The historical incarceration penalty in the U.S. *

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

September 2, 2024

Abstract

The United States has both the largest prison population and one of the highest incarceration rates worldwide. Furthermore, the U.S.'s reliance on incarceration in criminal justice 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, 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 the Census and state prison ledgers, we provide new descriptive statistics on the evolution of incarceration. We then estimate 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 6-14% over the period, with the penalty increasing over time as incarceration rates increased. Using samples linked back to their childhood Census 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 the year of prison admission for state prisoners linked to Census and show that labor market trajectories prior to incarceration 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 Virtual Economics of Racism Seminar 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 United States has both the highest incarceration rate among major countries and the largest prison population worldwide. This was also true historically, with only apartheid South Africa topping U.S. prison rates in the mid-20th century (see Figure 1). 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 70-year period from 1880 to 1950, 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 young men, conditional on a minimal set of geographic and demographic controls.

To do so, we use data on millions of historical inmates identified in full-count Censuses and administrative state prison and jail records between 1870-1950. To identify incarcerated individuals in the 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 group quarters fields. This generates one of our core samples of incarcerated individuals. We use this sample to produce new descriptive facts about the incarcerated population 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 inmates admitted into state prisons and jails over time. We link these records to Census and provide an estimate of the incarceration penalty for this sample relative to individuals who do not appear on state inmate ledgers.

Our main specification estimates the association between past incarceration and occupation-based income scores ("occupation scores") in a linked sample of men with controls for literacy, age, farm status, birth state, and county of residence. To control for family background prior to incarceration, 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 on occupation scores range from around -6.4% in 1880 to -14.2% by 1940. We find similar estimates and trends using the sample of inmates identified from state inmate

ledgers.

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 (available from 1940 onwards). 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 find that past incarceration is associated with a 27.8% fall in annual income in 1940 and a 26.5% fall for those with positive income. Our estimates of the penalty are not sensitive to the inclusion of controls for geography or parental income proxies, implying that neither geographic mobility nor our coarse measure of family background can explain the reduction in earnings associated with past incarceration.

An earnings penalty associated with past incarceration may reflect direct effects of the incarceration experience, from lost years of labor market experience to employer behavior towards individuals with prior criminal justice involvement. Alternatively, the penalty may reflect negative selection into incarceration based on earnings potential. Leveraging the advantages of Census-linked historical data on incarcerated individuals, we test for negative selection along two dimensions: family background and the labor market trajectory of incarcerated individuals prior to the incarceration event.

To assess the role of negative selection based on family background, we construct double-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 occupation 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. We also construct a double-linked sample of brothers, at least one of whom ends up incarcerated and one of whom does not. Although a much smaller sample, these brother pairs allow us to estimate the penalty while including childhood home fixed effects that account for unobserved family background differences between the incarcerated and non-incarcerated. We find that controlling for childhood home fixed effects does reduce the penalty. However, because the baseline penalty is larger in the double-linked sample of brothers, we recover penalty estimates that are strikingly close to our original estimates from the single-linked sample. Thus, even when conditioning on identical childhood home, there remains a significant penalty associated with past incarceration, particularly in the later years of our period of study where it amounts to 15% of occupation-based earnings.

To test for selection into incarceration by prior labor market trajectory, we implement a pseudo-event study around the year of an inmate's admission into prison. Specifically, we link prisoners admitted between Censuses to the Census prior to incarceration and estimate the penalty in their

earnings scores separately by year *until* prison admission, relative to those not incarcerated. We also create a sample linked to the Census immediately following the incarceration event and estimate the penalty in their earnings scores, separately by year *since* prison admission. The results show that future incarceration does not have a significant association with occupation scores, and this relationship does not vary over the years leading up to prison admission. After prison admission, incarceration status is associated with a large and immediate drop in occupation scores, which attenuates over time but converges to a stable penalty relative to the pre-incarceration years. A nearly identical picture emerges when examining the labor market participation margin. These results are consistent with a lack of selection based on prior occupational and labor market participation trajectories prior to incarceration.

We discuss alternative interpretations of the penalty, including the direct effects of the incarceration experience on earnings potential, changes in the return to work experience 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 penalty estimates–particularly for the later years in our sample–are in the ballpark of estimates of the penalty for a much later period (the 1980s and 1990s).

Our work relates to a growing quantitative literature on historical incarceration in the U.S. Eriksson (2019) pioneers the approach of identifying correctional inmates via enumerator entries into a variety of fields and showed that migrating North during the first Great Migration increased Black men's probability of incarceration relative to brothers who did not migrate.[1] By contrast, the construction of schools for Black children in the Jim Crow South lowered incarceration rates of exposed cohorts of Black men (Eriksson, 2020). Moehling and Piehl (2009) and Moehling and Piehl (2014)–the latter of which constructs a large sample of historical inmates via manual digitization of Census returns for correctional institutions–document differences in the incarceration probabilities of immigrants vs. the U.S.-born in the early 20th century, finding convergence with the incarceration rates of the U.S.-born. Abramitzky, Boustan, Jácome, et al. (2023) study these differences over a 150-year period, showing that from the late 19th century onwards, incarceration rates for immigrants have been lower than those of the U.S. born. 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. 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

[1] 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.

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.[2] Studying the labor market effects of incarceration requires cross-agency record linkages that are only feasible in a subset of jurisdictions in the U.S.[3] Mueller-Smith (2015) links criminal court records from Harris County, Texas to earnings records from the state's unemployment insurance records and uses exogenous courtroom assignment to document negative effects of incarceration on future labor market outcomes. 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 incarceration and future labor market outcomes during the emergence of the modern prison system in the U.S. A great advantage of our setting is breadth: we are able to study millions of incarcerated individuals across all U.S. states and institution types over a 70-year period. Linking these individuals to Censuses before and after their incarceration episode, we are able to explore the short-, medium, and long-run associations between incarceration and labor market outcomes and to document the evolution of the incarceration penalty during a period of rapid changes to the prison system and the U.S. economy. The picture that emerges is one of a sizable gap in labor market outcomes between the formerly incarcerated and non-incarcerated that increases to levels comparable to modern estimates by the end of our study period.

Constraints of the historical data limit our ability to examine the effects of a marginal year of incarceration or to use courtroom- or judge-based instruments for incarceration spells. Yet we are able to speak to two of the most important selection-based explanations for lower labor market outcomes for the formerly incarcerated: socioeconomic background and labor market trajectories prior to prison admission. The labor market penalty we estimate is perhaps all the more noteworthy against a backdrop of widespread use of prison labor, meaning a far greater share of the incarcerated would have worked during their prison spell, and the lack of centralized background check

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

[3] The number of jurisdictions that can be studied has been rapidly expanded by the advent of the Criminal Justice Administrative Records System–see Finlay, Mueller-Smith, and Papp (2022). The linkage of these records to IRS data and the Census and ACS has expanded the range of outcomes that can be studied in conjunction with criminal justice history–from earnings, to self-employment, to safety net outcomes (Finlay and Mueller-Smith, 2021; Finlay, Gross, et al., 2023; Finlay, Mueller-Smith, and Street, 2023).

systems which give employers easy access to prospective employees' criminal history. Consistent with modern evidence on the inefficacy of some criminal record expungement programs (Agan et al., 2022), our findings may point to a role for reduced years of civilian labor market experience in driving the penalty.

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 The evolution of incarceration in the U.S., 1870-1940

Quantitative accounts of historical incarceration in the U.S. tend to focus on specific states, localities, or even correctional institutions due to the difficulty in obtaining consistent and harmonizable data for the country as a whole. A major exception is Cahalan, 1979, which provides a unified picture of incarceration rates for the country as a whole by drawing from reports by the Census Bureau and the Department of Justice. One of the contributions of our study is to provide new descriptive facts on historical incarceration in the U.S. for the population as a whole, as well as for different age, gender, and racial groups and by geography.

2.1 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 1950. A core part of our methodology depends on identifying incarcerated individuals in any given Census year[4] 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

[4]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."[5] Such adjustments make sure we include as many incarcerated people as possible while still being conservative with regard to

[5]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 large margin.[6] 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 2 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.

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

Figure 2 Panel (a) presents our counts of the incarcerated population from 1870 to 1940 (dashed red line) based on the methodology described in Section 2.1. 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.

[6]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-to-household-head 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 2 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 3 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 men 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 detailed 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 5 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.[7] In 1870 and 1880 rates were also high in parts of the South.

[7]Sabol (1989) shows that Northern rates of incarceration for Black Americans did not converge to Southern ones until 1980.

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 incarcerated individuals identified in the Census via the method described in Section2.1 and from administrative state inmate ledgers to the next Census and observe the gap in their labor market status relative to people who were not incarcerated.[8] We probe robustness and heterogeneity in our penalty estimates across different specifications, alternative linking methods, different samples of the incarcerated population, and to various labor market outcomes.

3.1 Data, linked samples, and key outcomes

Data on incarcerated individuals Our first source of information on the incarcerated population is from the universe of incarcerated individuals identified and described in Section 2.1. An advantage of this sample of incarcerated individuals is that it covers all states and types of correctional institutions, including jails, prisons, and juvenile facilities. However, we are unable to identify individuals as formerly incarcerated if their incarceration spells fell fully between Censuses.

Our second source of data on incarcerated individuals comes from historical state inmate ledgers. These administrative records cover over 2 million individuals admitted into 22 state and 4 federal prisons between 1800 and 1970.[9] The information available on inmates varies by state. For 10 states, we have the requisite information to link individuals to Census data using the approach from Abramitzky, Platt Boustan, et al., 2021: first and last name, birth year, and place of birth. This limits our sample to 686,000 individuals in California, Florida, Iowa, Missouri, Nevada, New Mexico, New York, Pennsylvania, Texas, and Washington. We further restrict our analysis of this sample to the years 1910 onwards so that we have a large enough sample to estimate the penalty.

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 (80-90% over our period of study). 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), instead focusing on the U.S.-born. Third, we follow individuals who were between the ages 16 and 26 when

[8] We should note that individuals incarcerated between Censuses are not captured by our method of identifying correctional inmates in a given Census and thus enter our comparison group of "non-incarcerated" individuals. Our data from state inmate ledgers allows us to identify prisoners in the intercensal years; however, we largely do not observe the jailed population, who are likely in the comparison group in this analysis. In either case, given the relatively small size of the incarcerated population, we believe the bias from including these individuals in the non-incarcerated comparison group is small, and if anything, would lead to an understatement of the penalty.

[9] Records on inmates of New York City jails are also included.

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 this age group makes up a disproportionate share of the incarcerated population, and labor market outcomes during this period tended to stabilize after age 25. This choice enables us to follow a key segment of the incarcerated population and to observe representative long-run outcomes in the following Census.

Linked samples and representativeness To trace individuals' labor market outcomes after incarceration, we link the record where we identify an individual's incarceration status to the nearest next Census in which we observe individuals' labor market outcomes. For our Census sample of incarcerated individuals, we use links from the Census Linking Project (Abramitzky, Boustan, Pérez, et al., 2020) to link them to the subsequent Census ten years later. This name-based method links individuals from one Census to the next whenever there is a unique match based on the phonetic sounds of their names, year of birth, and their state of birth, conditional on their racial group. 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 Appendix C, we utilize alternative linking methods to identify individuals across Censuses. For individuals we identify in state inmate ledgers, we use the Abramitzky-Boustan-Eriksson ("ABE") method (Abramitzky, Platt Boustan, et al., 2021) to link these individuals to the Census.

Because information used in Census linking, such as names and race, are not randomly distributed in the population, selection into the linked sample is potentially important. We examine how representative the linked sample is compared to the general population along three key socioeconomic variables: age, literacy, and immigrant status in Appendix C. The linked sample 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 is relatively stable over time. We condition all our estimates of the penalty on controls for age and literacy. In Section 3.5, we further use alternative linking methods to address the concern of the bias introduced by linking Censuses.

Linking rates for the incarcerated population tend to be lower than those for the non-incarcerated population. The linkage rate for incarcerated individuals identified in Census is about 10%, about a half to a third of the linking rate for the general population. The linkage rate for state prison inmates is higher, ranging from 16-22%.

Labor market outcomes Our primary outcomes are occupation- and demographic-based income scores. We present the incarceration penalty with respect to occupation-based income scores, which reflect the median total income of a person in that occupation in 1950 and is a common measure for socioeconomic status in historical Censuses. We supplement analysis of the occupation score with analysis of Lasso Industry Demographic and Occupation ("LIDO") scores, which predicts earnings not just with occupation but with individuals' age, state, industry, race, and sex (Saavedra and Twinam, 2020). Starting in 1940, we observe annual wages for wage and salaried workers. In Section 3.4, we examine a variety of other labor market outcomes, including labor market participation and farm-ladder status.

3.2 Main Specification

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

$$Y_{it} = \alpha + \beta Incarcerated_{i,t_0} + X_{it}'B + \epsilon_{it} \tag{1}$$

where i denotes the individual and t denotes the Census year. Y_{it} is the individual's labor market outcome observed in Census t.[10] $Incarcerated_{i,t_0} \in [0, 1]$ is the individual's incarceration status in Census $t - 10$ for the Census sample or, for the state inmate sample, in year $t_0 \in (-10, 0)$ relative to Centsus year t. X_{it}' represents a vector of control variables related to labor market outcomes, including a quadratic in age, literacy, residence on a farm, county of residence in Census year t, and birth state. Note that the inclusion of county of residence in Census year t allows us to compare individuals in the same labor market, but conditions on a possible outcome if migration is a response to incarceration. We show sensitivity to including this control in Section 3.3. For all results except for extensive margin results, we normalize the incarceration penalty to be in percentage terms relative to the average occupation score of the non-incarcerated population. This allows for consistent comparison of our penalty estimates across decades and groups.

In some of our specifications, we also proxy for family background by including a control for average parental occupation score using individuals' first names and birth cohorts, following the methodology in Olivetti and Paserman (2015). The key insight is that children's first names provide information about parents' socioeconomic status at the time of the individuals' birth, reflecting naming practices that are stratified across different income groups. We construct this measure by taking the average of parental occupation scores by first name and age in Census $t - 20$, when the individuals we study were between the ages of 6 and 16. This age range ensures a high likelihood of co-residence with parents, allowing us to capture average parental occupation scores for the

[10] For people who were incarcerated in Census t, we assign zero to their occupation-based income score.

vast majority of individuals in our sample. In Section 4, we link our Census sample back to the individuals' childhood Censuses to directly control for family background. However, by using first-name-based proxies for parental income, we are able to control for family background in a much larger sample than one that links across three Censuses.

3.3 Baseline results

Figure 6 shows the estimated incarceration penalty separately for each Census year t. The incarceration penalty in our Census sample increased after 1900 from around 6% to 10-14% of the average occupation score of the non-incarcerated population. Incarceration penalties estimated from our state inmate ledgers sample are close to our Census sample estimates in 1910 and 1920 and larger in 1930 and 1940, potentially pointing to divergent outcomes for prison-based populations vs. the full incarcerated population (jails and prisons). The impact on LIDO scores tracks the effect on occupation scores closely until 1940 when the penalty on LIDO scores is slightly smaller than the penalty on occupation scores. Finally, the impact on income in 1940 is much larger than the impact on occupation or LIDO scores, with a drop of 26.5 across all wage and salary earners and a drop of 27.8 among those with positive income.

Dropping current county fixed effects and proxies for parental income from the model does not meaningfully change our estimates of the penalty, as shown in Figure 7. The stability across specifications suggests that the penalty is not attributable to differences in our proxy for family background or to potential location responses to incarceration.

Incarceration Penalty by Race There exist stark and persistent racial disparities in incarceration rates in the U.S. as we show in Section 2.2 and attested to repeatedly in the literature on historical incarceration (Muller, 2018; Poyker, 2019; Wu, 2024). Much smaller samples of linked non-white prisoners limits our ability to investigate differences in the penalty by racial group in detail. However, in Appendix C.1, we estimate the penalty separately for white and non-white linked samples. Appendix Figure C1 shows that in our Census sample, the two racial groups exhibit similar trends in incarceration penalty over time, but the incarceration penalty is higher for the white population than for the non-white population in most years.

3.4 Other Labor Market Outcomes

Our analysis of occupation scores includes those predicted to have zero earnings. Thus, our baseline estimates of the penalty combine associations between past incarceration with labor force participation as well as occupation-based earnings. We examine a variety of other labor market outcomes to study the extensive margin. We also study the effects on farm ladder position for

those in farming as well as direct measures of income for wage and salaried workers, which is available in 1940.

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 having a gainful occupation, and 0 if they reported a non-occupational response (e.g. homemaker/housewife, student, "other non-occupational response").

Figure 8 mirrors Figures 6, showing results from estimating Equation 1 using our constructed labor force participation variable. Results are very similar when we use a more precise measure of labor force participation status for the years in which this information is available (see Figure C9).

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.

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, the figure shows that the penalty conditional on positive occupation score is very similar to the overall penalty; however, there is no additional deepening of the penalty 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.

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 occupation-based income score are not driving the results.

3.5 Robustness and Heterogeneity in Incarceration Penalty

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 er-

rors systematically take place, the estimated incarceration penalty could be biased. To address this issue, we explore the robustness of our estimates to alternative linking approaches.

First, we use Census Tree links from Price et al. (2021) and Buckles et al. (2023). These links use 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). The authors develop a two-step probabilistic model that uses both demographic information as well as cohabitation within the same household across Censuses to link individuals across Censuses. Figure C6 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, particularly as the linked samples grow larger over time.

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 C2, 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 C3 further shows the changes in incarceration penalty by state. At this more granular level, the regional trends identified in Figure C2 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 C4 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. Thus, the penalty appears uncorrelated with state incarceration rates.

Our results thus far point to a significant gap in earnings between individuals with incarceration history and those without. This gap could be attributable to a number of different factors. Past incarceration may be directly affecting individuals' labor market status through the incarceration experience itself, through missed years of civilian labor market experience, or by the treatment of

criminal history by employers. Alternatively the penalty we estimate may simply reflect negative selection on earnings potential into incarceration. In the next section, we turn to an investigation of two plausible selection channels—selection based on family background and selection based on labor market trajectories prior to the incarceration event.

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. Some studies document significant causal effects of incarceration on labor market participation and earnings (see, for example, Mueller-Smith (2015)) while others find no effects on earnings beyond the incapacitation effects that manifest during the incarceration spell (see, for example, Garin et al. (2023)). To date there is virtually no information on the historical causal effects of incarceration on labor market outcomes. Although our present data do not allow us to test the causal effect of a marginal year of incarceration. Instead we focus on the importance of plausible types of selection in driving the penalty estimates we presented in Section 3.3.

First, we explore the extent to which the incarceration penalty simply reflects poorer socioeconomic backgrounds of incarcerated individuals. We do this by further linking individuals back to their childhood Census to directly observe parental labor market status. We further use a sample of double-linked brothers to control for family unobservables that may explain the penalty.

Next, we leverage state prison admissions data across 10 states and 4 decades to test for declines in labor market status prior to incarceration episodes. If individuals are on a downwards trajectory in labor market status just prior to incarceration, then the penalty we estimate may simply reflect this pre-incarceration slide in occupation-based income or labor market participation.

Results from both exercises show that the incarceration penalty is unlikely to be explained by these two selection mechanisms detailed above.

4.1 Double-Linked sample to control for selection

Building on the linked sample in Section 3.1, we use the same name-based methods to link individuals to their childhood Census when they were 6-16 years old. Although this further limits our sample size due to low match rates, these double-linked samples offer 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 prior to 1880 has too small an incarcerated group to conduct meaningful analysis.

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.

Directly Controlling for Family Background We estimate the incarceration penalty while directly controlling for parental occupation scores in the specification:

$$OccScore_{it} = \alpha + \beta Incarcerated_{i,t-10} + X'_{it}B + \gamma ParentalOccScore_{i,t-20} + \epsilon_{it} \quad (2)$$

where i denotes the individual and t denotes the Census year. $ParentalOccScore_{i,t-20}$ denotes the highest occupation score of the indivdual's parents as observed in their childhood 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.

Figure 11 shows the estimated incarcerated penalty with and without direct controls of parental occupation-based 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 families in ϵ_{it} that are not captured by parents' occupation scores. If these unobserved characteristics are correlated with incarceration probabilities in young adulthood, our interpretation of the estimated β would be a combination of the selection on these unobservables and the causal impact of incarceration. To account for selection on these family-level unobservables, we further limit the sample to brothers within the same household, at least one of whom ends up incarcerated and at least one of whom does not.

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

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

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.

Figure 12 shows the estimated incarcerated penalty using the brothers' sample across different

specifications—our baseline specification with a proxy for parental occupation score, the specification where we directly control for parental occupation score, and finally, the specification where we include childhood home fixed effects. The results suggest that including childhood home fixed effects absorbs part of the incarceration penalty. These results have two interpretations: first, they may document some degree of negative selection on family unobservables into incarceration. 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.

In the latter period of our analysis, the penalty, even after controlling for childhood home fixed effects, grows. By 1940, the estimate based solely on the within childhood home comparison is roughly the same magnitude of our baseline penalty estimate for 1940. This is due to the fact that the baseline penalty in the brothers' sample is higher than that in the single-linked sample.

4.2 Pseudo event study around year of prison admission, inmate ledgers sample

In the full-count Census sample of incarcerated individuals, we examine the issue of selection by linking to the individuals' childhood Census. However, because Censuses occur only every ten years, the individuals in our sample are too young in the prior Census for us to observe their own socioeconomic characteristics. It is possible that the penalty we observe in post-incarceration Censuses reflect declines in labor market status just before the incarceration event. To explore selection on labor market trajectories just prior to the incarceration event, we use our sample of incarcerated individuals identified in state inmate ledgers and link them to both the Census prior to and after prison admission to examine the dynamics of their occupation-based income scores in the years leading up to incarceration.

We create two linked samples for each group of state prisoners admitted between Census years. A sample linked to the last Census taken before the incarceration event and another linked to the most recent Census following the incarceration event. We estimate the following equation:

$$Y_{it_0} = \alpha + \beta Incarcerated_{i,t_0+t} + X'_{it_0}B + \epsilon_{it_0} \tag{4}$$

where i denotes the individual and t_0 denotes the Census year. Yit_0 is the individual's labor market outcome in year t_0, and $Incarcerated_{i,t_0-t}$ 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 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 Census 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:

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

where $Incarcerated_{i,t_0-t}$ indicates whether the individual was admitted into prison in $t_0 - t$, and all other specifications are the same as in Equation 1. 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 Equation 4 in the same graph as the dynamics after incarceration as estimated in Equation 5. Figure C12 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 occupation-based 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 4 and Equation 5 on the same double-linked sample. Figure C13 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. Results in Figure 14 show that the penalty in labor market participation exhibits similar patterns as for

occupation-based 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 do not fully recover even ten years later. When we limit the sample to individuals with non-zero occupation-based income scores, there is no recovery in income as shown in Appendix Figure C11, suggesting that the recovery in income after incarceration is driven by the extensive margin, or participation in the labor market.

4.3 Interpretations of the penalty and comparison with modern estimates

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, prison 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.[11]

A third 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. Given the lack of centralized background checks during this period, and the fact that employer liability was a novel concept in the law as late as 1940, we hypothesize that the role for employer discrimination based on criminal history was likely small. However, our ability to directly probe this mechanism is limited due to data constraints.

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. Our findings are similar when examining percentiles of occupation-based income scores, and we also show a growing impact on the extensive margin or employment. Thus, we do not believe the deepening of the incarceration penalty over time to be driven solely by changes to the overall earnings distribution.

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

[11] Manufacturing, and textile manufacturing specifically, made up a plurality of prison labor during this time period (Poyker, 2019). Given the relative concentration of these industries spatially, the potential for mismatch with skills in one's civilian local labor market were large.

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. Mueller-Smith (2015) finds that individuals convicted of felonies and misdemeanors earn .34-.42 less log income for every year incarcerated, though the effect is not statistically significant for misdemeanors. As discussed previously, studies using judge IV designs typically find smaller or null effects on formal earnings or employment (Garin et al., 2023; Kling, 2006; Loeffler, 2013).

Studies that compare the earnings of formerly-incarcerated individuals to the never incarcerated are more comparable to our approach. Our estimates for the 1930s and 1940s 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.

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 (Mueller-Smith, 2015; 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. In the latter part of the period we study, we find that labor market participation plays a large role in explaining the penalty and particularly the dynamics in the penalty around the incarceration event, with a large drop in the years where the individual is likely still incarcerated and a gradual recovery that stabilizes at a significant penalty.

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).

We use data on millions of historical inmates to study the evolution of labor market gaps between formerly incarcerated and non-incarcerated men between during the rise of the modern prison system in the U.S., from 1880 to 1940. Building on existing recent literature, we 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 institu-

tions. This information allows us to identify the near universe of individuals incarcerated in every historical Census from this period. We supplement these data with administrative state prison ledgers, allowing us to follow intercensal prison admissions for 10 states, including some of the most populous–California, Texas, Florida, and New York.

In addition to describing the evolution of incarceration in the U.S. with these detailed data, we document the penalty in labor market outcomes associated with for formerly incarcerated individuals relative to the non-incarcerated population. We do so by linking historical inmates and the non-incarcerated across Census decades and estimating differences in labor market outcomes in the subsequent Census with basic controls such as age, literacy, geography and proxies for family background. We show that an incarceration penalty was present from the earliest period of our data and that it widened over time, reaching nearly 15% of occupation-based income scores by 1940 and over 25% of annual wages by 1940.

We then test two potential selection-based explanations for the penalty—first, that it reflects differences in family backgrounds between those who do and do not end up incarcerated; and second, that individuals were on a declining labor market trajectory prior to their admission into prison. The penalty is robust to direct controls for parental income and family fixed effects in samples of individuals additionally linked back to their childhood Census, including in samples of double-linked brothers. These results suggest that selection into incarceration based on family characteristics is unlikely to explain the penalty we estimate. Turning to data from state inmate ledgers linked to the Census, which allow us to capture flows into state prisons in between Census decades, we examine whether selection on labor market trajectories prior to the incarceration event can explain our results, using a "pseudo-event study" around individuals' year of prison admission. We find no evidence that inmates were on a negative trajectory prior to the incarceration event.

Our penalty estimates, particularly towards the end of our period of study are comparable to modern estimates. We hypothesize that the penalty reflects the persistent consequences of missed civilian labor market experience for young men, a mechanism consistent with recent evidence on the long-run employment and earnings effects of first-time criminal charges (see Agan et al. (2022)). 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.

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 (1979). "Trends in incarceration in the United States since 1880: A summary of reported rates and the distribution of offenses". In: *Crime & Delinquency* 25.1, pp. 9–41.

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.

Finlay, Keith, Matthew Gross, Carl Lieberman, Elizabeth Luh, and Michael G Mueller-Smith (2023). *The impact of criminal financial sanctions: A multi-state analysis of survey and administrative data*. Tech. rep. National Bureau of Economic Research.

Finlay, Keith and Michael Mueller-Smith (2021). "Justice-involved individuals in the labor market since the Great Recession". In: *The ANNALS of the American Academy of Political and Social Science* 695.1, pp. 107–122.

Finlay, Keith, Michael Mueller-Smith, and Jordan Papp (2022). "The criminal justice administrative records system: A next-generation research data platform". In: *Scientific Data* 9.1, p. 562.

Finlay, Keith, Michael Mueller-Smith, and Brittany Street (2023). "Criminal Justice Involvement, Self-Employment, and Barriers in Recent Public Policy". In: *Journal of Policy Analysis and Management* 42.1, pp. 11–34.

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

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.

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.

Moehling, Carolyn and Anne Morrison Piehl (2009). "Immigration, crime, and incarceration in early twentieth-century America". In: *Demography* 46.4, pp. 739–763.

Moehling, Carolyn M and Anne Morrison Piehl (2014). "Immigrant assimilation into US prisons, 1900–1930". In: *Journal of Population Economics* 27, pp. 173–200.

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.

Sabol, William J (1989). “Racially Disproportionate prison populations in the United States: An overview of historical patterns and review of contemporary issues”. In: *Contemporary Crises* 13, pp. 405–432.

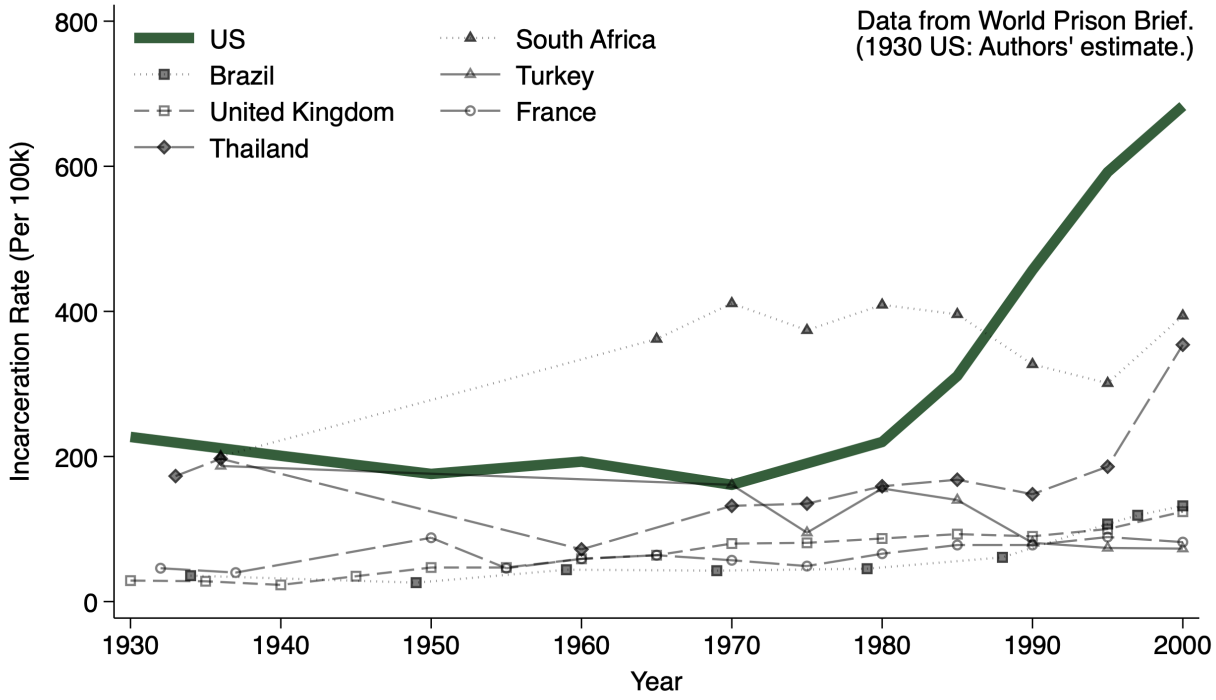
Waldfogel, Joel (1994). “The effect of criminal conviction on income and the trust" reposed in the workmen"”. In: *Journal of Human Resources*, pp. 62–81.

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*.

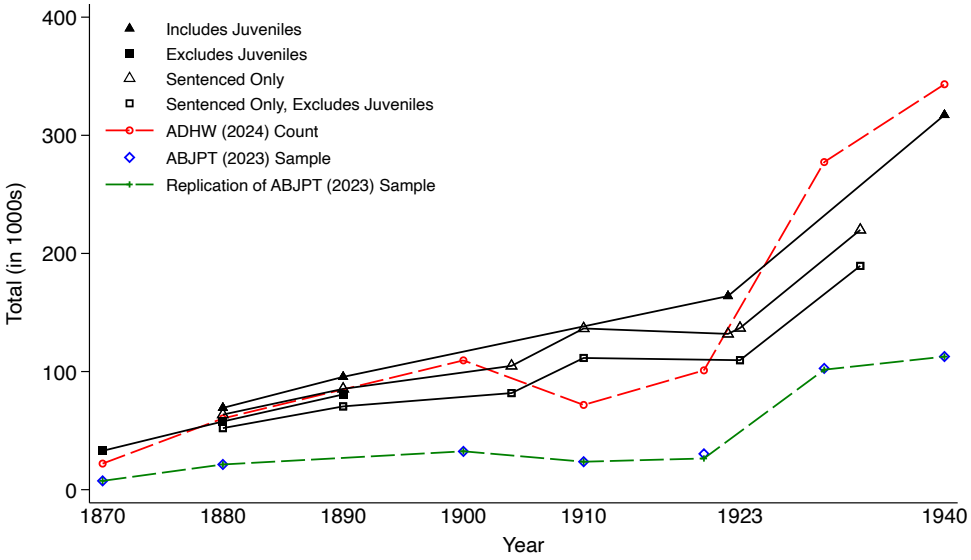
Figure 1: Incarcerated per 100,000: U.S. vs. other major countries, 1930-2000



Notes: This figure plots incarceration rates per 100,000 between 1930 and 2000 for countries in the top 25 of world population and top 100 of incarceration rates today. Additionally, incarceration rates are shown for the U.K. and France.

Sources: Data on incarceration rates come from the World Prison Brief (World Prison Brief, Institute for Crime & Justice Policy Research, 2024) with the exception of the 1930 U.S. estimate, which is the authors' own, based on the methodology described in Section 2.1 applied to full-count Census data from 1930 from Ruggles, Flood, et al. (2021).

Figure 2: Incarcerated population in the U.S., 1870-1940



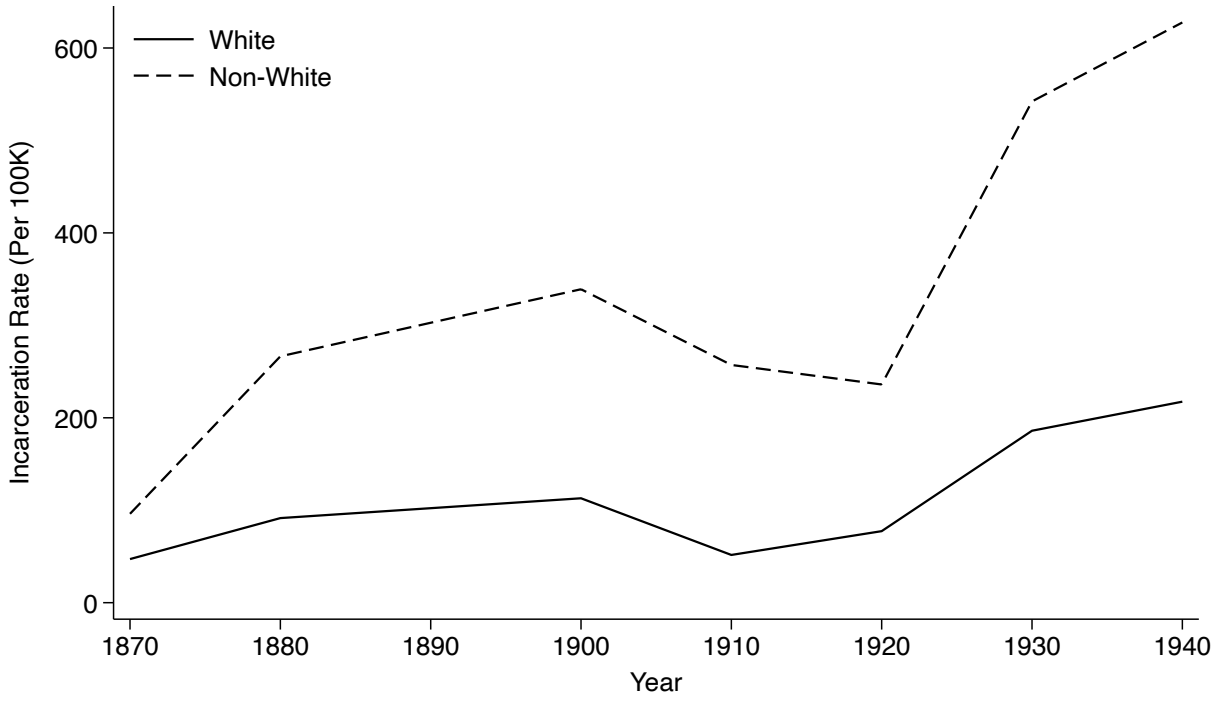
(a) Incarcerated Population Across Sources



(b) Incarceration Rate Estimates

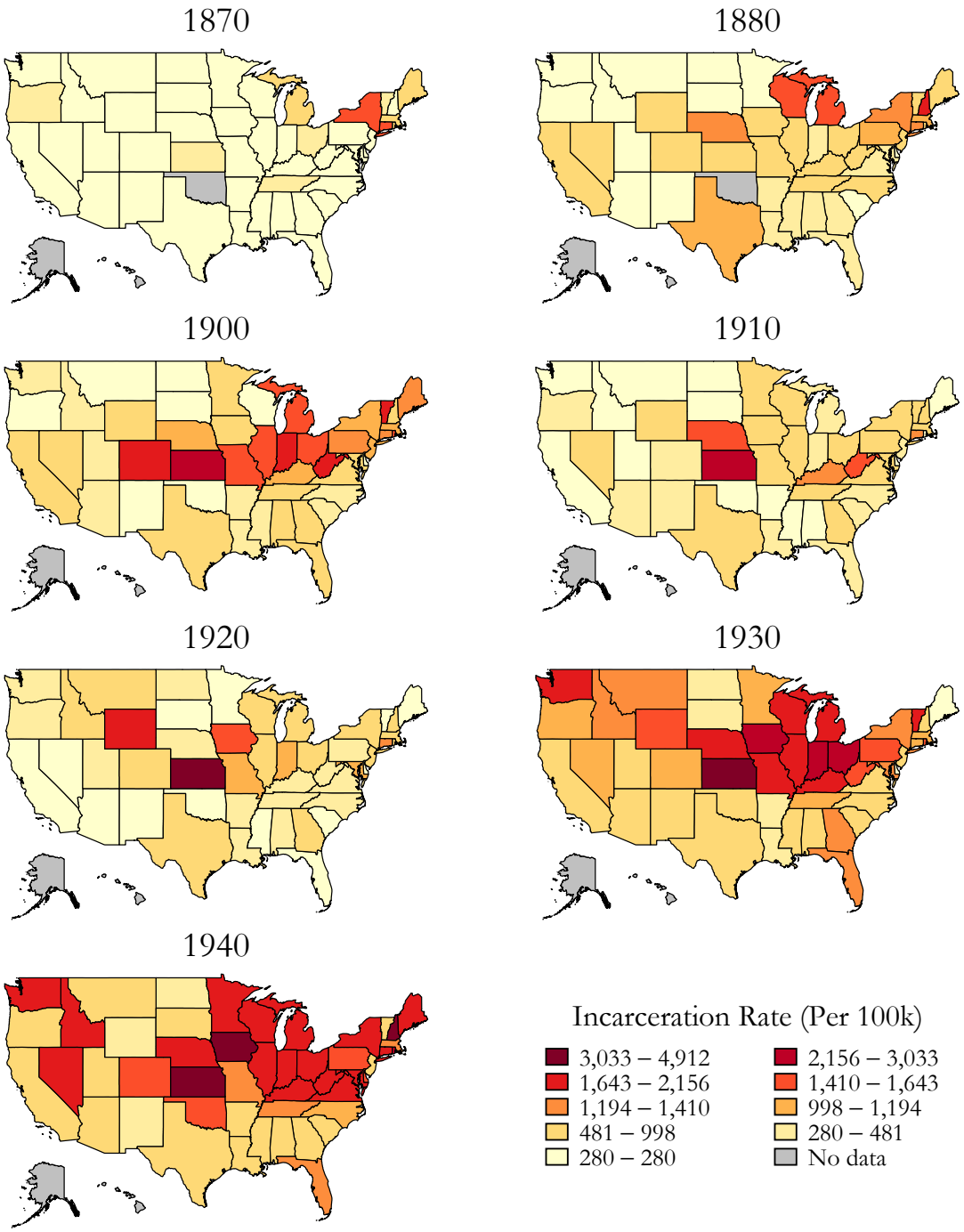
Notes: Panel (a) of this figure plots various estimates of the number of incarcerated individuals over time. Panel (b) plots the authors' estimates of incarceration rates per 100,000. Counts by sentence and juvenile status come from an aggregation of special Census reports by the Bureau of Justice Statistics (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 3: Incarceration rates in the U.S., 1870-1940



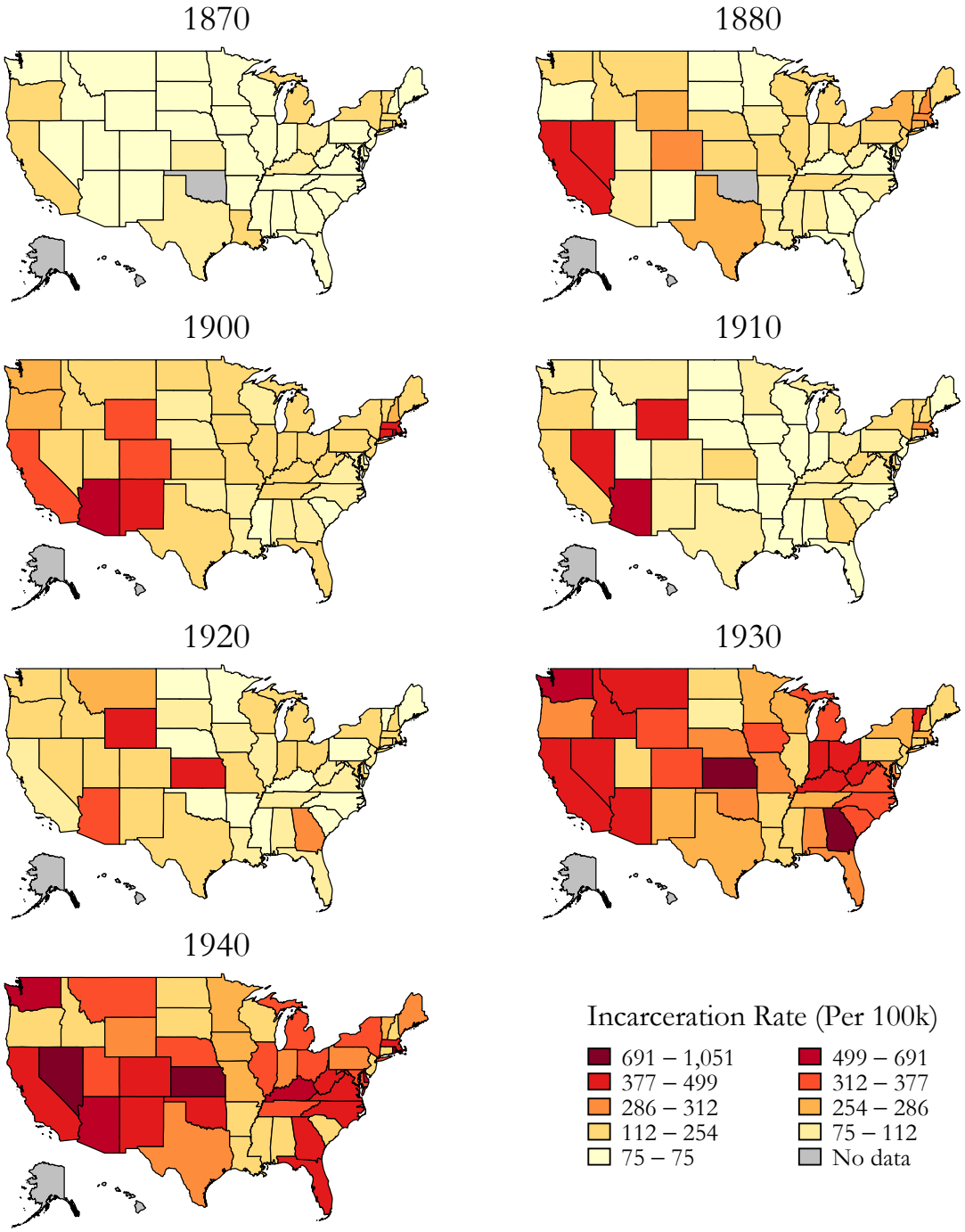
Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time by race using historical full-count Census data.
Sources: NBER restricted full count U.S. Census: 1870-1940

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



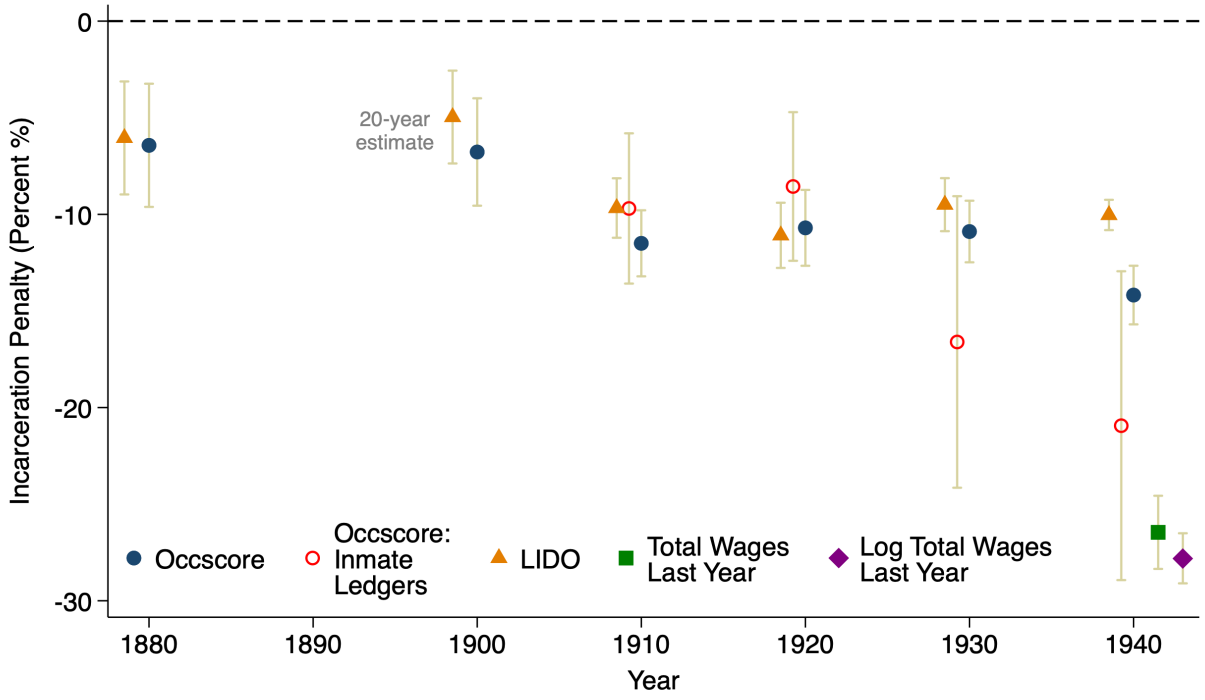
Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time and by state for the male, non-white population using historical full-count Census data.
Sources: NBER restricted full count U.S. Census: 1870-1940

Figure 5: White Men's Incarceration Rates by State, 1870-1940



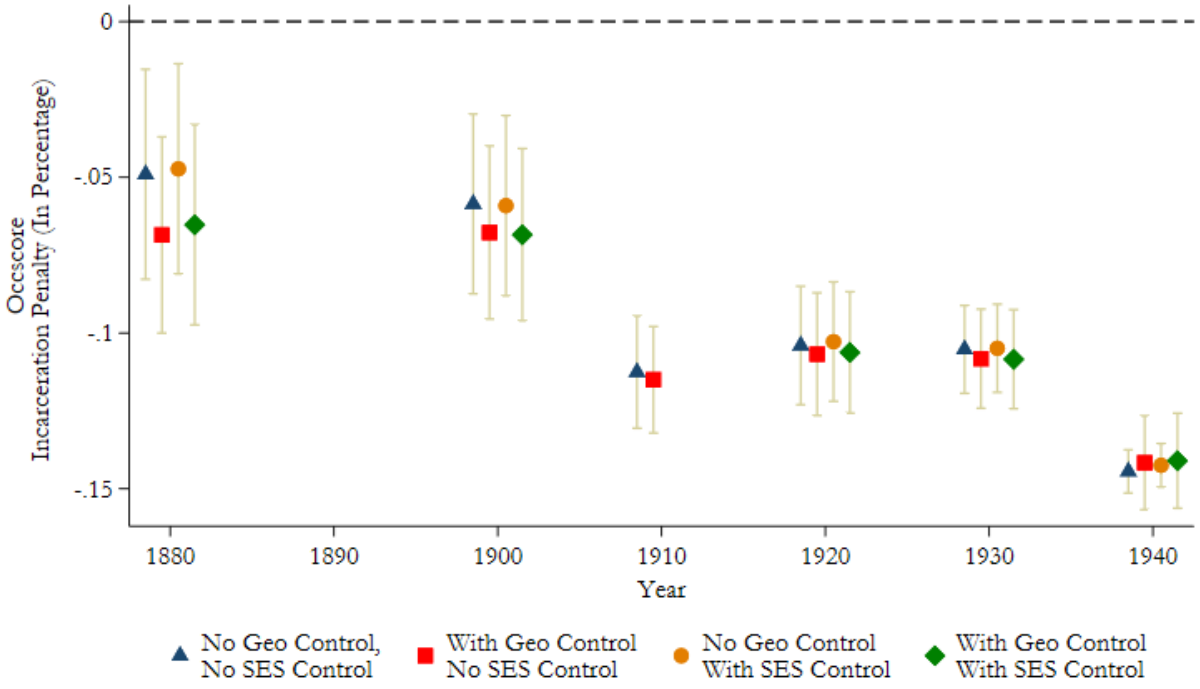
Notes: This figure plots the authors' estimates of incarceration rates per 100,000 over time and by state for the male, white population using historical full-count Census data.
Sources: NBER restricted full count U.S. Census: 1870-1940

Figure 6: Incarceration penalty, 1880-1940



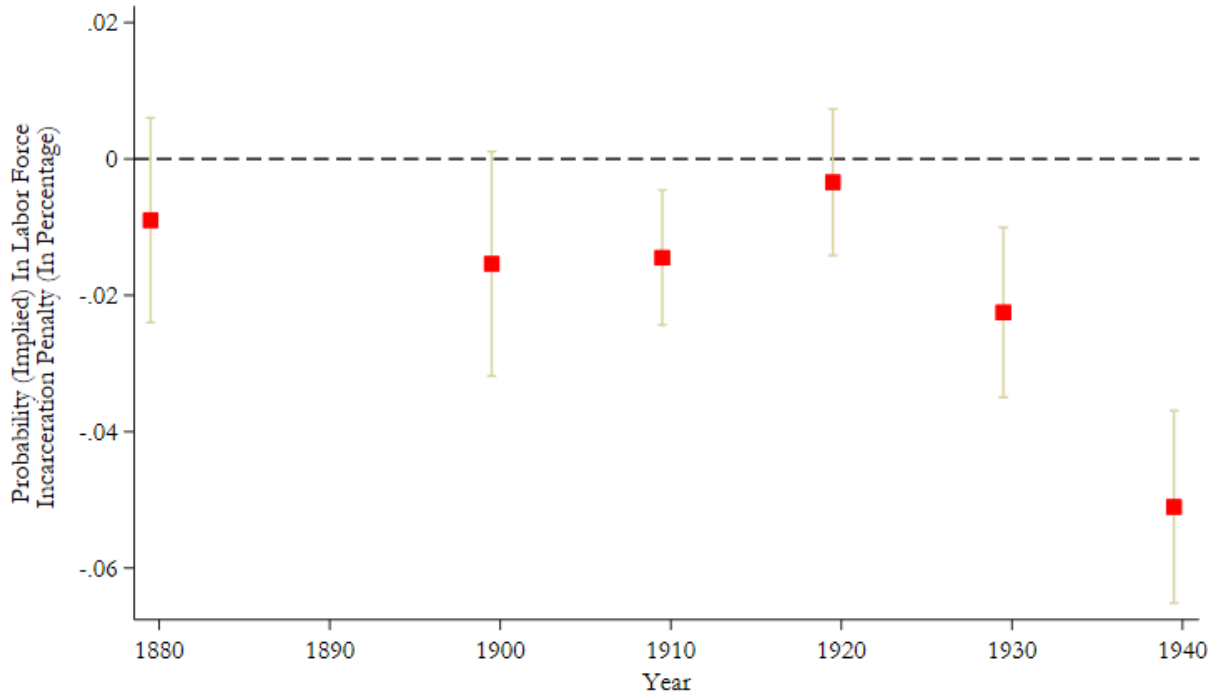
Notes: This figure plots β from Equation 1, which we interpret as the "incarceration penalty," estimated using data from the historical Census and state inmate ledgers. The coefficient is normalized by the mean of the dependent variable in Census year t among all linked men who were not incarcerated in Census year $t-10$, and represents the percent difference in income score or income in Census year t between men incarcerated in the previous Census compared to men not incarcerated in that Census. Each Census year regression include county of residence fixed effects, controls for farm residence, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1870-1940; inmate ledgers; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 7: Incarceration penalty, alternative specifications



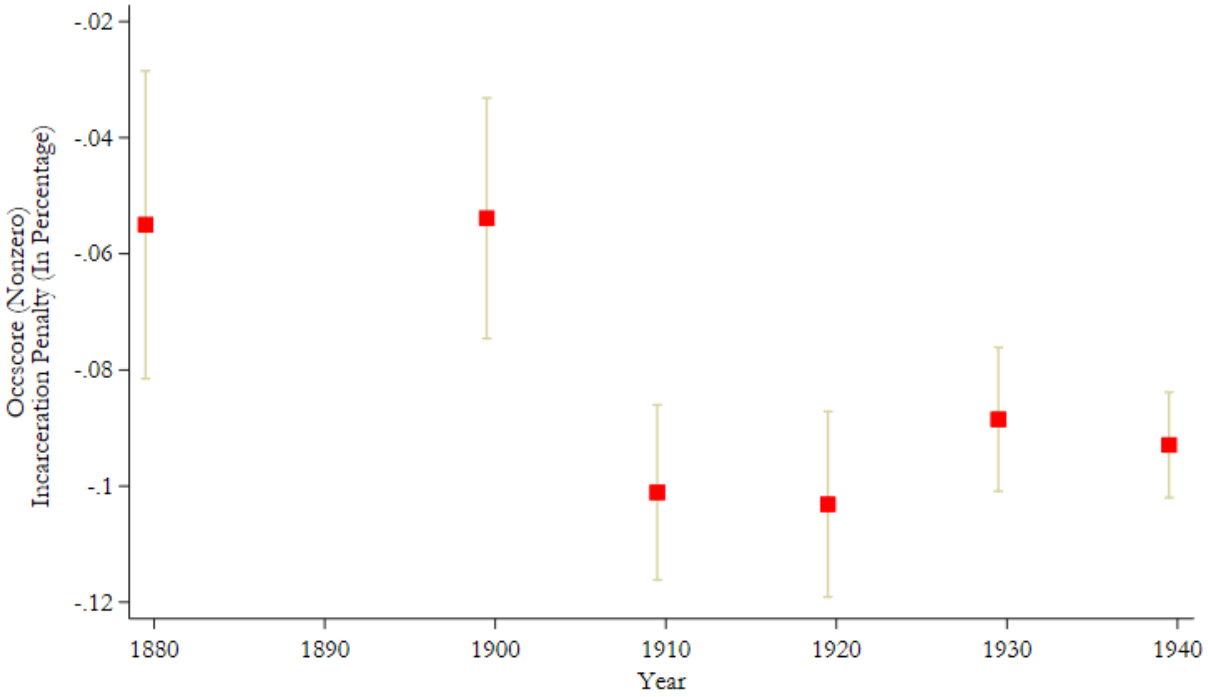
Notes: This figure plots β from Equation 1, which we interpret as the "incarceration penalty," estimated using data from the historical Census. The coefficient is normalized by the mean of occscore in decade t among all linked males that were not incarcerated in decade $t-1$, and represents the percent 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 occupation-based 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: Full count U.S. Census data accessed on the NBER server: 1870-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 8: Incarceration penalty on having a gainful occupation, 1910-1940



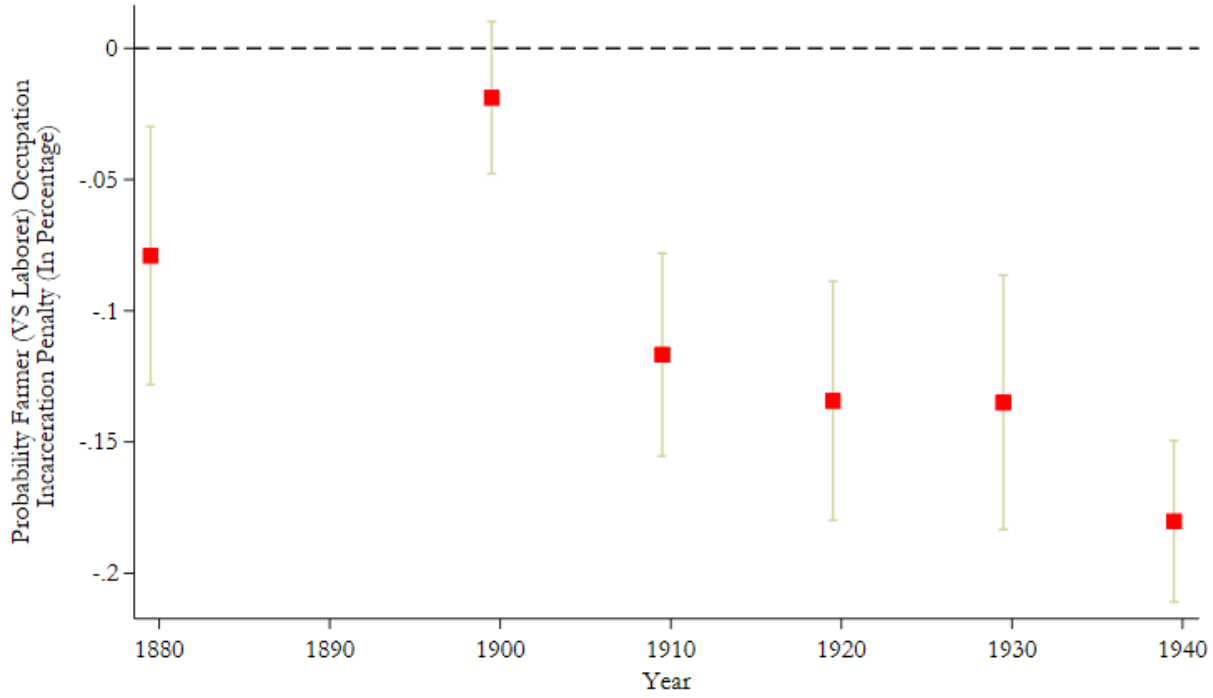
This figure calculates the incarceration penalty using our constructed labor force participation indicator (whether an individual reported a gainful occupation) as the outcome. See figure notes for Figure 6 for more information on the specification and sample. Each Census year regression include county of residence fixed effects, controls for farm residence, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1870-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 9: Incarceration penalty, conditional on positive occupation-based income score, 1910-1940



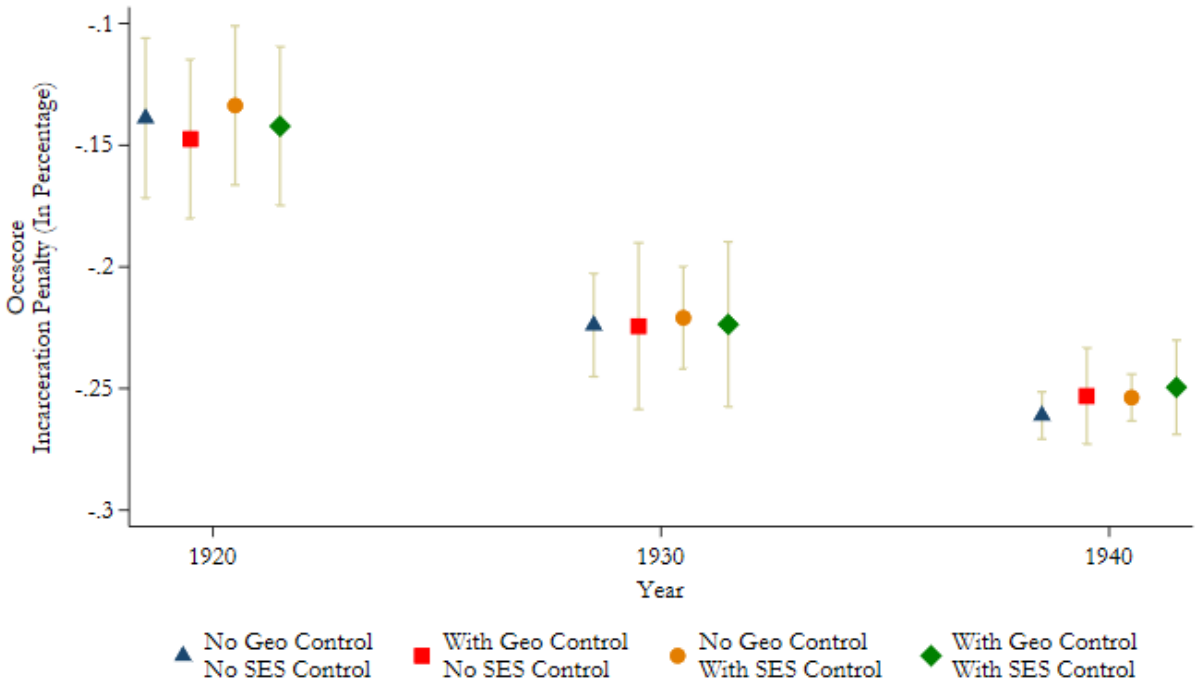
Notes: This figure plots the estimated incarceration penalty on occupation-based income scores, restricting the sample to those that have a positive score. See figure notes for Figure 6 for more information on the specification and sample. Each Census year regression include county of residence fixed effects, controls for farm residence, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1870-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 10: Farm ladder incarceration penalty, 1880-1940



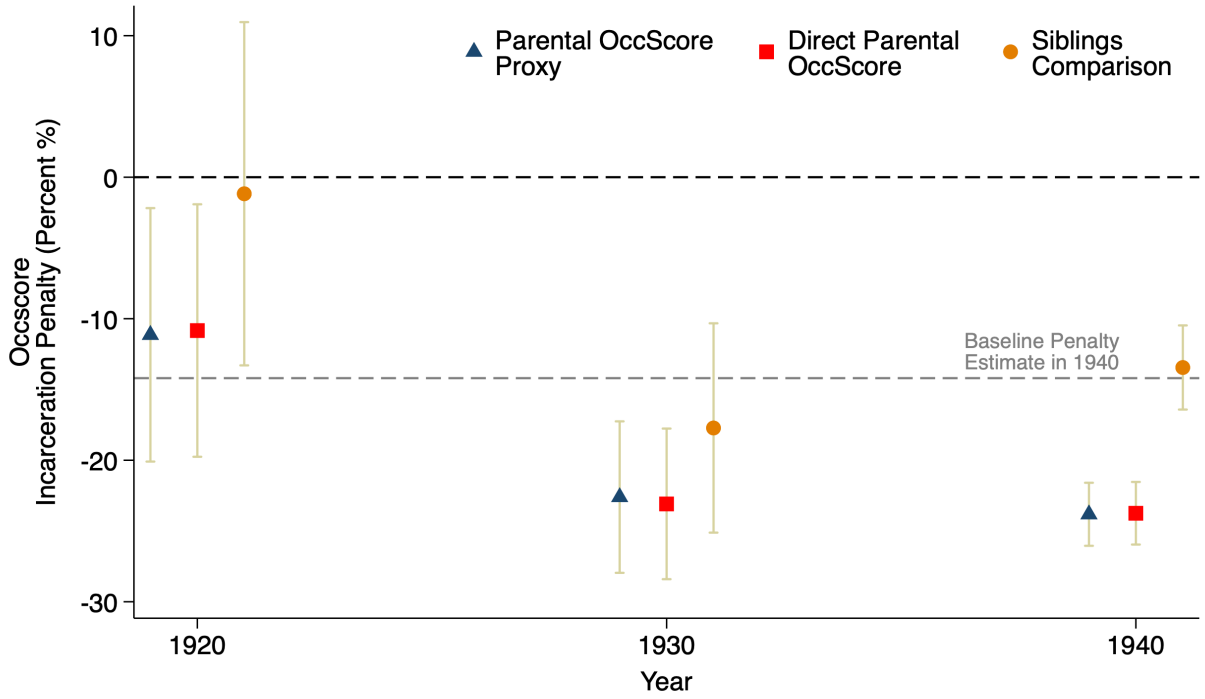
Notes: This figure plots the estimated incarceration penalty on an indicator variable for having a "Farmer" occupation as opposed to a "Farm Laborer" occupation, conditional on reporting a farm-related occupation. See figure notes for Figure 6 for more information on the specification and sample. Each Census year regression include county of residence fixed effects, controls for farm residence, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1870-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 11: Incarceration penalty in double-linked sample with control for parent occupation-based income score, 1920-1940



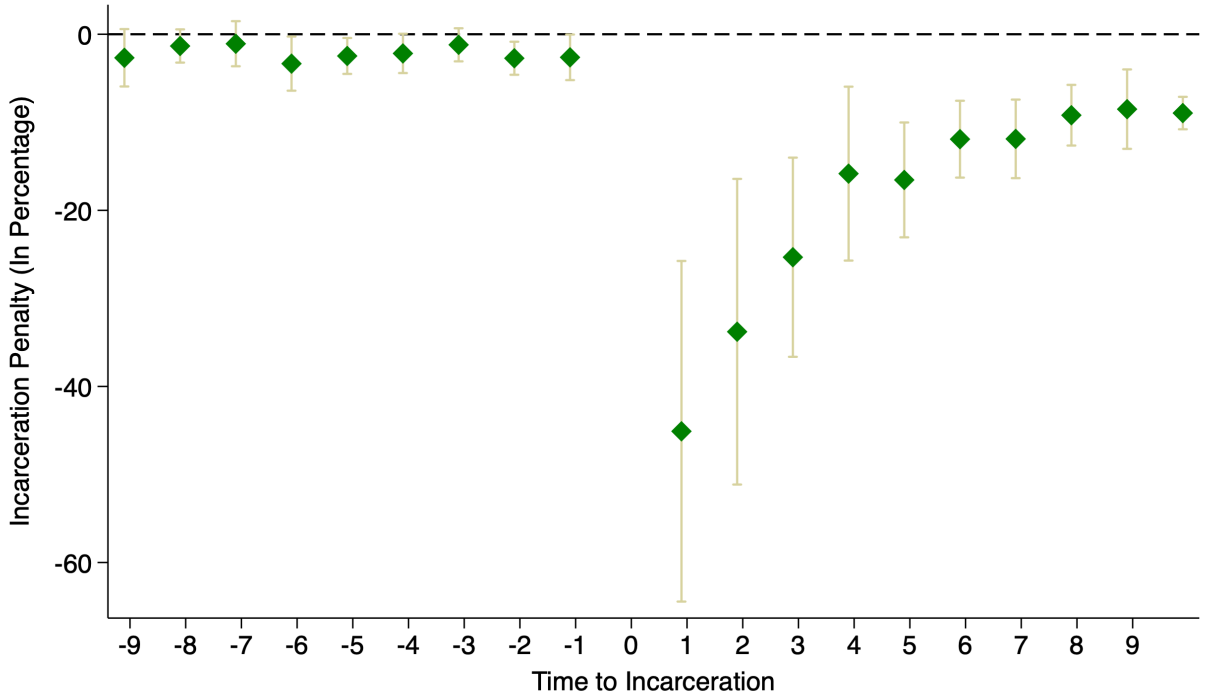
Notes: This figure plots estimates of the incarceration penalty on occscore from Equation 2, which restricts the sample to men aged 16-26 in decade $t-10$ who link back to their childhood Census in $t-20$ in addition to the Census where we measure occupation-based income score in year t. The figure shows 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; "SES Control" indicates the inclusion of individual level controls for the largest occupation-based income score of the individual's parents, identified in their household in Census year $t-20$. PAll regressions include controls for farm residency, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1910-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 12: Incarceration penalty in double-linked sample of brothers, 1920-1940



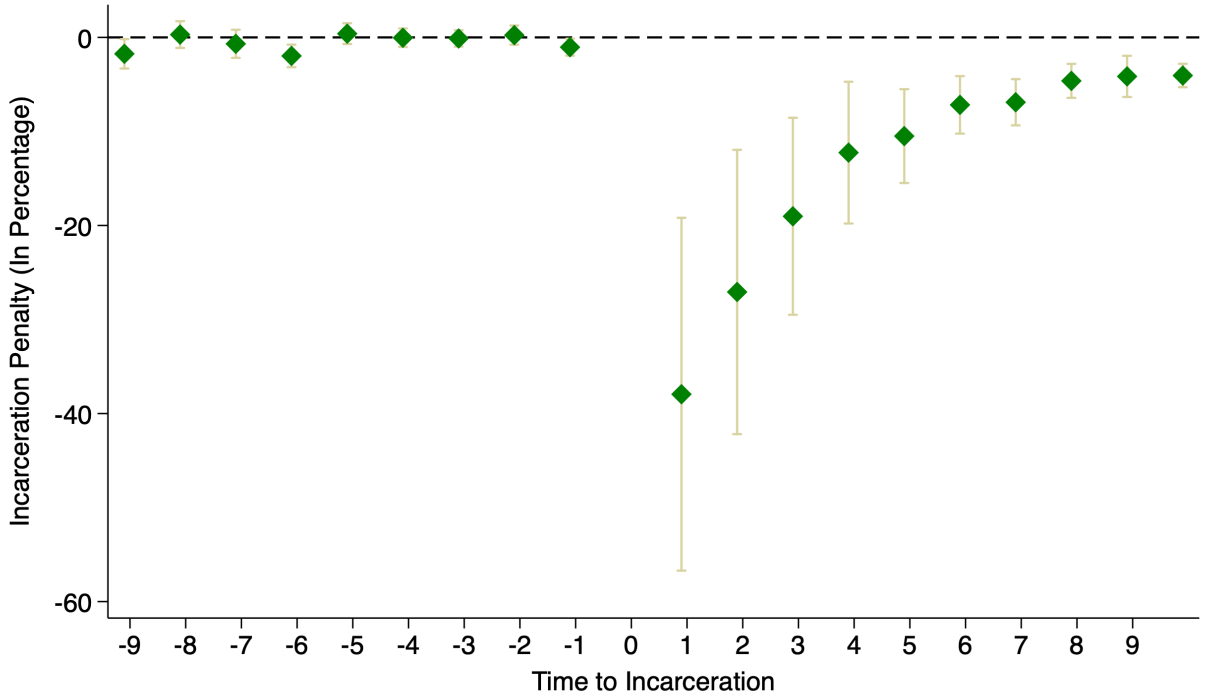
Notes: This figure plots estimates of the incarceration penalty on occupation-based income scores from Equations 1 (Parental OccScore Proxy), 2 (Direct Parental OccScore), and 3 (Siblings Comparison), while holding fixed the sample, namely the double-linked brothers sample described in Section 4.1, which 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: Full count U.S. Census data accessed on the NBER server: 1910-1940; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

Figure 13: Incarceration penalty by year since prison admission, 1910-1940



Notes: This figure plots β for occupation-based income scores from Equation 4 for every number of years before the prison admission and β from Equation 5 for every number of years after the prison admission. The coefficient is normalized by the mean of the occupation-based 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: Full-count decennial Censuses from 1910-1940 and inmate ledger records.

Figure 14: Dynamics of labor market participation penalty, 1910-1940



Notes: This figure plots β for implied labor force participation from Equation 4 for every number of years before the prison admission and β from Equation 5 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: Full-count decennial Censuses from 1910-1940 and inmate ledger records.

Table of Contents

Appendix A Identifying the incarcerated using full count Census data **40**
A.1 Incarceration classification in 1900, 1920-1940 Censuses 40
A.2 Incarceration classification in 1880, 1900, 1920-1940 Censuses 41
A.3 Incarceration classification in 1910 Censuses 42
A.4 Incarceration classification in 1850-1870 Censuses 42

Appendix B Summary statistics for linked samples **43**
B.1 Double Linked Sample 43
B.2 Double Linked Sample 43

Appendix C Additional results **43**
C.1 Penalty estimates by racial group 43
C.2 Penalty estimates by region and state 43
C.3 Robustness to Linking Methods 53
C.4 Alternative Penalty Measures 53

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.[12]

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 Penalty estimates by racial group

C.2 Penalty estimates by region and state

[12] 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)

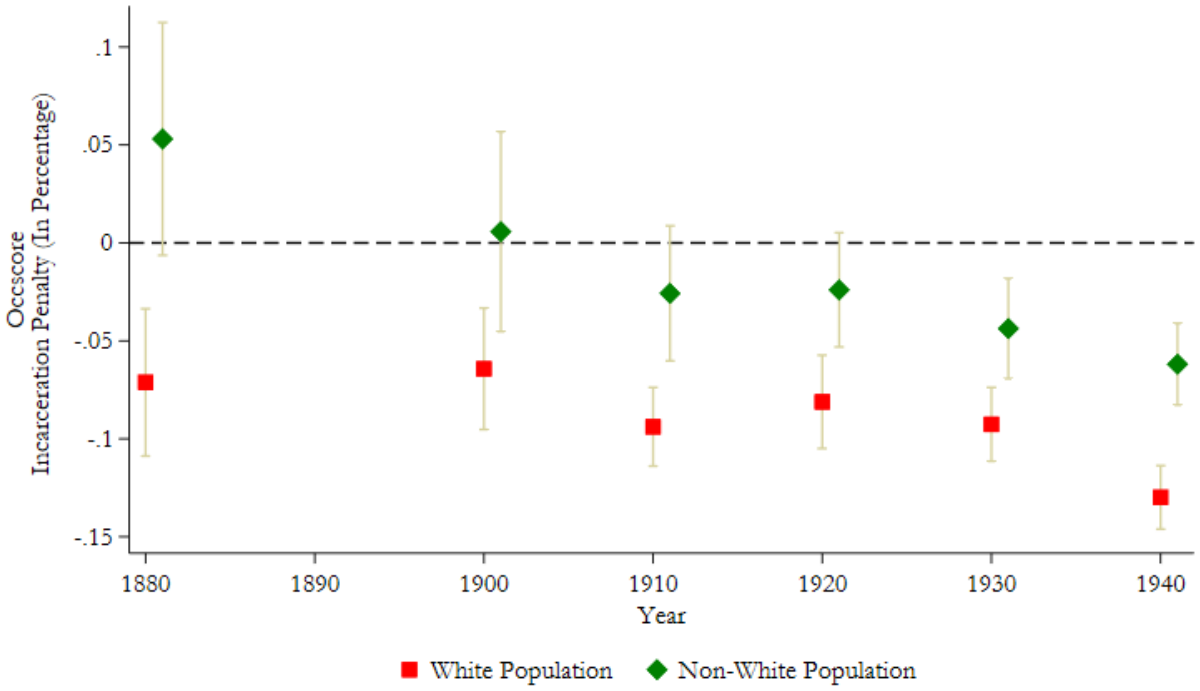
Table B6: Summary Statistics of 1930-1940 Linked Sample

	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

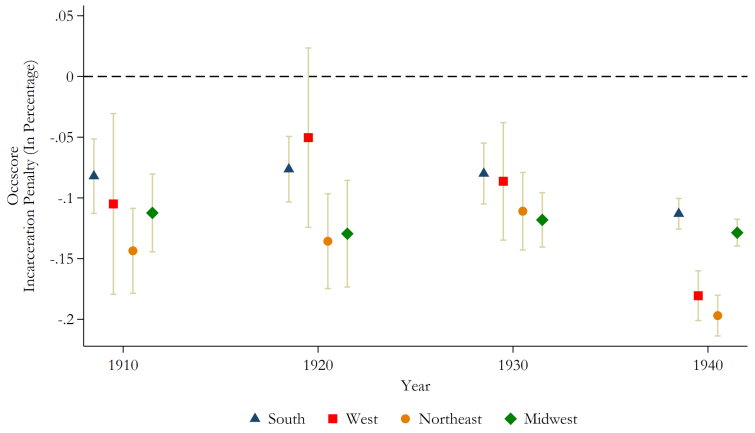
Notes: This table compares the means of the seven demographic variables for men in 1930, broken down by race (white vs. non-white), whether they were linked to 1940 using the links provided in Abramitzky, Boustan, Pérez, et al., 2020 (All vs Matched). Panel A restricts to the non-incarcerated population 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 penalty, 1880-1940

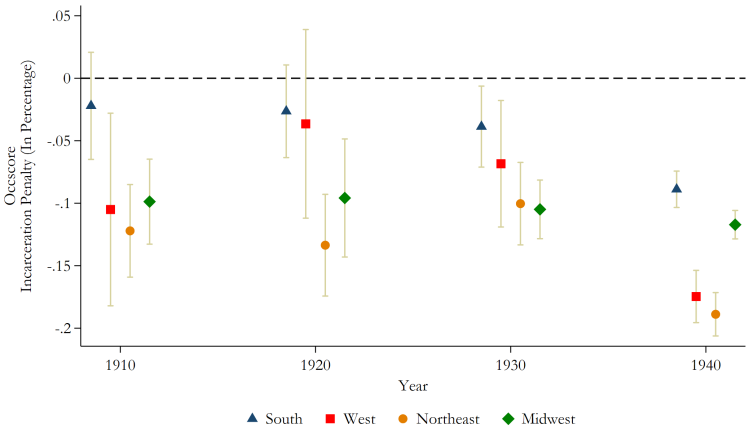


Notes: This figure plots β from Equation 1, which we interpret as the "incarceration penalty," estimated using data from the historical Census, separately by racial group. The coefficient is normalized by the mean of the dependent variable in Census year t among all linked men who were not incarcerated in Census year $t-10$, and represents the percent difference in income score or income in Census year t between men incarcerated in the previous Census compared to men not incarcerated in that Census. Each Census year regression include county of residence fixed effects, controls for farm residence, literacy, and a quadratic in age.
Sources: Full count U.S. Census data accessed on the NBER server: 1870-1940; inmate ledgers; Census Linking Project links from Abramitzky, Boustan, Pérez, et al. (2020).

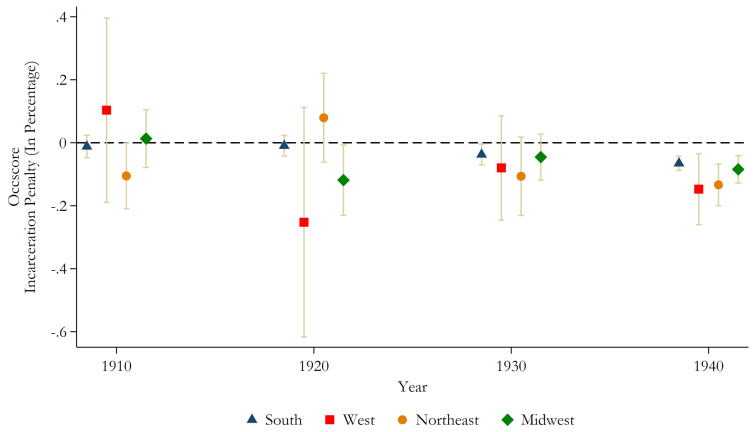
Figure C2: Incarceration Penalty by Region, 1910-1940



(a) All Population



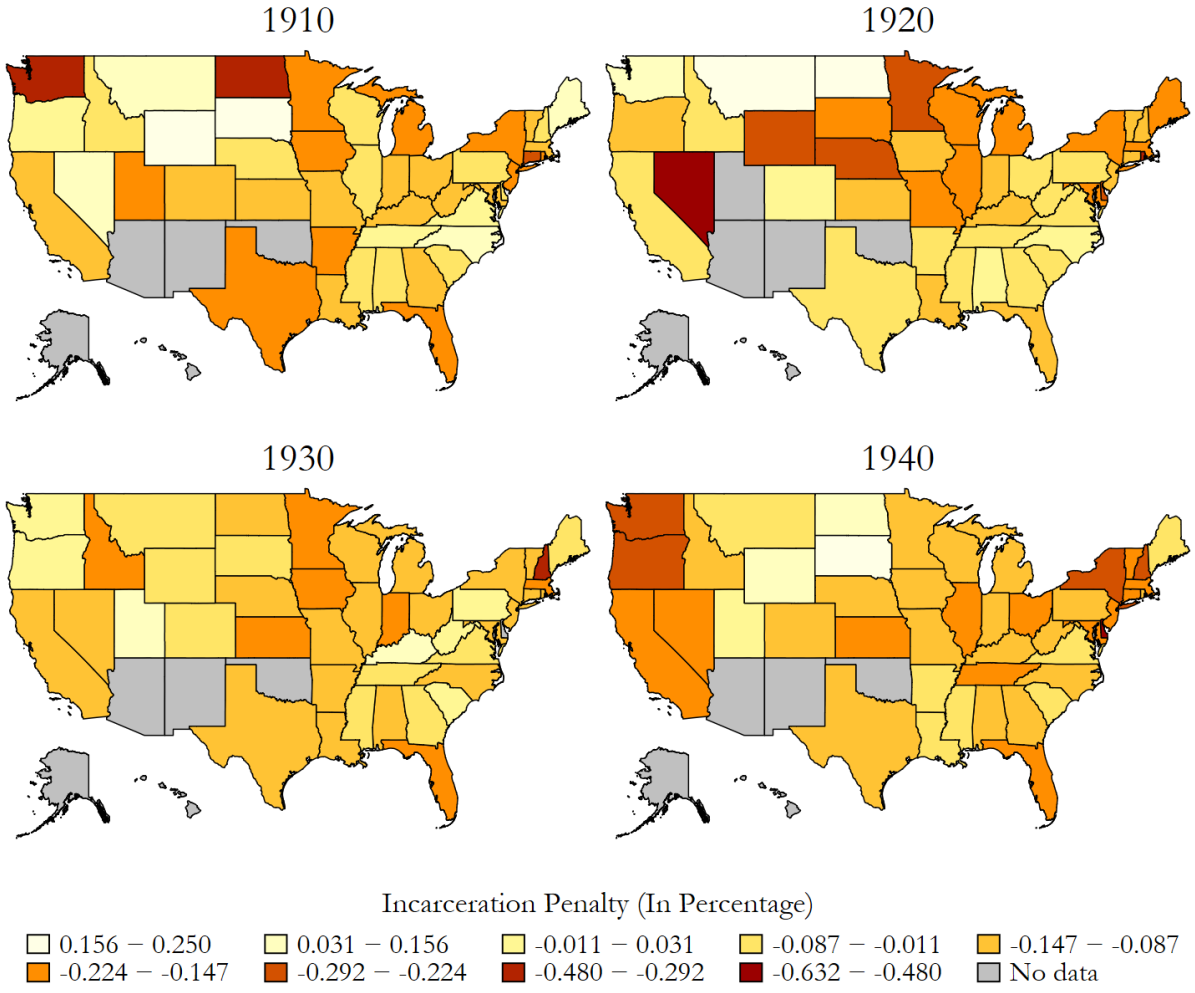
(b) White Population



(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 occupation-based 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 C3: Incarceration Penalty by State, 1910-1940

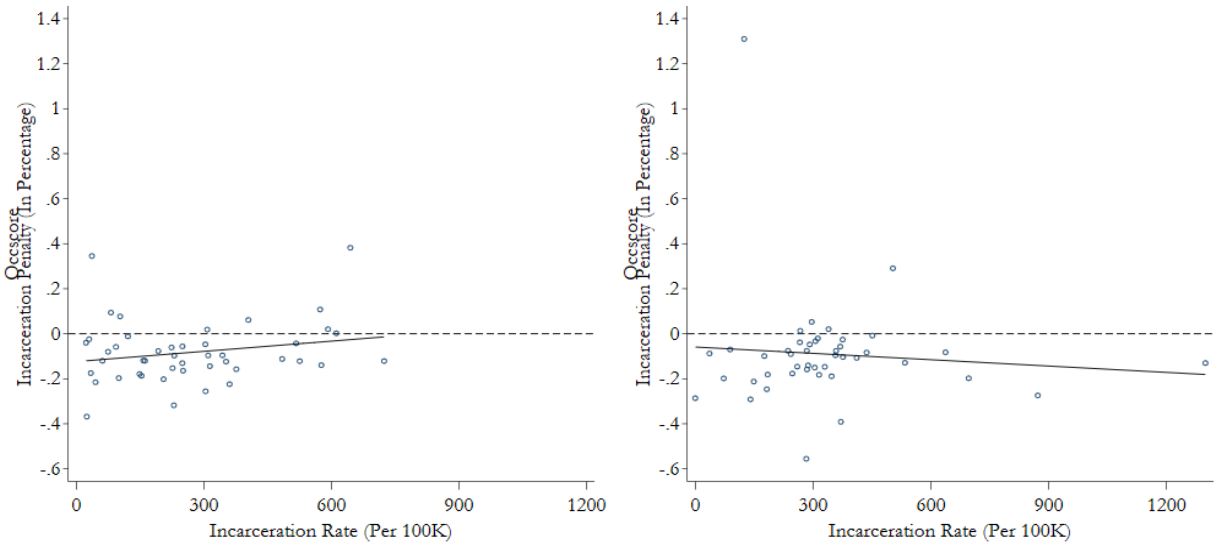


Notes: This figure plots ranges of estimates of the incarceration penalty on occscore for each state over time. There are no county of residence fixed effects, nor average parental occupation-based income score controls. All regressions include controls for farm residency, literacy, and a quadratic in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

C.3 Robustness to Linking Methods

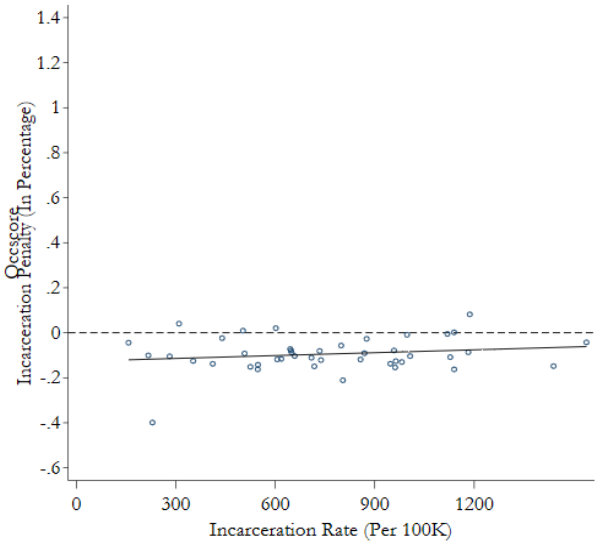
C.4 Alternative Penalty Measures

Figure C4: Correlation Between Incarceration Penalty and Rate, 1910-1940

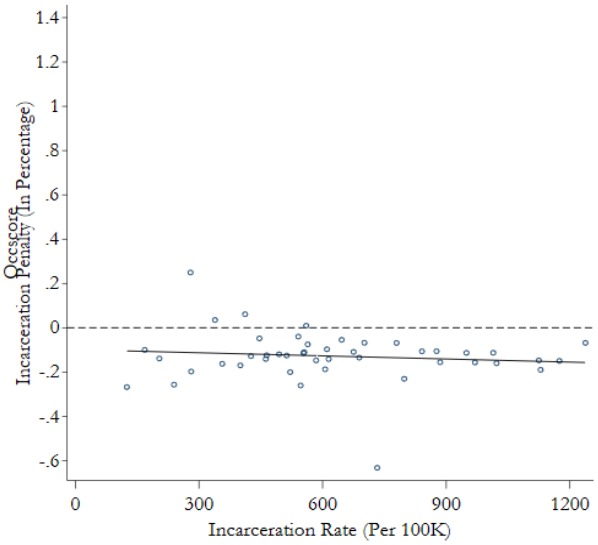


(a) 1910

(b) 1920



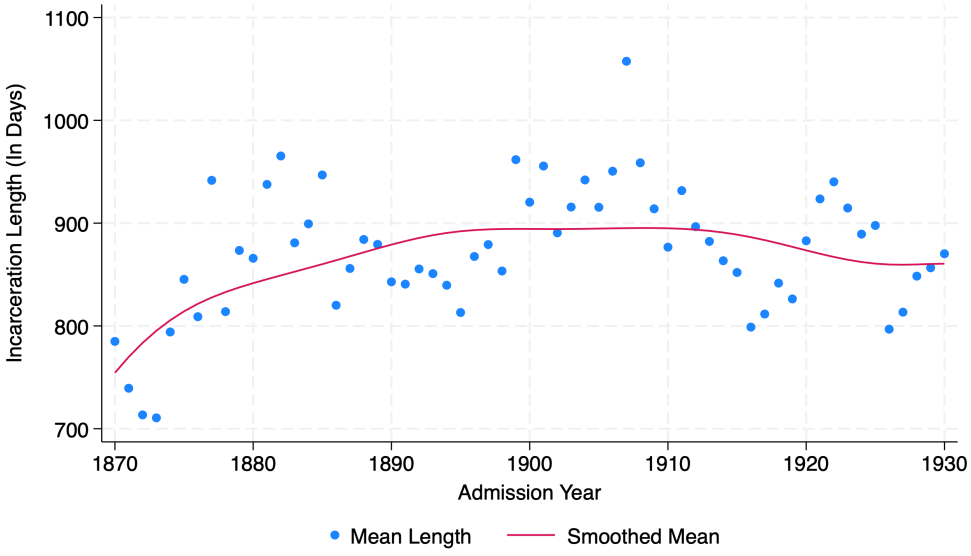
(c) 1930



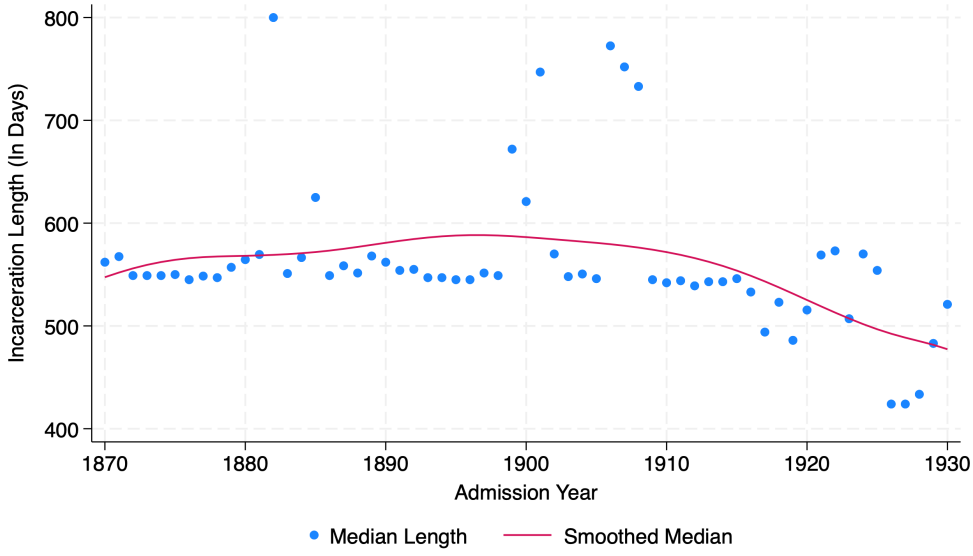
(d) 1940

Notes: This figure presents a scatter plot of state-level penalties (which are mapped out in Figure C3) 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 C5: Incarceration Length in Missouri, 1870-1930



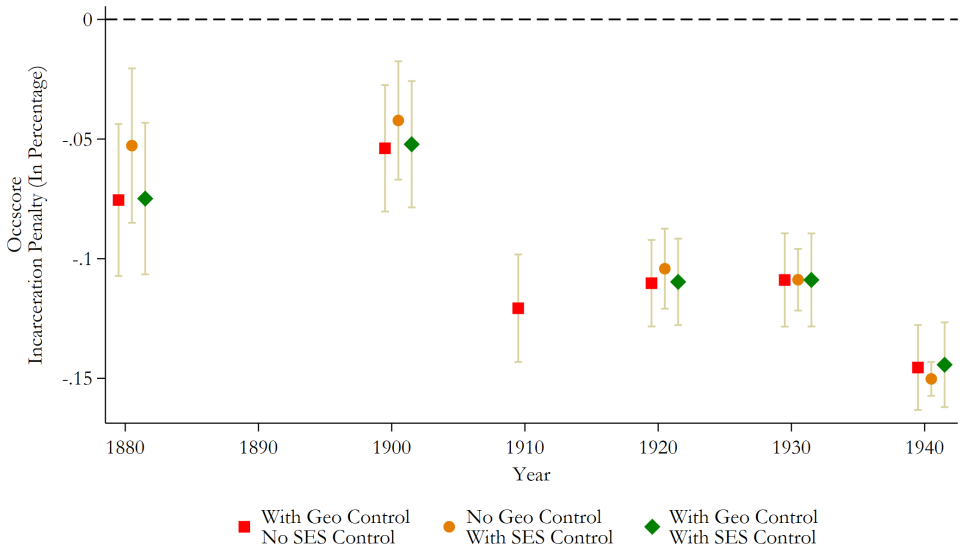
(a) Average



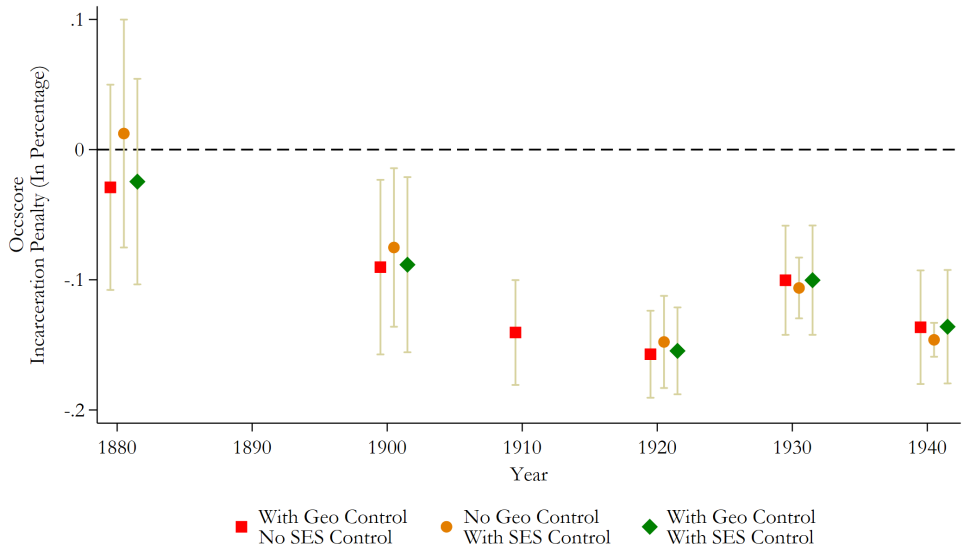
(b) Median

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 C6: Occscore Incarceration Penalty With Different Linking Methods, 1880-1940



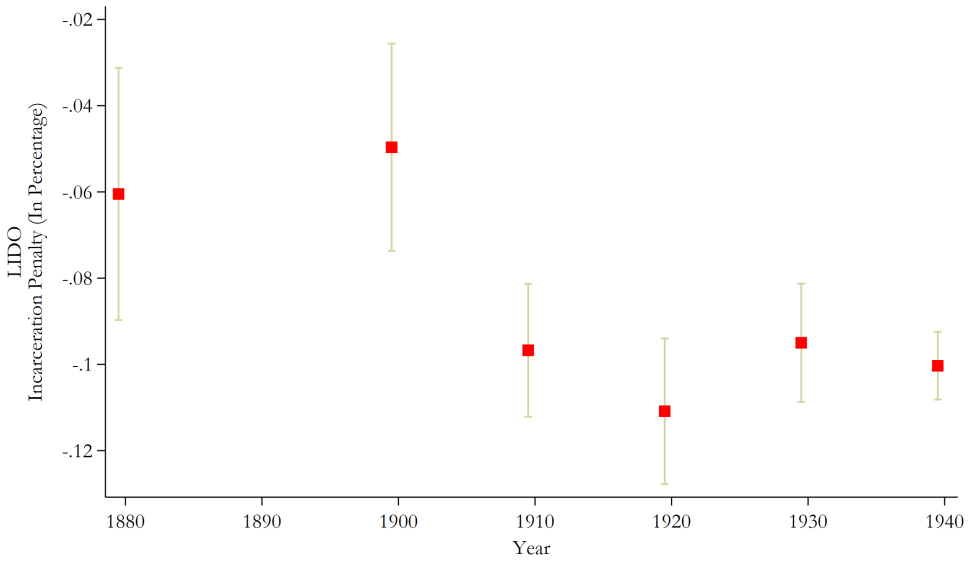
(a) Census Tree Linking



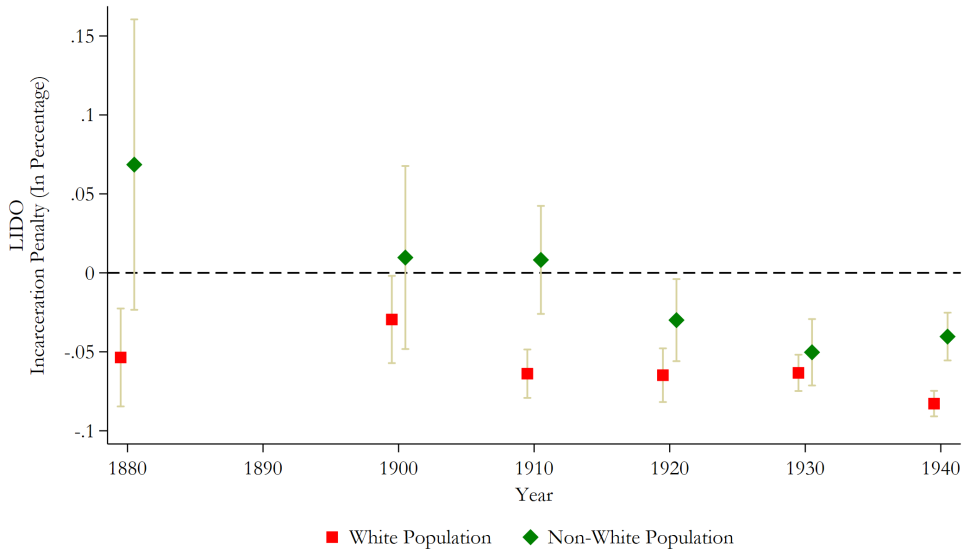
(b) MLP Linking

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, Fitch, 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 occupation-based 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, Fitch, et al. (2019)

Figure C7: LIDO Incarceration Penalty, 1880-1940



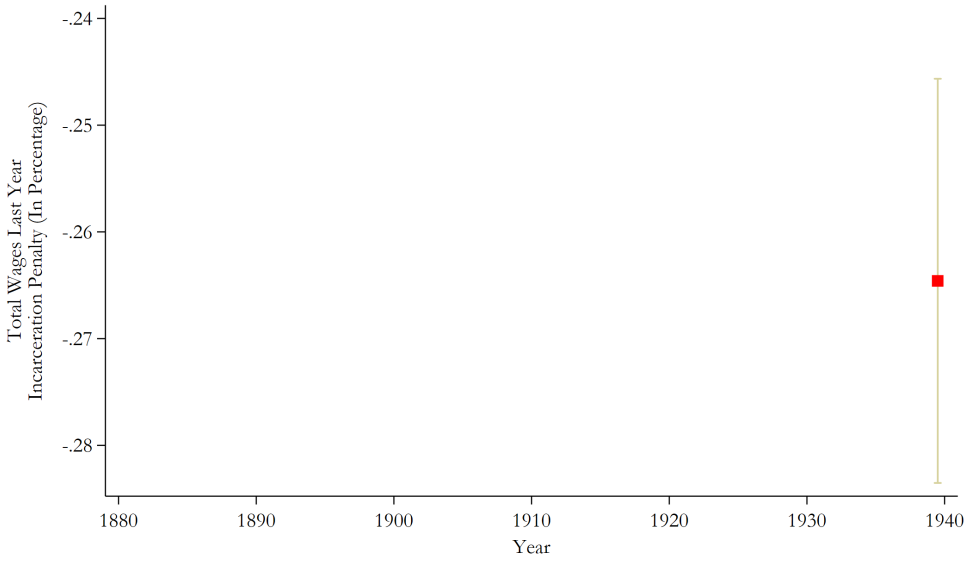
(a) Full Sample



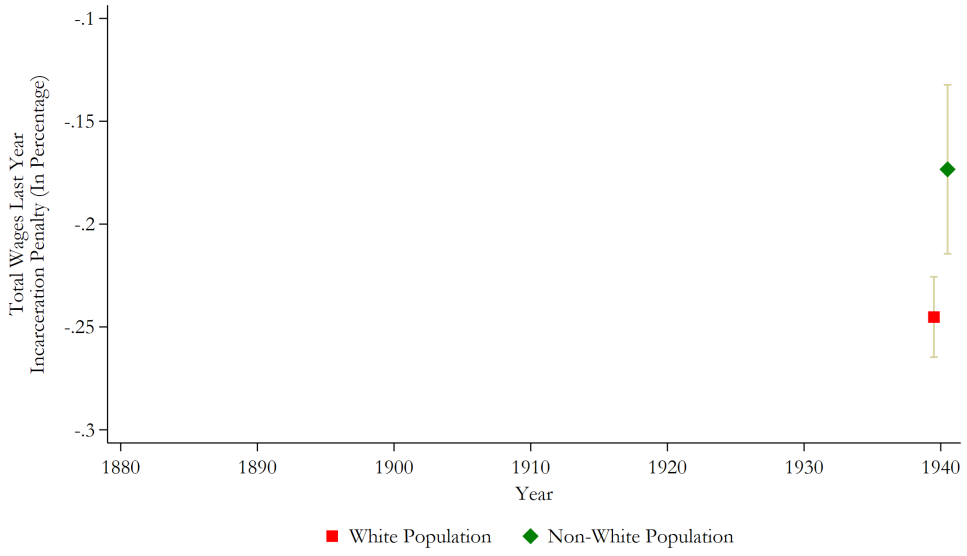
(b) By Race

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 occupation-based 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 C8: Annual Earnings Last Year Incarceration Penalty, 1880-1940



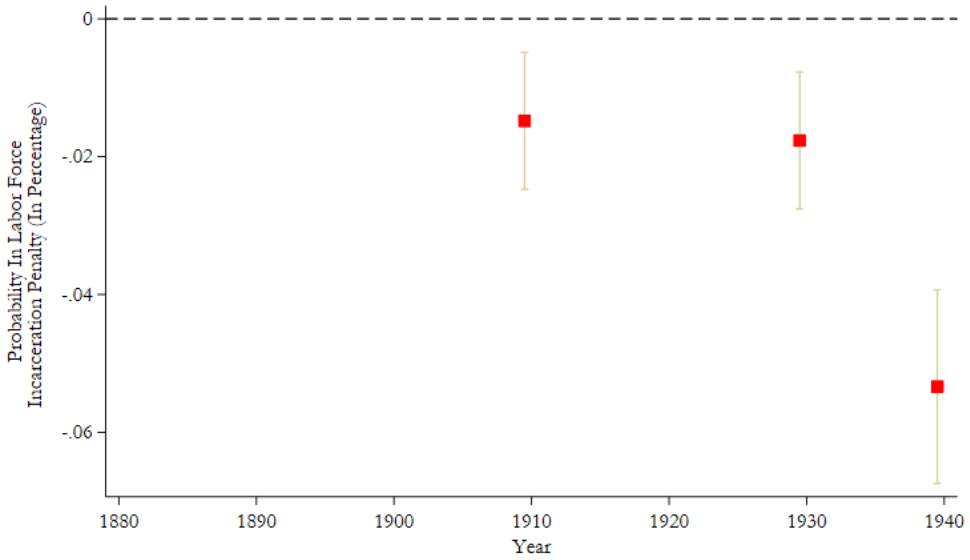
(a) Full Sample



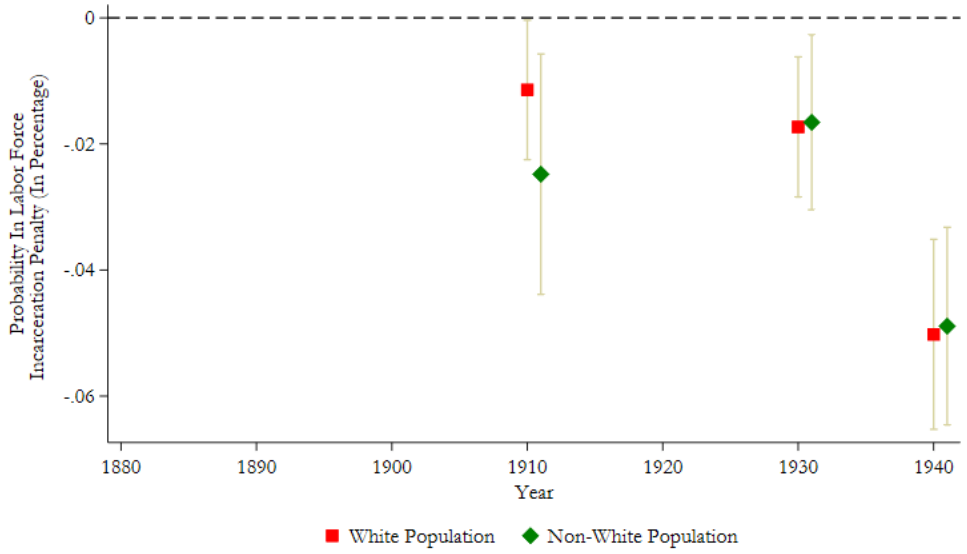
(b) By Race

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 occupation-based 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 C9: Official Labor Force Participation Incarceration Penalty, 1910-1940



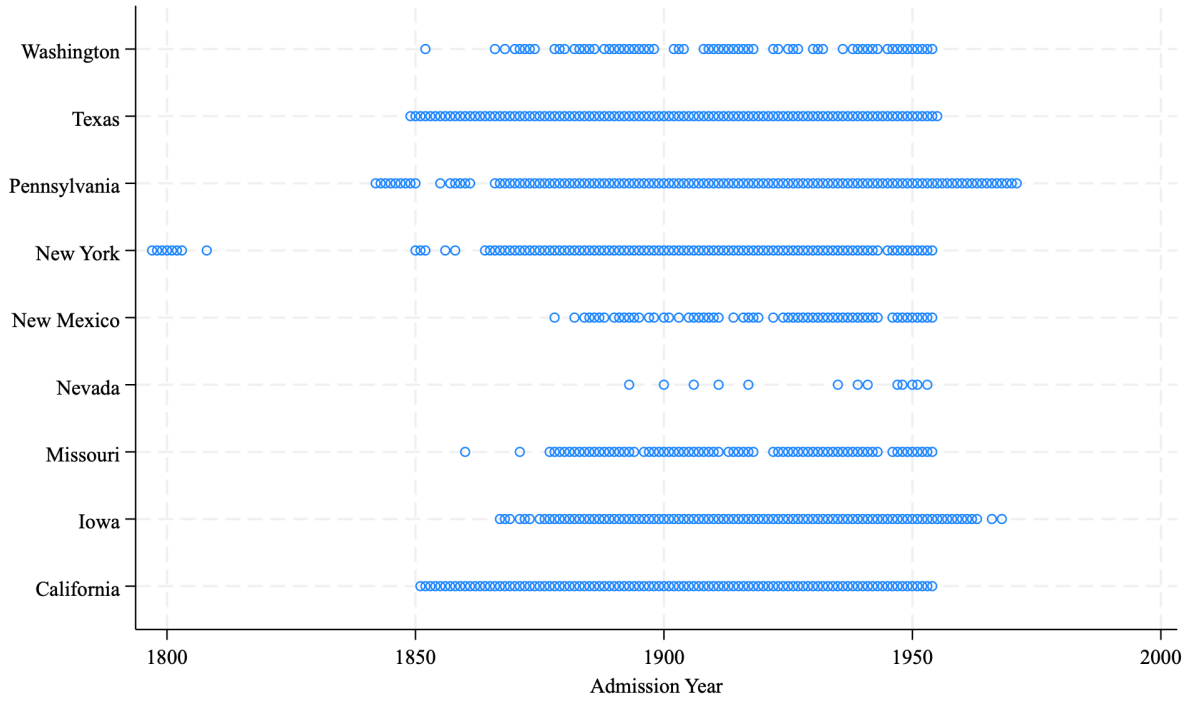
(a) Full Sample



(b) By Race

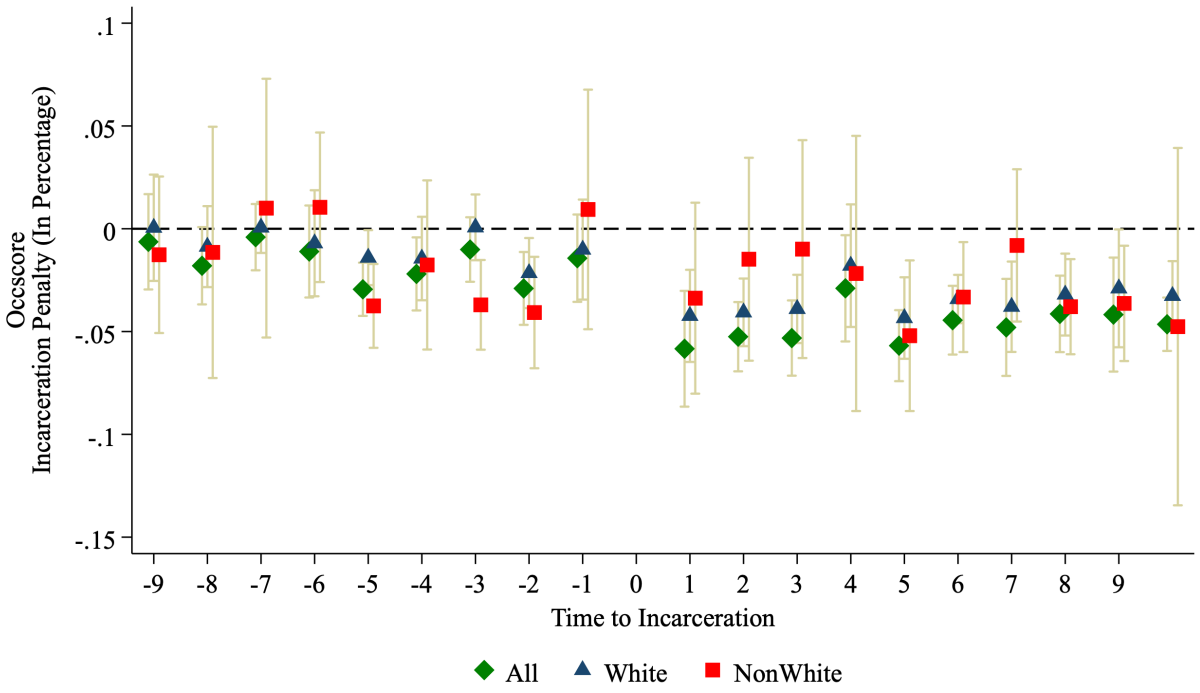
Notes: This figure calculates the incarceration penalty using the labor force participation variable included in the 1910, 1930, and 1940s Censuses. See figure notes for Figure 6 for more information on the specification and sample. Panel a plots the penalty for the entire population, while Panel (b) 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 occupation-based 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 C10: Coverage of Inmate Legder Data



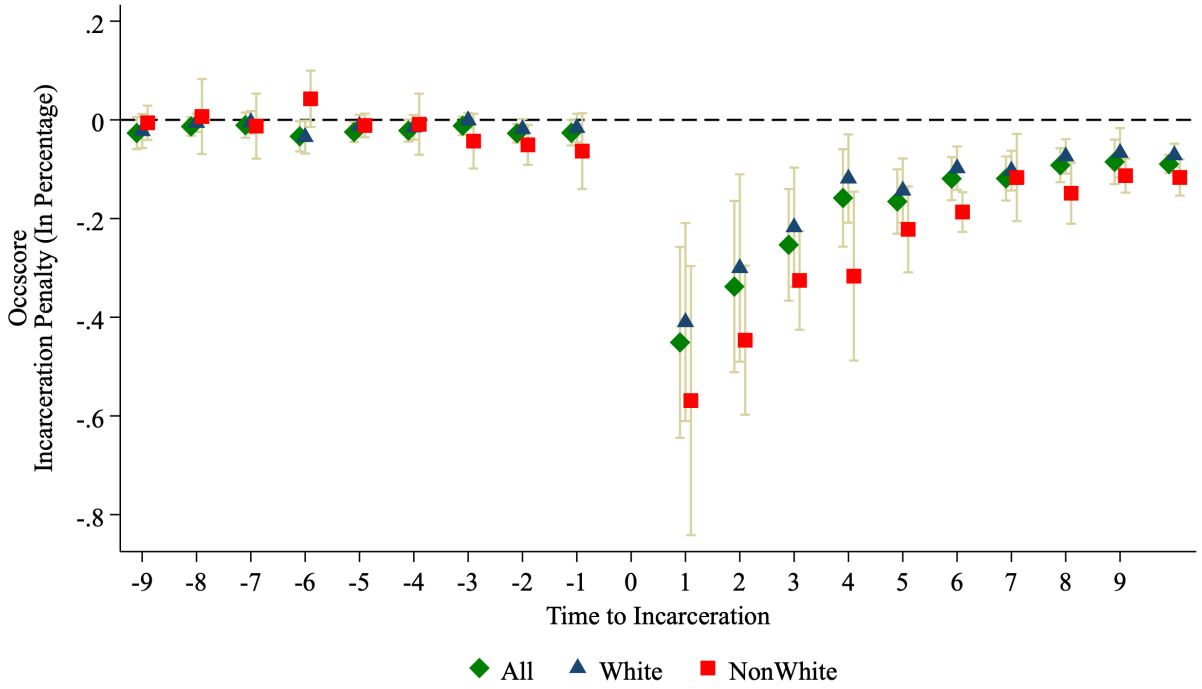
Notes: This figure plots the time coverage by state in the inmate ledger data linkable to Censuses. Each dot denotes a year in which the state is covered by the data.
Sources: Inmate ledger records. Abramitzky, Boustan, Pérez, et al. (2020).

Figure C11: Dynamics of penalty conditional on non-zero occupation-based income score, 1910-1940



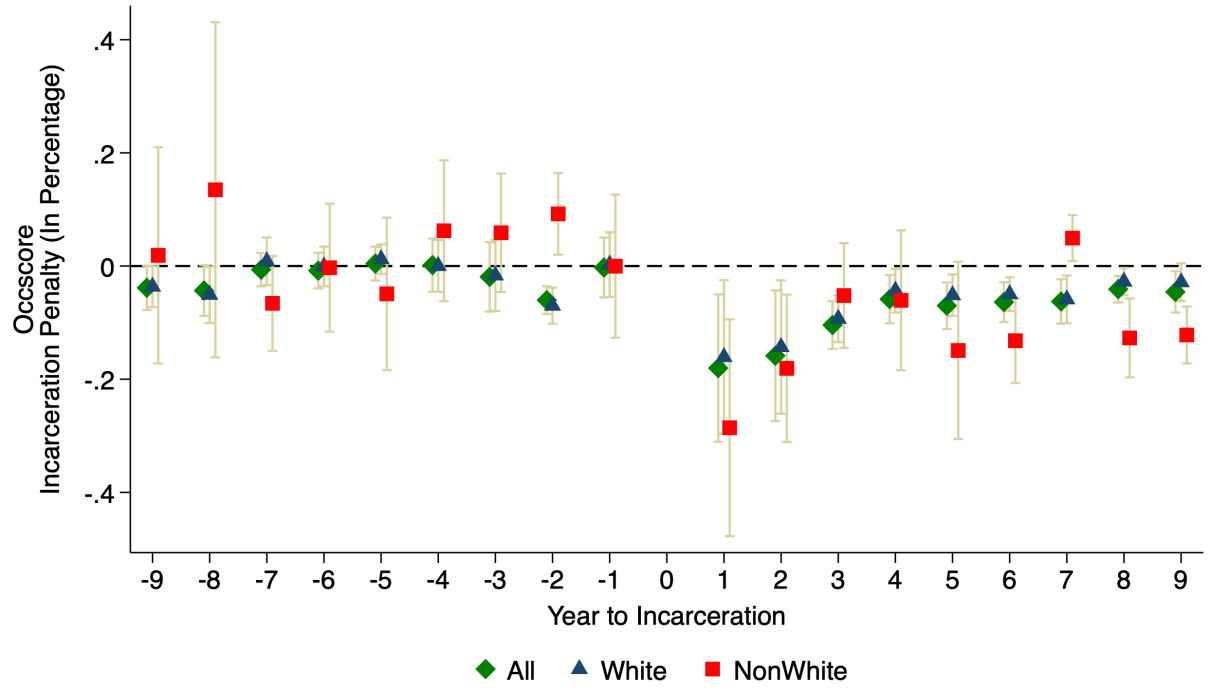
Notes: This figure plots β for occupation-based income score conditional on non-zero income from Equation 4 for every number of years before the prison admission and β from Equation 5 for every number of years after the prison admission. The coefficient is normalized by the mean of occupation-based income 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 C12: Incarceration penalty dynamics around year of prison admission, separately by racial group, 1910-1940



Notes: This figure plots β from Equation 5 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).

Figure C13: Dynamics of Incarceration Penalty on Double Linked Inmate Ledger, 1910-1940



Notes: This figure plots β for occupation-based income score from Equation 4 for every number of years before the prison admission and β from Equation 5 for every number of years after the prison admission. The coefficient is normalized by the mean of occupation-based 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).