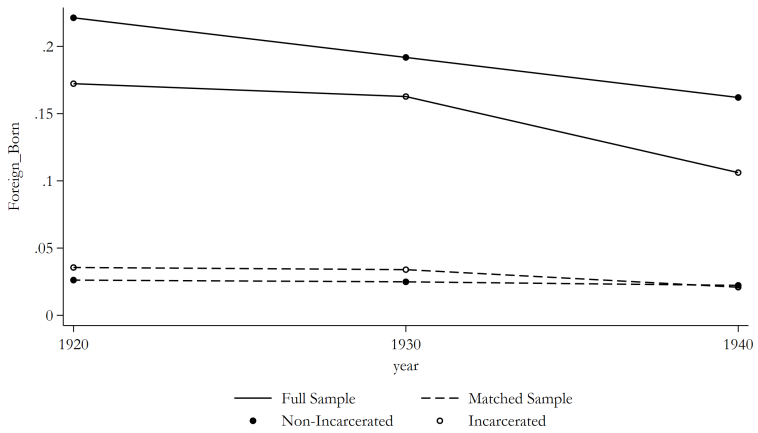
Figure 15: Characteristics of Double-Linked Sample



(a) Average Age



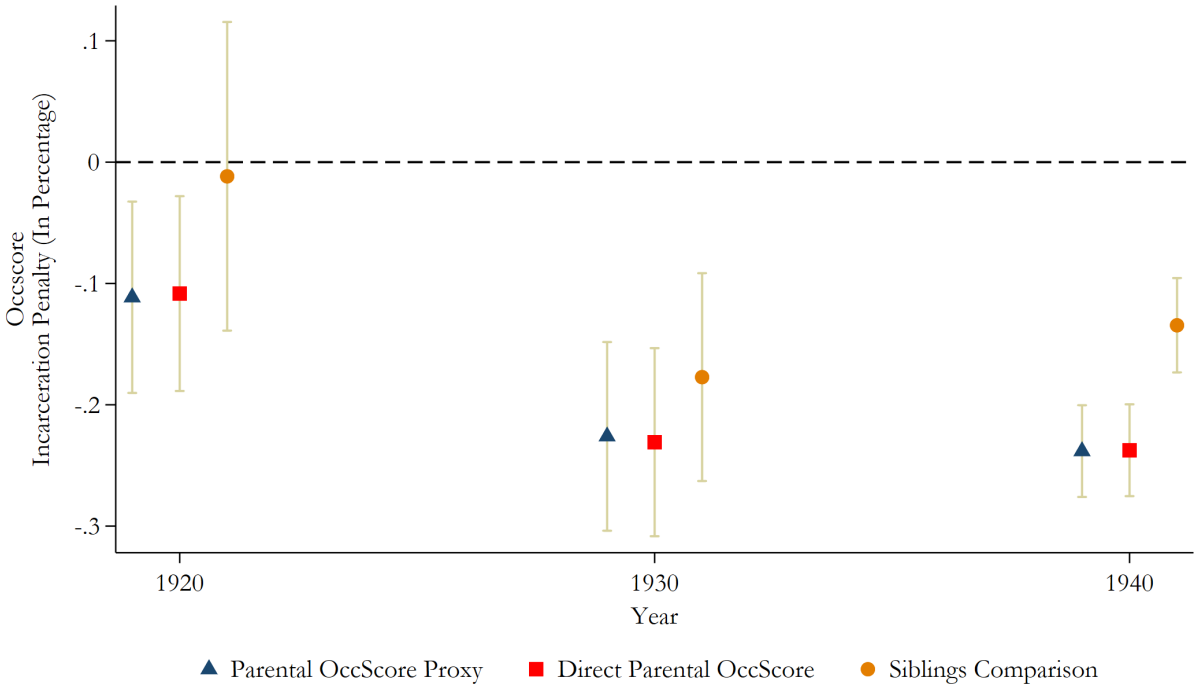
(b) Literacy Rates



(c) Share of Foreign-Born

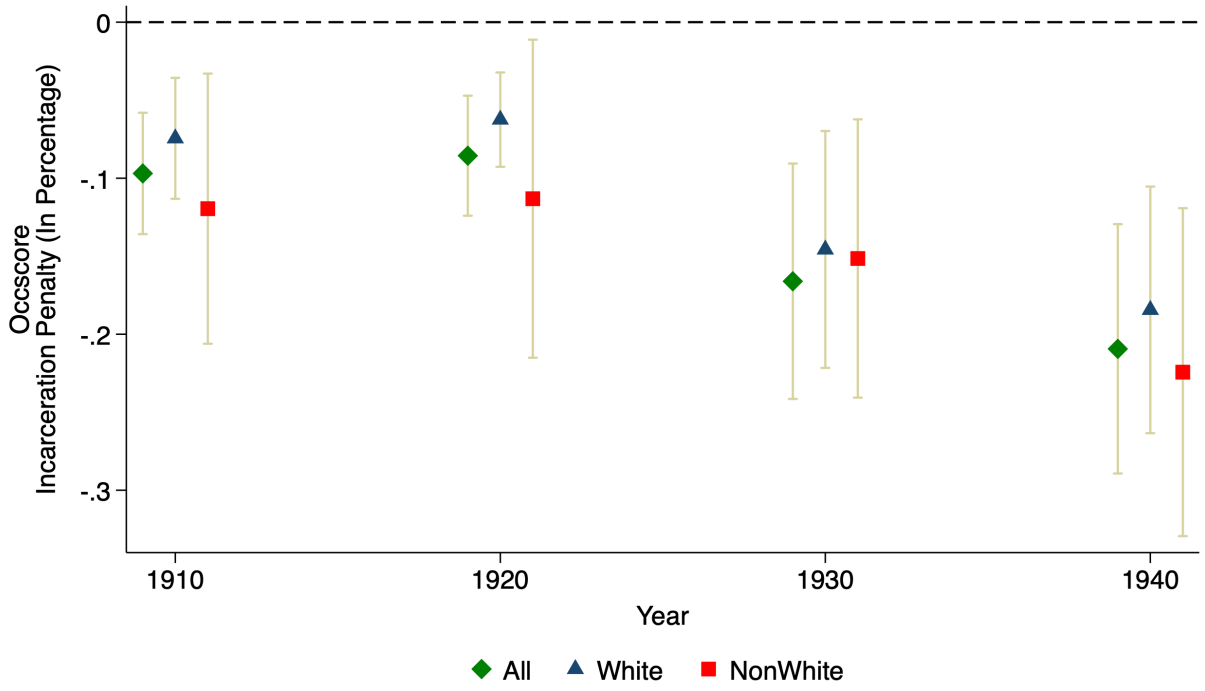
Notes: This figure compares the means of the age, literacy rate, and share of foreign-born by decade for all men in decade $t-1$ (Full Sample) versus the subset of men linked to both decade $t-2$ and t using the links provided in Abramitzky et al. (2020) (Matched Sample), with additional subgroup analysis conditioning on whether individuals were identified as incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1910-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 16: Method Comparison



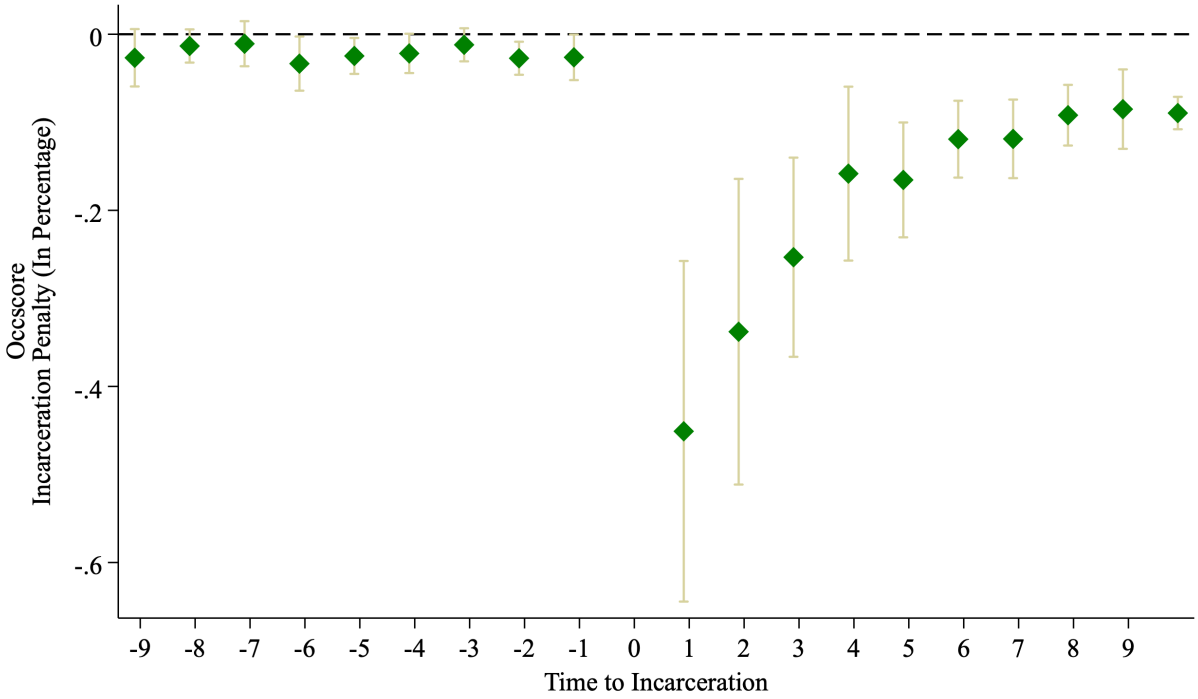
Notes: This figure plots estimates of the incarceration penalty on occscore from Equations 1 (Parental OccScore Proxy), 3 (Direct Parental OccScore), and 4 (Siblings Comparison), while holding fixed the sample, namely the double-linked brothers sample described in Figure 14 Panel (b), which restricts to brothers growing up in the same household who were both 26-36 years old in decade $t-1$. All regressions include county of residence in decade t fixed effects and controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1910-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure 17: Incarceration Penalty on Inmate Ledger Sample, 1910-1940



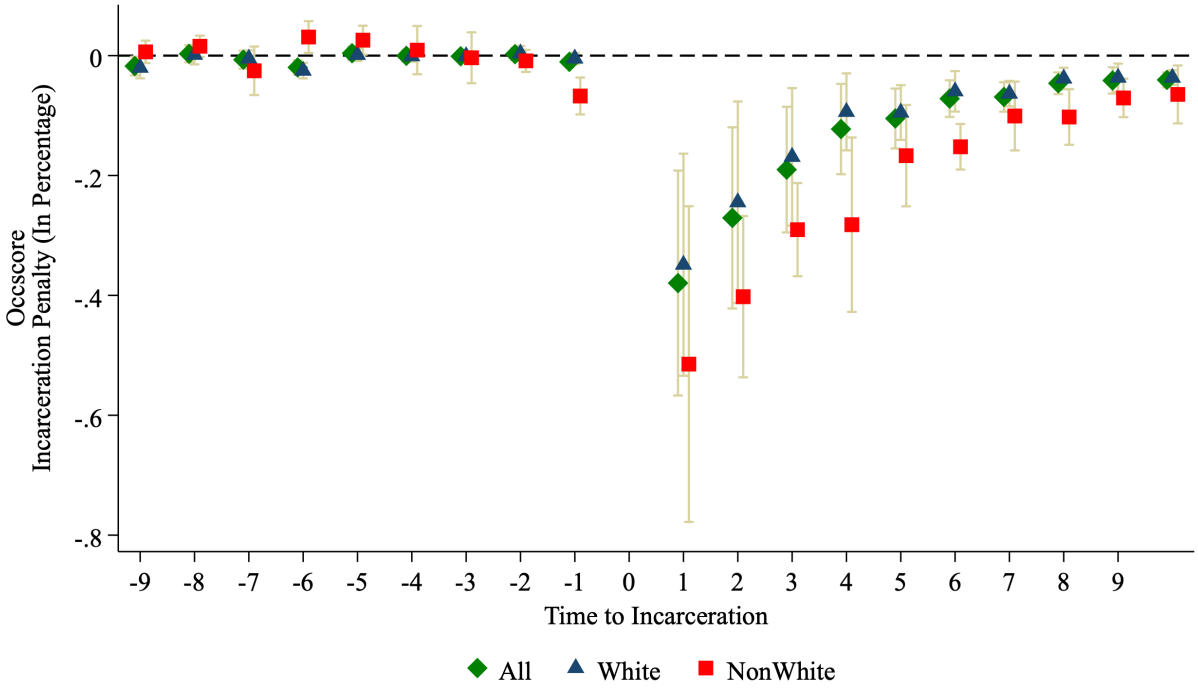
Notes: This figure plots β from Equation 2 for the full sample as well for different racial subgroups. The coefficient is normalized by the mean of occscore in decade t_0 for those males who were aged 26-36 and not linked to the inmate ledger in the previous decade. All regressions include controls for farm residency, literacy, a quadratic in age, and county fixed effects.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).

Figure 18: Dynamics of Incarceration Penalty, 1910-1940 Pooled



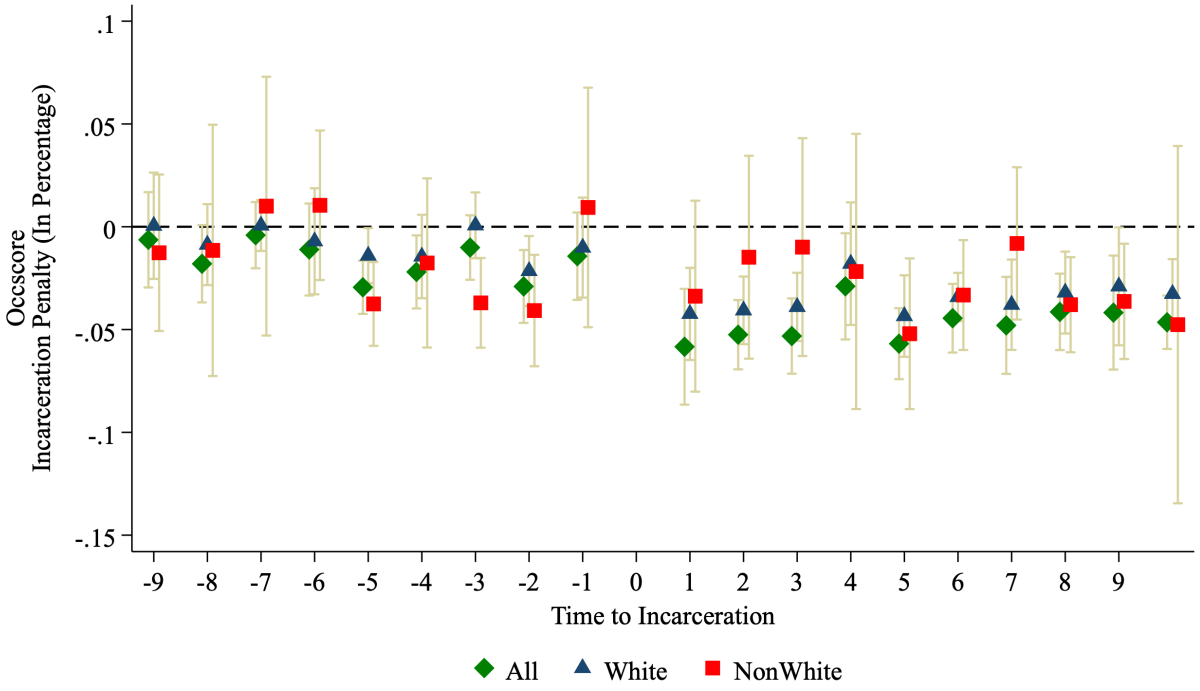
Notes: This figure plots β for occupational income scores from Equation 5 for every number of years before the prison admission and β from Equation 6 for every number of years after the prison admission. The coefficient is normalized by the mean of the occupational income score in the same Census for those males who were aged 26-36 in Census t_0 and not linked to the inmate ledger in the previous decade. All regressions include controls for farm residency, literacy, a quadratic in age, and county fixed effects.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).

Figure 19: Dynamics of Labor Market Participation Penalty, 1910-1940



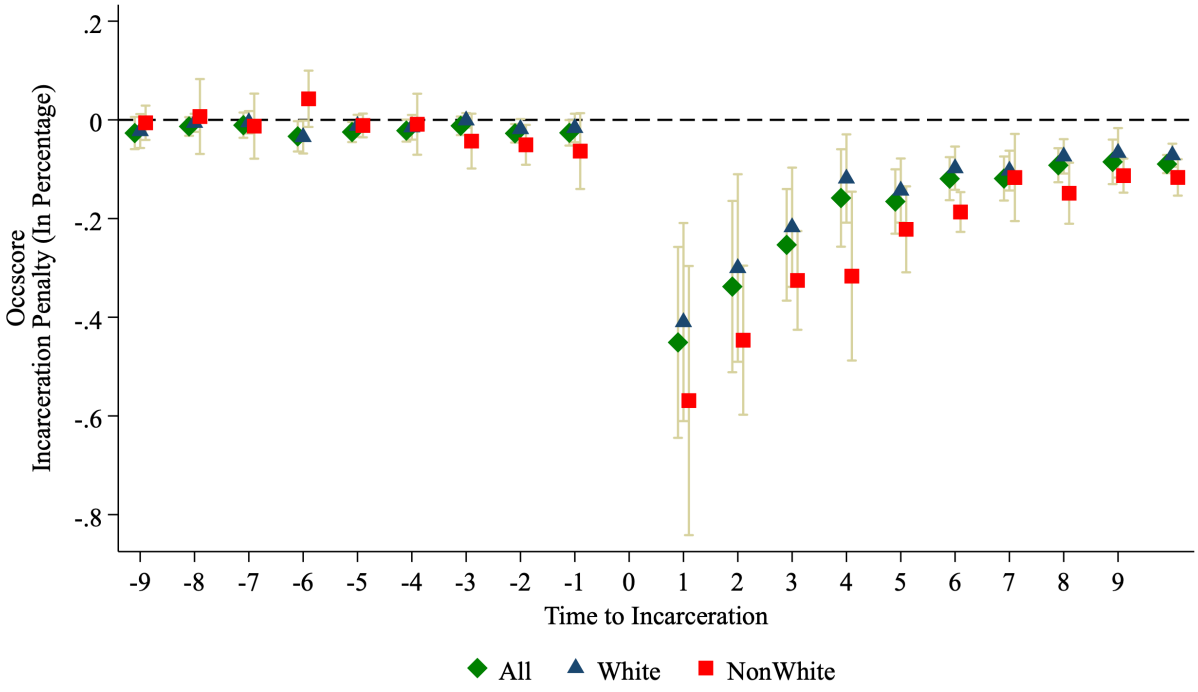
Notes: This figure plots β for implied labor force participation from Equation 5 for every number of years before the prison admission and β from Equation 6 for every number of years after the prison admission. The coefficient is normalized by the mean of implied labor market participation in the Census for those males who were aged 26-36 in Census t_0 and not linked to the inmate ledger in the previous decade. All regressions include controls for farm residency, literacy, a quadratic in age, and county fixed effects.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).

Figure 20: Dynamics of Penalty Conditional on NonZero OccScore, 1910-1940



Notes: This figure plots β for occupational income score conditional on non-zero income from Equation 5 for every number of years before the prison admission and β from Equation 6 for every number of years after the prison admission. The coefficient is normalized by the mean of occupational income scores conditional on nonzero earnings in Census for those males who were aged 26-36 in Census t_0 and not linked to the inmate ledger in the previous decade. All regressions include controls for farm residency, literacy, a quadratic in age, and county fixed effects.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).

Figure 21: Dynamics of Incarceration Penalty, 1910-1940



Notes: This figure plots β from Equation 6 for every number of years between income measures in the Census and the following prison admission in the inmate ledger. The coefficient is normalized by the mean of occscore in decade t_0 for those males who were aged 26-36 and not linked to the inmate ledger in the following decade. All regressions include controls for farm residency, literacy, a quadratic in age, and state fixed effects.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).

Table of Contents

Appendix A Identifying the incarcerated using full count census data **48**
A.1 Incarceration classification in 1900, 1920-1940 Censuses 48
A.2 Incarceration classification in 1880, 1900, 1920-1940 Censuses 49
A.3 Incarceration classification in 1910 Censuses 50
A.4 Incarceration classification in 1850-1870 Censuses 50

Appendix B Summary statistics for linked samples **51**
B.1 Double Linked Sample 51
B.2 Double Linked Sample 51

Appendix C Additional results **51**
C.1 Robustness to Linking Methods 51
C.2 Alternative Penalty Measures 51

A Identifying the incarcerated using full count census data

A.1 Incarceration classification in 1900, 1920-1940 Censuses

First, we include individuals whose relationship to the household head indicates that they were incarcerated. This includes people who have the following words in their relationship string variable: "Prisoner," "Convict," or "Jail." We also exclude individuals whose relationship string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." This is due to the fact that many staff at correctional facilities live within the prison but in separate housing in the historical period we study (Morris and Rothman, 1998). We further exclude individuals whose relationship indicates a familial relationship, such as "Wife" "Daughter," and "Son." Second, We include individuals whose relationship to the household head string includes both "inmate" and words that indicate a correctional institution: "Jail," "Prison," "Penitentiary," "Reformatory," or "Correction." This is to distinguish incarcerated people from inmates in other forms of group quarters such as mental institutions. Third, we turn to the group quarter variable if the information from the relationship to the household head is not sufficient. For individuals whose relationship to the household head is "inmate" (without any additional words) or a missing value, "X," or a "*", we classify them as incarcerated only if their group quarter string variable contains words that indicates correctional facilities, such as "Prison," "Jail," "Penitentiary," "Reformatory," "Correction," "Convict," "Delinquent," and "Penal." We also classify them as incarcerated if their group quarter type corresponds to a correctional institution and their relationship code indicates inmates.

We repeat a similar procedure using the occupation string variable instead of the relationship to household head to identify more incarcerated individuals. This includes people who have the following words in their occupation string variable: "Prisoner," "Convict," or "Jail." We also exclude individuals whose occupation string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." Then, We include individuals whose occupation string includes both "inmate" and words that indicate a correctional institution: "Jail," "Prison," "Penitentiary," "Reformatory," or "Correction." Lastly, if the occupation string indicates the lack of occupation ("No occupation," "without occupation," "None," or "Nothing") or if the occupation string is missing ("X," "*"), we classify them as incarcerated if the group quarter status indicates correctional institutions.

Finally, we classify individuals as incarcerated when it can be inferred by combining more than one variable. For people whose relationship to the household head string is a single word "Inmate", we classify them as incarcerated if their group quarter type indicates correctional institutions. For people whose relationship to the household head string is missing, we classify them as incarcerated if their relationship to the household head type is "Institutional inmates." Finally, for people whose

occupation string is a single word "Inmate," we classify them as incarcerated if their group quarter type indicates correctional institutions, and their relationship to the household head string doesn't suggest otherwise by including words other than "Inmate."

A.2 Incarceration classification in 1880, 1900, 1920-1940 Censuses

First, we include individuals whose relationship to the household head indicates that they were incarcerated. This includes people who have the following words in their relationship string variable: "Prisoner," "Convict," or "Jail." We also exclude individuals whose relationship string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." This is because many staff at correctional facilities live within the prison but in separate housing in the historical period we study (Morris and Rothman, 1998). We further exclude individuals whose relationship indicates a familial relationship, such as "Wife" "Daughter," and "Son." Second, We include individuals whose relationship to the household head string includes both "inmate" and words that indicate a correctional institution: "Jail," "Prison," "Penitentiary," "Reformatory," or "Correction." This is to distinguish incarcerated people from inmates in other forms of group quarters such as mental institutions. Third, we turn to the group quarter variable if the information from the relationship to the household head is not sufficient. For individuals whose relationship to the household head is "inmate" (without any additional words) or a missing value, "X," or a "*", we classify them as incarcerated only if their group quarter string variable contains words that indicates correctional facilities, such as "Prison," "Jail," "Penitentiary," "Reformatory," "Correction," "Convict," "Delinquent," and "Penal." We also classify them as incarcerated if their group quarter type corresponds to a correctional institution and their relationship code indicates inmates.

We repeat a similar procedure using the occupation string variable instead of the relationship to household head to identify more incarcerated individuals. This includes people who have the following words in their occupation string variable: "Prisoner," "Convict," "Jail," or "Penitentiary." We also exclude individuals whose occupation string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." Then, We include individuals whose occupation string includes both "inmate" and words that indicate a correctional institution: "Jail," "Prison," "Penitentiary," "Reformatory," or "Correction." Lastly, if the occupation string indicates the lack of occupation ("No occupation," "without occupation," "None," or "Nothing") or if the occupation string is missing ("X," "*"), we classify them as incarcerated if the group quarter status indicates correctional institutions.

Finally, we classify individuals as incarcerated when it can be inferred by combining more than

one variable. For people whose relationship to the household head string is a single word "Inmate", we classify them as incarcerated if their group quarter type indicates correctional institutions. For people whose relationship to the household head string is missing, we classify them as incarcerated if their relationship to the household head type is "Institutional inmates." Finally, for people whose occupation string is a single word "Inmate," we classify them as incarcerated if their group quarter type indicates correctional institutions and their relationship to the household head string doesn't suggest otherwise by including words other than "Inmate." We further make the restriction so that if an individual whose relationship string is missing is designated as incarcerated only under one of the two methods, we only recognize them as incarcerated if their "relate" variable code categorizes them as institutional inmates or if the "relate" variable string corresponds to the term "Inmate." For 1880, we further restrict this criterion to exclude people whose relationship code indicates other non-relatives from the incarcerated group.

A.3 Incarceration classification in 1910 Censuses

Because there is no group quarter string variable in 1910 Census records, we adjust the classification method accordingly.

First, we include individuals whose relationship to the household head indicates that they were incarcerated. This includes people who have the following words in their relationship string variable: "Prisoner," "Convict," or "Jail." We then exclude individuals whose relationship string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." This is because many staff at correctional facilities live within the prison but in separate housing in the historical period we study (Morris and Rothman, 1998). We further exclude individuals whose relationship indicates a familial relationship, such as "Wife" "Daughter," and "Son."

We repeat a similar procedure using the occupation string variable. This includes people who have the following words in their occupation string variable: "Prisoner," "Convict," "Jail," or "Penitentiary." We also exclude individuals whose occupation string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent."

A.4 Incarceration classification in 1850-1870 Censuses

Because there is no relationship to the household head string variable in the 1850-1870 Censuses, we adjust the classification method accordingly.

First, we use the occupation string to identify incarcerated individuals. This includes people who have the following words in their occupation string variable: "Prisoner," "Convict," "Jail," or

"Penitentiary." We again exclude individuals whose occupation string variable includes words that indicate some relationship to the incarcerated people, such as "Guard," "Jailer," "Chief," "Helper," "Officer," "Manager," "Charge," or "Superintendent." If there are words such as "Reformatory" or "Correction" in the occupation string, we classify them as incarcerated if the occupation string also includes "Inmate" and the group quarter status doesn't indicate other types of group quarters.[7]

We further use the group quarter string to refine the classification. If the group quarter string includes words such as "Jail," "Prison," "Penitentiary," "Prison," "Penal," "Correction," "Convict," and "Delinquent," as well as their spelling variants, we classify them if one of the two following criteria is satisfied:

- If the occupation string is missing or indicates no occupation, such as "No occupation," "None," "No," "Without occupation," and "Nothing."
- If the occupation string is "Inmate" and nothing else.

We finally use group quarter type to supplement the classification. We classify an individual as incarcerated if one of the two following criteria is met:

- If the group quarter type is "Correctional institution" and the occupation string is "Inmate" and nothing else.
- If the group quarter type is "Correctional institution" and the relationship to the household head type is "Institutional inmates."

B Summary statistics for linked samples

B.1 Double Linked Sample

B.2 Double Linked Sample

C Additional results

C.1 Robustness to Linking Methods

C.2 Alternative Penalty Measures

[7] This includes any of the following: mental institutions, institutions for the elderly, handicapped, and poor, military, rooming houses, and other non-institutional group quarters.

Table B1: Summary Statistics of 1870-1880 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	33.5	33.1	30.9	32.0
Male	1.00	1.00	1.00	1.00
Occ Score	16.5	16.6	12.0	12.2
Literate	0.93	0.91	0.28	0.29
Urban	0.25	0.29	0.12	0.14
Farm	0.48	0.43	0.25	0.24
Foreign Born	0.18	0.26	0.02	0.05
Observations	2,266,829	9,954,971	177,887	1,326,868

	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	29.2	29.2	26.4	26.7
Male	1.00	1.00	1.00	1.00
Occ Score	15.2	15.4	12.5	12.0
Literate	0.88	0.86	0.36	0.41
Urban	0.76	0.73	0.65	0.62
Farm	0.00	0.00	0.00	0.00
Foreign Born	0.26	0.32	0.02	0.05
Observations	1,687	14,046	407	4,191

Notes: This table compares the means of the seven demographic variables for men in 1870, broken down by race (white vs non-white), whether they were linked to 1880 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel (A) restricts to the non-incarcerated population in 1870, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1870-1880, Abramitzky, Boustan, Pérez, et al. (2020)

Table B2: Summary Statistics of 1880-1900 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	31.4	33.4	29.6	32.1
Male	1.00	1.00	1.00	1.00
Occ Score	17.9	18.5	13.9	14.3
Literate	0.96	0.94	0.44	0.42
Urban	0.25	0.30	0.14	0.16
Farm	0.49	0.43	0.41	0.36
Foreign Born	0.17	0.24	0.02	0.06
Observations	2,663,247	13,234,313	204,319	1,781,882

	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	29.5	30.3	27.0	27.7
Male	1.00	1.00	1.00	1.00
Occ Score	18.8	18.7	13.8	13.3
Literate	0.92	0.91	0.56	0.56
Urban	0.59	0.61	0.53	0.50
Farm	0.02	0.01	0.03	0.02
Foreign Born	0.22	0.25	0.01	0.04
Observations	3,533	35,015	1,295	16,164

Notes: This table compares the means of the seven demographic variables for men in 1880, broken down by race (white vs. non-white), whether they were linked to 1900 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population in 1880, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1880-1900, Abramitzky, Boustan, Pérez, et al. (2020)

Table B3: Summary Statistics of 1900-1910 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	33.3	34.0	31.5	32.4
Male	1.00	1.00	1.00	1.00
Occ Score	17.9	18.2	13.9	14.0
Literate	0.97	0.96	0.70	0.67
Urban	0.37	0.42	0.24	0.26
Farm	0.38	0.33	0.45	0.40
Foreign Born	0.16	0.23	0.03	0.04
Observations	4,729,897	21,393,249	323,736	2,607,062
	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	31.2	31.7	27.4	27.4
Male	1.00	1.00	1.00	1.00
Occ Score	10.9	10.9	7.7	8.0
Literate	0.93	0.93	0.66	0.64
Urban	0.61	0.63	0.50	0.49
Farm	0.01	0.01	0.01	0.01
Foreign Born	0.20	0.23	0.02	0.02
Observations	6,338	56,793	2,345	26,063

Notes: This table compares the means of the seven demographic variables for men in 1900, broken down by race (white vs. non-white), whether they were linked to 1910 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population in 1900, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1900-1910, Abramitzky, Boustan, Pérez, et al. (2020)

Table B4: Summary Statistics of 1910-1920 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	33.2	34.3	31.8	32.9
Male	1.00	1.00	1.00	1.00
Occ Score	19.9	20.2	14.9	15.0
Literate	0.97	0.95	0.77	0.72
Urban	0.45	0.50	0.30	0.32
Farm	0.33	0.28	0.43	0.40
Foreign Born	0.15	0.24	0.03	0.05
Observations	6,400,344	27,471,015	410,031	3,152,437
	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	32.2	32.9	28.8	28.8
Male	1.00	1.00	1.00	1.00
Occ Score	16.9	16.4	13.7	13.5
Literate	0.92	0.92	0.67	0.66
Urban	0.58	0.60	0.40	0.40
Farm	0.00	0.00	0.01	0.01
Foreign Born	0.21	0.26	0.01	0.02
Observations	4,914	41,943	2,310	24,219

Notes: This table compares the means of the seven demographic variables for men in 1910, broken down by race (white vs. non-white), whether they were linked to 1920 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population in 1910, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1910-1920, Abramitzky, Boustan, Pérez, et al. (2020)

Table B5: Summary Statistics of 1920-1930 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	33.9	35.1	32.6	34.2
Male	1.00	1.00	1.00	1.00
Occ Score	20.5	19.9	15.4	15.5
Literate	0.98	0.96	0.81	0.77
Urban	0.52	0.54	0.38	0.40
Farm	0.29	0.24	0.43	0.40
Foreign Born	0.15	0.21	0.03	0.04
Observations	8,821,511	31,258,217	510,451	3,411,596
	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	28.3	29.5	28.6	29.4
Male	1.00	1.00	1.00	1.00
Occ Score	13.8	13.9	12.0	11.7
Literate	0.94	0.93	0.75	0.74
Urban	0.55	0.57	0.42	0.43
Farm	0.01	0.00	0.01	0.01
Foreign Born	0.18	0.23	0.01	0.02
Observations	7,853	47,638	2,340	21,455

Notes: This table compares the means of the seven demographic variables for men in 1920, broken down by race (white vs. non-white), whether they were linked to 1930 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population in 1920, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1920-1930, Abramitzky, Boustan, Pérez, et al. (2020)

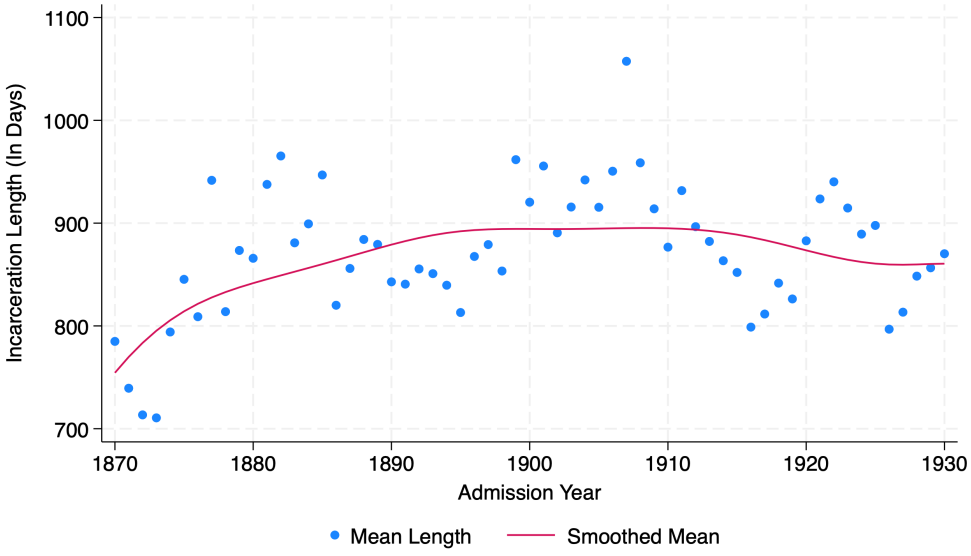
Table B6: Summary Statistics of 1930-1940 Linked Sample

	Panel A: Non-Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	34.2	35.6	32.6	34.0
Male	1.00	1.00	1.00	1.00
Occ Score	20.8	20.5	15.4	15.4
Literate	0.98	0.97	0.85	0.81
Urban	0.57	0.60	0.46	0.47
Farm	0.24	0.22	0.36	0.34
Foreign Born	0.13	0.17	0.08	0.11
Observations	10,938,677	36,226,728	691,401	4,412,204

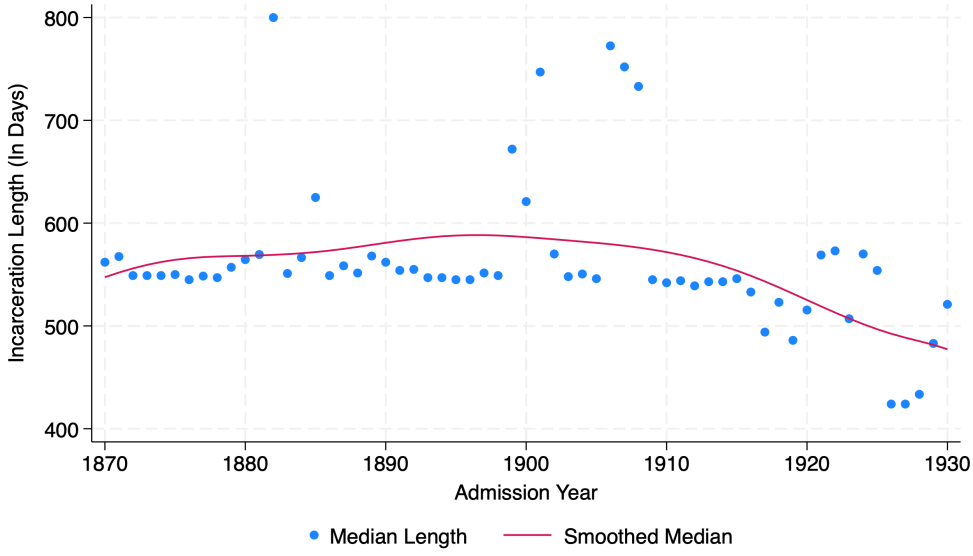
	Panel B: Incarcerated			
	Matched White	All White	Matched Non-White	All Non-White
Age	29.5	30.9	29.5	30.0
Male	1.00	1.00	1.00	1.00
Occ Score	11.1	10.7	10.2	9.8
Literate	0.95	0.94	0.83	0.80
Urban	0.51	0.52	0.43	0.43
Farm	0.00	0.00	0.00	0.00
Foreign Born	0.08	0.12	0.05	0.08
Observations	29,314	152,505	8,169	66,126

Notes: This table compares the means of the seven demographic variables for men in 1930, broken down by race (white vs. non-white), whether they were linked to 1940 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population in 1930, and Panel (B) restricts to the incarcerated population, as identified incarcerated using the authors' method described in the text.
Sources: NBER restricted full count U.S. census: 1930-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure C1: Incarceration Length in Missouri, 1870-1930



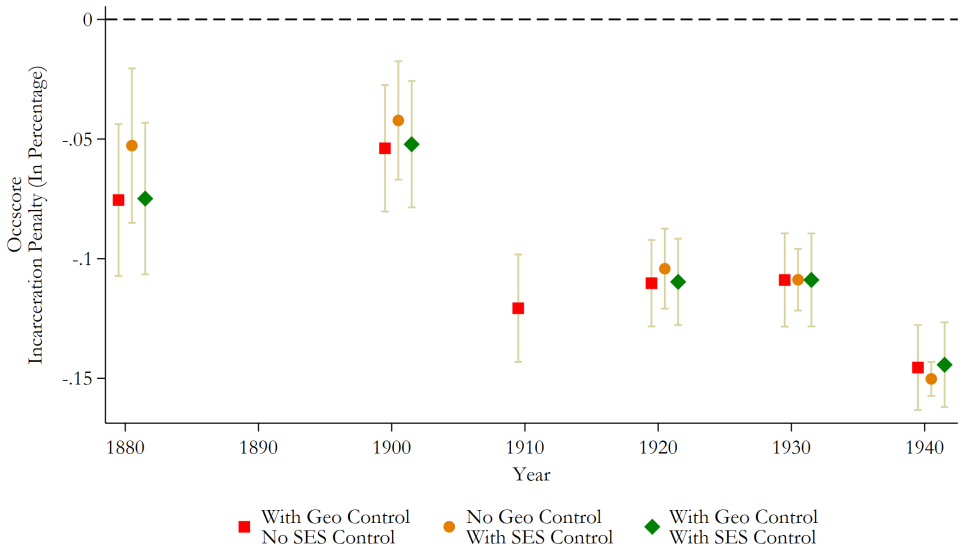
(a) Average



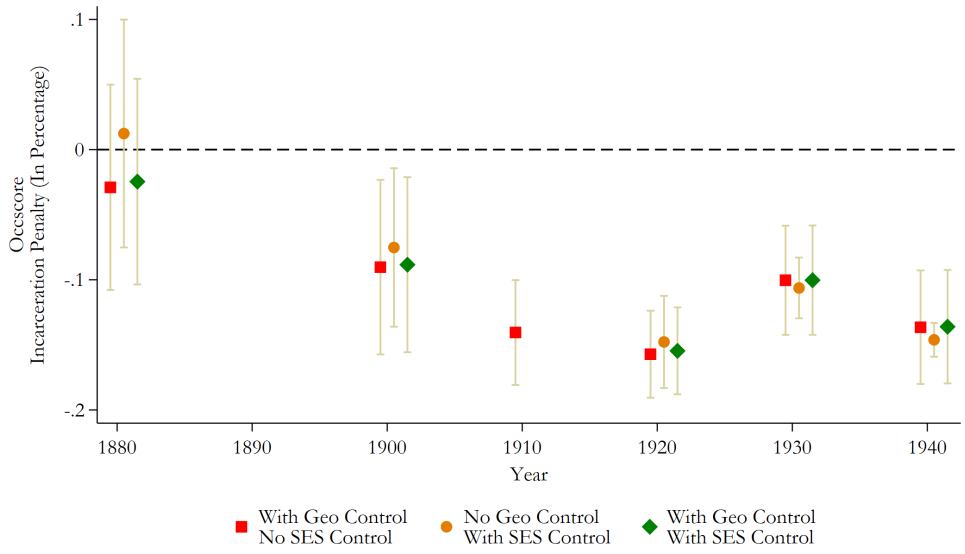
(b) Median

Notes: This figure plots the average and mean of the incarceration length in Missouri between 1870 and 1930, using Missouri State Penitentiary records
Source: Wu (2024)

Figure C2: Occscore Incarceration Penalty With Different Linking Methods, 1880-1940



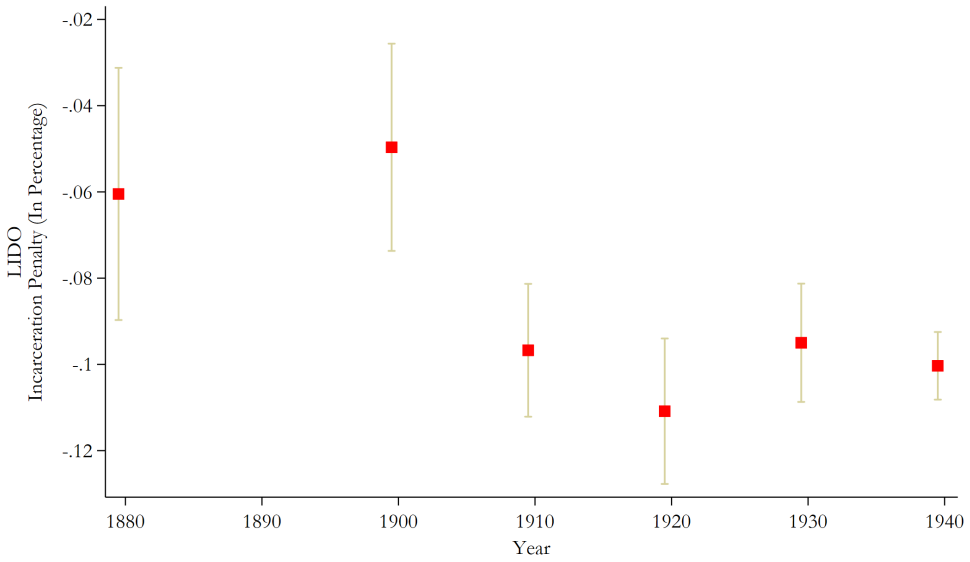
(a) Census Tree Linking



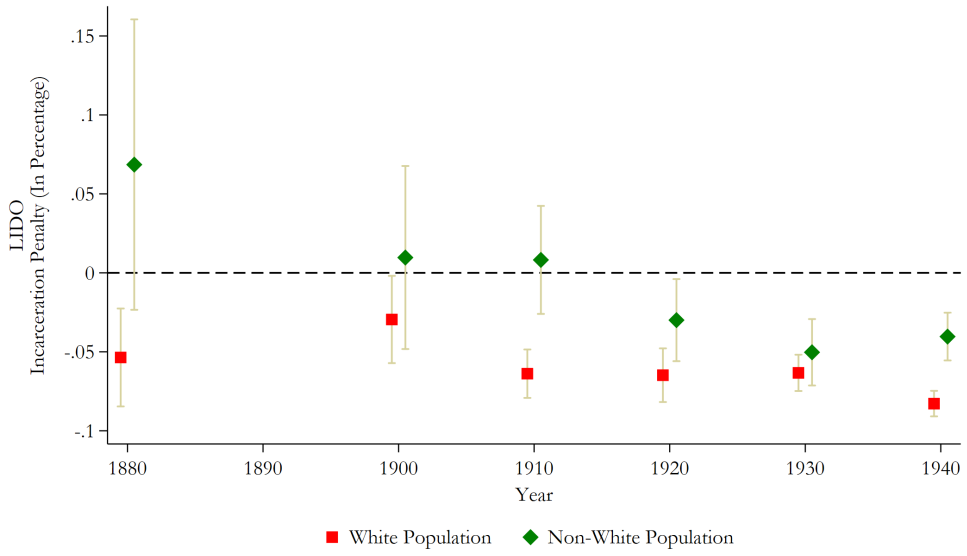
(b) MLP Linking

Notes: This figure plots reproduces 6, except it uses different linking procedures to construct the sample. Panel (a) uses Census Tree linking developed in Price et al. (2021) and Buckles et al. (2023). Panel (b) uses links created by the Multigenerational Longitudinal Panel Ruggles et al. (2019). "Geo Control" indicates the inclusion of county of residence in decade t fixed effects, while "SES Control" indicates the inclusion of a control for average parental occupational income score based on the name of an individual's parent, as described in the text. All regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Price et al. (2021), Buckles et al. (2023), and Ruggles et al. (2019)

Figure C3: LIDO Incarceration Penalty, 1880-1940



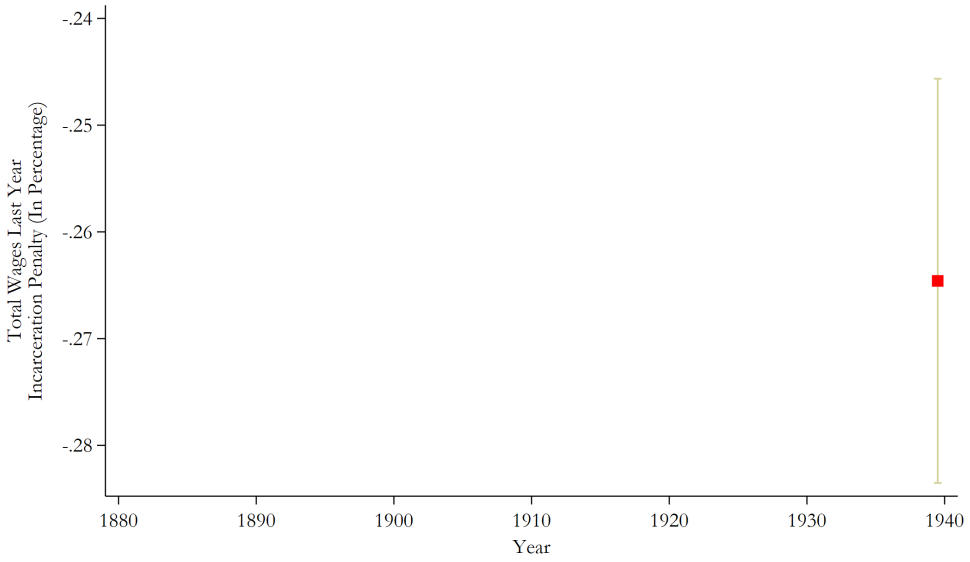
(a) Full Sample



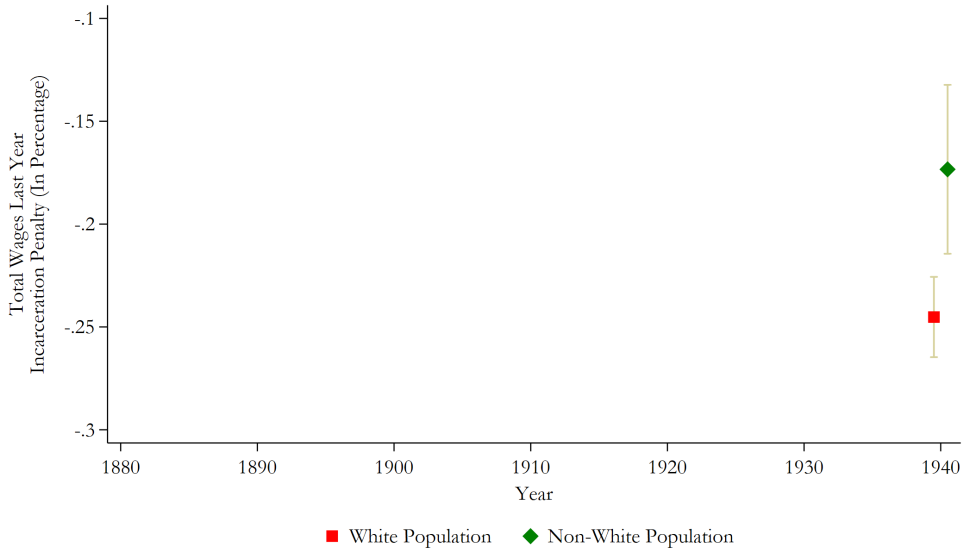
(b) By Race

Notes: This figure calculates the incarceration penalty using the LIDO income measure developed by Saavedra and Twinam (2020). See figure notes for Figure 6 for more information on the specification and sample. Panel (a) plots the penalty for the entire population, while Panel (b) shows results from regressions that were ran separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls, as well as controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020) and Saavedra and Twinam (2020)

Figure C4: Annual Earnings Last Year Incarceration Penalty, 1880-1940



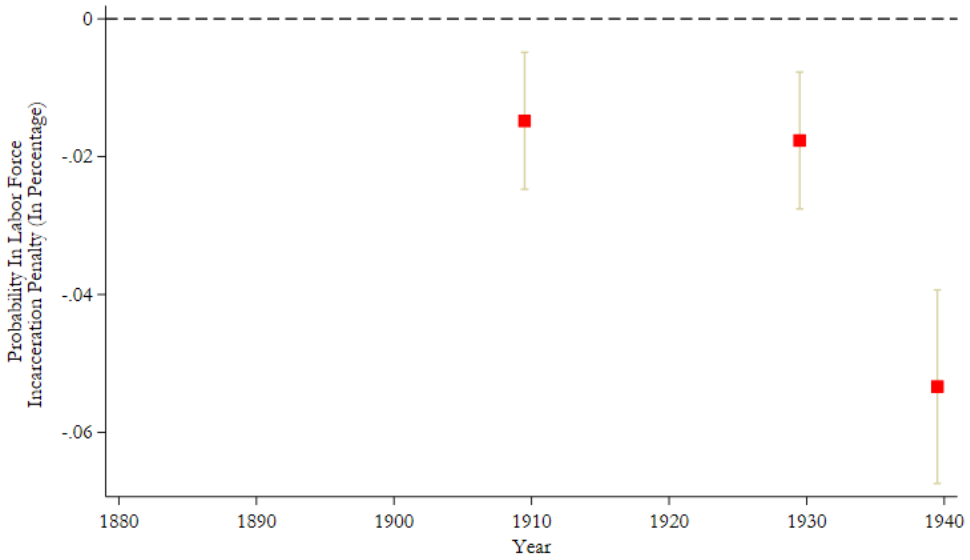
(a) Full Sample



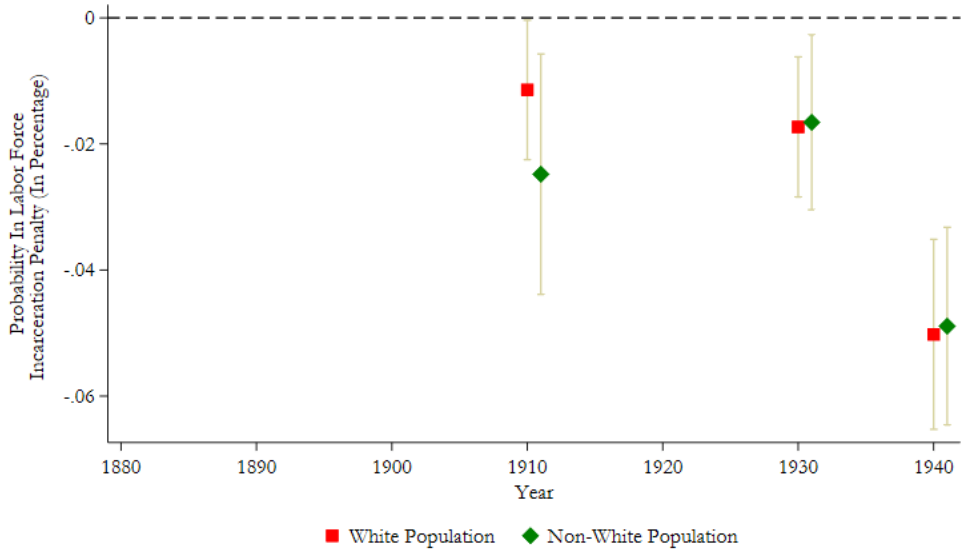
(b) By Race

Notes: This figure calculates the incarceration penalty using the "incwage" variable present in the 1940 census: last years earnings (which includes zeros). See figure notes for Figure 6 for more information on the specification and sample. Panel (a) plots the penalty for the entire population, while Panel (b) shows results separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls, as well as controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1870-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure C5: Official Labor Force Participation Incarceration Penalty, 1910-1940



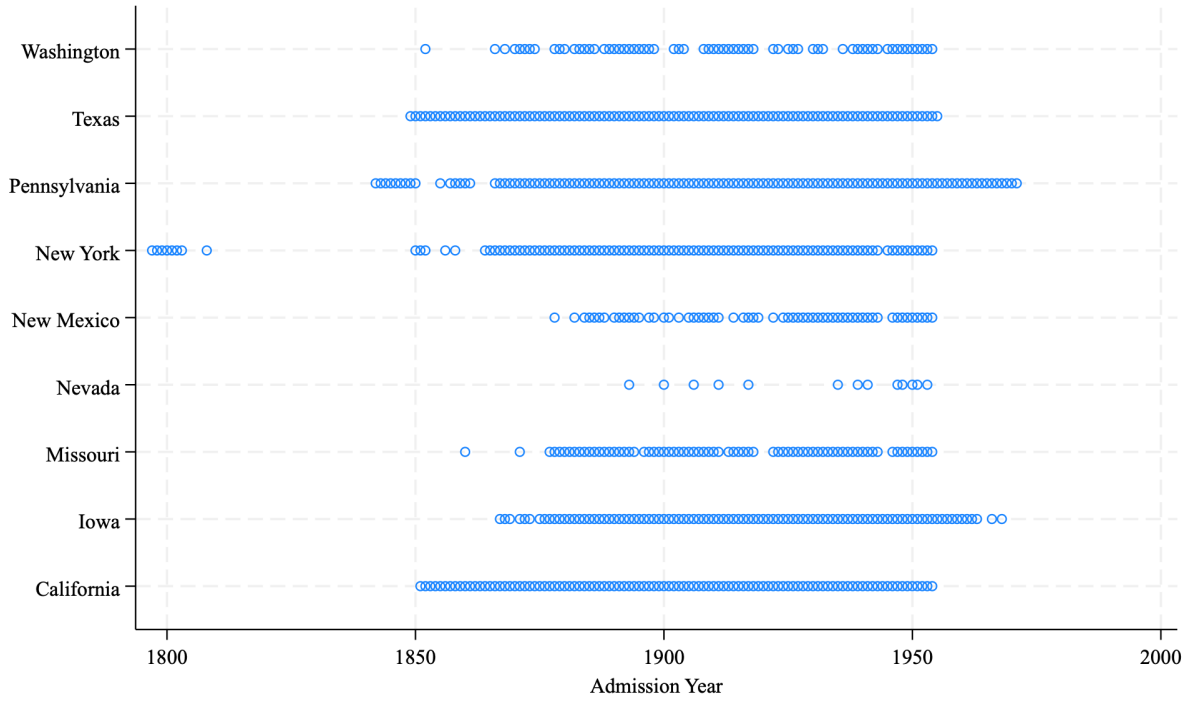
(a) Full Sample



(b) By Race

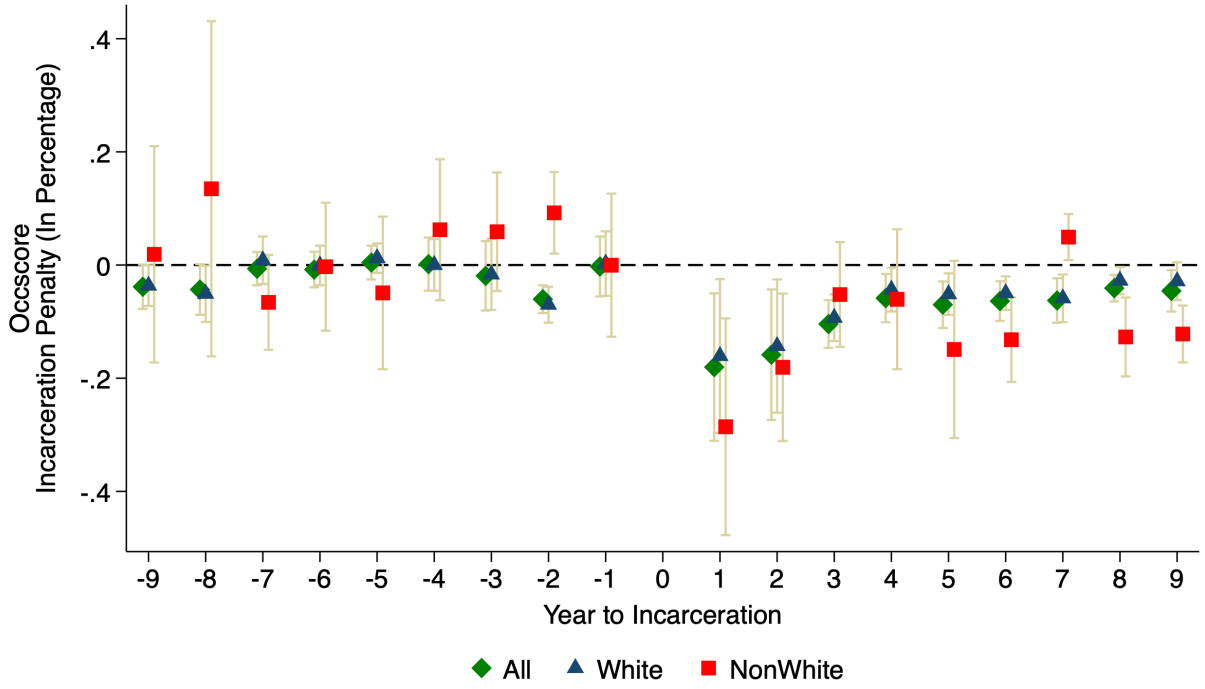
Notes: This figure calculates the incarceration penalty using the labor force participation variable included in the 1910, 1930, and 1940s censuses. See figure notes for Figure 6 for more information on the specification and sample. Panel a plots the penalty for the entire population, while Panel (b) shows results from regressions that were ran separately by race. Both panels' regression estimates include county of residence in decade t fixed effects, but do not include average parental occupational income score controls, as well as controls for farm residency, literacy, and a quadratic in age.
Sources: NBER restricted full count U.S. census: 1900-1940, Abramitzky, Boustan, Pérez, et al. (2020)

Figure C6: Coverage of Inmate Legder Data



Notes: This figure plots the time coverage by state in the inmate ledger data linkable to Censuses. Each dot denotes a year in which the state is covered by the data.
Sources: Inmate ledger records. Abramitzky, Boustan, Pérez, et al. (2020).

Figure C7: Dynamics of Incarceration Penalty on Double Linked Inmate Ledger, 1910-1940



Notes: This figure plots β for occupational income score from Equation 5 for every number of years before the prison admission and β from Equation 6 for every number of years after the prison admission. The coefficient is normalized by the mean of occupational income scores in Census for those males who were aged 26-36 in Census t_0 and not linked to the inmate ledger in the previous decade. All regressions include controls for farm residency, literacy, a quadratic in age, and county fixed effects. The estimation is based on the inmate ledger sample that is linked to Censuses both before and after prison admissions.
Sources: NBER restricted full count U.S. Census: 1910-1940 and inmate ledger records, Abramitzky, Boustan, Pérez, et al. (2020).