Essays on Economic Inequality and Mobility

Hugo Reichardt

A thesis presented for the degree of Doctor of Philosophy.

Department of Economics
London School of Economics
London, United Kingdom

May 29, 2024

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

This thesis consists of 45,281 words.

Chapter 2 is joint work with Lukas Althoff and Chapter 3 is joint work with Lukas Althoff and Harriet Brookes Gray. In these chapters, I contributed a half and a third of the work, respectively.

Abstract

This dissertation studies three key drivers of economic inequality and mobility.

Chapter I shows that scale bias, the extent to which technical change increases the productivity of large relative to small firms, is important for inequality. I develop a tractable framework where people choose to work for wages or earn profits as entrepreneurs and where entrepreneurs choose from a set of available production technologies that differ in their fixed and marginal cost. Large-scale-biased technical change lowers entrepreneurship rates and increases top income inequality, primarily by concentrating business income. Small-scale-biased technical change does the opposite. I show the empirical relevance of scale bias by identifying the causal effects of adoption of two general purpose technologies that vary in scale bias, but are otherwise similar: steam engines (large-scale-biased) and electric motors (small-scale-biased). Using newly collected data from the United States and the Netherlands and a range of identification strategies, I show that these two technologies had the effects predicted by the theory: steam engines increased firm sizes and inequality, while electric motors decreased both.

In Chapter II, we study the long-run effects of slavery and restrictive Jim Crow institutions on Black Americans' economic outcomes. We track individual-level census records of each Black family from 1850 to 1940, and extend our analysis to neighborhood-level outcomes in 2000 and surname-based outcomes in 2023. We show that Black families whose ancestors were enslaved until the Civil War have considerably lower education, income, and wealth than Black families whose ancestors were free before the Civil War. The disparities between the two groups have persisted, not because of slavery per se, but because most families enslaved until the Civil War lived in states with strict Jim Crow regimes after slavery ended. In a regression discontinuity design based on ancestors' enslavement locations, we show that Jim Crow institutions sharply reduced Black families' economic progress in the long run.

Chapter III studies the role of women in historical intergenerational mobility in the US. Previous research has focused on father-son income correlations. We build a new linked census panel to include daughters (1850-1940). To also incorporate the role of mothers, we propose a mobility measure that considers parental human capital alongside income (R^2) and a semi-parametric latent variable method to estimate this measure from historical data. Our approach reveals increasing mobility, overturning conclusions based on income alone. Mothers' human capital was more predictive than fathers' and accounted for the increase in mobility. Aligning with their historical role in homeschooling, mothers were especially important when school access was limited.

Acknowledgements

I am extremely grateful to my advisers, Ethan Ilzetzki, Camille Landais, and Ricardo Reis for their guidance, mentorship, and support. I also thank my co-authors, Lukas Althoff and Harriet Brookes Gray, without whom this dissertation would have been much worse.

I also gratefully acknowledge the financial support of the Economic and Social Research Council (ESRC), the Centre for Macroeconomics (CFM), and the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD).

Special thanks also go to my colleagues and friends Loughlan O'Doherty, Arnaud Dyevre, Andres Felipe Fajardo Ramirez, Peter Lambert, Junyi Liao, Bernardo Mottironi, Veronica Salazar Restrepo, Yannick Schindler, Maria Ventura, and many others for making the six years I spent writing this dissertation so much fun.

To my parents, my siblings, and Caterina.

Contents

I Scale-Biased Technical Change and Inequality

Hugo Reichardt (LSE)

1 Introduction

Income and wealth inequality have significantly increased in many countries in recent decades. Between 1980 and 2014, top-decile incomes in the United States rose more than twice as fast as below-median incomes (Piketty et al., 2018).

Skill-biased technical change is a frequently cited explanation for increases in wage inequality: if new technologies more strongly complement high-skilled labor—or tend to automate low-skilled jobs—, this can increase wage inequality (Katz and Murphy, 1992; Krusell et al., 2000; Violante, 2008; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018, 2022). But wages are not the only source of income. For those at the top of the distribution, business income is the dominant source of income and most of it accrues to entrepreneurs that own large shares of their own business (e.g. Smith et al., 2019; Kopczuk and Zwick, 2020).[1]

Can technical change affect the concentration of business income too and, if so, how and when? The answer I provide is: yes, with the direction of the effect depending on the *scale bias* in technical change. I define scale bias as the extent to which technical change differentially affects the productivity of large versus small firms. Large-scale-biased technical change skews productive resources and profits towards larger firms. Given that firm ownership tends to be concentrated, this shift in profits across firms implies a redistribution of income across households. In other words, I argue that the firm size distribution constitutes a channel through which technical change can affect income inequality.

First, to formalize the theory of scale-biased technical change and inequality, I develop a simple and tractable model where households that are heterogeneous in entrepreneurial productivity can choose to either work for wages or be an entrepreneur. Entrepreneurs have access to a set of available technologies—defined by a marginal and a fixed cost—and adopt the one that maximizes profits. I show that technical change is large-scale-biased if it increases fixed costs relative to previously adopted technologies. If technical change is large-scale-biased, it lowers entrepreneurship rates and leads to larger firms on average. With fewer and larger firms, top entrepreneurs are capturing a larger share of the profits which increases top income inequality. If technical change is small-scale-biased, it has the opposite effects.

Second, to empirically test the theory, I estimate and compare the causal effects of the adoption of steam engines and electric motors. Steam engines became the dominant power source in manufacturing in the second half of the 19$^{\text{th}}$ century. Electric motors began to be widely used around 1900, and in the first half of the 20$^{\text{th}}$ century purchased

[1] See also Atkeson and Irie (2022) that argue for the importance of undiversified business ownership in accounting for wealth mobility and changes in wealth inequality.

electricity and steam engines were each other's substitute in providing power to the factory. These two general purpose technologies provide an appropriate and useful comparison because i) their adoption was sufficiently widespread and transformative to have a meaningful impact on the overall economy ii) they were similar in their capability and purpose—converting energy into rotary motion in manufacturing—, iii) their cost structure induced technical change with strongly different scale bias.

Steam engines entailed much higher fixed costs of purchase and operation than electric motors. The annualized cost, exclusive of fuel, of a 50 horsepower (hp) steam engine was equal to the *yearly* wage of around 3 to 4 unskilled workers.[2] For an electric motor run by purchased electricity with the same capacity, these costs were only around 2 percent of a yearly wage, two hundred times lower than for steam engines.[3] Also, for reasons of technological efficiency, steam engines came in much larger sizes than electric motors.[4] As a result, the adoption rates of the two technologies across the firm size distribution were different. Large establishments were more likely to adopt steam engines than small establishments (see also Atack et al., 2008). I show that, in contrast, electric motors driven by purchased electricity were adopted uniformly across the firm size distribution. Some electric motors in manufacturing were not driven by purchased electricity, but by electricity generated in the plant using steam engines. As this required incurring the high fixed cost of steam engine operation, such systems were skewed to large firms too.

To measure the effect of scale-biased technical change, I construct a rich data set on steam engine and electric motor adoption, firm sizes, and inequality through digitization of various archival sources from the Netherlands and the United States. For the United States, I draw on the Census of Manufactures that provides information such as the number of establishments, employment, value added, and power adoption by state and industry. I digitize and compile these data for each decade year between 1850 and 1940 and 1947. The industry classification in the Census of Manufactures was highly granular, yielding over 50 thousand state-industry observations. Using these data, I investigate the role of steam engines and electric motors in shaping the firm size distribution in manufacturing in the United States.

The first theoretical prediction is on establishment sizes: large-scale-biased technical change increases the average number of workers per establishments and small-scale bias decreases it. In line with these predictions, I find that steam engines increased establishment sizes while electric motors decreased them. To identify these effects, I use variation in natural resources across the United States that affected the costs of using the technologies. Specifically, I use access to historical coal resources and hydropower potential as instruments for steam engine and electric motor adoption, respectively.[5] I estimate how this natural variation affected within-industry firm size differences over

[2]Computations based on the United States in 1874. The total annualized cost was $1404 (see Table IV.11 in Appendix 1.5) and the *yearly* wage of an unskilled worker was around $400 Abbott (1905).

[3]Computations based on the United Kingdom, around 1925. Total annualized cost of an electric motor of 50 hp in 1925 was £2.46 (see Table IV.11 in Appendix 1.5) and the weekly wage was around £2.00 (Bank of England, 2017).

[4]In the United States in 1910, the average steam engine had a capacity of 93.4 horsepower, more than 10 times that of the average electric motor (8.5 hp).

[5]Various other authors have used hydropower potential as an instrument for electricity adoption (e.g. Leknes and Modalsli, 2020; Gaggl et al., 2021). Data to construct the instruments are from the Coal Resources Data System (coal resources) and Young (1964) (hydropower potential).

time across states. I find that high-coal access states experienced a growth in establishment sizes relative to 1850, when steam engines started to be adopted. In contrast, after the introduction of electric motors around 1900, high-hydropower states experienced a decrease in establishment sizes. Using this variation, I estimate the effect of a 1% increase in steam engine capacity in horsepower to be a 1.1% increase in firm size. For electric motors, I estimate this elasticity to be -0.4.

The second prediction is that large-scale-biased technical change increases the ratio between average profits and average wages, while technical change has the opposite effect if it is small-scale-biased. The profit-wage ratio is a measure of inequality between workers and entrepreneurs in the model, where each entrepreneurs owns exactly one firm. I compute profits in the Census of Manufactures using data on output, cost of raw materials, cost of labor, the capital stock, and other expenses. I test this second prediction of the theory using the same geographic instruments and econometric specification with which I test the first prediction. I show that steam engines and electric motors indeed had opposite effects on profit-wage ratios in the direction predicted by the theory.

Is the profit-wage ratio a good measure of inequality between workers and entrepreneurs in practice? And, more generally, does the profit distribution across firms affect the distribution of income across people? The answers to these question depend on the degree of firm ownership concentration. The stronger firm ownership concentration is, the more the distribution of profits are reflected in the personal income distribution. Empirically, firm ownership is highly concentrated, both in the past and the present, even for large publicly traded firms. For example, Goldsmith et al. (1940) reports that in 1940 only 13 families held over 8 percent of the equity in the largest 200 corporations and that each family "has shown a strong tendency to keep its holding concentrated in the enterprise in which the family fortune originated". Similarly, Anderson and Reeb (2003) finds that in the 1990s founding families accounted for 18 percent of outstanding equity in Fortune 500 firms, the largest US firms by revenue. Unsurprisingly, firm ownership concentration is even stronger—and almost perfect—in non-publicly traded firms (e.g. Smith et al., 2019). I show using US census data on wealth from 1860 and 1870, that, as a consequence of ownership concentration, profit-wage ratios are strongly correlated with inequality across people by state and industry ($\rho = 0.67$).

The verification of the theoretical predictions on the effects of scale-biased technical change on profit-wage ratio, coupled with the strong correlation between profit-wage ratios and wealth inequality, already offers suggestive evidence that scale-biased technical change affects income and wealth inequality across people, too. However, granular data on income or wealth in the United States during steam engine and electric motor adoption is not available after 1870. To study the two technologies' effects on inequality, I therefore turn to the Netherlands, for which I collect unique data on income and wealth inequality over the course of industrialization. The dataset I build includes micro-level information on names, demographics, occupation, and, importantly, wealth of each decedent between 1878 and 1927 in five major provinces in the Netherlands, covering over a million decedents and more than half of the national population. It is, to the best of my knowledge, the largest dataset on inequality in any country during the period of steam engine and electric motor adoption.

Using the Dutch dataset, I verify the third prediction of the theory: that the effect of technical change on inequality depends on its scale bias. Using municipality-by-industry level data from the Dutch Census of Companies in 1930, I compute the share of employees that work in establishments with steam engines, with electric motors, and without power for each municipality. I then show how wealth inequality evolved in municipalities that saw strong steam-engine adoption, controlling for municipality fixed effects. I find that municipalities that adopted engines became significantly more unequal over time, especially from around 1910 onward. In contrast, municipalities with high electric motor adoption saw a slight decrease in inequality after 1900. Furthermore, I use an industrial census from 1816—long before industrialization—to create an industry-based measure of "exposure" to steam engines and electric motors. Municipalities whose industrial composition in 1816 exposed them to steam engines showed a strong increase in inequality between 1880 and 1930, while those exposed to electric motors experienced a slight decrease in wealth inequality. The effects on inequality are primarily driven by the very top of the distribution, while the rest of the distribution was not much affected.

Lastly, I show that the effects of scale-biased technical change on top wealth inequality manifests themselves through entrepreneurs that adopt the technology. To test this prediction, I zoom into the major industrializing city of Enschede, in the east of the Netherlands. The pre-existing textile industry made this city particularly exposed to the introduction of the steam engine. Even though wealth inequality decreased in most areas, it increased sharply in Enschede. I find that the rise in top inequality was driven by the textile entrepreneurs that adopted the technology. I do not find any meaningful increase in inequality after excluding the textile entrepreneurs and their spouses from the sample. This finding shows that the rise in inequality was driven by entrepreneurial income—not wages—so that it can not be explained by skill-biased technical change. The proposed theory of *scale*-biased technical change does offer an explanation: the large-scale-biased technical change in textile manufacturing meant that firm concentration increased strongly, which concentrated business income in the hands of a few entrepreneurs.

Related literature. First and foremost, this paper contributes to our understanding of the effect of technical change on income and wealth inequality. Scale-biased technical change offers a view on the distributional effects of technology that complements existing theories on skill bias (e.g., Katz and Murphy, 1992; Acemoglu and Autor, 2011). The case of electricity illustrates the differences.

First, the two theories highlight different features of electric motors as relevant for inequality. Goldin and Katz (1998) argue that electric motor adoption increased the relative demand for skilled workers by facilitating a shift to continuous process and batch methods. Electric motors enabled this shift mostly because they improved the efficiency of "unit drive" systems.[6] I argue that electric motor adoption constituted small-scale-biased technical change because it allowed to "separate the place of generation from the place of use" (Helpman, 1998), reducing the fixed costs of power usage. This shows that technical change can be skill- and scale-biased simultaneously. To nonetheless distinguish scale from skill, I study the role of the primary source of power—generated or

[6]Unit drive refers to a power distribution method where each machine is run by its own electric motors.

purchased—not the system that delivers the power. Importantly, the technological advantages of electric motors in batch and continuous processes (the source of skill bias) exist regardless of whether the electricity is purchased or generated in the plant.

Second, skill and scale bias may imply opposing distributional effects. Because the adoption of electric motors was biased to skilled workers, it exerted upward pressure on wage inequality Goldin and Katz (1998).[7] I claim that its adoption was biased to small firms and therefore pushed inequality between entrepreneurs and workers *down*. Of course, these statements do not contradict each other. Since the top of the distribution tends to be dominated by entrepreneurs, top income inequality may be particularly strongly affected by scale-biased technical change. During the first half of the twentieth century, the time of electric motor adoption, almost every industrialized country witnessed a large decline in the income shares of the top 1 percent (Lindert and Williamson, 2016, p. 194). The findings in this paper suggest that electrification contributed to that trend.

Another large literature relates increased firm concentration to technical change, especially a move toward high fixed cost technologies (e.g. Poschke, 2018; Autor et al., 2020; Hsieh and Rossi-Hansberg, 2023; Kwon et al., 2023). Intangible inputs such as software have been posited as an example of this (Brynjolfsson et al., 2008; Lashkari et al., 2023; De Ridder, 2023). So far, it has been hard to establish credible causal evidence of the effect of technical change on the firm size distribution. Furthermore, because most modern technologies vary on many dimensions other than their cost structure, it is difficult to isolate the role of specific characteristics in driving their concentrating effect. A contribution of this paper is that it studies two technologies that were similar except for their cost structure, allowing to single out the role of fixed costs in shaping the firm size distribution. The theory of scale-biased technical change also provides an additional motive to study business patterns: their implications for economic inequality.[8]

This paper also relates to studies highlighting the role of entrepreneurship in shaping income and wealth inequality (Quadrini, 2000; Cagetti and De Nardi, 2006; Buera and Shin, 2013; Atkeson and Irie, 2022; Albuquerque and Ifergane, 2023). Accounting for entrepreneurship in models of wealth accumulation allows to match the high concentration of wealth observed in the data. In contrast to previous work, I focus on the role of the production technology in shaping inequality and the entrepreneurship decision. For this purpose, I provide a simple and tractable framework in which entrepreneurs face a technology adoption decision. The tractability of the model allows to characterize in closed form how entrepreneurship and the income distribution depend on the set of technologies available in the economy.

Lastly, this paper speaks to the patterns of inequality during industrialization. Kuznets (1955) hypothesized that inequality rises in the early stage of industrialization and later decreases, because of a shift away from the agricultural sector to the more productive, but potentially more unequal, manufacturing sector. Interestingly, he explicitly related inequality to scale: "inequalities [in manufacturing] might be assumed to be far wider than those for the agricultural population which was organized in relatively small in-

[7] Goldin and Katz (1998) argue, however, that an increase in the supply of high-school graduates kept the skill premium in check.

[8] See De Loecker et al. (2022) for other reasons to study the firm size distribution.

dividual enterprises." This paper provides a theoretical foundation and empirical evidence for that argument.

The remainder of the paper is organized as follows. Section 3 lays out the theory of scale-biased technical change and inequality formally. Section 3 describes the historical background of, and differing scale bias between, steam engines and electric motors. In Section 2, I discuss how the data is constructed. The methodology and results on the effect of technology on scale and inequality are shown in Sections 5 and 6, respectively. Section 7 shows evidence that inequality between workers and entrepreneurs was the main channel through which steam engines increased inequality. Section 7 concludes.

2 Model

There is a continuum of households with unit measure that differ in their entrepreneurial productivity ψ. I assume that ψ has a probability density function $f(\cdot)$ with semi-infinite support on $\mathbb{R}^+$, i.e., $\{\psi \mid f(\psi) > 0\} = [\psi_m, \infty)$ for some $\psi_m \geq 0$.[9] In a first stage, before observing their entrepreneurial productivity ψ, each household decides whether to be a worker or to be an entrepreneur (Lucas, 1978). A household knows that by choosing entrepreneurship, it is foregoing the wage w.

Once this opportunity cost is sunk, in the second stage, entrepreneurs observe their productivity ψ and choose whether to enter business or not.

An entrant chooses, in a third stage, chooses from an exogenous set of available production technologies $T \equiv \{t_1, .., t_J\}$. Each technology $t_j \in T$ is a tuple $\{\alpha_j, \kappa_j\}$ where α_j is a parameter that affects marginal labor cost and $\kappa_j > 0$ is its fixed cost in terms of the final good.[10] I assume that T does not contain trivially dominated technologies. That is, if $t_j, t_k \in T$ and $\alpha_j < \alpha_k$, then $\kappa_j > \kappa_k$.[11] Technologies are arranged in order of increasing fixed costs ($\kappa_1 < .. < \kappa_J$).

Finally, in stage four, after adopting technology j, entrepreneurs maximize profits given their productivity ψ, yielding $\pi_j(\psi)$. Figure I.1 visualizes the decision process and pay-offs. I characterize optimal behavior and derive equilibrium conditions by backward induction.

Stage 4: Profit maximization

Each entrepreneur produces a differentiated good. Given technology t_j and entrepreneurial productivity ψ, their production function is

$$y_j(\psi) = \frac{\psi l}{\alpha_j} \tag{I.1}$$

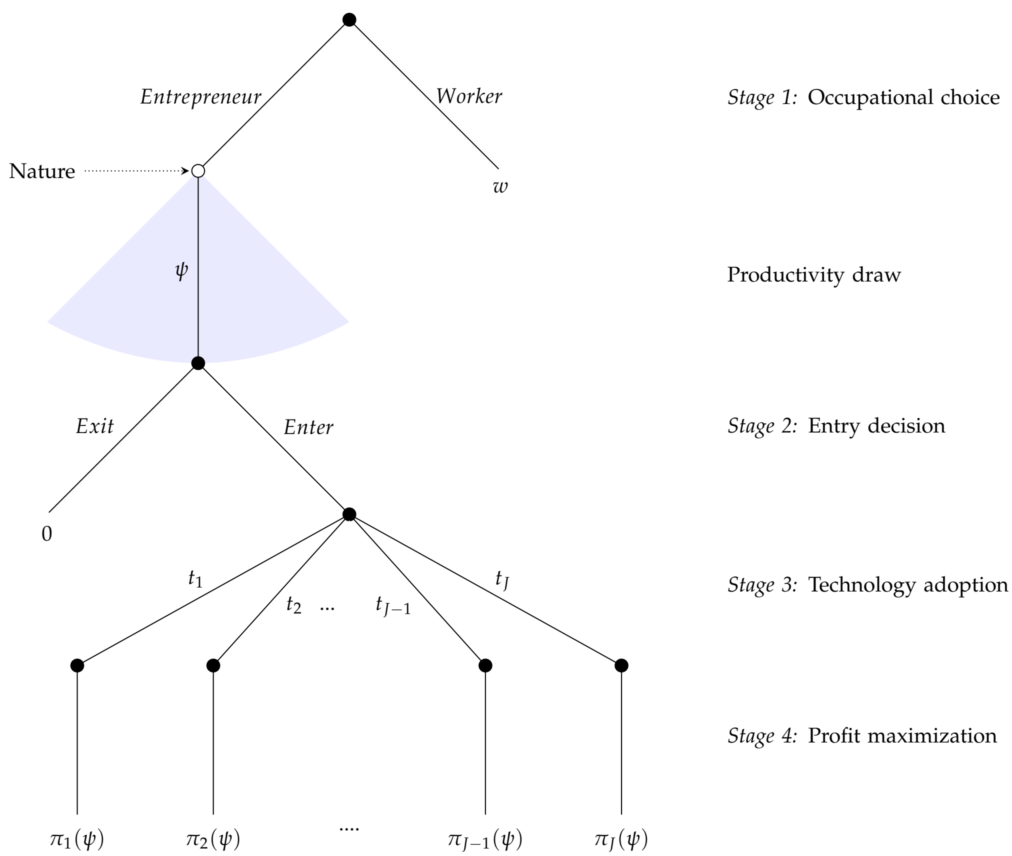
where l is labor and α_j is the marginal labor cost for technology t_j. The total cost to produce y given t_j and ψ is $C_j(y \mid \psi) = \frac{\alpha_j w}{\psi} y + \kappa_j$ where κ_j is the fixed cost in terms

[9] To derive a closed-form solution of the equilibrium, I will later assume that $\psi \sim \text{Pareto}(\psi_m, \xi)$.

[10] This can be seen as a generalization of the binary technology choice in (Yeaple, 2005; Bustos, 2011), who are concerned with the connection between trade and technology adoption.

[11] This assumption does not affect any equilibrium outcome as such trivially dominated technologies would not be adopted.

FIGURE I.1: Pay-off tree



of the final good. Each household's utility is characterized by a constant elasticity of substitution σ over a continuum of these differentiated goods indexed by ω (Dixit and Stiglitz, 1977; Melitz, 2003):

$$U \equiv Y = \left[\int_{\omega \in \Omega} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}}. \tag{I.2}$$

The demand for good ω is thus $y(\omega) = Y \left(\frac{p(\omega)}{P} \right)^{-\sigma}$ where $p(\omega)$ is the price of good ω and $P \equiv \left[\int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}$. Hereafter, I use the normalization that $P = 1$. Profit maximization conditional on technology and productivity then yields the pricing rule

$$p_j(\psi) = \frac{\alpha_j w}{\rho \psi} \tag{I.3}$$

where $\rho \equiv \frac{\sigma-1}{\sigma}$. This pricing rule is standard (e.g., Melitz, 2003, eq. (3)), except that the production technology may vary across producers. In equilibrium, this yields (conditional) profits $\pi_j(\psi)$ equal to

$$\pi_j(\psi) = \frac{Y}{\sigma} \left(\frac{\rho \psi}{\alpha_j w} \right)^{\sigma-1} - \kappa_j. \tag{I.4}$$

Stage 3: Technology adoption

An entrepreneur that chooses to produce can use any of the J available technologies in the set T. She therefore adopts the technology j that yields largest profits, so the profits of an entrepreneur with productivity ψ are:

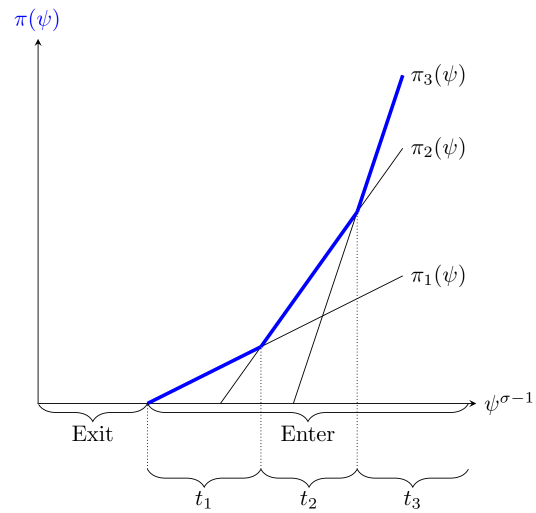
$$\pi(\psi) = \max_{j \in \{1,2,..,J\}} \{\pi_j(\psi)\}. \tag{I.5}$$

An important implication of this profit function is that more productive entrepreneurs choose higher fixed costs technologies. To see this, note that for an entrepreneur with productivity ψ, the difference in profits between technologies t_j and t_k are:

$$\Delta\pi_{jk}(\psi) \equiv \pi_j(\psi) - \pi_k(\psi) = \frac{Y}{\sigma} \left(\frac{\rho \psi}{w} \right)^{\sigma-1} \left(\alpha_j^{1-\sigma} - \alpha_k^{1-\sigma} \right) - (\kappa_j - \kappa_k). \tag{I.6}$$

Recall that since $j > k$, $\kappa_j > \kappa_k$ and $\alpha_j < \alpha_k$. It then follows from the expression that $\Delta\pi_{jk}(\psi)$ is strictly increasing in ψ. That is, the more productive an entrepreneur is, the larger their profits under technology j (higher fixed, lower marginal cost) relative to technology k (lower fixed, higher marginal cost). A corollary of this result is that prices are strictly decreasing in ψ (see equation (I.3)), such that entrepreneurs with higher productivity face more demand and, hence, produce more.

FIGURE I.2: Profit $\pi(\psi)$ and productivity ψ in case of three adopted technologies



Notes: The braces indicate the optimal action in Stage 2 and 3 given productivity ψ. The elasticity of substitution σ is larger than one so that $\psi^{\sigma-1}$ is increasing in ψ.

Stage 2: Entry decision

After observing their entrepreneurial productivity ψ, each entrepreneur decides whether or not to exit or enter. Since the opportunity cost is zero (as the opportunity cost of not working is already sunk), they decide to enter if and only if $\pi(\psi) \geq 0$.

There is a unique $\bar{\psi} > 0$ such that an entrepreneur enters if and only if $\psi \geq \bar{\psi}$. To see this, note that equation (I.4) implies that $\pi_j(\psi)$ is strictly increasing in ψ for each $j \in \{1, 2, .., J\}$. Therefore, $\pi(\psi)$ is the maximum of J strictly increasing functions and is thus also strictly increasing. Finally, $\pi(0) = -\kappa_1 < 0$ and $\pi(\psi) \to \infty$ as $\psi \to \infty$. It thus follows that there is a unique $\bar{\psi}$ implicitly defined by

$$\pi(\bar{\psi}) = 0. \tag{I.7}$$

To solve for this threshold, note that profits under each technology are strictly increasing in $\pi_j(\psi)$. Therefore, each technology j has itself a zero profit cut-off $\bar{\psi}_j$ above which profits are positive. From equation (I.4), this threshold is defined by

$$\bar{\psi}_j = \alpha_j \kappa_j^{\frac{1}{\sigma-1}} \left(\frac{\sigma}{Y}\right)^{\frac{1}{\sigma-1}} \frac{w}{\rho}.$$

Since an entrepreneur enters if and only if at least one technology yields positive profits, the entry decision is governed by the technology for which the entry threshold $\bar{\psi}_j$ is lowest. Combining equations (I.4), (I.5), (I.7) gives a solution for $\bar{\psi} > 0$:

$$\bar{\psi} = \min_{j \in 1,2,..,J} \bar{\psi}_j = \min_{j \in 1,2,..,J} \left\{ \alpha_j \kappa_j^{\frac{1}{\sigma-1}} \right\} \left(\frac{\sigma}{Y}\right)^{\frac{1}{\sigma-1}} \frac{w}{\rho}. \tag{I.8}$$

Figure I.2 shows the profit function $\pi(\psi)$ and the optimal decision in Stage 2 and 3. It illustrates that the entry cut-off $\bar{\psi}$ is the productivity level for which the technology with the lowest entry threshold gives positive profits.

Stage 1: Occupational choice

Free entry into entrepreneurship (and risk-neutrality) implies that in equilibrium the expected profits of entering must be equal to the wage. That is,

$$\int_{\bar{\psi}}^{\infty} \pi(\psi) dF(\psi) = w. \tag{I.9}$$

Defining average profits of producing entrepreneurs as $\bar{\pi} \equiv \frac{1}{1-F(\bar{\psi})} \int_{\bar{\psi}}^{\infty} \pi(\psi) dF(\psi)$, equation (I.9) can be written as

$$(1 - F(\bar{\psi}))\, \bar{\pi} = w.$$

The probability of entry times the average profits after entry should equate the wage. Were the wage lower (higher) than the expected profits, no one would decide to work (be an entrepreneur).

2.1 Which technologies are adopted?

Answering this question requires defining some notation. First, it follows from optimal behaviour in Stages 2 and 3 that a technology is adopted in equilibrium if there is a set of entrepreneurs that both i) decides to enter and ii) finds it profit-maximizing to produce with that technology. I define the *adopting set* for technology j as the set of productivity levels for which both conditions are satisfied:

$$\Psi_j \equiv \{\psi \mid \pi(\psi) \geq 0\} \cap \left\{\psi \mid \pi_j(\psi) = \max_{k \in \{1,2,..,J\}} \pi_k(\psi) \equiv \pi(\psi)\right\}. \tag{I.10}$$

A technology j is adopted if the probability measure of the adopting set Ψ_j is strictly positive. Let $T^* \subseteq T$ be *the set of adopted technologies*, so that

$$t_j \in T^* \iff \Pr\left(\psi \in \Psi_j\right) > 0 \text{ for any } j = 1, 2, ..J.$$

Proposition 1 shows which technologies are adopted in equilibrium.

Proposition 1 (Adopted technologies)**.** *Let $t_j^* = \{\alpha_j^*, \kappa_j^*\}$ be the technology in T^* with the jth-lowest fixed cost κ_j^* and let $J^* \equiv |T^*|$. Then, the set of technologies adopted in equilibrium, $T^* = \left\{t_1^*, .., t_{J^*}^*\right\}$, is such that*

(a) the adopted technology with the highest marginal (lowest fixed) cost $t_1^* = (\alpha_1^*, \kappa_1^*)$ is such that

$$\alpha_1^* (\kappa_1^*)^{\frac{1}{\sigma-1}} = \min_{j \in 1,2,..,J} \left\{\alpha_j \kappa_j^{\frac{1}{\sigma-1}}\right\} \textit{ and;}$$
$$\alpha_1^* = \min_{j \in 1,2,..,J} \left\{\alpha_j \mid \alpha_j \kappa_j^{\frac{1}{\sigma-1}} = \min_{l \in 1,2,..,J} \left\{\alpha_l \kappa_l^{\frac{1}{\sigma-1}}\right\}\right\}$$

(b) the adopted technology with the lowest marginal (highest fixed) cost $t_{J^*}^* = (\alpha_{J^*}^*, \kappa_{J^*}^*)$ is

such that

$$\alpha^*_{J^*} = \min_{j \in 1,2,..,J} \{\alpha_j\} \text{ and;}$$
$$\kappa^*_{J^*} = \min_{j \in 1,2,..,J} \left\{ \kappa_j \mid \alpha_j = \min_{l \in 1,2,..,J} \{\alpha_l\} \right\}$$

(c) *any technology with fixed cost* $\kappa^*_1 < \kappa_j < \kappa^*_{J^*}$ *is adopted if and only if for any* $k \in \{1, .., j-1\}$ *and* $l \in \{j+1, .., J\}$

$$\frac{\alpha_l^{1-\sigma} - \alpha_j^{1-\sigma}}{\alpha_j^{1-\sigma} - \alpha_k^{1-\sigma}} < \frac{\kappa_l - \kappa_j}{\kappa_j - \kappa_k}.$$

Proof of Proposition 1. See Appendix 1.3. □

Proposition 1(a) indicates which technology is the adopted technology with highest marginal cost (and thus lowest fixed cost). Since the profit gain of a marginal cost reduction is increasing in productivity ψ, this is the technology that is adopted by the marginal entrepreneur ($\psi = \bar{\psi}$). Also, the marginal entrepreneur must use the technology j with the lowest *entry threshold* $\bar{\psi}_j$ (in Figure I.2, the technology with the leftmost intersection with the zero-profit axis). The first condition in Proposition 1(a) then follows from equation (I.8). The second condition in Proposition 1(a) states that—in knife-edge cases where there is more than one technology that minimizes the entry threshold—only the technology with the lowest marginal cost among those that minimize the entry threshold are adopted because all but the marginal entrepreneur would strictly prefer that technology.

Proposition 1(b) shows that the technology with the lowest marginal cost is always adopted, regardless of its fixed cost. The result follows from the unbounded support of the productivity distribution. Since the gains from lowering marginal cost are strictly increasing in productivity, the gains from lowering marginal cost are unbounded. Therefore, no matter how high the fixed cost, there is always a strictly positive measure of entrepreneurs willing to incur it to reduce marginal cost. Of course, if there are multiple technologies that minimize marginal cost, only the technology with lowest fixed cost among them is adopted. It follows from combining Propositions 1(a) and 1(b) that only one technology is adopted in equilibrium if and only if the technology in T with the lowest marginal cost also comes with the lowest entry threshold. Th

Lastly, Proposition 1(c) covers all remaining adopted technologies, if any. Intuitively, for a technology to be adopted by an entrepreneur, their productivity must be *high enough* to make the technology more profitable than any other technology with higher marginal cost (and lower fixed cost), but also *low enough* to make the technology more profitable than adopting any other technology with lower marginal cost (and higher fixed cost). Proposition 1(c) sets out the conditions under which the set of productivities that satisfy these conditions has a strictly positive probability measure. To illustrate the condition, consider Figure I.2: there is an intermediate set of productivity levels, for which technology t_2 yields higher profits than both t_1 and t_3. For such a set of productivity levels to exist, the lower bound above which t_2 higher profits than t_1 must be smaller than the upper bound below which it yields higher profits than t_3.

2.2 Equilibrium

Definition (Competitive equilibrium)**.** Given an exogenous technology set $T = \{t_1, .., t_J\}$, a *competitive equilibrium* consists of a price w, profits $\{\pi(\psi)\}$, output Y, productivity threshold $\bar{\psi}$, adopting sets $\{\Psi_j\}_{j=1}^{J}$, and a share of entrants L such that

- profits $\pi(\psi)$ are as defined in (I.4) and (I.5);
- the adopting set of technology j, Ψ_j, is as defined in (I.10);
- the free entry condition in (I.9) holds;
- the labor and goods markets clear, so that

$$L = (1-L)Y\left(\frac{\rho}{w}\right)^{\sigma}\sum_{j=1}^{J}\alpha_j^{1-\sigma}\int_{\psi\in\Psi_j}\psi^{\sigma-1}dF(\psi), \tag{I.11}$$

$$Y = Lw + (1-L)\left(\sum_{j=1}^{J}\kappa_j\int_{\psi\in\Psi_j}dF(\psi) + \sum_{j=1}^{J}\int_{\psi\in\Psi_j}\pi(\psi)dF(\psi)\right); \tag{I.12}$$

- the pricing by entrepreneurs is consistent with a price index equal to 1, so that

$$1 = (1-L)\left(\frac{w}{\rho}\right)^{1-\sigma}\sum_{j=1}^{J}\alpha_j^{1-\sigma}\int_{\psi\in\Psi_j}\psi^{\sigma-1}dF(\psi). \tag{I.13}$$

Having defined the equilibrium in general, in order to get more concrete results, form now on I assume that the distribution of productivity ψ is Pareto. With this assumption, the model has closed-form analytical solutions reported in Appendix 1.3.

Proposition 2 (Closed-form equilibrium)**.** *Suppose that the distribution of productivity ψ is Pareto with shape parameter ξ and a minimum productivity level of $\psi_m > 0$ such that $\xi > 1$ and $\xi > \sigma - 1$. Then, the closed-form solutions to the competitive equilibrium for L, $\bar{\psi}$, Y, w, and $\bar{\pi}$ are given by equations* (IV.2), (IV.3), (IV.4), (IV.5), *and* (IV.6) *in Appendix 1.3.*

Proof of Proposition 2. See Appendix 1.3. □

Proposition 1 and 2 together fully characterize the equilibrium in closed form. In the next subsection, I use these results to study the effect of scale-biased technical change on entrepreneurship, firm concentration, wages, output, profits, and inequality.

2.3 Scale bias and testable implications

To formalize scale-biased technical change, I first define the *total factor productivity* of a firm as the idiosyncratic productivity of the entrepreneur ψ divided by the marginal cost parameter of the technology in T that it adopts:

$$TFP\left(\psi \mid T\right) = \begin{cases} \frac{\psi}{\alpha(\psi|T)} & \text{if } \psi \geq \bar{\psi}\left(T\right) \\ 0 & \text{otherwise} \end{cases}$$

where $\bar{\psi}(T)$ and $\alpha(\psi \mid T)$ are the entry threshold (derived in closed-form in Proposition 2) and the marginal cost parameter of the optimally adopted technology given technology set T. I set total factor productivity to zero for entrepreneurs that do not produce to ensure that changes on the extensive margin (in and out of production) are reflected in TFP changes.

Technical change is an addition of a new technology, say t_{new}, to the technology set T_{old} such that $T_{new} = T_{old} \cup \{t_{new}\}$. From there, I define scale-biased technical change formally.

Definition (Scale-biased technical change)**.** Technical change is *large-scale-biased* if and only if there exists some $k > \min\{\bar{\psi}(T_{new}), \bar{\psi}(T_{old})\}$ such that it increases TFP for $\psi > k$ and does not increase it for $\psi < k$:

$$\begin{aligned} &TFP\left(\psi \mid T_{new}\right) > TFP\left(\psi \mid T_{old}\right) \quad \forall \psi > k \text{ and;} \\ &TFP\left(\psi \mid T_{new}\right) \leq TFP\left(\psi \mid T_{old}\right) \quad \forall \psi \in \left(\min\left\{\bar{\psi}(T_{new}), \bar{\psi}(T_{old})\right\}, k\right). \end{aligned} \tag{I.14}$$

It is *small-scale-biased* if and only if

$$\begin{aligned} &TFP\left(\psi \mid T_{new}\right) \leq TFP\left(\psi \mid T_{old}\right) \quad \forall \psi > k \text{ and;} \\ &TFP\left(\psi \mid T_{new}\right) > TFP\left(\psi \mid T_{old}\right) \quad \forall \psi \in \left(\min\left\{\bar{\psi}(T_{new}), \bar{\psi}(T_{old})\right\}, k\right). \end{aligned} \tag{I.15}$$

In other words, technical change is large-scale-biased if it increases the productivity of firms above some level of entrepreneurial productivity, while it does not increase the productivity of other firms. I do not consider cut-off levels k below $\min\{\bar{\psi}(T_{new}), \bar{\psi}(T_{old})\}$ because for those levels of productivity people do not choose to be entrepreneurs under either technology set.

The definition is similar to that of *skill*-biased technical change as increasing skilled workers' productivity relative to unskilled labor (Katz and Murphy, 1992; Violante, 2008). Krusell et al. (2000) provide a micro-foundation for skill-biased technical change by considering that the relative productivity changes could be caused by capital-skill complementary. In the same vein, I provide an explicit mechanism for relative productivity increases of large firms in terms of the available technologies. That is, I derive the conditions on the technological parameters under which technical change is large-scale-biased in equilibrium. Proposition 3 lays out these conditions.

Proposition 3 (Scale-biased technical change)**.** *Suppose that the assumptions in Proposition 2 (Pareto distribution) hold, that $\sigma > 2$, and that $T^*_{new} = T^*_{old} \cup \{t_{new}\}$ (the new technology is adopted alongside the previously adopted technologies). Then,*

(a) the technical change is large-scale-biased if and only if

$$\kappa_{new} > \max_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \kappa_j;$$

(b) and the technical change is small-scale-biased if and only if

$$\kappa_{new} < \min_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \kappa_j.$$

Proof. See Appendix 1.3. □

Proposition 3 shows that the addition of a technology constitutes large-scale-biased technical change if and only if the new technology comes with highest fixed cost. Conversely, it is small-scale-biased if the new technology has lowest fixed cost. Since no technology can strictly dominate another adopted technology, the result implies that technical change is large-scale-biased if and only if the new technology has lowest *marginal* cost.

The intuition behind the "if" is that a technology on the extreme end of the technology set would be adopted by the most productive or least productive entrepreneurs. Also, under the assumptions in Proposition 3, if a new technology is adopted, it *reduces* profits when using the other technologies. Therefore, entrepreneurs that do not adopt the new technology do not reduce marginal cost through a change to a third technology. If anything, some may find it optimal to use a technology with higher marginal and lower fixed costs than before in response to other entrepreneurs using the new technology. Thus, if a new technology has largest fixed cost, it increases the productivity of the top entrepreneurs, but not the rest. Vice versa, if it comes with lowest fixed cost, it increases the relative productivity of small entrepreneurs.

If a technology is adopted that has neither the highest nor the lowest fixed cost, it will be used by a set of intermediate entrepreneurs. This means that both the largest and the smallest firms do not adopt this technology. Hence, by the same reasoning as above, this type of technical change does not increase the productivity of either small or large firms and is thus neither large- nor small-scale-biased.

The condition that $\sigma > 2$ is the empirically relevant case for at least three reasons. First, it is consistent with estimates of σ around 6 for US manufacturing data (Bernard et al., 2003) and with the calibration of $\sigma = 4$ by Melitz and Redding (2015). Second, $\sigma \leq 2$ implies a labor share of a half or lower, while the labor share has been consistently larger than a half in the US and other countries. Third, if $\sigma \leq 2$, the implied mark-up (i.e., the ratio of price to marginal cost) is larger than 2.

Proposition 3 covers all cases where the new technology is adopted, but does not make any existing technologies "obsolete". It is however possible that a (subset of) previously adopted technologies are no longer adopted after a new technology is introduced. In Proposition 3A (in Appendix 1.3), I derive the technological conditions for large- and small-scale-biased technical change in such cases.

Using Propositions 2 and 3, I generate three main predictions of the theory. First, large-scale biased technical change increases average firm sizes, while small-scale-biased technical change decreases them. Second, large-scale biased technical change increases income inequality between workers and entrepreneurs. Third, large-scale biased technical change increases top income inequality.

Proposition 4 (Theoretical implications of scale-biased technical change)**.** *Suppose the assumptions in Proposition 3 hold. Then, large-scale-biased technical change*

(a) increases the average firm size as measured by employment;

(b) increases income inequality between active entrepreneurs and workers;

(c) *increases the income share of the top k% income earners for any k below some* $\bar{k} \in (0, 100)$.

Small-scale-biased technical change has the opposite effects.

Proof of Proposition 4. See Appendix 1.3. □

The remainder of the paper is devoted to testing the theoretical predictions above. I will use the case of steam engines and electric motors. In the next section, I show that steam engine adoption is large-scale-biased and electric motor adoption small-scale-biased technical change.

3 Scale bias in steam engines and electric motors

To test the theory of scale-biased technical change, I compare the effects of steam engine and electric motor adoption. I argue that the comparison of these two technologies is uniquely appropriate to test the theory for three main reasons. First, the steam engine and the electric motor are two of the most important general purpose technologies in human history (Bresnahan and Trajtenberg, 1995). Second, they served a similar purpose: the conversion of energy into rotary motion in manufacturing. Third, as I will argue in this section, they varied crucially on scale bias: steam engine adoption constituted large-scale-biased technical change, while electric motor adoption constituted small-scale-biased technical change.

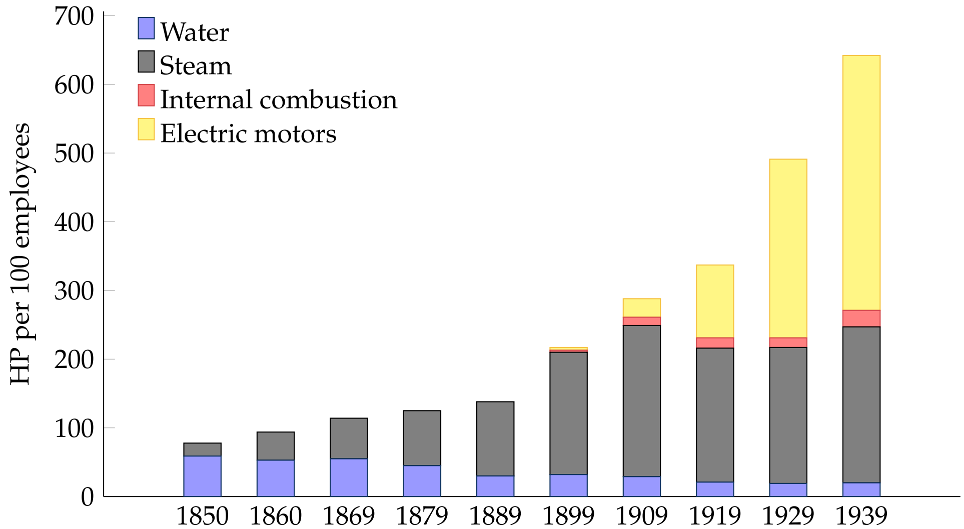
I first briefly describe the history of steam engine and electric motor adoption. Figure I.3 illustrates the timing and degree of adoption of each type of primary power. Three main patterns jump out. First, the waterwheel was slowly replaced by the steam engine in the second half of the 19th century. Second, steam engines, and later the electric motor, were the dominant power source from around 1870 onward. Third, electric motors were adopted from around 1900 and their superiority meant that internal combustion engines were never adopted on a large scale (Du Boff, 1967). Fourth, electric motors driven by purchased electricity started to become dominant around the 1930s, but steam engines remained an important source of primary power until at least 1939. Figure IV.1 shows the same patterns for the Netherlands.[12] Below, I lay out the features of the technologies that make steam engine adoption large-scale-biased and electric motor adoption small-scale-biased.

First, steam engines come with much higher fixed costs of purchase, renewal, and operation than electric motors. The price of a steam engine (including boiler) of average capacity was around \$5331 in 1874, more than 13 times the yearly wage of an unskilled manufacturing worker (Emery, 1883; Abbott, 1905).[13] On top of that, it required an

[12] A distinction can be made between the primary source of power (from the perspective of the plant) and the system to deliver that power. Many electric motors in manufacturing were not driven by purchased electricity, but by electricity generated in the plant. Such "secondary movers" are excluded from Figure I.3 to avoid double counting of capacity. The share of non-electric primary power, such as steam engines, that served to generate electricity for intra-plant use grew strongly over time: from 14.8% percent in 1909 to 65.8% in 1939 (Du Boff, 1979, Table 15). Hence, electricity as a system of power delivery was more dominant than suggested by considering only the primary source of power. In this paper I focus on the primary source of power as the key distinction between "steam engines" and "electric motors".

[13] The average steam engine in the United States in 1889 had a capacity of 50.1 horsepower (Du Boff, 1979). The daily wage of an unskilled worker was \$1.29 Abbott (1905), which I multiplied by 309 days as in (Emery, 1883).

FIGURE I.3: Capacity of primary power by type in horsepower per 100 employees in manufacturing in the United States



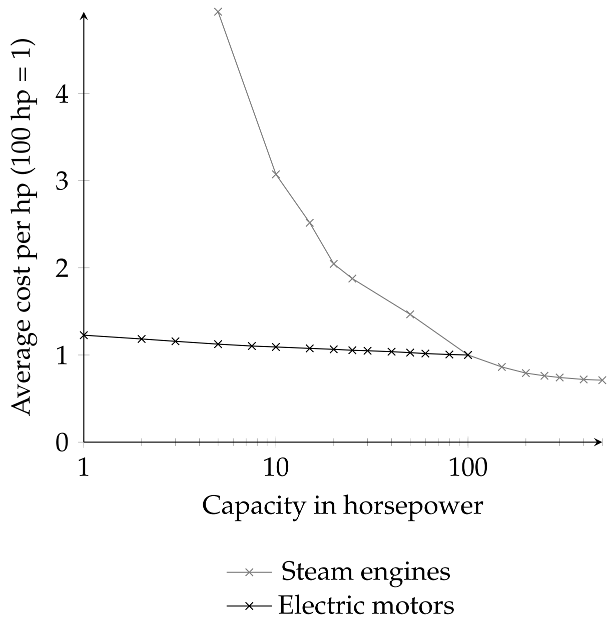
Notes: Electric motors refer to primary electric motors, i.e., electric motors driven by purchased electricity, only. Electric motors driven by energy generated in the plant are covered under steam engines. *Sources*: (Atack, 1979, Table 1) for the number of steam engines and waterwheels in 1850 and 1860; (Atack et al., 1980, p. 285) for their average size (21 and 15 hp, respectively); Census of Manufactures 1860 for the total number of employees in 1850 and 1860; Census of Manufactures 1939, Power equipment and energy consumption, Table 3 for all years after 1860.

engineer and a firemen, supplies, oil, and repairs. In total, I estimate the annualized cost of purchase, renewal, maintenance, and operation of a 50 horsepower steam engine to be around $1378, about 3 to 4 times the yearly unskilled wage. In other words, for the cost of operating an average-sized steam engine excluding fuel, one could hire around 3 to 4 unskilled workers. In comparison, the equivalent annualized fixed costs of an electric motor of that size were negligible: the fixed cost amounted to only 2 percent of the yearly wage of an unskilled worker (Bolton, 1926). In Appendix 1.5, I provide more details on computations and sources.

Second, larger steam engines were considerably more efficient in converting energy into motion than small ones (Atack, 1979; Devine, 1983). In contrast, electric motors' efficiency does not vary nearly as much with size. In the words of the contemporaneous engineer Bell (1891): "With the electric motor the case is very, very different [from steam engines]; an eight horse-power motor may be as completely worked out in detail as one of a hundred times its power, and may be only slightly less efficient." Figure IV.2 illustrates the efficiency of steam engines and electric motors for different sizes (horsepower capacity) relative to a 100 hp equivalent based on estimates by Emery (1883) and Bolton (1926). A steam engine of 10 hp required more than twice as much coal per horsepower of energy output than a 100 hp steam engine. Coal-efficiency was an important consideration given that coal accounted for between a half and two-thirds of the total operating costs for the larger engines.

The marginal and fixed costs of steam engines and electric motors can be combined to estimate an average cost curve by rated capacity for the electric motor and the steam engine. Figure I.4 shows the results.[14] Clearly, steam engines were much more cost-efficient on a large scale. For electric motors, scale was close to irrelevant as almost all costs were marginal, coming from the purchase of electricity, and the efficiency loss of small motors was minor.

FIGURE I.4: Average cost per horsepower per year of steam engines and electric motors of different capacities relative to its 100-horse power equivalent



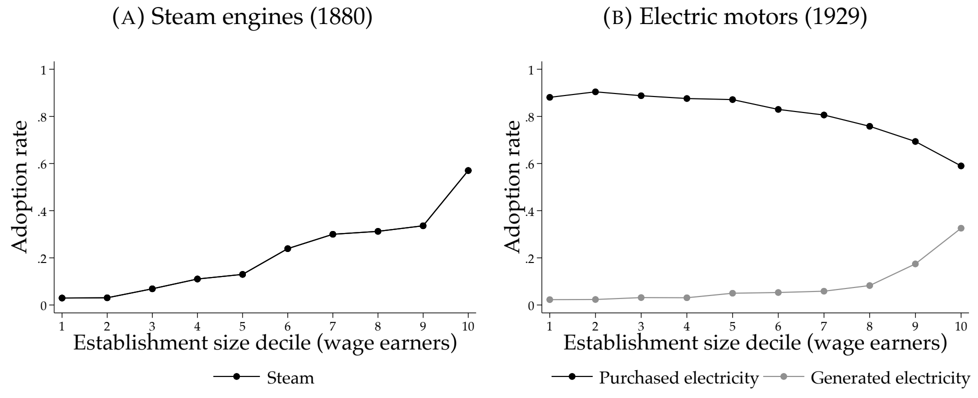
Notes: Author's computation based on contemporaneous price and efficiency data. *Sources*: (Emery, 1883) for steam engines and coal; (Bolton, 1926; Hannah, 1979) for electric motors and electricity. See Appendix 1.5 for further details.

Lastly, there were reasons for steam engine adoption to be skewed to large establishments that are less easily quantified. A steam engine occupied a large amount of space and fuel storage, water supply, and mitigation of fire hazard further increased the fixed costs of operating steam engines (Hunter and Bryant, 1991, p. 56). Also, the "notoriously wasteful" steam engine had to be run at full capacity even if only small doses of power were required, a feature likely to be specifically uneconomical for small establishments (Du Boff, 1967).

The adoption rates by plant size reflect the considerations above. Figure I.5(A) shows that large plants are much more likely to adopt steam engines, as documented before by Atack et al. (2008). In contrast, Figure I.5(B) indicates that electric motors were almost uniformly adopted across the establishment size distribution. However, small firms tended to rely solely on purchased electricity while large firms were more likely to use self-generated electricity. This further confirms that, for the purpose of studying scale bias, the relevant distinction is the primary source of power, not the system of delivery.

[14] I have assumed an interest rate of 5 percent, depreciation rates as estimated by Emery (1883); Bolton (1926) and a price of electricity as reported by Hannah (1979) and of coal as Emery (1883). In Appendix 1.5, I explain the assumptions and computations underlying Figure I.4 in further detail. Consistent with my estimates based on Emery (1883), (Kapp, 1894, p. 234) reports that the cost per horsepower hour of a "small" steam engine was about four times the cost of that of a "large" engine.

FIGURE I.5: Adoption rates by establishment size



Notes: This figure indicates the share of establishments using steam engines in 1880 (panel A) and electric motors driven only by purchased electricity vs. generated electricity in 1929 (panel B) by establishment size as computed from micro-samples of the Census of Manufactures. *Sources*: for 1880, the national random sample of the Census of Manufactures (Atack and Bateman, 1999); for 1929, the Census of Manufactures for selected industries (Vickers and Ziebarth, 2018). I left out the concrete industry as data on electric motors driven by generated electricity is not available for that industry.

4 Data construction

This paper uses newly collected and digitized data from the United States as well as the Netherlands. In this section, I discuss the sources and construction of the data for both countries.

4.1 United States

For the United States, I most heavily rely on the tabulations of the decennial Census of Manufactures by state and industry. I digitized and compiled these data for each decade year between 1850 and 1940 and 1947. The information in the Census of Manufactures varied somewhat from year to year, but key variables such as the number of establishments, employment, and value added are always available. Furthermore, from 1870 onward, the tabulations reported the adoption of power technologies such as water wheels, steam engines, and, later, electric motors. The industry classification is detailed; in the average year, there are around three to four hundred different manufacturing industries. In total, the data comprise of 51,263 state-industry-year observations.

Since industry classifications changed over time, I created two crosswalks that allow to compare industries over time. The first covers all industries between 1860 and 1900, the period of most rapid steam engine adoption, and consists of 182 industries. This crosswalk is an extension of the 1860 to 1880 crosswalk published by Hornbeck and Rotemberg (2021). The second crosswalks consists of 206 harmonized industries across the six censuses between 1890 and 1940. To create this second crosswalk, I used tabulations by industries over time published in the Census of Manufactures.[15] The final crosswalks can be found in Appendix 1.4.2. I also coded each Census of Manufactures industry to the 1950 Census Bureau industrial classification system to allow matching

[15]In particular, I mostly used "comparative summaries" and descriptions of industry classifications in the appendices in the Census of Manufactures.

with the IPUMS USA population censuses between 1850 and 1940.

To construct instrumental variables for technology adoption, I use data on coal resources and hydropower potential by state. Data on historical coal resources by county are taken from the National Coal Resources Data System from the United States Geological Survey (USGS).[16] The dataset contains information on the "rank" (i.e., type) of coal, the estimated tonnage available, the thickness of the field, and the "overburden" (i.e. the depth of the material that lies above the coalfield). Using this information, I compute the total coal resources in British thermal units (Btu) for each county.[17] Recognizing that coal was traded across counties, I compute a measure of "coal access" by county similar to the measure of market access used by Donaldson and Hornbeck (2016). That is, for destination county c in state s, coal access is given by

$$\mathrm{COAL}_c^s = \sum_o \tau_{oc}^{-\theta} \mathrm{BTU}_o \tag{I.16}$$

where $\tau_{oc} \geq 1$ is the "iceberg cost" of transporting coal between counties o and c in 1830, θ is the trade elasticity, and BTU_o is the total amount of coal resources in county o measured in Btu.[18] Intuitively, the coal resources in county o more strongly count towards county c's coal access if the transportation costs between these counties is low. Importantly, I use transportation costs before the introduction of the railroads to avoid capturing infrastructure investments. I similarly use estimates of coal resources prior to mining to avoid contamination by selective mining. Figure IV.3 shows the spatial distribution of coal access on the county-level.

Hydropower potential is defined as the total horsepower of energy that can be feasibly generated by waterpower given the topographic characteristics of the area. Importantly, it covers both developed and undeveloped sites. Estimates of hydropower potential of each state were published by USGS at various points in time. I use the estimates of hydropower potential published in (Young, 1964, Table 10).[19] Figure IV.4 shows a map of hydropower potential across the United States.

4.2 Netherlands

For the Netherlands, I assemble a large micro-database that contains the names, occupation, residence, birth place, and wealth at death for all individuals who died in selected provinces between 1879 and 1927. The provinces cover around a half to two-thirds of the national population. Furthermore, I collected data on manufacturing on the local level for selected years. In all data, each municipality is coded to their "Amsterdamse code",

[16] The source file can be downloaded from `https://www.usgs.gov/media/files/uscoal`.

[17] Following Averitt (1975), I convert the tonnage of coal of different ranks to Btu using the following ratios: Anthracite, 12,700 Btu per pound; bituminous coal, 13,100 Btu per pound; subbituminous coal, 9,500 Btu per pound; lignite, 6,700 Btu per pound. I include the coal resource only if the overburden is less than 3,000 feet and the thickness is more than 14 inches for anthracite and (sub)bituminous coal or more than 28 inches for lignite (Averitt, 1975).

[18] Specifically, as in (Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021), $\tau_{oc} = 1 + t_{oc}/\bar{P}_{coal}$. I set $\bar{P}_{coal} = 6.08$ to the average dollar per ton anthracite coal price in 1830, Philadelphia (Chandler, 1972, Table 2). t_{oc} is the transportation cost per ton-mile between counties o and c in 1830 as estimated by Donaldson and Hornbeck (2016). The trade elasticity θ is set to 8.22 as estimated by (Donaldson and Hornbeck, 2016).

[19] Since water flow can vary seasonally, hydropower potential may not be constant within a year. I use estimates of hydropower potential available 50 percent or more of the time.

an identifier for each historical Dutch municipality.[20]

4.2.1 Wealth

The data on wealth derive from the inheritance tax administration. The tax was levied nationally since 1818. All source data up to 1927 is publicly available in regional archives in the Netherlands. Before 1878, the inheritances were only subject to tax if not all recipients were descendants in the direct line. After 1878, all inheritances above ƒ1000 (a thousand Dutch guilders) were taxed. However, the value of many estates worth less than ƒ1000 were assessed and recorded. The source files are printed tables that were filled in by hand indicating decedent's name, occupation, place of residence, marital status, date of death, and importantly, the value of their estate. The tables were referred to contemporaneously as "Tafels V-bis". Figure IV.14 is an example of a source image. It also contains decedents whose inheritance were not subject to taxation. De Vicq and Peeters (2020) have digitized the Tafels V-bis for decedents who were subject to taxation in 1921. For more information on the source, I refer to their paper.

I cover the entire period between 1879 and 1927. I included all areas for which the source files were available online as scanned images, namely the provinces Noord-Holland, Zuid-Holland, Noord-Brabant, Gelderland, and Overijssel.[21] In 1900, these five provinces contained 70 percent of the population.[22] For Zuid-Holland, scanned images were only available up to around 1900. The source files are printed tables that were filled in by hand indicating decedent's name, occupation, place of residence, marital status, date of death, and importantly, the value of their estate. Figure IV.14 is an example of a source image. The tables were digitized using Transkribus, an AI-powered platform specialized in digitization of historical records.[23] In total, I digitized more than 130 thousand images.

I mitigate noise coming from automatic digitization of the data in two ways. First, the wealth of all observations with wealth recognized to be larger than ƒ100,000 (19,178 observations) were checked by hand. Second, I link the digitized dataset to existing high-quality hand-collected information from the civil death registry by (fuzzy) matching based on name, place and date of death, and age.[24] Around 80 percent of the observations can be linked to a record in the civil death registry.

Using the data, I create a panel data on the local wealth distribution. I use the smallest geographical unit, the municipality, as the unit of analysis. To ensure a sufficient amount of observations per time period, I compute the distributional statistics by decade.[25] As reported above, all estates worth more than the taxable threshold of ƒ1000 were assessed and taxed, but many estates were assessed to be below the threshold. Which estates were assessed may have varied somewhat across tax offices and over time: the exact criteria

[20] See Huijsmans (2020) for a database of all historical municipalities.

[21] The archival sources are: Noord Hollands Archief, record group 178 (for Noord-Holland); Nationaal Archief, record group (i.e. "inventarisnummer") 3.06.05 (for Zuid-Holland); Brabants Historisch Informatie Centrum, record group 82 (for Noord-Brabant); Gelders Archief, various record groups (for Gelderland); Collectie Overijssel, record group 136.4 (for Overijssel).

[22] See http://www.volkstelling.nl for data on population by province. The four provinces for which the entire period is covered contained 47 percent of the population in 1900.

[23] For more information, see https://readcoop.eu/transkribus/.

[24] The civil registry data can be downloaded in bulk at https://www.openarch.nl/exports/csv/.

[25] Since the dataset starts in 1879, I assign that year to the 1880s too.

under which an estate was assessed are to my knowledge unknown. The need to avoid that variations in assessments affect the measures of inequality, would suggest to only include decedents with an assessed wealth above *ƒ*1000 (as they should always have been assessed). However, including as many people as possible reduces variance in the measures of inequality. I balance these interests by including every decedent with an assessed wealth above *ƒ*300 in the sample on which measures of the wealth distribution are computed.

The resulting dataset on wealth over the period of industrialization is unique in its size and geographic scope. The existing literature has focused on documenting national trends in the wealth distribution. For instance, Lindert (1986) (UK) samples 12,581 estates across four regions and five dates between 1670 and 1875, Piketty et al. (2006) (France) cover a random sample of Parisian estates in selected years in the 19th century, and Bengtsson et al. (2018) (Sweden) collect information on samples of around 5000 probate inventories between 1750 and 1900. This dataset is an illustration of the value of using newly available technologies for scalable digitization of handwritten historical records. With more than 1.5 million decedents—of which 550,966 had their wealth assessed and recorded—and coverage across the country, it allows for a detailed look on the wealth distribution. Furthermore, and importantly for the purpose of this paper, it provides complete coverage between 1879 and 1927, the period where first steam engines and then electric motors were adopted in the Netherlands.

I assess the reliability of the data by comparing the measures of inequality with data from two other sources that I have digitized. First, I uncovered a parliamentary document that recorded in large detail the distribution of income by municipality in 1883 for 79 municipalities.[26] These data were derived from local income tax administrations. I also collected data on income distributions of 8 additional cities with a local income tax whose distribution was not included in the parliamentary study.[27] The second source of the data are income and wealth distributions derived from national taxation for the largest 45 municipalities for 1926 in (Centraal Bureau voor de Statistiek, 1928). Table I.1 shows that the correlations are strong, and importantly, they are strongest for the relevant time period. For instance, the top decile share of income in 1883 correlates strongly with the top decile wealth share in 1880, but much less strongly with that in 1920. These correlations provide evidence that the data provide accurate measures of inequality both in the cross-section and over time. Furthermore, Table I.1 shows that wealth inequality among decedents (as measured by the inheritance data) correlates strongly with wealth (and income) inequality among the living population.

Lastly, I use newly digitized data on the income distribution in every municipality in 1946, the first year for which this is available (Statistics Netherlands, 1952).[28] Since over 85 percent of households were subject to income tax, I treat the taxed units as the target population for which I estimate the distribution of income. To estimate the distribution of income from the tabulations, I use the generalized Pareto interpolation method

[26] Tweede Kamer (*House of Representatives*) 1883-1884 kamerstuknummer (*document number*) 172.13. The source file can be found on `https://zoek.officielebekendmakingen.nl/0000397139`.

[27] The cities are: Breda (1880), Vlissingen (1883), Enschede (1880), Utrecht (1888), Delft (1893), Eindhoven (1885), Hilversum (1880), Nijmegen (1880). The sources for these extra cities are documented in Appendix 1.4.5.

[28] See Figure IV.15 for an image of the original source file.

TABLE I.1: Correlations between top decile shares based on inheritance data and alternative data sources

	Wealth, inheritance data				
	1880	1890	1900	1910	1920
Income, 1883	**0.86**	0.77	0.73	0.62	0.54
Income, 1926	0.38	0.33	0.54	0.60	**0.71**
Wealth, 1926	0.48	0.56	0.66	0.72	**0.76**

Notes: This table shows the correlations between the measures of municipality-level top wealth inequality for each decade derived from the inheritance data and measures of income and wealth inequality from other sources. Observations are weighted by the number of individuals on which the inheritance wealth inequality measure is based. *Sources*: local income tax data for income inequality in 1883; national income (wealth) tax data for income (wealth) inequality in 1926.

(Blanchet et al., 2022).[29]

4.2.2 Manufacturing

I use newly digitized data on manufacturing by municipality for the years 1816-1819 and 1930. The first official Census of Companies ("Bedrijfstelling") in the Netherlands was performed in 1930. It offers a high-quality snapshot of manufacturing by industry by municipality.[30] This source provides information on the number of establishments and workers by size class by industry by municipality and the adoption of motive power (in horsepower).[31] Importantly, it breaks down motive power by electric motors driven by purchased energy and other motive power (i.e., steam engines or electric motors driven by steam engines in the plant). Figure IV.16 provides an example of a source page. In total, the data consists of 33,134 municipality-by-industry observations.

The data for the years 1816-1819 derive from two government surveys from which the results are compiled and published in print by (Brugmans, 1956; Damsma et al., 1979).[32] I digitized the data from that source and coded the establishment types to a 2-digit ISIC industry code.[33] Where data is available for both 1816 and 1819, I use the data for 1819. Furthermore, I added the results for the municipality of Rotterdam and neighbouring municipalities—which were excluded by (Brugmans, 1956; Damsma et al., 1979)—from (Korteweg, 1926). The inquiry contains, by municipality, information on the number of establishments for each type of establishment (e.g. tannery or cotton factory) and the number of workers. Brugmans (1956); Damsma et al. (1979) were not able to retrieve the survey results of all municipalities in three out of eleven provinces (Zuid-Holland, Overijssel, and Groningen). The final data contain 3,658 municipality-by-industry observations in 539 distinct municipalities.[34] The data includes nearly all

[29] The R-package gpinter implements the method.

[30] While it also provides information on non-manufacturing firms, I have digitized the data only for manufacturing firms. Source images can be downloaded from https://doi.org/10.17026/dans-xqs-5q6e.

[31] The establishments are broken down by those employing none or one person, 2 to 5 persons, 6 to 10 persons, or 11 or more persons.

[32] The source images can be downloaded from https://resources.huygens.knaw.nl/nijverheid.

[33] Specifically, I coded the establishment types to the International Standard Industrial Classification of All Economic Activities, Rev. 4.

[34] Around 1200 municipalities existed at the time. For eight out of eleven provinces, (Brugmans, 1956; Damsma et al., 1979) retrieved the complete returns of the surveys so that any "missing" municipalities are likely to not have had any significant manufacturing presence. For the remaining three provinces, some mu-

large cities and other places with a strong manufacturing presence.

For comparability across years, I coded each industry or establishment type to its relevant 2-digit ISIC industry code.

5 The effect of scale-biased technical change on firm size

This section documents the impact of the adoption of steam engines—large-scale-biased technical change—and the adoption of electric motors—small-scale-biased technical change—on establishment sizes. The first prediction of the theory is that steam engine adoption causes an increase in the average establishment size, while electric motor adoption decreases it. I verify the prediction using exogenous geographical variation within the United States in the costs of the two technologies. Specifically, I use differences in access to natural coal reserves and hydropower potential across the United States as instrumental variables to identify the causal effects of adoption.

First stage. Figure IV.5 shows that "coal access" strongly affected coal prices ($\rho = -0.58$ on the state-level). I test the hypothesis that, as a result, coal access affected the adoption of steam engines. In 1890, the Census of Manufactures reported steam engine and other power use for each state-industry combination. For that year, I estimate

$$\text{STEAM}_{ist} = \delta_i + \theta \ln\left(\text{COAL}_s\right) + \epsilon_{ist} \tag{I.17}$$

where the subscripts i, s, and t refer to industry, state, and year, respectively. STEAM_{ist} refers to measures of steam engine adoption, i.e., steam engines' horsepower per employee and the share of steam engines in total horsepower. COAL_s is the measure of state s's coal access, computed as the average coal access of the counties in state s as given by equation (I.16). Standard errors are clustered at the state-level and the regression is weighted by the total number of establishments in industry i, state s, and year t. Table I.2 shows that coal resources strongly predicted steam engine adoption, both relative to employment and relative to other power sources (mostly water wheels), even within narrow industries. This relationship is robust to—and if anything strengthened by—controlling for hydropower potential and market access in state s.

nicipalities may be missing despite some manufacturing industry.

TABLE I.2: The effect of coal access on steam engine adoption (1890)

	Steam hp per worker (asinh)			Steam as share of total hp		
Coal access (logs)	0.022***	0.022***	0.023***	0.031***	0.031***	0.035***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)
Hydro-potential (logs)		-0.006**	-0.006*		-0.007	-0.006
		(0.003)	(0.003)		(0.007)	(0.005)
Market access (logs)			X			X
Observations	4237	4237	4237	3395	3395	3395

Notes: This table shows the estimated effect of coal access (in logs) on horsepower of adopted steam engines per employee and as fraction of total horsepower. Standard errors in parentheses are clustered at the state-level. Industry fixed-effects included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The price of electricity depended strongly on the "hydropower potential" that a state had to offer. Figure IV.6 shows the correlation between hydropower potential and electricity prices in 1929 on the state-level ($\rho = -0.56$). Coal access and hydropower potential are not correlated (Figure IV.7, $\rho = 0.03$). I estimate the effect of the instrument (hydropower potential) on the use of purchased electric energy, first reported in 1939. That is, I estimate for the year 1939:

$$\text{ELECTRICITY}_{ist} = \delta_i + \theta \ln\left(\text{HYDRO}_s\right) + \boldsymbol{\lambda}' \boldsymbol{X}_{ist} + \epsilon_{ist}.^{35} \tag{I.18}$$

ELECTRICITY_{ist} refers to two measures of electric motor adoption: the total megawatt hour of purchased electric energy per employee and the cost of purchased electric energy as a share of total fuel costs.[36] $\ln\left(\text{HYDRO}_s\right)$ refers to the logarithm of the hydropower potential of state s. Table I.3 shows the results. Hydropower potential caused firms to use more electric energy, relative to employment and relative to other fuels.

[35] For simplicity, I chose notation identical to (I.17). Of course, the parameters in (I.17) and (I.18) are different.

[36] The megawatt hour of purchased electric energy per employee is obtained by dividing the cost of purchased electricity by the average price of electricity per MWh for manufacturers in the state in 1939. The average price was, in turn, computed by dividing the total cost of purchased electric energy in the state (Census of Manufactures 1939, Volume 1, Ch. VII, Table 3) by the quantity purchased in MWh. (Census of Manufactures 1939, Volume 1, Ch. VI, Table 6).

TABLE I.3: The effect of hydropower potential on purchased electric energy use (1939)

	MWh per worker (asinh)			Electricity as share of fuel		
Hydro-potential	0.110***	0.116***	0.120***	0.020***	0.018***	0.017***
	(0.029)	(0.024)	(0.021)	(0.004)	(0.003)	(0.003)
Coal access		0.022	0.015		-0.007**	-0.005*
		(0.017)	(0.017)		(0.003)	(0.002)
Market access (logs)			X			X
Observations	5031	5031	5031	5010	5010	5010

Notes: This table shows the estimated effect of hydropower potential (in logs) on megawatt hour of purchased electricty per employee of adopted steam engines per employee and as fraction of total horsepower. Standard errors in parentheses are clustered at the state-level. Industry fixed-effects included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results. I estimate the reduced form effects of coal access and hydropower potential on the firm size using the following regression equation:

$$\ln(y_{ist}) = \alpha_s + \eta_{it} + \sum_{k \in T} [\beta_k \ln(\text{COAL}_s) D_{tk} + \gamma_k \ln(\text{HYDRO}_s) D_{tk}] + \boldsymbol{\lambda}' \mathbf{X}_{st} + \varepsilon_{ist} \quad \text{(I.19)}$$

where the subscripts i, s, and t refer to industry, state, and year, respectively. D_{tk} is a dummy that is 1 if $t = k$ and 0 otherwise and T contains all but one reference census year. y_{ist} is the average firm size (in terms of employment). Standard errors are clustered at the state-level and the regression is weighted by the total number of establishments in industry i, state s, and year t. $\mathbf{X}_{st}$ is a vector of controls on the state-year level: it contains the density of the population in state s at time t and interactions between time and "market access" in state s.[37] Controlling for market access ensures that the estimated effect of access to coal does not reflect low-cost access to consumer markets.

Figure I.6 shows the estimates and 95% confidence intervals for the effects of coal access and hydropower potential across years. I find that firm sizes in states with high coal access—adopting more steam engines—grew from 1850 onward relative to other states. In contrast, states with high hydropower potential—adopting more electric motors—experienced relative reductions in average firm sizes. Importantly, as depicted in Figure I.6(B), there were no differential trends in firm size based on hydropower potential prior to the electric motor's introduction between 1890 and 1900, providing evidence for the validity of the instrument.

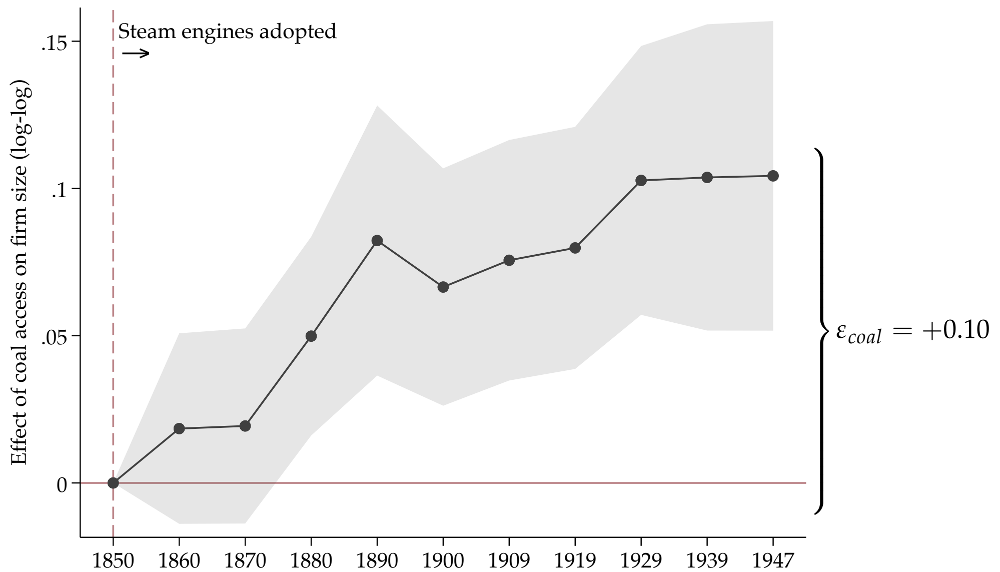
Consistent with the exclusion restriction that coal access affects firm sizes only through steam engine adoption, I show that firm sizes in industries that used little power nationally in 1890 were barely affected by coal (see Figure IV.8). Specifically, I estimate equation (I.19) for the years between 1860 and 1900, now including state × industry fixed effects using the 1860 to 1900 industry crosswalk in Appendix 1.4.3. I estimate this equation separately for a set of "placebo" industries—industries in the bottom quartile of power usage in 1890—and the remaining "treated" industries.[38] Similarly, hydropower poten-

[37] I compute market access by county for the year 1830 (before railroads) as in (Donaldson and Hornbeck, 2016) and average it to the state-level.

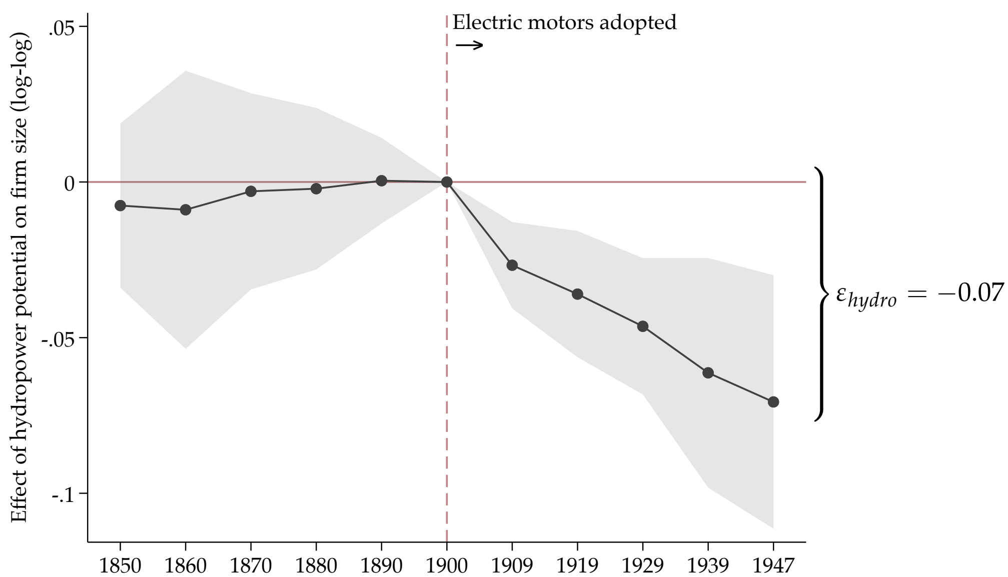
[38] Power usage is defined as the share of establishments reporting any power use.

FIGURE I.6: Effects of coal access and hydropower potential on firm sizes

(A) Coal access



(B) Hydropower potential



Notes: Panel (A) and (B) of this figure show estimates of the reduced form effects of coal access and hydropower potential on firm sizes relative to the base year, accounting for industry and state fixed effects. Estimates in Panel (A) and (B) are jointly estimated in one specification (see equation (I.19) for the econometric specification), the only difference being the base year relative to which the estimates are estimated. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the state-level.

tial only affected firm sizes in industries that used electric motors (see Figure IV.9). To test this, I run the same procedure for the years between 1890 and 1939 using the crosswalks in Appendix 1.4.4. For electric motors, I define placebo industries as those in the bottom quartile of the share of purchased electricity in overall fuel costs.

I estimate the effect of steam engine and electric motor adoption on the firm size using IV in two . Specifically, I regress state-by-industry firm size growth on adoption, instrumented by hydropower potential and coal access. That is, I estimate

$$\ln\left(y_{is,1890}\right) - \ln\left(y_{is,1860}\right) = \alpha_1 + \beta_1 \text{STEAM}_{is,1890} + \boldsymbol{\lambda}_1' \boldsymbol{X}_{is} + \varepsilon_{is} \tag{I.20}$$

$$\ln\left(y_{is,1939}\right) - \ln\left(y_{is,1900}\right) = \alpha_2 + \beta_2 \text{ELECTRICITY}_{is,1900} + \boldsymbol{\lambda}_2' \boldsymbol{X}_{is} + \eta_{is} \tag{I.21}$$

where $\text{STEAM}_{is,1890}$ and $\text{ELECTRICITY}_{is,1939}$ are steam engine horsepower per worker in 1890 and megawatt hour of purchased electricity per worker in 1939. Both are transformed using the inverse hyperbolic sine function.

Table I.4 shows the results of the instrumental variable regressions in equations (I.20) and (I.21). The estimate in the first column suggest that a 1% percent increase in steam engine use led to an increase in average firm size of about 1.1%. The second and third columns explore the sensitivity of the estimates to changes in the set of controls. While steam engines increased firm size, column four to six show that electric motor adoption decreased it with an elasticity around -0.4.

TABLE I.4: The effect of steam engine and electric motor adoption on firm sizes

	$\Delta \ln(\text{firm size}_{is})$					
	1860-1890			1900-1940		
$\text{STEAM}_{is,1890}$	1.058**	1.152**	1.089**			
	(0.450)	(0.465)	(0.483)			
$\text{ELECTRICITY}_{is,1939}$				-0.386***	-0.383***	-0.353***
				(0.094)	(0.104)	(0.113)
$\Delta \ln(\text{population density}_s)$		X	X		X	X
$\Delta \ln(\text{income/wealth p.c.}_s)$			X			X
Observations	1900	1900	1900	2117	2117	2117
Kleibergen-Paap F-stat.	42.9	33.4	24.7	16.8	14.1	13.3

Notes: This table shows the estimated effects of steam engine and electric motor adoption on the change in log firm size in a given state and industry. The explanatory variables are the inverse hyperbolic sine of steam engine horse power in 1890 and megawatt-hour of purchased electricity per worker in 1939. The adoption variables are instrumented with coal access (first three columns) and hydropower potential (last three columns). Observations are weighted by the number of establishments in the base year. Standard errors in parentheses are clustered at the state-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 The effect of scale-biased technical change on inequality

The previous section's results demonstrated that large-scale-biased technical change increases establishment sizes, while small-scale-biased technical change does the opposite. In this section, I study the second and third prediction of the theory.

The second prediction is that large-scale biased technical change increases the profit-wage ratio, a measure of income inequality between workers and entrepreneurs. I use data from the Census of Manufactures in the United States—and the same geographic variation as in the previous section—to show that steam engines increased the profit-wage ratio, while electric motors decreased it. Furthermore, I find that profit-wage ratios are, as suggested by the theory, a good proxy for economic inequality between households. Using data from the 1860 and 1870 US Census of Population, I find a remarkably strong correlation between profit-wage ratios and top wealth inequality ($\rho = 0.67$).

The third prediction of the theory is that steam engines and electric motors had opposite effects on income inequality. I use the Dutch panel data on local wealth inequality for this purpose. Local wealth inequality, besides being a measure of economic inequality in its own right, was strongly correlated with local income inequality (see Section 2). I show that wealth inequality rose in municipalities with high steam engine adoption, while it did not in those with high electric motor adoption. For identification of causal effects, I exploit that some municipalities were more exposed to the use of these technologies given their industry composition within manufacturing in 1816, long before the widespread adoption of either technology.

6.1 Profit-wage ratio

In the model in Section 3—where each entrepreneur owns one firm—the ratio between the average profits and the wage is a perfect measure of income inequality between workers and entrepreneurs. The free entry condition in equation (I.9) suggests that this ratio is proportional to the average firm size. Specifically, it implies

$$\ln\left(\frac{\bar{\pi}_{is}}{w_{is}}\right) = \text{constant} + \ln\left(\text{firm size}_{is}\right). \tag{I.22}$$

That is, the larger is the average firm size, the larger is the average profit of an establishment relative to the wage.

To test whether the free entry condition holds empirically, I estimate average profits and wages from the Census of Manufactures. Atack and Bateman (2008) estimate profits in the 1890 Census of Manufactures using information on output, wage costs, raw materials, the capital stock, and other expenses. Unfortunately, such detailed information is not available for all years. In particular, estimates of the capital stock were only reported up to 1919 and "miscellaneous expenses" only between 1890 and 1909. I therefore approximate average profits as output minus cost of raw materials and labor costs per establishment, which can be computed for all years. The correlation between this measure of average profits and the measure used by Atack and Bateman (2008) is high:

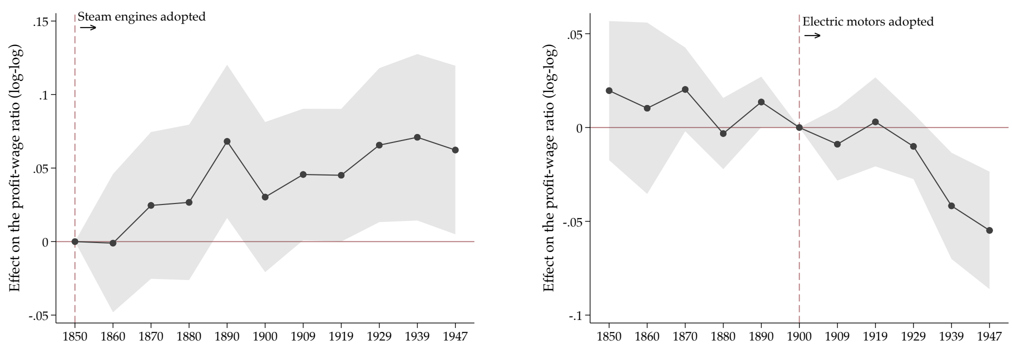
0.75 in levels and 0.96 in logs.[39] I estimate the wage as the total wage bill divided by the total number of workers. For 1940, this measure of wage income corresponds closely with the average reported wage income by state and industry in the population census, with a correlation of 0.93 in levels and 0.94 in logs.

Figure IV.10 shows that the relation between firm sizes and profit-wage ratios in equation (I.22) holds strongly in the data ($\rho = 0.87$). Because the previous section showed that firm sizes were affected by steam engine and electric motor adoption, it is natural to test whether profit-wage ratios were too. I do this by re-estimating the reduced-form effect of coal access and hydropower potential on the profit-wage ratio. Specifically, I estimate equation (I.19) where the outcome variable y_{ist} is now the profit-wage ratio in industry i, state s, and year t.

FIGURE I.7: Effects of coal access and hydropower potential on the profit-wage ratio

(A) Coal access

(B) Hydropower potential



Notes: Panel (A) and (B) of this figure show estimates of the reduced form effects of coal access and hydropower potential on the ratio between average profits and average wages relative to the base year, accounting for industry and state fixed effects. Estimates in Panel (A) and (B) are jointly estimated in one specification (see equation (I.19) for the econometric specification), the only difference being the base year relative to which the estimates are estimated. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the state-level.

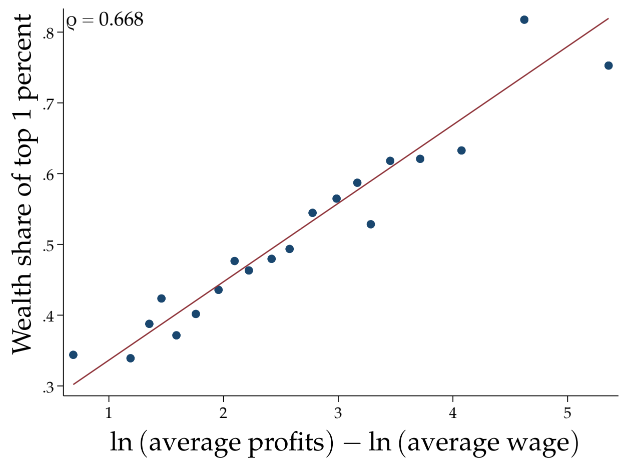
I find that the reduced form effects of coal access and hydropower potential on profit-wage ratio are qualitatively and quantitatively similar to the effects on firm size (Figure I.7). Steam engines increased the profit-wage ratio, while electric motors decreased it. Table IV.2 shows the IV estimates of the elasticity of the profit-wage ratio to steam and electric motor adoption. The point estimates are very similar to those found for the firm size in Section 5.

Under the model's assumptions, this finding is sufficient to conclude that large-scale-biased technical change—in the form of steam engine adoption—increases income inequality between workers and entrepreneurs. When technical change is large-scale-biased, fewer entrepreneurs operate in equilibrium, and the surviving entrepreneurs capture a larger share of profits than they did before. Of course, in practice, firm ownership is less concentrated than it is in the model. People may own shares in one or multiple firms, diluting the relation between the profit distribution across firms and inequality between households quantitatively.

[39]Specifically, for manufacturing censuses between 1890 and 1909, I compute profits as output minus cost of raw materials, labor costs, capital costs, and miscellaneous expenses per establishment. I compute capital costs as 4.33 percent of the capital stock. Atack and Bateman (2008) assumed a different capital cost rates for plants (2%) than for equipment (6.67%); I choose 4.33 percent as the average of these two rates.

Using data on wealth from the Census of Population in 1860 and 1870, I show that profit-wage ratios strongly correlate with measures of wealth inequality. That is, I compute top wealth inequality by year, state and 1950 industry in the Census of Population. I compute profit-wage ratios in the Census of Manufactures by the same industry classification using newly created crosswalks. Figure I.8 illustrates the strong relationship between wealth inequality (as measured by the share of wealth held by the top 1 percent) and the profit-wage ratio. This shows that the profit-wage ratio is a good proxy for inequality.

FIGURE I.8: The profit-wage ratio correlates strongly with wealth inequality



Notes: This figure shows a bin scatter visualizing the correlation of wealth inequality and the profit-wage ratio by state and industry. Each dot is an industry-state-year combination in 1860 and 1870. Wealth inequality is computed from the Census of Population. Average profits are approximated by dividing total output minus cost of raw materials and labor costs by the number of establishments. The wage rate is approximated by dividing total wage costs by employment.

The finding that steam engines increased profit-wage ratios and electric motors decreased them, coupled with the strong correlation between profit-wage ratios and inequality, suggests that steam engines increased inequality, while electric motors decreased it. Direct evidence on income or wealth is, however, not available for the United States after 1870. Therefore, to test whether scale-biased technical change affects inequality in the personal income and wealth distribution, I use data from the Netherlands for which detailed information on wealth and income over a long horizon is available.

6.2 Wealth and income inequality

I use the digitized Dutch inheritance tax data to create various measures of local inequality for the period between 1879 and 1927. With this dataset, I first study how wealth inequality evolved across municipalities with varying rates of adoption of steam engines and electric motors. I use wealth inequality, rather than income inequality, primarily for reasons of data availability. Table I.1 shows, however, that income and wealth inequality are strongly correlated. Furthermore, I also estimate the effects on income inequality for a subset of municipalities for which data is available.

As a measure of adoption, I use the share of local manufacturing employment that works in establishments using the technologies. I measure this using the newly digitized 1930 Census of Dutch Companies. Particularly, I divide establishments in three groups: 1) those using prime movers run by energy generated in the plant (steam engines), 2) those only using prime movers run by purchased electricity (electric motors), and 3) those not using any prime movers at all. The measure of local steam engine adoption is the share of workers in the first type of establishments. Similarly, electric motor adoption is measured as the share of workers in the second group of establishments, so that:

$$\text{STEAM}_{1930,m} = \frac{\text{Employment in plants using prime movers run by generated energy in } m}{\text{Total employment in } m} \tag{I.23}$$

$$\text{ELECTR}_{1930,m} = \frac{\text{Employment in plants using prime movers run by purchased electricity in } m}{\text{Total employment in } m}. \tag{I.24}$$

The main specifications are as follows:

$$\text{INEQUALITY}_{mt} = \alpha_{1m} + \eta_{1t} + \sum_{k \in T \setminus \{1880\}} \beta_{1k} \left(\text{STEAM}_{1930,m} \times D_{tk} \right) + \varepsilon_{1,mt} \tag{I.25}$$

$$\text{INEQUALITY}_{mt} = \alpha_{2m} + \eta_{1t} + \sum_{k \in T \setminus \{1880\}} \beta_{2k} \left(\text{ELECTR}_{1930,m} \times D_{tk} \right) + \varepsilon_{2,mt} \tag{I.26}$$

where the subscript $t \in T = \{1880, 1890, 1900, 1910, 1920\}$ refers to the decade, m to the municipality and D_{tk} is a dummy that 1 if $t = k$ and 0 otherwise. INEQUALITY$_{mt}$ is the share of wealth held by the top 1% of decedents with wealth. The coefficients β_{1k} and β_{2k} capture the association between steam engine and electric motor adoption and the change in wealth inequality from 1880, the reference year, to year k.

Figure IV.11(a) plots the coefficients of β_t for each decade relative to 1880. The coefficient suggest that a 1 percentage point increase in the share of employment exposed to steam engines leads to an increase in the top 1% wealth share of about 0.2 percentage points. This effect is statistically and economically significant. Local steam engine adoption varied strongly: around 10 percent of municipalities adopted no steam engines at all, while in some municipalities more than 90 percent of manufacturing employment was in steam-powered establishments. A one standard deviation increase in steam engine adoption (0.19) increases the top 1% wealth share by around 4 percentage points in 1920. The average top 1% wealth share across municipalities was 21 percent.

The estimated effects of electric motor adoption on wealth inequality are shown in Figure IV.11(b). The figure shows that electric motor adoption did not increase wealth inequality. If anything, it decreased it. However, the size of the estimated effect is smaller than for steam engines and not statistically significant on the 95% confidence level.

The coefficients in Figure IV.11 reflect the different evolution of wealth inequality in municipalities along one dimension of power usage (steam engine adoption or electric motor adoption). When electric motor adoption is low, this could be because mostly steam engines were used or because there was little use of power of any sort. To directly compare the effect of steam engine adoption and electric motor adoption, I also estimate equation (I.25) while controlling for the share of employment in establishments that do not use any power in 1930 (similarly interacted with time dummies).[40] Since $\text{STEAM}_{1930,m}$, $\text{ELEC}_{1930,m}$, and $\text{NOPOWER}_{1930,m}$ sum to one by construction, the coefficient of interest in this regression reflects the increase in wealth inequality associated with a 1 percentage point increase in steam engine adoption and a 1 percentage point *decrease* in electric motor adoption. The results are shown in Figure IV.12. It shows that holding total power usage constant, when more steam engines were used—and thus less electric motors—wealth inequality increased relative to 1880.

Instrumental variable analysis. The municipality-fixed effects specifications in equations (I.25) and (I.26) control for any time-invariant unobserved heterogeneity across municipalities. Time-varying heterogeneity is a potential remaining threat to causal interpretation of the coefficients in Figure IV.11. For instance, it is a priori conceivable that changes in local inequality between 1880 and 1920 also affected technology adoption, leading to reverse causality. To assess the quantitative importance of such threats to identification, I employ an instrumental variable strategy.

The identification strategy uses that the local industry composition in manufacturing in 1816 (see Section 4.2.2 for details on the data) is predictive of the local adoption rates of steam engines and electric motors. I assign 2-digit ISIC industry codes to each industry in the manufacturing data in 1930 and 1816. Then, using the 1930 data, I compute industry i's adoption of steam engines and electric motor adoption. The adoption rates are computed analogously to $\text{STEAM}_{1930,m}$ and $\text{ELECTR}_{1930,m}$ in equations (I.23) and (I.24), only changing the unit of analysis from municipality m to industry i.

Table IV.3 shows the adoption rates for each manufacturing industry. The textile industry, together with the much smaller beverage industry, was the largest adopter of steam engines, with half of employment in establishments using steam. On the other hand, the leather, apparel, tobacco, and printing industries almost did not use any steam engines at all. Using these adoption rates in 1930, I then compute the exposure to steam engines and electric motors in municipality m in 1816 as:

$$\text{STEAM\_EXP}_{1816,m} = \sum_{i \in I} \frac{\text{Employment in industry } i \text{ in } m \text{ in 1816}}{\text{Total employment in } m \text{ in 1816}} \times \text{STEAM}_{1930,i} \quad \text{(I.27)}$$

$$\text{ELECTR\_EXP}_{1816,m} = \sum_{i \in I} \frac{\text{Employment in industry } i \text{ in } m \text{ in 1816}}{\text{Total employment in } m \text{ in 1816}} \times \text{ELECTR}_{1930,i}. \quad \text{(I.28)}$$

The exposure measure is a strong predictor of actual adoption in 1930 (see Table IV.4 for the correlation).

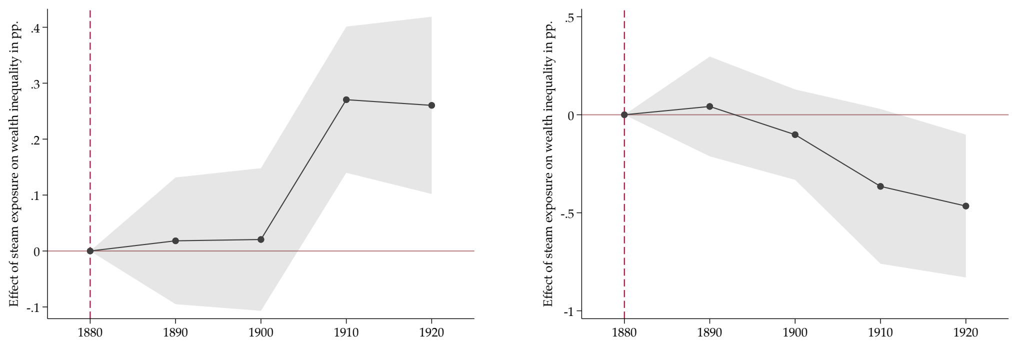
I estimate the "reduced form" of the instrumental variable analysis equivalently to equations (I.25) and (I.26) except that the actual adoption rates are changed for the predicted rates in equations (I.27) and (I.28). That is, I estimate how wealth inequality evolved between 1880 and 1927 across municipalities that were more or less exposed to the two technologies.

[40] That is, I estimate:

$$\text{INEQUALITY}_{mt} = \alpha_{3m} + \eta_{3t} + \sum_{k \in T \setminus \{1880\}} [\beta_{3k} (\text{STEAM}_{1930,m} \times D_{tk}) + \gamma_{3k} (\text{NOPOWER}_{1930,m} \times D_{tk})] + \varepsilon_{3,mt}.$$

FIGURE I.9: Steam engine adoption increased wealth inequality, electric motors did not

(A) Effect of steam engines (B) Effect of electric motors



Notes: This figure shows the estimated effects in percentage points of pre-industrial exposure to steam engine (in panel A) and electric motor adoption (in panel B) on within-municipality top wealth inequality (top 1% share) for each decade relative to 1880. The instrumental variable is exposure to the respective technology which is computed on the basis of the local industry composition in 1816 and adoption rates by industry in 1930. Observations are weighted by the number of individuals on which the inequality measure is based. Shaded areas represent 95% confidence intervals.

Figure I.9 shows that places more exposed to steam engines became more unequal, while places more exposed to electric motors became more equal, providing further evidence that steam engines and electric motors had a causal effect on inequality as predicted by the theory.

Further evidence using income data. The model of scale-biased technical change proposed in this paper relates technical change to income inequality. Since wealth inequality is strongly correlated with income inequality (see Table I.1) and consistent time-series data are available for local wealth inequality (but not for income inequality), I use wealth inequality as the dependent variable for the main analysis. I nonetheless assess the robustness of the results to using income inequality as the outcome variable.

As described in Section 4.2.1, I uncovered and digitized data on the income distribution in 1883 for 87 (mostly large) municipalities and for all municipalities in 1946. From there, I compute the percentage point change in income inequality (as measured by the income share of the top percentile) between 1946 and 1883. I regress the growth in income inequality on $\text{STEAM}_{1930,m}$ and $\text{ELECTR}_{1930,m}$ defined in equations (I.23) and (I.24), using ordinary least squares as well as using the respective instrumental variables. Table IV.5 shows the results. It verifies the results obtained using wealth inequality as the dependent variable: steam engine adoption increased inequality, while electric motors had a marginal negative effect.

7 Who gains from large-scale-biased technical change?

Section 6 showed that steam engine adoption led to increased inequality, while electric motor adoption did not. The last question is then: how did steam engines increase inequality? In this section, I zoom in to Enschede—the major Dutch textile city—to understand *who* was capturing the rents from large-scale-biased technical change. I find that the increased inequality was predominantly due to the textile factory owners amassing wealth at a much higher rate than other households. This finding confirms the prediction of the theory of scale-biased technical change that the concentration of business income, not of wages, was the key driver of increased inequality.

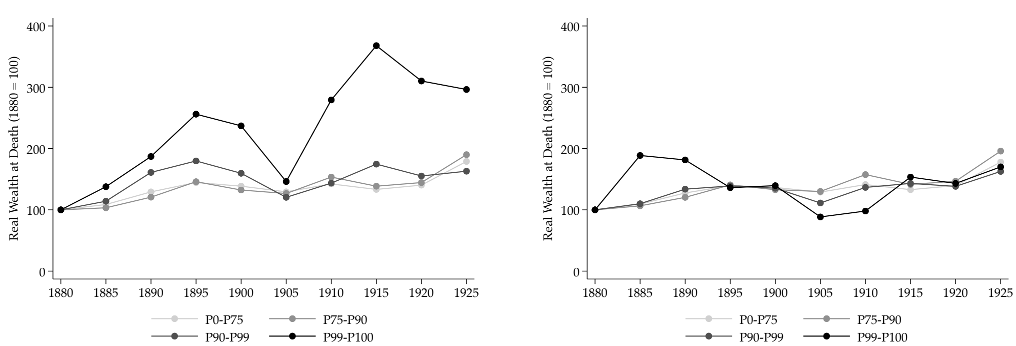
I selected Enschede for this case study because, being a major textile producer, it heavily depended on steam engines and witnessed a strong increase in wealth inequality. Figure IV.13 charts the wealth share of the top 1% over time. Another advantage of studying Enschede is that the history of its textile industry is well documented and the identities of the factory owners are known.

The foundations of the textile industry in Twente, the region around Enschede, already had been laid in the 16th century. At the time, many Flemish entrepreneurs had their linen woven in Twente, due to its attractive position between Amsterdam and North Germany (Schot et al., 2003). In 1728, Enschede had acquired the right to produce *bombazijn*, a textile woven from a combination of linen and cotton threads, and it became the largest producer of this textile halfway into the 18th century (Stroink, 1962). By 1750, 40% of the labor force was occupied in the textile industry. Since textile manufacturing was the industry most exposed to steam engines (see Table IV.3), Enschede's rate of steam engine adoption was among the highest in the country.

The theory predicts that large-scale-biased technical change impacts inequality through the profits accrued by entrepreneurs. Therefore, one should expect the see that wealth inequality is driven mostly by them. To test this prediction, I compute the evolution of average wealth in different parts of the wealth distribution on samples including and excluding textile owners. Specifically, I exclude people from the sample if they belong to one of 22 families that are considered the "core" and "inner circle" of textile owners by Willink (2015). I use the last name as a proxy for family membership.[41]

FIGURE I.10: Wealth inequality is driven by entrepreneurs adopting steam engines

(A) Including textile owners (B) Excluding textile owners



Notes: This figure shows the evolution of the top 1 percent wealth share in Enschede when this measure is estimated on the full population (in panel A) and when measured on the sample excluding textile owners (in panel B). For each year, wealth inequality is computed from the sample of decedents in a 10-year window around it.

Figure I.10(A) shows the mean wealth at death for different percentile groups. It illustrates that wealth inequality increased through a divergence of the top 1 percent from

[41] The last names are: Blijdenstein, Ten Cate, Van Heek, Jannink, Ter Kuile, Scholten, Stork, Van Delden, Elderink, Van Gelderen, Gelderman, Hofkes, Ter Horst, Jordaan, Ledeboer, Menko, De Monchy, Palthe, Salomonson, Spanjaard, Stroink, Willink Cromhoff, Jannink, Gelderman, Heek, Ledeboer, Kuile, and Scholten.

the rest of the distribution. However, panel (B) indicates that wealth inequality among everyone except the textile families Figure I.10(B) did not go up. These pattern indicate the importance of studying inequality in the overall population, not only among wage earners. Scale-biased technical change primarily affects the concentration of business income. Therefore, it most strongly affects the income of top business owners relative to the rest of the distribution.

8 Conclusion

In this paper, I highlight a new channel through which technical change can affect inequality: scale bias, the degree to which technical changes increases the relative productivity of large firms. I show that technical change is large-scale-biased if it increases fixed costs. When fixed costs of a new technology are sufficiently high, only the largest firms opt to incur the fixed cost to reduce marginal cost, while smaller firms keep using the existing technology or even go out of business. As a result, profits concentrate into a smaller set of firms. With fewer and larger firms, top entrepreneurs are capturing a larger share of the profits, pushing top income inequality up.

I showed that the adoption of steam engines and electric motors offer a unique opportunity to test the theory: while the two technologies are otherwise similar, the fixed costs of steam engines were an order of magnitude larger. I then tested the theoretical predictions on the effects of steam engine adoption (large-scale-biased) and electric motor adoption (small-scale-biased). I found that the effects of these technologies were in line with the theory's prediction: steam engine adoption increased firm sizes and inequality while electric motor adoption reduced it.

While this research shows that entrepreneurs and their incomes are key for shaping and understanding inequality, existing work primarily focuses on the impact of technical change on wage inequality, not overall income inequality.[42] The effect of technical change on the distribution of business income and inequality between workers and entrepreneurs has, to the best of my knowledge, so far not been studied. This is an important omission, because business income is a large source of income, especially at the top of the distribution. In the US, more than half of total income for the top 0.1 percentile is business income (Smith et al., 2019). Similarly, 81 percent of individuals in the top 1 percent of the wealth distribution was a business owner or self-employed (Cagetti and De Nardi, 2006).

Even today, the concentration of firm ownership is high, so that the distribution of profits across firms matters for the distribution of income across people. In the US, "pass-through" businesses account for 51 percent of all business income in 2013 (Nelson, 2016).[43] The typical such business is owned by one to three people (Smith et al., 2019) and 69% of its income accrues to the top 1% (Cooper et al., 2016). The great bulk of the remaining income is earned by a small share of publicly traded firms (Clarke and

[42] As a notable exception, Moll et al. (2022) recently expanded the scope beyond wage inequality by studying automation's effect on income (and wealth) derived from both wages and capital: by raising the returns to capital, automation increases income and wealth inequality.

[43] Pass-through businesses are businesses that are not subject to corporate tax and whose income instead "pass through" to their owners to be taxed under individual income tax. Specifically, they comprise S-corporations, sole proprietorships, and partnerships.

Kopczuk, 2017). While ownership of publicly traded firms is less concentrated, it is not as diffuse as commonly thought.[44] Even for firms in the Fortune 500, the 500 largest US firms by revenue, founding families alone accounted for 18 percent of outstanding equity between 1992 and 1999 (Anderson and Reeb, 2003).[45]

Trends in the last three decades are consistent with the implications of large-scale-biased technical change. First, firm sizes and concentration are increasing and entrepreneurship is in decline (Autor et al., 2017, 2020; Salgado, 2020; Jiang and Sohail, 2023; Kwon et al., 2023). A large and growing theoretical literature relates these patterns to technical change, specifically the growing importance of scale advantages arising from intangible capital and information technology (Brynjolfsson et al., 2008; De Ridder, 2023; Hsieh and Rossi-Hansberg, 2023; Kwon et al., 2023; Lashkari et al., 2023). Unger (2022) shows that specifically customized software (large fixed adoption cost) is highly skewed to large firms, while pre-packaged software (low fixed adoption cost) is used by small and large firms alike. Second, top income and wealth inequality has increased sharply. For example, between 1980 and 2014, the United States experienced 21% growth in the incomes of the bottom half of the distribution, while the top 10 percent saw their incomes more than double during the same period (Piketty et al., 2018). Third, since the 1990s, business income—not wage income—accounts for the largest part of the rise of top incomes in the United States (Smith et al., 2019, Figure IX). This paper provides a unified framework to understand all these trends.

This paper leaves several important questions for future research. First, in the stylized model presented, technical change and its direction is exogenous. While this assumption is reasonable in the case of steam engine and electric motor adoption in the US and the Netherlands, modelling technical change as the outcome of a directed research effort could provide further useful insights. A concentrated firm size distribution may further incentivize large-scale-biased technical change, similar to how the skill distribution may induce innovation in technologies that complement the more abundant factor (Acemoglu, 2002). Another important simplification of the model is that while technology adoption matters for inequality, inequality does not matter for technology adoption . A useful, more quantitative, model could include risk aversion or liquidity constraints. In such models, entrepreneurship is skewed towards high wealth individuals because they are more equipped to take risk and can afford larger up-front investments (Quadrini, 2000; Cagetti and De Nardi, 2006; Buera and Shin, 2013). High fixed cost technologies may further reduce entry of low-wealth individuals and can thus worsen aggregate productivity (Buera et al., 2011). Lastly, the on-going development of artificial intelligence technologies raises important questions on its distributional effects. Research shows that large firms tended to be the early adopters of the technology (McElheran et al., 2023). More research into the cost structure of these technologies is necessary to understand whether this will remain the case as the technologies mature.

[44] For instance, among a random sample of US publicly traded firms, 96 percent had shareholders that own at least 5% of the stock, and in 53 percent of firms, the largest shareholder is a family (Holderness, 2009).

[45] Peter (2021) shows evidence on concentrated ownership of European firms.

Bibliography

ABBOTT, E. (1905): "The wages of unskilled labor in the United States 1850-1900," *Journal of Political Economy*, 13, 321–367.

ACEMOGLU, D. (2002): "Directed technical change," *The review of economic studies*, 69, 781–809.

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

ACEMOGLU, D. AND P. RESTREPO (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 108, 1488–1542.

——— (2022): "Tasks, automation, and the rise in US wage inequality," *Econometrica*, 90, 1973–2016.

ALBUQUERQUE, D. AND T. IFERGANE (2023): "The Racial Wealth Gap: the Role of Entrepreneurship," .

ANDERSON, R. C. AND D. M. REEB (2003): "Founding-family ownership and firm performance: evidence from the S&P 500," *The journal of finance*, 58, 1301–1328.

ATACK, J. (1979): "Fact in fiction? The relative costs of steam and water power: a simulation approach," *Explorations in Economic History*, 16, 409–437.

ATACK, J. AND F. BATEMAN (1999): "Nineteenth-century US industrial development through the eyes of the census of manufactures a new resource for historical research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 32, 177–188.

——— (2008): "Profitability, firm size, and business organization in nineteenth-century US manufacturing," in *Quantitative Economic History*, Routledge, 72–95.

ATACK, J., F. BATEMAN, AND R. A. MARGO (2008): "Steam power, establishment size, and labor productivity growth in nineteenth century American manufacturing," *Explorations in Economic History*, 45, 185–198.

ATACK, J., F. BATEMAN, AND T. WEISS (1980): "The regional diffusion and adoption of the steam engine in American manufacturing," *The Journal of Economic History*, 40, 281–308.

ATKESON, A. G. AND M. IRIE (2022): "Rapid dynamics of top wealth shares and self-made fortunes: What is the role of family firms?" *American Economic Review: Insights*, 4, 409–424.

AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. V. REENEN (2017): "Concentrating on the Fall of the Labor Share," *American Economic Review*, 107, 180–185.

AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2020): "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135, 645–709.

AVERITT, P. (1975): *Coal Resources of the United States, January 1, 1974*, 1410-1412, US Government Printing Office.

BANK OF ENGLAND (2017): "A millennium of macroeconomic data," .

BELL, L. (1891): "Electricity as the rival steam," *Journal of the Franklin Institute*, 131, 212–227.

BENGTSSON, E., A. MISSIAIA, M. OLSSON, AND P. SVENSSON (2018): "Wealth inequality in Sweden, 1750–1900," *The Economic History Review*, 71, 772–794.

BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and productivity in international trade," *American economic review*, 93, 1268–1290.

BLANCHET, T., J. FOURNIER, AND T. PIKETTY (2022): "Generalized Pareto curves: theory and applications," *Review of Income and Wealth*, 68, 263–288.

BOLTON, D. (1926): "The variation of efficiency with size, and the economic choice of electrical machinery," *Journal of the Institution of Electrical Engineers*, 64, 337–348.

BRESNAHAN, T. F. AND M. TRAJTENBERG (1995): "General purpose technologies 'Engines of growth'?" *Journal of econometrics*, 65, 83–108.

BRUGMANS, I. J. (1956): *Statistieken van de Nederlandse nijverheid uit de eerste helft der 19e eeuw*, vol. 98, M. Nijhoff.

BRYNJOLFSSON, E., A. MCAFEE, M. SORELL, AND F. ZHU (2008): "Scale without mass: business process replication and industry dynamics," *Harvard Business School Technology & Operations Mgt. Unit Research Paper*.

BUERA, F. J., J. P. KABOSKI, AND Y. SHIN (2011): "Finance and development: A tale of two sectors," *American economic review*, 101, 1964–2002.

BUERA, F. J. AND Y. SHIN (2013): "Financial frictions and the persistence of history: A quantitative exploration," *Journal of Political Economy*, 121, 221–272.

BUSTOS, P. (2011): "Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms," *American economic review*, 101, 304–340.

CAGETTI, M. AND M. DE NARDI (2006): "Entrepreneurship, frictions, and wealth," *Journal of political Economy*, 114, 835–870.

CENTRAAL BUREAU VOOR DE STATISTIEK (1928): "Statistiek der Rijksinkomsten over het jaar 1926," The Hague, Netherlands: Drukkerij Trio.

CHANDLER, A. D. (1972): "Anthracite coal and the beginnings of the industrial revolution in the United States," *Business History Review*, 46, 141–181.

CLARKE, C. AND W. KOPCZUK (2017): "Business income and business taxation in the United States since the 1950s," *Tax Policy and the Economy*, 31, 121–159.

COOPER, M., J. MCCLELLAND, J. PEARCE, R. PRISINZANO, J. SULLIVAN, D. YAGAN, O. ZIDAR, AND E. ZWICK (2016): "Business in the United States: Who Owns It, and How Much Tax Do They Pay?" *Tax Policy and the Economy*, 30, 91–128.

DAMSMA, D., J. DE MEERE, AND L. NOORDEGRAAF (1979): *Statistieken van de Nederlandse nijverheid uit de eerste helft der 19e eeuw: supplement*, vol. 3, Nijhoff.

DE LOECKER, J., T. OBERMEIER, AND J. VAN REENEN (2022): "Firms and inequality," .

DE RIDDER, M. (2023): "Market power and innovation in the intangible economy," *American Economic Review*.

DE VICQ, A. AND R. PEETERS (2020): "Introduction to the Tafel v-bis Dataset: Death Duty Summary Information for The Netherlands, 1921: Social and Economic History," *Research Data Journal for the Humanities and Social Sciences*, 5, 1–19.

DEVINE, W. D. (1983): "From shafts to wires: Historical perspective on electrification," *The Journal of Economic History*, 43, 347–372.

DIXIT, A. K. AND J. E. STIGLITZ (1977): "Monopolistic competition and optimum product diversity," *American Economic Review*, 67, 297–308.

DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 131, 799–858.

DU BOFF, R. B. (1967): "The introduction of electric power in American manufacturing," *The Economic History Review*, 20, 509–518.

——— (1979): *Electric Power in American Manufacturing, 1889-1958*, New York, NY: Arno Press.

EMERY, C. E. (1883): "The Cost of Steam Power," *Transactions of the American Society of Civil Engineers*, 12, 425–431.

GAGGL, P., R. GRAY, I. MARINESCU, AND M. MORIN (2021): "Does electricity drive structural transformation? Evidence from the United States," *Labour Economics*, 68, 101944.

GOLDIN, C. AND L. F. KATZ (1998): "The origins of technology-skill complementarity," *The Quarterly journal of economics*, 113, 693–732.

GOLDSMITH, R. W. ET AL. (1940): "The distribution of ownership in the 200 largest nonfinancial corporations," *Washington DC: Securities and Exchange Commission.*

HANNAH, L. (1979): *Electricity before nationalisation*, Springer.

HELPMAN, E. (1998): *General purpose technologies and economic growth*, MIT press.

HOLDERNESS, C. G. (2009): "The myth of diffuse ownership in the United States," *The Review of Financial Studies*, 22, 1377–1408.

HORNBECK, R. AND M. ROTEMBERG (2021): "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies," .

HSIEH, C.-T. AND E. ROSSI-HANSBERG (2023): "The industrial revolution in services," *Journal of Political Economy Macroeconomics*, 1, 000–000.

HUIJSMANS, D. (2020): "HSN Gazetteer," .

HUNTER, L. C. AND L. BRYANT (1991): *A History of Industrial Power in the United States, 1780-1930, Volume 3: The Transmission of Power*, The MIT Press.

JIANG, H. AND F. SOHAIL (2023): "Skill-biased entrepreneurial decline," *Review of Economic Dynamics*, 48, 18–44.

KAPP, G. (1894): *Electric Transmission of Energy and Its Transformation, Subdivision, and Distribution: A Practical Handbook*, Whittaker & Company.

KATZ, L. F. AND K. M. MURPHY (1992): "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 107, 35–78.

KOPCZUK, W. AND E. ZWICK (2020): "Business incomes at the top," *Journal of Economic Perspectives*, 34, 27–51.

KORTEWEG, S. (1926): *Rotterdams welvaartsbronnen in 1816*, Brusse.

KRUSELL, P., L. E. OHANIAN, J.-V. RÍOS-RULL, AND G. L. VIOLANTE (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 68, 1029–1053.

KUZNETS, S. (1955): "Economic growth and income inequality," *The American economic review*, 45, 1–28.

KWON, S. Y., Y. MA, AND K. ZIMMERMANN (2023): "100 years of rising corporate concentration," *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

LASHKARI, D., A. BAUER, AND J. BOUSSARD (2023): "Information technology and returns to scale," *American Economic Review*.

LEKNES, S. AND J. MODALSLI (2020): "Who Benefited from Industrialization? The Local Effects of Hydropower Technology Adoption in Norway," *The Journal of Economic History*, 80, 207–245.

LINDERT, P. H. (1986): "Unequal English wealth since 1670," *Journal of Political Economy*, 94, 1127–1162.

LINDERT, P. H. AND J. G. WILLIAMSON (2016): *Unequal Gains: American Growth and Inequality since 1700*, Princeton University Press.

LUCAS, R. E. J. (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, 508–523.

MCELHERAN, K., J. F. LI, E. BRYNJOLFSSON, Z. KROF, E. DINLERSOZ, L. FOSTER, AND N. ZOLAS (2023): "AI Adoption in America: Who, What, and Where," *NBER Working Paper*.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.

MELITZ, M. J. AND S. J. REDDING (2015): "New trade models, new welfare implications," *American Economic Review*, 105, 1105–1146.

MOLL, B., L. RACHEL, AND P. RESTREPO (2022): "Uneven growth: automation's impact on income and wealth inequality," *Econometrica*, 90, 2645–2683.

NELSON, S. C. (2016): *Paying Themselves: S Corporation Owners and Trends in S Corporation Income, 1980-2013*, Department of the Treasury.

PETER, A. (2021): "Equity Frictions and Firm Ownership," .

PIKETTY, T., G. POSTEL-VINAY, AND J.-L. ROSENTHAL (2006): "Wealth concentration in a developing economy: Paris and France, 1807-1994," *American economic review*, 96, 236–256.

PIKETTY, T., E. SAEZ, AND G. ZUCMAN (2018): "Distributional national accounts: methods and estimates for the United States," *The Quarterly Journal of Economics*, 133, 553–609.

POSCHKE, M. (2018): "The firm size distribution across countries and skill-biased change in entrepreneurial technology," *American Economic Journal: Macroeconomics*, 10, 1–41.

QUADRINI, V. (2000): "Entrepreneurship, saving, and social mobility," *Review of economic dynamics*, 3, 1–40.

SALGADO, S. (2020): "Technical change and entrepreneurship," *Available at SSRN 3616568*.

SCHOT, J. W., H. W. LINTSEN, A. RIP, AND A. A. DE LA BRUHÈZE (2003): *Techniek in Nederland in de twintigste eeuw*, Walburg Pers.

SMITH, M., D. YAGAN, O. ZIDAR, AND E. ZWICK (2019): "Capitalists in the Twenty-first Century," *The Quarterly Journal of Economics*, 134, 1675–1745.

STATISTICS NETHERLANDS (1952): "Gegevens per gemeente betreffende de inkomensverdeling 1946," The Hague, Netherlands.

STROINK, L. A. (1962): *Stad en land van Twente*, Smit.

UNGER, G. (2022): "Scale-Biased Technological Change," .

VICKERS, C. AND N. L. ZIEBARTH (2018): "United States Census of Manufactures, 1929-1935," .

VIOLANTE, G. L. (2008): "Skill-biased technical change," *The new Palgrave dictionary of economics*, 2, 1–6.

WILLINK, B. (2015): *De textielbaronnen: Twents-Gelders familisme en de eerste grootindustrie van Nederland 1800-1980*, Walburg Pers.

YEAPLE, S. R. (2005): "A simple model of firm heterogeneity, international trade, and wages," *Journal of international Economics*, 65, 1–20.

YOUNG, L. L. (1964): *Summary of Developed and Potential Waterpower of the United States and Other Countries of the World, 1955-62*, vol. 483, US Government Printing Office.

II Jim Crow and Black Economic Progress After Slavery

Lukas Althoff (Stanford) & Hugo Reichardt (LSE)

1 Introduction

Black Americans have faced a long history of economic oppression in the United States. Throughout the country's early history, slavery was legal—until around 1800 in the Northern states and until the end of the Civil War (1861–1865) in the South. Soon after slavery ended, Southern states created racially oppressive regimes that limited the economic progress of newly freed Black families—a set of institutions collectively known as Jim Crow. States' Jim Crow regimes instituted racial segregation, Black voter disenfranchisement, and restrictions to Black Americans' economic and geographic mobility.[1] The Jim Crow era persisted for almost 100 years and only ended with the passage of the Civil Rights legislation in the 1960s, which outlawed racial discrimination.

This paper studies the extent to which Black Americans' economic status continues to be shaped by their ancestors' historical exposure to racial oppression. Our results reveal that such exposure continues to impact Black families, primarily because it increased their likelihood of facing continued oppression under subsequent regimes. Specifically, we find that Black families whose ancestors were enslaved until the Civil War still have far lower economic status than those who were free before the Civil War. However, the importance of differential exposure to slavery per se in contributing to these disparities dissipated over the early 20th century.[2] Instead, the gap faced by families formerly enslaved until the Civil War persists due to their disproportionate exposure to continued oppression under Jim Crow. The rapid southern expansion of the US plantation economy meant that the longer a family was enslaved, the more likely they were to be concentrated in the southernmost states—later the epicenter of Jim Crow. The severe and long-lasting impact of Jim Crow institutions thus perpetuated the economic disadvantage faced by formerly enslaved families to the 21st century.

We develop new methods to overcome the challenge of measuring families' historical exposure to slavery and Jim Crow. First, we infer if a family was free before the Civil War based on their ancestors' presence in the 1850 or 1860 census, which only enumerated free Black people. We then trace enslavement status across generations using 1) automated record-linkage (Abramitzky et al., 2021) and 2) a new surname-based approach (Ager et al., 2021). Second, we measure a family's exposure to Jim Crow by combining their ancestors' location, traced through automated record-linkage, with proxies for each state's Jim Crow intensity. Finally, we relate our exposure measures to the outcomes of Black prime-age men. Our linking-based approach uses individual-level census data (1850–1940) and neighborhood-level proxies for the late-life economic status of individ-

[1] Throughout this paper, we use the term "Jim Crow" to refer to state-level institutions that limited Black Americans' civil rights. Examples include school segregation, vagrancy laws, and poll taxes.

[2] To quantify differences in exposure to slavery, we estimate that the average free Black family was free 50 years before the Civil War—around 1815. We do so by using aggregate counts of the Black population starting in 1790 and assuming that free Black families' fertility equaled that of white families.

uals who experienced both the Jim Crow era and its aftermath, derived from mortality records (1988–2007) linked to the 1940 census. The surname-based approach extends the coverage from the linked sample to the entire historical census population and real-time credit bureau data (2023).[3]

Our first result is that today, Black families enslaved until the Civil War continue to have lower education, income, and wealth than Black families freed before the Civil War. These Free-Enslaved gaps are almost half as large as the corresponding Black-white gaps. While the Free-Enslaved gaps were even larger immediately after slavery, their narrowing has been much slower than one would expect under standard rates of intergenerational mobility. We demonstrate the robustness of our results to measurement error in ancestors' enslavement status by combining our surname- and linking-based measures in an instrumental variable strategy.

Second, we find that the Free-Enslaved gap persisted because families enslaved until the Civil War were disproportionately concentrated in states that harmed Black economic progress after slavery. We use plausibly exogenous variation from enslavement locations to estimate each Southern state's effect on the descendants of those freed from slavery there. We find that these effects were large and drive the Free-Enslaved gap's persistence. Conditional on their ancestor's location, the economic status of Black Americans ceased to depend on their ancestor's enslavement status by 1940. Importantly, our results capture only the additional disadvantage faced by those enslaved until the Civil War, not the broader impact of slavery on all Black Americans regardless of when they gained freedom.

Our third result is that Jim Crow institutions underlie the severely limiting effects of certain states on Black economic progress. To isolate the impact of these state institutions from other factors, such as economic activity, culture, or climate, we use a regression discontinuity design that compares the outcomes of Black families freed across state borders. We find that with the onset of the Jim Crow era, Black economic progress began to diverge sharply across state borders. For example, families freed in Louisiana attained 1.2 fewer years of education by 1940 compared to families freed just a few miles away in Texas. Notably, the long-run border discontinuity estimates, which capture the effects of institutions, are nearly identical in magnitude to the overall long-run state effects, which encompass both institutional and non-institutional factors. Moreover, these border differences increase with the difference in the intensity of states' Jim Crow regimes. These findings implicate state-level Jim Crow institutions as a central factor shaping the geography of Black economic progress and perpetuating the disadvantages faced by families enslaved until the Civil War.

We extensively validate our empirical strategy. For the border discontinuity design, we show that 1) gaps in the economic status of formerly enslaved people only arise with the beginning of Jim Crow (circa 1880), 2) those gaps only exist for borders where states' Jim Crow regimes differ and increase with those differences, 3) before Jim Crow there are no border gaps in counties' economic, agricultural, political, or demographic characteristics, 4) with the beginning of Jim Crow, large border gaps emerge in key county-level outcomes targeted by those regimes, including votes cast per adult male

[3] Due to data-sharing agreements, we cannot disclose the name of the credit bureau.

and Black school quality, and 5) Jim Crow regimes did not harm white families' economic outcomes. Basing our design on ancestor location before 1865—rather than the current location—leaves little room for selection, given that enslaved people had no say in their place of residence. Both historical and new empirical evidence support our main identifying assumption that an enslaved person's birthplace is exogenous to future generations' potential economic outcomes. Because of high migration costs, partly due to Jim Crow's institutional barriers to mobility, a family's enslavement location is a strong indicator of their exposure to Jim Crow. However, as many families did migrate despite those barriers, we assess the role of migration in shaping place effects using a standard framework of random assignment with imperfect compliance.

We explore potential mechanisms of how Jim Crow regimes slowed Black economic progress using a newly compiled dataset on state-level Jim Crow laws. We first classify Jim Crow laws by topic and find that the largest number pertains to education. Education is the target of 283 laws—one-third of all Jim Crow laws passed throughout the South. Those laws racially segregated schools, reduced educational resources allocated to Black children, shortened term lengths for Black schools, and prevented Black Americans from participating in the local bodies that governed education. Indeed, we find that the quality of Black schools drops sharply across borders with states that have more oppressive Jim Crow regimes. In addition, our main regression discontinuity estimates are similar when using educational Jim Crow laws or Black school quality, rather than more comprehensive measures of Jim Crow intensity. Statements from leading historians confirm that educational restrictions were likely a key factor in Jim Crow's negative impact on Black economic progress.

This paper makes several contributions. First, leveraging new methods to link families' data across generations (Abramitzky et al., 2020), we generate new evidence on the mechanisms behind institutions' persistent effects (Acemoglu et al., 2002; Dell, 2010; Donaldson, 2018; Dell and Olken, 2019). Second, we design methods to identify descendants of enslaved people, uncovering important economic differences among Black Americans based on ancestral enslavement status. Third, by analyzing exposure to Jim Crow, we find that systemic discrimination—the higher exposure to ongoing discrimination *because of past discrimination* (Bohren et al., 2022)—is central to the enduring legacy of racial oppression in the US. We find that Black economic progress was rapid where conditions allowed, consistent with seminal works (Du Bois, 1935; Woodward, 1955; Ransom and Sutch, 2001; Aaronson and Mazumder, 2011; Naidu, 2012; Wright, 2013). Last, despite the recognized impact of location on upward mobility, its underlying causal mechanisms remain unclear (Olivetti and Paserman, 2015; Chetty et al., 2014; Chetty and Hendren, 2018). Our results show that institutions can play a key role in shaping upward mobility.

2 Historical Context

This section provides historical context for the evolution of anti-Black institutions in the US—from slavery to Jim Crow and beyond.

2.1 Free Black Americans before 1865

In 1860, just before the Civil War (1861–1865) that led to the abolition of slavery, 4 million enslaved and 0.4 million free Black people lived in America. Enslaved people had existed on American soil since the country's colonial origins (Sowell, 1978). The roots of the free Black population may trace back to 1619 when settlers in Virginia purchased the first 20 Black people. Little is known about their fate, but it is likely that some of them were treated as servants who had to work for a fixed term and gained freedom afterward (Frazier, 1949). Around 1660, both law and practice had changed, implying that virtually all Black individuals who arrived in the colonies were enslaved for life (Galenson, 1981). From 1662 onwards, the law also mandated that a child would inherit their legal (i.e., free or enslaved) status from their mother regardless of race.

For some enslaved people, the Revolutionary War (1775–1783) provided a road to freedom. Responding to a need for troops and laborers, both the British and American leadership promised freedom to enslaved people willing and able to serve. It is estimated that up to 100,000 enslaved people ran away from plantations to do so (Schama, 2006). After the war, many remained in the US as free persons. As a result, the free Black population in some states increased dramatically.

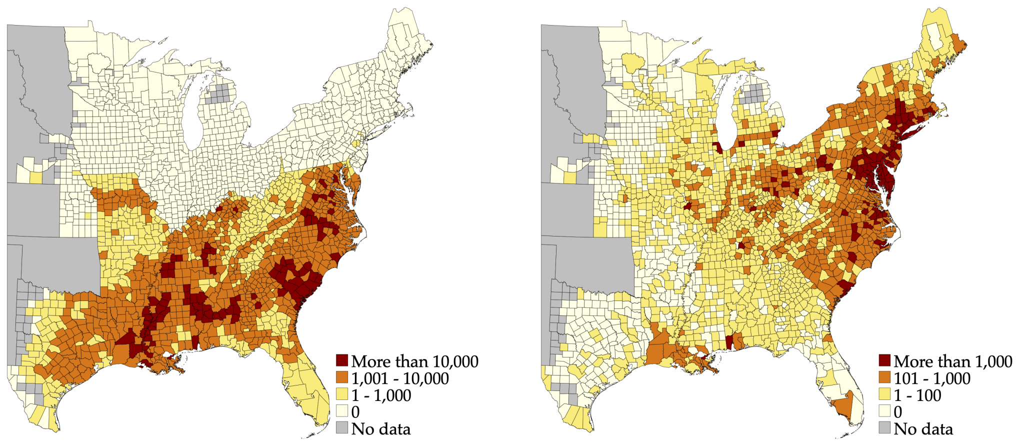
The Revolutionary War also spread a spirit of egalitarianism, challenging the institution of slavery in some regions. In the North, the abolitionist movement grew quickly after the war. While only a few Black people lived free of slavery before the Revolutionary War, most Northern states adopted gradual emancipation laws after the war. New Jersey was the last Northern state to do so in 1804.

In the South, the path to freedom was narrow, especially in the Lower South.[4] All Southern states except North Carolina allowed masters to free ("manumit") their enslaved people by 1790, but the practice was employed to different degrees across regions. In the Upper South, the first wave of manumissions occurred between 1783 and 1793, the first decade after the Revolutionary War. Motivated by anti-slavery beliefs, most manumitters freed all their enslaved people at once. However, manumission gradually became more selective and turned into a reward system designed to uphold slavery (Wolf, 2006). By 1860, 0.2 million of the 1.8 million Black Americans in the Upper South were free (11.1 percent). The Lower South did not see a similar manumission wave after the war, as manumissions there were usually limited to masters' "illicit offspring, special favorites, or least productive slaves" (Berlin, 1974). The free Black population of the Lower South mainly originated from refugees who fled from Saint-Domingue (now Haiti) and the purchase of Louisiana from France, which had a sizable free Black population. By 1860, 40,000 of the 2.5 million Black Americans in the Lower South were free (1.6 percent).

[4]The Lower South comprises Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, South Carolina, and Texas. The Upper South comprises Delaware, Washington, DC, Kentucky, Maryland, Missouri, North Carolina, Tennessee, Virginia, and West Virginia. The North comprises all other states.

FIGURE II.1: Population by County in 1860

(A) Number of Enslaved (B) Number of Free



Notes: This figure shows the population sizes of enslaved Black Americans (panel A) and free Black Americans (panel B) in the 1860 census. The maps are truncated to omit the western half of the country, which at the time was only sparsely populated. Appendix Figure IV.38 shows the maps for 1790.

The legal and economic status of free Black Americans varied greatly across locations and over time before 1865 (Sowell, 1978). In most states, free Black Americans were deprived of the right to vote and to hold political office. However, their legally protected property rights were respected in most cases. With the limited freedom they enjoyed, some free Black families could accumulate modest wealth and social status. Most of them, however, lived in poverty "under conditions barely distinguishable from those of the mass of slaves" (Berlin, 1974). Their economic status varied considerably across the country and, perhaps surprisingly, tended to be better further South (Berlin, 1976). In the North, free Black families were concentrated in cities where they suffered from competition with and hostility from white laborers (Frazier, 1949). Most free Black families in the South lived in rural areas, working as farmhands and casual laborers (Berlin, 1974).

By the beginning of the Civil War (1861–1865), the enslaved population was concentrated in the Lower South (see Figure II.1). The free Black population, in contrast, was concentrated in the North and the Upper South. These differences in geographic location exposed them to different institutional regimes after slavery.

2.2 Freedom of All Black Americans after 1865

The Civil War led to the emancipation of enslaved families, giving all Black Americans the same legal status. The average free Black family had likely already been free for around 50 years. For the first 12 years after the Civil War—the Reconstruction era (1865–1877)—the Union Army occupied the South. Black Americans experienced unprecedented economic progress under Reconstruction (Foner, 2014; Frieden et al., 2023). New schools and colleges were built to educate Black Americans throughout the South. Black men participated politically, casting their votes in high numbers and serving in public office (Logan, 2020). Throughout Reconstruction, Black economic and political progress was met with violent opposition from white Southerners (Du Bois, 1935; Foner, 1963; Blackmon, 2009).

In 1877, the Union Army left the South, abandoning the project of Reconstruction. The disenfranchisement of Black people through legal and extra-legal means led to massive declines in Black political participation (Kousser, 1974; Wright, 1986; Perman, 2001; Naidu, 2012). Many free Black Americans lost their higher social status and some left the South (Woodson, 1918).

Black Americans who remained in the South after Reconstruction faced increasing oppression through the rise of Jim Crow (1877–1964). Jim Crow regimes governed almost every aspect of Black life. Schools, workplaces, public transport, medical facilities, and parks were racially segregated (Murray, 1950). Poll taxes, literacy tests, and other rules limited Black suffrage (Naidu, 2012; Walton et al., 2012). Enticement laws, contract enforcement laws, and emigrant-agent laws prevented Black workers from seeking economic opportunities with new employers or in states outside the South (Roback, 1984; Naidu, 2010). Vagrancy laws criminalized the unemployment of Black people (Blackmon, 2009). In addition to legal factors, various extra-legal means of excluding Black Americans spread through the South and beyond.

From 1910 to 1940, many Black Americans started to leave the (Upper) South in the first wave of the Great Migration. Black families from the Lower South participated less in this first wave, both because Jim Crow limited their geographic mobility and because migration was more costly for them (Roback, 1984; Naidu, 2010; Carrington et al., 1996).

After almost 100 years, the Civil Rights Movement successfully fought oppression starting in the mid-1950s and eventually ended Jim Crow—"one of the most significant legislative achievements in American history" (U.S. Senate, 2019). The Great Migration continued until the end of the movement in the late 1960s. By then, six million Black Americans had left the South (Boustan, 2016). However, many Black families still faced challenges in capitalizing on available opportunities in the North (Collins, 1997; Akbar et al., 2020; Derenoncourt, 2022). In addition, even after the achievements of the 1960s, old forms of racial oppression persisted, and new forms—such as mass incarceration and "color-blind" voter suppression—have arisen since (Western, 2006; Alexander, 2010; Bonilla-Silva, 2015; Darity et al., 2016). The narrowing of racial disparities has slowed substantially since the 1960s (Bayer and Charles, 2018; Althoff, 2021; Derenoncourt et al., 2022).

3 Data and New Methods to Measure a Family's Exposure to Slavery and Jim Crow

A major empirical challenge we overcome in this paper is to measure a Black family's exposure to slavery and Jim Crow. We construct family histories for Black Americans in the historical censuses and develop new methods to measure two critical components of a family's historical exposure to institutionalized oppression: *how long* a family was enslaved and *where* they were freed, determining the intensity of the Jim Crow regime under which they likely lived.

3.1 Measuring How Long a Family Was Enslaved

To measure how long a family was enslaved, we leverage that the pre-Civil War censuses of 1850 and 1860 did not record enslaved people.

Main method based on census linking. We identify Black Americans free before 1865 ("the Free") as those who were 1) recorded in the 1850 or 1860 census or 2) born in a state that had already abolished slavery; Black Americans who were born in slave states before 1865 and cannot be traced back to ancestors in the 1850 or 1860 census are classified as enslaved until 1865 ("the Enslaved").[5] We then carry this information forward to their descendants. To do so, we build family trees using the census's information on family interrelationships for members of the same household and by linking individuals' records across time.

This classification strategy accurately identifies whether a Black family's ancestor was enslaved until 1865. In principle, if a family cannot be linked back to the 1850 or 1860 census, this could either mean that they were enslaved until 1865 or that they could not be linked using automated methods—for example, because their name was misspelled in one census. Hence, in the South, we inevitably misclassify some Black families who were free before 1865. However, census records show that only 6 percent of the Southern Black population were free in 1860. Therefore, our comparison involves a group almost certainly free in 1860 against a group where at least 94 percent were enslaved until the Civil War, minimizing the potential for attenuation bias due to imperfect linking rates (see also Appendix 2.1.1). Record linkage helped us identify around 20 percent of free Black Americans in the 1870 census, 10 percent of whom we trace to descendants in 1940.

Our classification method has two critical advantages over previous research, which typically relied on birthplaces to identify how long a family was likely enslaved. First, because the census only provides information on birthplaces for a person and their parents, the effects of slavery cannot be studied beyond the second generation in the census cross-section. Our panel allows us to follow individual Black families' records until 2000. Second and most importantly, relying on a person's birthplace can only identify free Black families born in the North. However, 50 percent of all Black families free before 1865 lived in the South. Our method identifies a large number of those families. Measuring how long a family was enslaved and where it was freed is crucial to determining what role slavery, Jim Crow, and their interaction play in shaping the persistent effects of institutionalized racial oppression.[6]

The Free-Enslaved gap quantifies disparities based on a family's male ancestry. Due to women's surname changes upon marriage, accurately linking female ancestry is challenging. Focusing on the male lineage minimizes bias that could arise from selective marriage patterns, allowing us to accurately estimate the Free-Enslaved gap as we define it. However, this approach limits our ability to estimate another important measure: the variation in economic status based on the proportion of Free vs. Enslaved ancestors

[5] We refer to Black families free before 1865 as "the Free" even though they or their ancestors may have been enslaved in previous decades. We refer to those enslaved until 1865 as "the (formerly) Enslaved." We choose this terminology to avoid confusion engendered by the sometimes-used terms "Freemen" (Free) and "Freedmen" (formerly Enslaved). We avoid the term "slave" and capitalize "Free" and "Enslaved" when used as nouns to be respectful of the people we study.

[6] See Appendix Figure IV.30 for average socioeconomic outcomes among descendants of the Enslaved and the Free by region of origin.

across both maternal and paternal lines. Given the vast geographic and socioeconomic divides between Free and Enslaved families, intermarriage between these groups was likely limited by 1940. This is corroborated by quantitative evidence and historical narratives (see Appendix 2.1.2). However, we show that in the presence of intermarriage, even if limited, the Free-Enslaved gap serves as a lower bound for the disparities between families with exclusively versus no enslaved ancestors.[7]

Alternative method based on surnames. We develop a second strategy to identify descendants of the Free and Enslaved based solely on surnames, without requiring census linkage. We use the change in the distribution over surnames from before 1865 (pooling the 1850 and 1860 censuses), when the census included only free Black Americans, to after 1865 (pooling the 1870 and 1880 censuses), when it included all Black Americans.[8]

While some surnames were common among the Free and the Enslaved, others were characteristic of one group (see Appendix Table IV.18). For example, the surname "Du Bois" was relatively frequent among free Black families in the 1860 census. However, with the inclusion of the families newly freed in 1865 in the 1870 census, Du Bois became ten times less frequent—an indication that having this surname meant a person likely descended from the Free. In contrast, the surname "Freedman" did not exist in the 1860 census but appeared in the 1870 census after many newly freed families chose it as their new surname. Thus, Black families called Freedman were likely enslaved until 1865.

This surname-based approach allows us to measure the likelihood that one's ancestors were enslaved until the Civil War in any dataset that includes surnames, such as the full (not only the linked) sample of Black Americans in the historical censuses as well as real-time credit bureau data. The linking-based and the surname-based approaches yield highly correlated Free-Enslaved classifications (see Appendix Figure IV.31).

3.2 Measuring the Exposure to State-Led Oppression During Jim Crow

Black families' exposure to slavery and Jim Crow is highly correlated. Families enslaved until 1865 were also geographically concentrated in states that would become the epicenter of Jim Crow. In contrast, families freed earlier were concentrated in states that would adopt less intensive Jim Crow regimes. These different geographic distributions result from the rapid southern expansion of the US plantation economy. The longer a family was enslaved, the more likely they were to be freed in the Lower South.

To measure a family's likely exposure to Jim Crow, we use that record linkage allows us to observe the birthplace of their formerly enslaved ancestors. A family's enslavement location is generally a strong indicator of their exposure to Jim Crow over the subsequent 75 years. Black Americans whose ancestors were enslaved in the Lower South were likely exposed to the strict Jim Crow regimes in the region for decades. Appendix Figure IV.32 shows that prior to 1930, the share of Black families originating from the Lower South who migrated out of the region was less than 10 percent—significantly lower than the mobility rates experienced by Black families from the Upper South. Among families

[7] In Appendix 2.1.2, we derive this result theoretically. We estimate that for the first generation born after 1865, the gaps between Black Americans whose ancestors only descend from Enslaved vs. free Black ancestors could be 15 percent larger than the Free-Enslaved gap.

[8] Census pooling reduces the impact of imperfect coverage in any given decade.

enslaved until the Civil War, the propensity to migrate North was especially low compared to Black families free earlier. However, it is worth noting that many families migrated despite Jim Crow's institutional barriers to mobility (Roback, 1984; Wright, 1997; Naidu, 2010) and high migration costs (Carrington et al., 1996). We formally account for migration in our econometric analysis.

Our primary measure of the intensity of states' anti-Black institutions, including their Jim Crow regime, is a composite index of persistent state-level racial oppression—the Historical Racial Regime (HRR) index (Baker, 2022). This index is derived from four key components: a state's population share enslaved in 1860; its share of sharecroppers who were Black in 1930; its number of Jim Crow disfranchisement devices; and its share of congressional delegates that signed the Southern Manifesto.

To complement our analysis and validate our main findings, we consider alternative Jim Crow intensity measures. First, we create a new composite index that, in contrast to the HRR index, focuses on institutional factors and the Jim Crow era specifically. We derive this new "Jim Crow index" from five factors frequently referred to in the historical literature as reflections of Jim Crow regimes: 1) the anti-Black discriminatory share of a state's laws specific to race; 2) a state's number of disfranchisement devices; 3) the share of congressional delegates who signed the Southern Manifesto; 4) the Black-white disparity in schools' term lengths; and 5) the year minimum pay for teachers was introduced—legislation central to narrowing the large wage penalty historically suffered by Black teachers (Card et al., 2022; Cascio and Lewis, 2022). This Jim Crow index is highly correlated with the HRR index ($\rho = 0.99$).

Additionally, we consider a state's total number of Jim Crow laws. We analyzed over 800 laws from multiple sources, including newly digitized data from "States' Laws on Race and Color," which aimed to document all race-related state laws in 1950 (Murray, 1950). We categorized each law as discriminatory (Jim Crow) or not based on its content and context provided by the authors. We also incorporated additional laws on employment and suffrage not covered in the primary source (Roback, 1984; Cohen, 1991; Walton et al., 2012). The number of Jim Crow laws correlates with the HRR index ($\rho = 0.74$).

Another measure we consider is a new composite index of Black school quality, derived from three factors: teacher salaries, student-to-teacher ratios, and term lengths for Black children in 1940—sourced from (Card and Krueger, 1992). Black school quality negatively correlates with the HRR index ($\rho = -0.94$).

We acknowledge the challenge in quantifying the severity of Jim Crow regimes, which employed both legal methods (e.g., literacy tests) and extra-legal methods (e.g., voter intimidation) to marginalize Black Americans. As Woodward noted, "[t]here [was] more Jim Crowism practiced in the South than there [were] Jim Crow laws on the books" (p. 102 Woodward, 1955). While no single measure can fully capture this complexity, all of our different proxies are highly correlated (see Appendix Figure IV.33). We argue that a collective analysis of our proposed measures offers valuable insights into the nature and extent of Jim Crow institutions in different states.

3.3 Linked Data

We use full-count census data for all available decades between 1850 and 1940 (Ruggles et al., 2020) and link observations across adjacent and non-adjacent decades using the automated linking methodology provided by Abramitzky et al. (2020). A person is linked from one census to another if their name, year of birth, and state of birth match and if the match is *unique* conditional on race. We use a method that allows for misspellings by matching names based on their phonetic sound (NYSIIS). Allowing for misspellings tends to be a more conservative approach because it treats phonetically similar names as equivalent, yet maintains the requirement for uniqueness in establishing a match. Because women tend to change their surname upon marriage, only men can be linked over time (Althoff et al., 2024).

The census also contains information on the relationship between individuals in the same household. By observing a person in their parents' household during child- or adulthood, we can build family trees based on this information. We transfer parental data, such as Free-Enslaved status and county of residence, to subsequent census records of the individual and their descendants. These family trees allow us to study the evolution of a family's social, economic, and geographic mobility across generations. We study individuals' outcomes in census records between 1870 and 1940 (from the first census to include all Black Americans to the most recent full-count census available). Our primary outcomes include education, income, and wealth (Appendix 2.2.1 describes all outcome variables in detail). Over time, the census data provide increasingly rich information on those outcomes. Therefore, we focus particular attention on the 1940 census.

To extend our analysis to the 21st century, we link the 1940 census to administrative mortality records from the Social Security Administration (Goldstein et al., 2021).[9] Effectively, this sample contains individuals born before 1940 and deceased between 1988 and 2007. The mortality records contain a person's last neighborhood of residence (nine-digit ZIP code) at the time of death. We use National Historical Geographic Information System (NHGIS) data on each neighborhood's distribution of education, income, and wealth by race to proxy for a person's economic status (see Appendix 2.2.2 for details).

To extend our results to the present day, we combine our surname-based Free-Enslaved classification with real-time data from one of the primary US credit bureaus. The credit bureau merged our probabilistic classification with their universe of credit reports before removing personally identifying information. The main outcomes include predicted total income, predicted disposable income, and credit score. Because those predictions are based on data and models proprietary to the credit bureau, our ability to validate the accuracy of these predictions is limited. However, recent work using similar credit bureau data validate the accuracy of these predictions using payroll records (Mello, 2023). We subset the data to focus on Black prime-age men. The credit bureau does not observe a person's race directly and instead predicts it based on the person's first and last name as well as their neighborhood (nine-digit ZIP code).[10] We access a snapshot of this anonymous data from March 2023 through a secure server (see Appendix 2.2.3 for

[9] The linkage from 1940 to 2000 leverages automated methods based on a person's name, year of birth, and state of birth (Abramitzky et al., 2020), analogous to the linkage between 1850 and 1940.

[10] Using a separate dataset—our Social Security mortality records—we find that surnames and nine-digit ZIP codes combined capture 90 percent of the variation in whether a person is Black or not.

further details).

3.4 Sample

For our analysis, we focus on Black men aged 20 to 54 and limit our linked sample to individuals who can be traced back to their ancestors in 1880 or earlier. The latter restriction serves two purposes. First, our method for identifying families who gained freedom before 1865 requires linking them to their ancestors in 1850 or 1860. This requirement may introduce bias in the Free-Enslaved gap resulting from comparing families who can be linked back in time with those who cannot. By restricting the sample to Black Americans linkable to 1880 or earlier, we minimize this potential bias. Second, this restriction excludes families who immigrated to the US after 1880, as they may have experienced significantly different institutional contexts prior to their arrival, which could confound our analysis. Our results are not sensitive to this restriction.

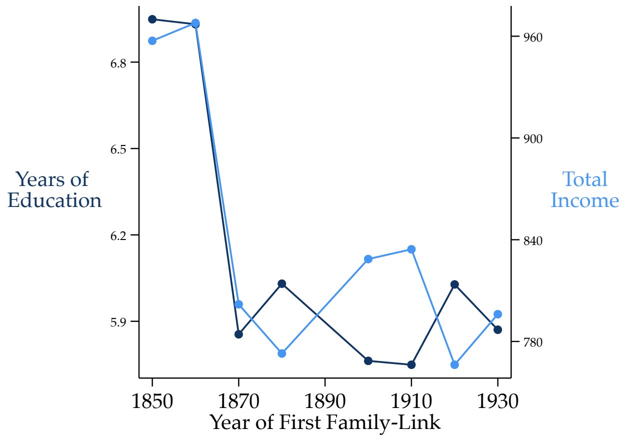
For 1940, our sample of Black prime-age men consists of 155,813 descendants of families enslaved until 1865 and 9,325 descendants of families freed before 1865. Linking a large number of descendants in 1940 to their Civil War-era ancestors is feasible for several reasons. First, to track an individual over time, we use links between both adjacent and non-adjacent census years. Second, we establish links between fathers and sons through their cohabitation. Third, the likelihood of establishing at least one link to a male descendant increases if an ancestor has multiple male descendants. On average, we make 3.7 links across different census decades to establish a 1870–1940 family tree. We link 10 percent of families in 1870 to at least one ancestor in 1940 (see Appendix Table IV.19). This statistic is essential because those links allow us to observe the state in which ancestors were freed from slavery via their birthplace in the 1870 census. Our data show a marginally higher match rate for free Black men compared to formerly enslaved men (18.5 vs. 17.1 percent, respectively, from 1870 to 1880).[11] From the 1940 census to administrative records in 2000, we link 21,059 descendants of enslaved and 1,591 descendants of free Black families.

Our sample is highly balanced on observable characteristics (see Appendix Table IV.12). For example, the literacy rate (20.4 percent) of those who we classify as formerly enslaved in our linked sample of 1870 matches the literacy rate of the 1870 Black census population—the vast majority of whom were enslaved until the Civil War. For free Black families in our linked sample of 1860, the literacy rate (65.1 percent) is also close to that of the 1860 Black census population (66.8 percent)—all of whom were free by definition of who was included in the census prior to 1865. The sample of individuals in 1940 linked to ancestors between 1850 and 1880 is also highly balanced compared to all Black men with US-born parents in 1940.

Potential Linking Bias. One may be concerned that linking procedures introduce mechanical differences between families enslaved until 1865 and those freed earlier. The most plausible concern is that a person's economic status depends on how many generations or decades they can be linked backward.

[11] To evaluate linking rates by Free-Enslaved status, we contrast Black Americans born in the North (Free) with those from the South (mostly Enslaved), rather than basing the Free-Enslaved status on linkability in earlier decades. The relatively lower linking rates for Southern-born Black Americans may stem partly from their larger population sizes, which decrease the likelihood of having unique names within their birth states.

FIGURE II.2: Average Outcomes in 1940



Notes: This figure shows the average outcomes of Black Americans in 1940 by the earliest year to which we can link them back to one of their ancestors. The dark blue line (left y-axis) shows the years of education; the light blue line (right y-axis) shows the total predicted income. The lines suggest no trend in outcomes outside of the break from 1860 to 1870. See Data Appendix 2.2 for details on the sample and data.

To examine the quantitative importance of this concern, we group Black Americans in 1940 by the earliest decade in which we can link them back to one of their ancestors and plot their average outcomes by group (see Figure II.2). In 1870, Black families enslaved until 1865 were included in the census for the first time. Consistent with that change in sample composition, we observe a significant drop in average income and education for people who can be linked to ancestors in 1870 but not 1860 or 1850. Aside from this drop, there are no trends in income or education, suggesting that individuals who can be linked further do not have a mechanically higher economic status. To err on the side of caution, we limit our sample to individuals who can be linked back to 1880 or earlier throughout this paper.

4 A Simple Model of Black Economic Progress After Slavery

We propose a simple econometric model of Black economic progress to guide our interpretation of the forces that shape the Free-Enslaved gap's long-run persistence. Our framework incorporates intergenerational mobility, the effects of exposure to location-specific factors, (selective) migration, and the effect of delayed freedom. We use this model to answer the following questions: What factors determine the gap's long-run persistence? How important was the differential exposure to location-specific factors among the Enslaved and the Free in shaping the gap? Is the persisting disadvantage faced by descendants of the Enslaved a causal effect of slavery or Jim Crow?

4.1 Model setup

Let $y_{i,t}$ denote the human capital—or any other outcome of interest—for person i at time t. For simplicity, let there be two time periods, $t \in \{0, 1\}$; the model is easily extendable to more time periods. We think of $t = 0$ as reflecting 1865, the year of Emancipation, and $t = 1$ as reflecting 1940, the last census year to which we can link families. We model $y_{i,t}$ to be determined by

$$y_{i,t} = \alpha_{i,t} + \gamma^{t}_{\ell(i,t)} + \rho y_{i,t-1} + \varepsilon_{i,t} \tag{II.1}$$

such that it depends on four factors: a factor capturing innate "ability" $\alpha_{i,t}$ with c.d.f. $F(\cdot)$, the family's previous human capital $y_{i,t-1}$, their location $\ell(i,t) \in \mathcal{L}$, and a random error term $\varepsilon_{i,t}$ that satisfies $\mathbb{E}[\varepsilon_{i,t} \mid s_i, \alpha_{i,t}, \ell(i,t)] = 0$. Last, we define γ^t_ℓ as the effect of being exposed to location ℓ at time t. We model $y_{i,0}$ (the starting condition) as

$$y_{i,0} = \alpha_{i,0} + \gamma^{0}_{\ell(i,0)} - \delta s_i + \varepsilon_{i,0}, \tag{II.2}$$

where s_i is an indicator for whether the family was enslaved until 1865. That is, in 1865, the outcomes depend on "ability," location, and whether a person had been free before the Civil War. The parameter $\delta \geq 0$ captures any direct advantage that free Black Americans had relative to the Enslaved, such as access to education during slavery.[12]

4.2 The Intergenerational Effect of Being Enslaved Until the Civil War

We define the effect of descending from ancestors who were enslaved until the Civil War ($s_i = 1$) as the expected difference between the two groups in the absence of differences in "ability" ($\alpha_{i,0}$). That is, we define the average treatment effect as

$$ATE \equiv \int \left(\mathbb{E}[y_{i,1} \mid s_i = 1, \alpha_{i,0}] - \mathbb{E}[y_{i,1} \mid s_i = 0, \alpha_{i,0}]\right) dF(\alpha_{i,0}). \tag{II.4}$$

Throughout the paper, this definition will guide the interpretation of our estimates.

In conceptual contrast to prior work (e.g., Sacerdote, 2005), we argue that one should not think of slavery's average treatment effect merely as an effect *conditional on location*. Descending from an enslaved person made a person much more likely to come from (and still live in) environments that were relatively harmful to their economic progress. Their enslavement status directly caused the location of enslavement, and the treatment effect should include its impact. From an econometric perspective, geographic location can be interpreted as a *bad control* since it is a mediating variable through which slave status affects future descendants (Angrist and Pischke, 2008).

[12] At time $t = 1$, the outcomes then become

$$y_{i,1} = (\lambda + \rho)\,\alpha_{i,0} + \rho\gamma^{0}_{\ell(i,0)} + \gamma^{1}_{\ell(i,1)} - s_i\rho\delta + \rho\varepsilon_{i,0} + \varepsilon_{i,1}, \tag{II.3}$$

where $\alpha_{i,1} = \lambda\alpha_{i,0}$ allows for transmission of "ability" over multiple generations. Thus, outcomes are determined by the "ability" of the initial generation through direct transmission of "ability" (λ) and through intergenerational advantage derived from "ability" in previous generations (ρ). The current location ($\gamma^{1}_{\ell(i,1)}$) shifts the level of a person's human capital. Through intergenerational transmission, human capital is also affected by 1) how previous generations were impacted by where they lived ($\gamma^{0}_{\ell(i,0)}$), 2) whether their ancestors were enslaved until 1865 (δ), and 3) their ancestors' idiosyncratic human capital shocks ($\varepsilon_{i,0}$).

5 Economic Gaps between Descendants of Free and Enslaved Families

This section documents the gaps in education, income, and wealth from 1870 to 2023 between descendants of families enslaved until the Civil War and those freed earlier. We find that these gaps are large and persist until today.

5.1 Evolution of the Free-Enslaved Gap until 1940

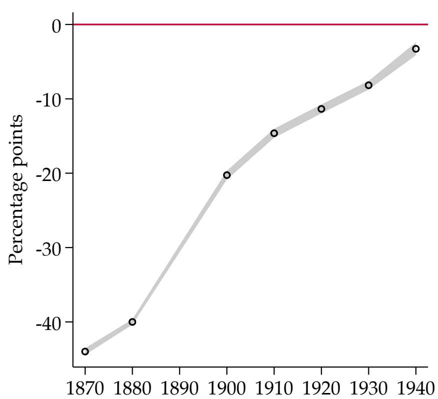
We estimate the Free-Enslaved gap (β_t) in economic outcomes ($y_{i,t}$) separately for each decade t in our linked sample from 1870 to 1940:

$$y_{i,t} \quad = \quad \alpha_t + \beta_t s_i + \phi_t' X_{i,t} + \varepsilon_{i,t}, \tag{II.5}$$

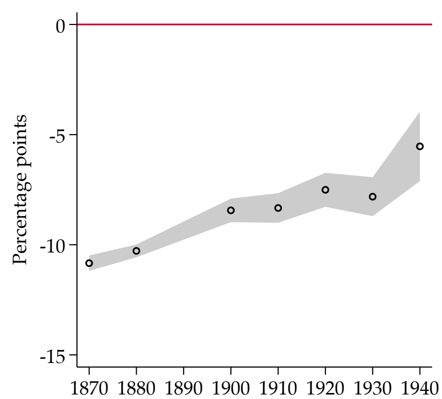
where s_i is equal to one if person i is classified as a descendant of the Enslaved and is zero otherwise. $X_{i,t}$ is a vector of controls that includes a quadratic term of age in our baseline specification. We cluster standard errors at the family level.[13]

FIGURE II.3: Free-Enslaved Gap (1870–1940)

(A) Literacy



(B) Occupational skill



Notes: This figure shows the gaps in literacy and occupation skill among prime-age (20-54) male descendants of enslaved vs. free Black Americans in each census decade. The sample includes both the South and North of the US. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. We assign "skilled" to occupations classified as "medium skilled workers" or above by the HISCLASS scheme (Leeuwen and Maas, 2011); and "unskilled" to others. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

We find that the economic differences between descendants of the Free and Enslaved are large and persistent. In 1870, the formerly Enslaved were 2 times (over 40 percentage points) more likely to be illiterate than free Black Americans (see Figure II.3). By 1940, the gap was still 1.8 times (5 percentage points). Descendants of the Enslaved worked in less skill-intensive occupations than descendants of the Free from 1870 to 1940. Consistent

[13]We define a family as a group of individuals with a common 1870 ancestor. In 1940, our linked sample comprises 49,876 families with an average of 1.6 prime-age male descendants each.

with this skill gap, descendants of the Enslaved earn lower incomes and are significantly less likely to own their homes (see Appendix Figure IV.42). Overall, we estimate the Free-Enslaved gap to be *smaller* than the gap between Black Americans born in the North vs. South before 1865—a comparison that Sacerdote (2005) uses as a proxy for the Free-Enslaved gap (see Appendix Figure IV.43). Our estimates capture the important fact that free Black Americans fared far worse in the South than in the North after slavery.

The rich information on education, income, and wealth provided by the 1940 census allows us to get a detailed picture of the Free-Enslaved gap 75 years after slavery ended. Using those outcomes, we find that descendants of the Enslaved are less educated, earn lower incomes, and have accumulated less wealth than descendants of free Black Americans in 1940 (see Table II.1).[14] The gap in education amounts to 1.6 years—more than one-quarter of the average years of education among Black men in 1940. The likelihood that a descendant of the Enslaved earned a high school or college degree was only half compared to descendants of the Free (see Appendix Table IV.23).

TABLE II.1: Free-Enslaved Gap (1940)

	Education (Years) Mean: 5.99	Wage Income (USD) Mean: 381.20	Homeownership (%) Mean: 29.25	House Value (USD) Mean: 1,371.95
Ancestor Enslaved until Civil War	**-1.59***** (0.05)	**-145.92***** (6.13)	**-7.24***** (0.62)	**-694.69***** (65.85)
Controls (age, age^2)	Y	Y	Y	Y
% of Black-white gap	42	29	36	37
Adjusted R^2	0.04	0.05	0.01	0.01
Observations	163,549	154,463	164,357	46,971
Ancestor Free	*9,078*	*8,551*	*9,070*	*3,227*

Notes: This table shows the gap in years of education, wage income, homeownership, and house value (conditional on ownership) among prime-age (20–54) male descendants of enslaved vs. free Black Americans in 1940. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The narrowing of the Free-Enslaved gap from 1870 to 1940 is slow relative to benchmark rates of intergenerational mobility among white Americans. To compare the convergence speed, we estimate economic gaps from 1870 to 1940 between white families whose ancestors had no measurable physical or human capital in 1870 and all other white families (see Appendix Figure IV.17). In only 30 years, the gap in literacy between those two groups of white Americans rapidly shrunk from over 90 percentage points to less than 10 (from twice the Free-Enslaved gap in 1870 to half the Free-Enslaved gap in 1900). The homeownership gap for the two groups was similar to the respective Free-Enslaved gap in 1870 but closed by 1900—while the Free-Enslaved gap changed very little until then.

Robustness. We re-estimate the Free-Enslaved gap based on the *full* population (rather than the linked sample) of Black Americans in 1940 using our surname-based approach, yielding results very similar to our preferred approach based on record linking (see Ap-

[14] Appendix Table IV.22 compares the Free-Enslaved gap across different income measures.

pendix Table IV.13). The gaps between Black families with surnames that convey high vs. low likelihoods of having been enslaved until the Civil War are -1.40^{***} (0.09) in years of education, -113.15^{***} (25.50) in wage income, -2.31^{**} (1.05) in homeownership, and $-1,098.68^{***}$ (282.83) in house values.

Next, to mitigate misclassification bias, we use our surname-based measure as an instrumental variable (IV) for the linking-based measure. The resulting IV estimates offer an unbiased assessment of the Free-Enslaved gap if the errors in the linking-based measure are uncorrelated with the errors in the surname-based measure (Ashenfelter and Krueger, 1994; Angrist and Pischke, 2008). This assumption is supported by the surname-based measure's independence from census-linking methods. These IV estimates suggest that measurement error reduces our initial estimates of the Free-Enslaved gap by an average of 9 percent across various outcomes (see Appendix 2.1.1). For example, based on our IV estimates, descendants of the Enslaved attained 1.67^{***} (0.15) years less in education in 1940 than descendants of the Free, compared to 1.59^{***} (0.05) via OLS.

We also conduct an array of placebo exercises to validate our empirical strategy (see Appendix 2.1.3). First, we use 1875 as a placebo year of Emancipation. Specifically, we classify Black families as descending from the Free or the Enslaved based on whether or not we can link them back to ancestors in 1870 (rather than 1860). This placebo exercise yields no economically significant gaps. For example, a small gap of less than 1 percent in education emerges (compared to 25 percent in our baseline). Second, we use white Americans as a placebo group. Specifically, we divide white families into two groups depending on whether or not we can link them back to ancestors in the 1860 census, similar to our Free-Enslaved classification. Again, this placebo exercise yields no economically significant gaps (at most 1.7 percent across all outcomes, most of them not statistically significant).

5.2 The Free-Enslaved Gap in the 21st Century

The Civil Rights Movement (1954–1968) ended Jim Crow, thereby instigating institutional change that held the promise to accelerate Black economic progress. Existing evidence indeed suggests that Black Americans' economic mobility temporarily surged around 1970 (Wright, 2013; Clark, 2014; Margo, 2016). How has the Free-Enslaved gap evolved since the end of Jim Crow?

We extend our analysis past 1940 using two methods. First, we merge data from a major US credit bureau with our surname-based probabilities of descending from ancestors enslaved until the Civil War. This approach lets us estimate the Free-Enslaved gap in real-time without needing record linkage. We use a snapshot of this data from March 2023, limiting the main sample to Black Americans as identified by the credit bureau through names and nine-digit ZIP codes. Second, we link 1940 census records for Black Americans to administrative mortality data, covering birth cohorts from 1910 to 1940. These records include a person's last residential nine-digit ZIP code, allowing us to infer neighborhood proxies for their income, wealth, and education circa 2000.

TABLE II.2: Free-Enslaved Gap (2023)

	Total income (USD) Mean: 92,068.48	Disposable income (USD) Mean: 52,773.74	Credit Score (from 300 to 850) Mean: 630.41	Hourly Job Mean: 0.72
Ancestor Enslaved until Civil War	**-12,487.72*****	**-11,623.44*****	**-33.15*****	**0.05*****
	(1,147.08)	(920.12)	(2.07)	(0.01)
Controls (age group-FE)	Y	Y	Y	Y
% of Black-white gap	23	26	40	69
Adjusted R^2	0.001	0.001	0.003	0.000
Observations	547,189	547,189	547,189	459,889

Notes: This table shows the Free-Enslaved gap in predicted total income, predicted disposable income, credit score, and hourly-wage employment among Americans as of March 2023. We estimate a person's likelihood to descend from free Black Americans via their surname, not requiring record linkage. We re-weight the sample to hold the distribution of surnames constant at the 1870 level. The sample's average likelihood of a person's ancestor to be free before the Civil War based on their surname is 9.6 percent—very close to the factual fraction. The sample includes both the South and North of the US. Credit scores (VantageScore® 3.0) reflect a person's credit health, ranging from 300 to 850 (scores above 700 are considered "good" and scores below 550 "very poor"). See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Using US credit bureau data from 2023, we find that descendants of the Enslaved have vastly lower predicted incomes and worse credit health than descendants of free Black Americans (see Table II.2). The Free-Enslaved gap in predicted disposable incomes is $11,620 (22 percent of Black Americans' average). The Free-Enslaved gap in credit scores is 33 points (one-fifth of the difference between "good" and "very poor" credit). Descendants of the Enslaved are also more likely to work in hourly-wage jobs, presumably leading to higher uncertainty in earnings compared to salaried jobs. These Free-Enslaved gaps amount to 23 to 69 percent of the corresponding Black-white gaps.

Using neighborhood-level data from mortality records linked to the 1940 census, we find that around 2000, descendants of the Enslaved resided in neighborhoods with lower education, income, and wealth than those of the Free descendants (see Appendix Table IV.24). Descendants of the Enslaved lived in neighborhoods where Black residents were 3.9 percentage points less likely to hold a high school degree and 2.6 percentage points less likely to hold a college degree. Black residents' expected incomes in those neighborhoods were $5,100 lower (17 percent of the median). Owner-occupied houses in those neighborhoods were worth $17,500 less (19 percent of the median).

In sum, our two strategies suggest that the present-day Free-Enslaved gaps in various economic outcomes amount to at least one-fifth of the corresponding Black-white gaps. This finding highlights the enduring impact of historical oppression on present racial disparities. Importantly, the Free-Enslaved gap only quantifies the *additional* disadvantage faced by those whose ancestors were enslaved until 1865 compared to those who gained freedom earlier. Most Black families, even those who were free before the Civil War, were enslaved in earlier periods, and all Black Americans faced discrimination due to slavery and Jim Crow, regardless of their specific family history. The sheer difference in intensity of their experiences yields economic gaps of such enormous magnitude. Next, we turn to the drivers of this persistence.

5.3 Interpreting the Free-Enslaved Gap

Using our model from Section 4, the Free-Enslaved gap measured as $\hat{\beta}_{1940}$ in equation (II.5), is a consistent estimator of

$$\begin{aligned}\mathbb{E}[y_{i,1} \mid s_i = 1, X_{i,t}] - \mathbb{E}[y_{i,1} \mid s_i = 0, X_{i,t}] = \\ (\lambda + \rho)\left(\mathbb{E}[\alpha_{i,0} \mid s_i = 1, X_{i,t}] - \mathbb{E}[\alpha_{i,0} \mid s_i = 0, X_{i,t}]\right) + \\ \mathbb{E}\left[\rho\gamma^0_{\ell(i,0)} + \gamma^1_{\ell(i,1)} \mid s_i = 1, X_{i,t}\right] - \mathbb{E}\left[\rho\gamma^0_{\ell(i,0)} + \gamma^1_{\ell(i,1)} \mid s_i = 0, X_{i,t}\right] - \rho\delta.\end{aligned}$$

Intuitively, the Free-Enslaved gap therefore reflects 1) any potential differences in "ability" between the two groups transmitted over generations, 2) different exposure to locations over time (as a result of slavery and potential selection), and 3) the inherited disadvantage of descending from an enslaved person conditional on environment and "ability." In the next section, we show that the two groups' differential exposure to locations due to slavery—not selection—accounts for virtually all of the Free-Enslaved gap.

6 The Importance of Geography in Shaping Black Economic Progress After Slavery

In this section, we use ancestors' enslavement locations as plausibly exogenous variation in where Black families lived to identify what fraction of the Free-Enslaved gap is caused by differential exposure to place-specific factors. We limit our sample to Black Americans whose ancestors were enslaved until the Civil War. We find that state-specific factors are the leading cause of the Free-Enslaved gap's persistence after 1940.

6.1 States' Effect on Black Economic Progress After Slavery

We estimate each state's causal effect on the long-run economic progress of Black families freed there in 1865 (excluding free Black Americans and their descendants). Our empirical strategy to identify the importance of exposure to location-specific factors builds on the following assumption, which we discuss in detail in Section 6.3.

Assumption 1 (Exogeneity of enslavement location)**.** The enslaved population was not selected into location. That is,

$$\alpha_{i,0} \perp\!\!\!\perp \ell(i,0) \;\text{ if }\; s_i = 1$$

where s_i is a dummy variable equal to 1 if one's ancestor was enslaved up to 1865, $\ell(i,0)$ is the birthplace of one's enslaved ancestor, and $\alpha_{i,0}$ is the innate "ability" of one's enslaved ancestor.

We limit our sample to families whose ancestors were enslaved until the Civil War and estimate the causal effect that the geographic distribution of formerly enslaved ancestors had on the Black economic progress of their descendants:

$$y_i = \eta_{\ell(i,1865)} + \phi' X_i + \epsilon_i, \tag{II.6}$$

where y_i are economic outcomes in 1940 and X_i is a vector of controls as defined in equation (II.5). In the context of the model introduced in Section 4,

$$\eta_\ell = \rho\gamma_\ell^0 + \mathbb{E}[\gamma^1_{\ell(i,1)} \mid s_i = 1, \ell(i,0) = \ell, X_i], \tag{II.7}$$

where γ_ℓ^0 and γ_ℓ^1 are the effects that location ℓ had on Black families during and after slavery respectively. Thus, η_ℓ reflects both the (inherited) effect the state of birth ℓ had on the ancestor during slavery and the expected effects of future locations of their descendants given the 1865 location. One can interpret η_ℓ as an intent-to-treat (ITT) effect of living in location l from before the Civil War to 1940, where the initial location is plausibly randomly assigned, but the post-1865 location is a result of endogenous (and potentially selective) migration decisions.

The effect of being freed in each state in 1865. We find a distinct geography of Black economic progress after slavery (see Appendix Figure IV.44). Gaining freedom in a state further south negatively affected Black families' economic outcomes in the long run. For example, a family freed in Louisiana would attain over two years more education had they instead been freed in Kentucky.[15] States affect other outcomes, such as literacy and income, with similarly large magnitudes. States' effects are substantial even in 2000 when, for example, families freed in Louisiana live in neighborhoods with average incomes lower by over one-quarter of the average income among Black Americans compared to those rooted in the Upper South.

Accounting for migration: the effect of living in each state between 1865 and 1940. Our estimates of the effect of being freed in each state in 1865 may partly reflect differences in migration opportunities. We formally assess the importance of post-slavery migration and recover the effect of *living* in each location ℓ between 1865 and 1940 on Black economic progress absent migration (γ_ℓ^1). We do so based on Assumption 1 and the additional assumption that place-specific experiences during slavery ceased to affect descendants in 1940 directly ($\rho\gamma_\ell^0 = 0$); we formalize this decomposition in Appendix 2.1.4. This problem is a standard case of multiple instruments (location assignment) and imperfect compliance (migration). Specifically, the intent-to-treat effect of initial location ℓ, η_ℓ, is the average of all potential future locations' treatment effects, $\gamma^1_{\ell'}$, weighted by the probability of migrating from ℓ to ℓ':

$$\eta_\ell = \sum_{\ell' \in \mathcal{L}} p_{\ell,\ell'} \cdot \gamma^1_{\ell'}.$$

We invert the migration probability matrix to recover the effect of living in each state until 1940, which is unaffected by selective migration under the assumption that the average innate "ability" of Black Americans in 1865 did not differ across enslavement locations.

Our results indicate that the effect of being freed in location ℓ closely approximates the treatment effect of living in ℓ from 1865 to 1940. The recovered treatment effects are almost identical to the intent-to-treat effects estimated using equation (II.6), except for

[15] Being freed in Louisiana has the strongest negative impact on education by 1940 (-0.84 years less than the average across Southern Black Americans)—followed by Georgia and South Carolina (-0.47 years). Missouri has the strongest positive impact (2.28 years), followed by Kentucky (1.66 years).

the border states of the Upper South. The effect of living in the border states is more negative than the effect of being freed there, suggesting that the relatively better conditions for Black Americans were partly due to greater migration opportunities. For those freed in the Lower South, benefits from Northern opportunities were more limited due to lower migration rates and a reduced likelihood of the North being their destination conditional on migration.

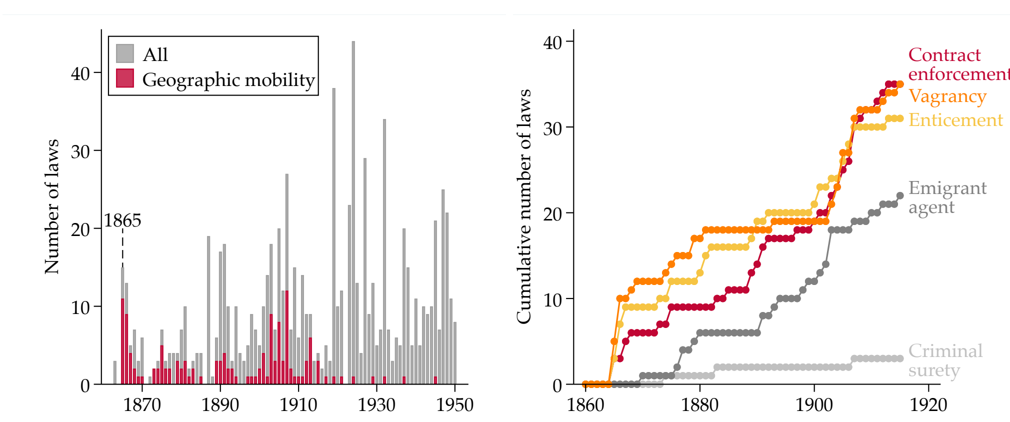
Early Black migration mostly consisted of movement within the South, often between states offering similarly limited opportunities for economic advancement. North-South migration was rare due to the isolation of the Southern labor market, particularly in the Deep South, which experienced "nearly complete isolation [...] before 1916" (Wright, 1997). Within the South, migration flowed mainly from the low-wage Southeast to the high-wage Southwest. Southwestern states such as Mississippi, Louisiana, and Arkansas attracted many Black migrants in the early post-slavery era, as they offered the potential for landownership and political participation. However, the intensification of Jim Crow around 1890 ultimately reversed the fortunes of these migrants.

With Black families freed in the Lower South faring so much worse than those freed elsewhere, it may seem puzzling why the region did not experience a larger exodus than the Upper South. For example, 75 percent of Black families enslaved in Louisiana still lived there in 1940; less than 10 percent reached the North (see Appendix Figures IV.34 and IV.35). Lower Southern white families were almost 30% more likely to migrate. Institutional and economic factors partly resolve this puzzle.

First, Jim Crow directly targeted the geographic mobility of Black people (Roback, 1984; Cohen, 1991; Naidu, 2010): enticement laws and contract enforcement laws limited Black workers' ability to terminate their employment contracts; vagrancy laws criminalized being out of employment; emigrant-agent laws prevented employers from seeking workers from other states; criminal surety laws created the possibility of involuntary servitude upon arrests for minor charges (see also Blackmon, 2009). These laws began emerging immediately after slavery (see Figure II.4).

FIGURE II.4: Number of Jim Crow Laws Across the South

(A) New laws by year (B) Laws on geographic mobility by type



Notes: This figure shows the number of new Jim Crow laws passed across all Southern states each year (panel A) and the cumulative number of laws pertaining to the geographic mobility or employment of Black Americans by type (panel B). See Data Appendix 2.2 for details on the data.

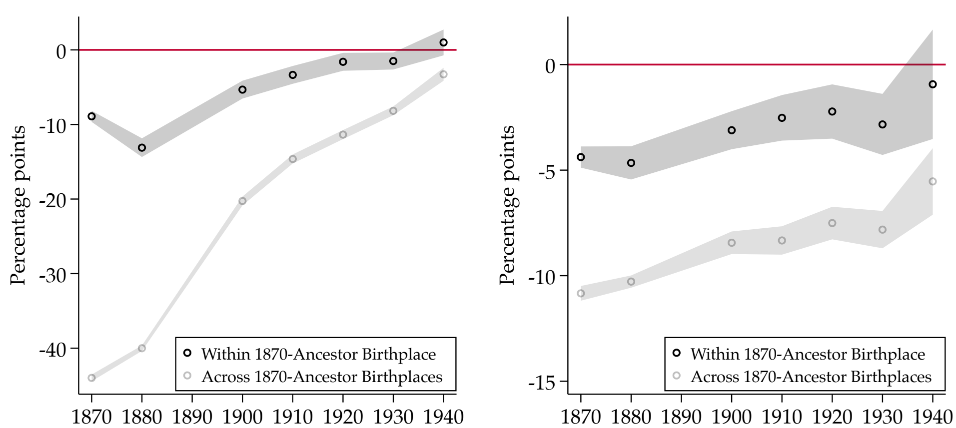
Second, moving to the North was costly, especially from the Lower South. Among families enslaved until the Civil War, the propensity to migrate North was especially low compared to Black families free earlier—some of whom may have used the resources they had accumulated by the end of the Civil War to leave the South. The region's geographic distance to the North limited the potential of social networks to lower the cost of migration (Carrington et al., 1996). Moreover, despite successful migration to the North, many Black families still faced challenges in capitalizing on available opportunities (Collins, 1997; Akbar et al., 2020; Derenoncourt, 2022).

6.2 The Free-Enslaved Gap is Driven by Geography

To explore the importance of differential exposure to state-specific factors, we first compute the Free-Enslaved gap conditional on ancestor location. To do so, we add fixed effects for the state of birth ℓ of a family's ancestor before 1865 to our baseline specification in equation (II.5). This exercise provides a back-of-the-envelope assessment of how important geography was in shaping the Free-Enslaved gap's long-run persistence. It does not account for free Black Americans' potential selection into states before 1865.

FIGURE II.5: Free-Enslaved Gap Conditional on Ancestor State (1870–1940)

(A) Literacy (B) Occupational skill



Notes: This figure shows the gaps in literacy and occupational skill before (light) and after (dark) including fixed effects for 1870 ancestor state of birth. The sample includes both the South and North of the US. The comparison is made between prime-age (20-54 years) male descendants of enslaved vs. free Black Americans in each census decade. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. We assign "skilled" to occupations classified as "medium skilled workers" or above by the HISCLASS scheme (Leeuwen and Maas, 2011); and "unskilled" to others. Both panels control for age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

We find that in contrast to the unconditional Free-Enslaved gap, the conditional gap was large in 1870 but shrunk to virtually zero after 1940 (see Figure II.5).[16] The 1940 gap in literacy, for example, fully closes after accounting for variation across ancestor states. Similarly, the conditional Free-Enslaved gap in 2000 is insignificant for all outcomes (see

[16]The 1940 gaps in almost any other outcome also shrink to zero after conditioning on the 1870 state of origin (see Appendix Figure IV.45 and Appendix Table IV.25).

Appendix Table IV.26). These results suggest that the Free-Enslaved gap persists mainly because the two groups were exposed to different state-specific factors after slavery.

We also assess the *causal* importance of state-specific factors (robust to free Black Americans' potential selection into states before 1865). Two counterfactual analyses (see Appendix 2.4) show that 1) had the Enslaved ancestors been distributed as the Free *within the South*, the Free-Enslaved gap would have been at least 67 percent smaller (lower bound),[17] and 2) had the Enslaved ancestors been distributed as the Free *within both the South and North*, the gap would have closed entirely by 1940 (see Appendix Table IV.28). Overall, our results show that group differences in initial location were the primary driver of the persistent Free-Enslaved gap.

In addition, we show that it is ancestor *states* that explain the Free-Enslaved gap, not other levels of ancestor geography (see Appendix Figure IV.46). The gap conditional on ancestor *region* is still large after 1940, suggesting that the Free-Enslaved gap is not merely a result of North-South differences. Adding ancestor *county* fixed effects does not further explain the Free-Enslaved gap, suggesting that it is not geographic granularity that makes states an important explanation.

With the ancestor state accounting for the vast majority of the Free-Enslaved gap, there is little room for other factors—such as differences in "ability" or the advantage of being free earlier—to drive the gap after 1940. State-specific factors compressed the economic status of Black Americans within states irrespective of their ancestors' enslavement status (see Appendix Figure IV.47). Their exposure to states that slowed Black economic progress after slavery placed descendants of the Enslaved at a disproportionate disadvantage.

Two exercises provide additional evidence in support of this interpretation. First, we consider free Black Americans who had no measured physical or human capital by the end of slavery. We find that even this group of free Black Americans had higher socioeconomic status than descendants of the Enslaved by 1940 (see Appendix Table IV.27). This result further supports the conclusion that the Free-Enslaved gap's persistence is unlikely to be driven by selection into freedom or the inherent advantage of being free earlier. Second, we estimate the Free-Enslaved gap controlling for skin tones. We find that the Free-Enslaved gap is almost identical with or without this control (see Appendix Figure IV.18). This result suggests that potential differences in discrimination of descendants of the Free and the Enslaved based on their skin tones is not a key driver of the gap's persistence (see also Abramitzky et al., 2023).

6.3 Location of Freedom and the Question of Exogeneity

Estimating the causal effect of place-specific factors requires that a person's location is orthogonal to their potential outcomes. Our empirical strategy relies on the immobility of the enslaved population. In particular, we build on the circumstance that the Enslaved did not have freedom of movement before 1865, leaving no room for self-selection into location. In contrast, past research typically relied on "mover designs" (e.g., Chetty et al.,

[17] We argue that the Enslaved's geographic disadvantage *within the South* provides a lower bound for the importance of group differences in location, as the Free in the North faced more favorable post-slavery conditions.

2016). In those studies, places' effects are estimated from the outcomes of families who move between them. Assumptions on the nature of their moves allow for a causal interpretation.

The lack of free movement among enslaved people lends plausibility to the key identifying assumption of an enslaved person's birthplace to be orthogonal to the potential outcomes of their (third-generation) descendants. The main threat to our identification assumption is the possibility of selective *forced* migration of enslaved people. Even though the Enslaved did not choose where they lived, owners' or traders' decisions may have induced selection into enslavement locations.

Slaveholder migration and the domestic slave trade contributed equally to the forced migration before 1865 (Fogel and Engerman, 1974; Tadman, 1979; Pritchett, 2001; Steckel and Ziebarth, 2013). Slaveholders were generally non-selective in moving all their enslaved people with them (Fogel and Engerman, 1974; Pritchett, 2001; Tadman, 2008; Pritchett, 2019). In principle, selection could also arise through differences in the slaveholders who choose to migrate. However, for selection to arise, the slaveholder's decision would need to be correlated with the potential outcomes of their enslaved people—a scenario we cannot rule out but deem unlikely. The domestic slave trade accounts for the remaining inter-regional slave mobility. Selective slave trade is only evident in the small sugar cultivation areas.[18] Sugar cultivation accounted for 6 percent of the rural enslaved population (Tadman, 1977, 1979).[19]

If anything, one can hypothesize that the selection into location based on physical traits has biased upward the estimates of states that supposedly selected positively on height and strength. In contrast, we find that such states—those in the Lower South in general and those in the sugar region of Louisiana in particular—were especially detrimental to Black economic progress.

The results from the following section strongly support our key identifying assumption. Because our estimated place effects vary sharply across state borders (and less within states), any relevant selection would need to occur sharply at the border. Such forms of selection are implausible given that enslaved people were—if anything—selectively forced to migrate to specific locations based on the crops cultivated there. We verify that crops do not discontinuously change across state borders. We also verify that the observable characteristics of enslaved people—such as their age in 1860 or their literacy in 1870—did not discontinuously vary across borders, ruling out selection on observable characteristics directly.

[18] In contrast to the sugar industry, the cotton and tobacco industries (accounting for around 87 percent of enslaved agricultural workers) were generally non-selective on age and sex (Tadman, 1977).

[19] By the nature of the work required, enslaved people there tended to be physically stronger and more likely to be male (Phillips, 1918). Traded enslaved people were found to be disproportionately likely to be young adults (e.g., Pritchett, 2019) and more likely to be male (Fogel and Engerman, 1974), but some of this evidence is nuanced by Tadman (1977, 1979). Pritchett (2001) finds that traded enslaved people were marginally taller than the average enslaved population, conditional on age and sex, but Steckel and Ziebarth (2016) contest this finding. Physical characteristics were also co-determined by environmental influences such as nutrition, illness, or stress (Steckel, 1979; Carson, 2008). There is no evidence that traders selected enslaved people on anything other than such basic physical characteristics. This is consistent with the dehumanization of Black people that characterized the slave trade, which "reduced people to the sum of their biological parts" (Smallwood, 2008, p. 43).

7 The Jim Crow Effect

Our analysis so far attributes the Free-Enslaved gap's persistence primarily to the two groups' differential exposure to place-specific factors. This section assesses whether state institutions, particularly Jim Crow regimes, underlie the importance of those place-specific factors. We find evidence that implicates state institutions as the main drivers: 1) places' effects on Black economic progress differ sharply across state borders and 2) observed non-institutional factors do not differ across state borders. Furthermore, our evidence suggests that Jim Crow regimes are key state institutions responsible: 1) the negative impact of state institutions was race-specific, largely leaving the economic status of white families unaffected, 2) the impact of state institutions can be statistically explained by various measures of states' Jim Crow intensity, and 3) the impact of state institutions emerged with the onset of the Jim Crow era.

7.1 State Institutions and Black Economic Progress After Slavery

Places may affect families' economic status for many reasons, be it cultural, climatic, economic, or institutional. We argue that only institutions change sharply at state borders, while other factors vary continuously. Therefore, to distinguish the effects of institutions from those of other factors, we decompose the location-specific parameters in equation (II.1):

$$\gamma^t_\ell = \gamma^t_{\boldsymbol{\epsilon}(\ell)} + \gamma^t_{s(\ell)}, \tag{II.8}$$

where $\gamma^t_{\boldsymbol{\epsilon}(\ell)}$ captures factors that vary continuously across state borders and $\gamma^t_{s(\ell)}$ captures factors that vary discontinuously across state borders. We can think of $\boldsymbol{\epsilon}(\ell)$ as the geographic coordinates of location ℓ, and $s(\ell)$ as the state that location ℓ is in.[20] In the next section, we propose a border discontinuity design to separate the effect of institutions, $\gamma^t_{s(\ell)}$, from the effect of non-institutional factors, $\gamma^t_{\boldsymbol{\epsilon}(\ell)}$.

7.2 Border Discontinuity Design

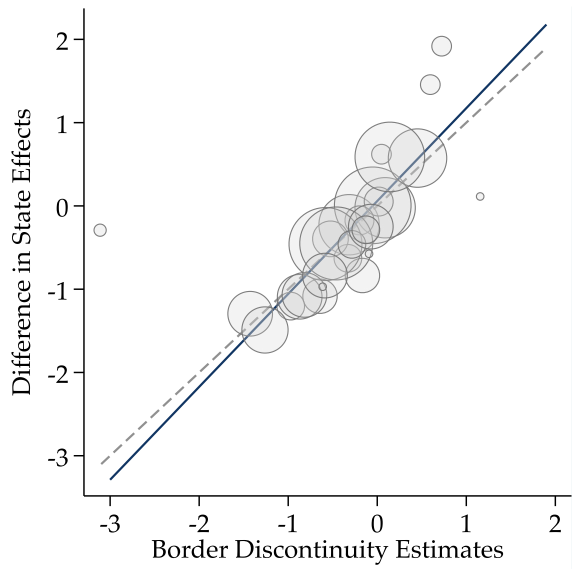
Our border discontinuity design compares the economic status of families in 1940 whose ancestors were freed on different sides of (but in close proximity to) state borders within the South in 1865. The border discontinuity design takes the following form:

$$y^{1940}_{i,b} \quad = \quad \alpha_b + \beta_b \cdot \text{High}^{1870}_{i,b} + \upsilon_b \cdot dist^{1870}_{i,b} + \psi_b \cdot dist^{1870}_{i,b} \cdot \text{High}^{1870}_{i,b} + \varepsilon_{i,b}, \tag{II.9}$$

separately for each border b in the South (see Appendix Figure IV.19), where $y^{1940}_{i,b}$ is the economic status of Black person i in 1940 whose ancestors were freed close to state-border b, $\text{High}^{1870}_{i,b}$ indicates whether i's 1870 ancestors lived on the side of border b that had a more intensive Jim Crow regime than the state on the other side of the border, and $dist^{1870}_{i,b}$ is the distance between border b and the county's centroid in which i's ancestors lived in 1870. The main coefficient of interest, β_b, captures the long-run effect of being freed on the more oppressive side of border b on a Black family's economic status.

[20] Formally, $||\boldsymbol{\epsilon}(\ell) - \boldsymbol{\epsilon}(\ell')|| \to 0 \;\Rightarrow\; |\gamma^t_{\boldsymbol{\epsilon}(\ell)} - \gamma^t_{\boldsymbol{\epsilon}(\ell')}| \to 0$, whereas $\gamma^t_{s(\ell)}$ only depends on which side of a border ℓ is on, not on the precise coordinates $\boldsymbol{\epsilon}(\ell)$: $\gamma^t_{s(\ell)} = \gamma^t_s$.

FIGURE II.6: Differences in Black Economic Progress Arise Sharply at State Borders

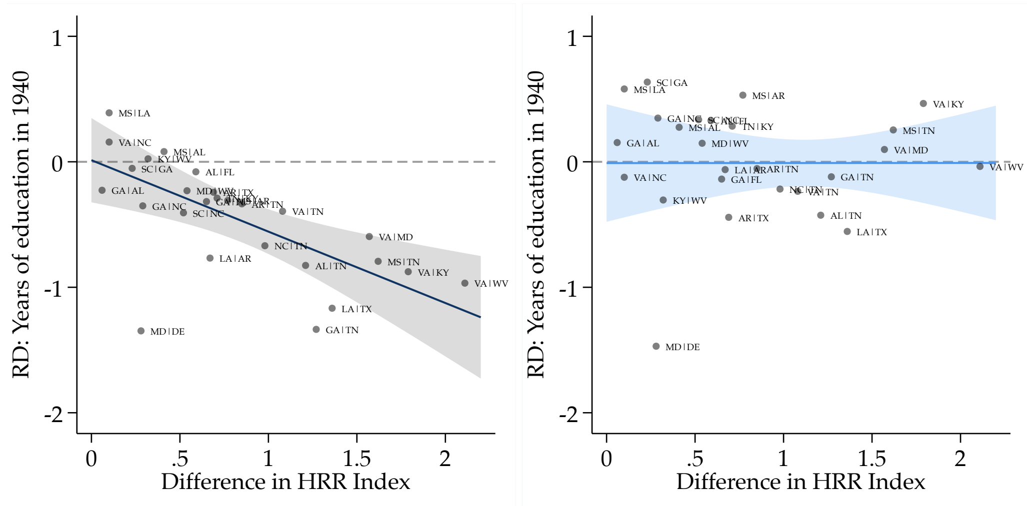


Notes: This figure relates each RD estimate (as shown in Figure II.7) to the difference in the two states' overall causal effect on 1940 years of education (as shown in panel A of Appendix Figure IV.44). Estimates are weighted by the minimum sample size underlying the difference in state effects. A gray dashed 45 degree line shows the benchmark of equal differences across two states and across the border counties of two states. The blue line shows the best weighted linear fit ($\widehat{\beta} = 1.12^{***}$, $R^2 = 0.77$). Findings are robust to excluding Louisiana and Virginia (results available upon request). See Data Appendix 2.2 for details on the sample and data.

To assess the extent to which institutions shaped the geography of Black economic progress, we compare the sharp differences in progress that emerge at state borders with the overall differences between states' effects (see Figure II.6). We find large border discontinuities, indicating that Black families freed in close proximity to each other but on opposite sides of state borders experienced vastly different economic trajectories. These border discontinuities account for a significant portion of states' overall long-run effects ($R^2 = 0.77$), suggesting that institutional factors, rather than factors that vary continuously across borders, are the primary drivers shaping the geography of Black economic progress. While institutional factors play a predominant role, there is residual variation that may be attributable to differences in economic activity, culture, or climate.

FIGURE II.7: Regression Discontinuity Estimates and Jim Crow

(A) Black Americans (B) White Americans



Notes: Panel A of this figure shows each separate RD estimate in 1940 years of education for Black families whose ancestors were freed on different sides of state borders in 1865. Panel B shows the same for white families depending on where their ancestors lived in 1870. Each label shows the more oppressive before the less oppressive state. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Negative estimates reflect lower education in more oppressive states. Lines show the best linear fit between RD estimates and the differences in Jim Crow intensity, weighted by the inverse of the estimates' standard error. Shaded areas represent robust 95 percent confidence bands. For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data.

Having established the importance of state institutions, we next examine whether it was Jim Crow institutions specifically that slowed Black economic progress. To do so, we correlate our border discontinuity estimates $\widehat{\beta}_b$ with differences in Jim Crow intensity, using that Jim Crow regimes differ more drastically across some borders than others. To quantify Jim Crow severity—which encompasses both de jure and de facto tactics (Woodward, 1955; Acemoglu and Robinson, 2008)—we employ a range of proxies that, despite their differing natures, are highly correlated. For example, the HRR index and the Jim Crow index have a correlation of $\rho = 0.99$; the HRR index and Black school quality have a correlation of $\rho = -0.94$ (see Appendix Figure IV.33). Across these measures, we consistently arrive at the same key finding.

We find that states' intensity of Jim Crow regimes predicts border discontinuities in Black economic progress. Specifically, families freed in states with more severe regimes experienced significantly lower rates of progress, starting from the Jim Crow era (see panel A of Figure II.7). These gaps widen as the difference in Jim Crow severity increases across a border. For example, consistent with Louisiana's more severe Jim Crow regime compared to Texas's, families freed in Louisiana attained 1.2 fewer years of education by 1940 than those freed just miles away in Texas. Similarly, residing in states with more severe Jim Crow regimes led to a greater likelihood of working as a farmer in 1940 but did not significantly affect wage incomes (see Appendix Figure IV.48). No differences emerge for families freed across borders where states have comparable institutions. Incorporating extensive controls for 1860 local demographics, characteristics of slaves, crop suitability, and economic activity further strengthens these findings (see

Appendix Figure IV.20).

We also find that, as expected, families who left their enslavement state before the Jim Crow era were unaffected by their origin state's Jim Crow regime (see Appendix Figure IV.49). However, if a family stayed and became exposed to the Jim Crow regime, the exposure had a persistent effect even for families who migrated in later decades. For instance, families freed in states with severe Jim Crow regimes who stayed there until 1920 were still strongly impacted by their pre-1920 experiences in 1940. The longer a family was exposed, the larger the effect on their economic status.

In principle, Jim Crow could also have affected white Americans, not only Black Americans. First, some Jim Crow laws may have directly harmed poor white Americans. For example, poll taxes aimed at disenfranchising Black voters also disenfranchised some poor white voters. Second, Jim Crow may have benefited white elites. For example, vagrancy and emigrant-agent laws depressed farm workers' wages, potentially increasing land-owning families' profits.

We find that in contrast to Black families, the economic status of white families was not negatively affected by the Jim Crow intensity of the state in which their ancestors lived in 1870 (see panel B of Figure II.7). The same is true even for poor white Americans whose ancestors had no measurable human or physical capital in 1870 (see panel A of Appendix Figure IV.50). Our findings are consistent with existing evidence of Black Americans being the main beneficiaries of ending Jim Crow through the Civil Rights legislation (Wright, 2013).

We do, however, find *positive* effects for the white land-owning elite. We find that the more oppressive a Jim Crow regime, the more economically significant the gains by the border region's wealthiest ten percent of white families (see panel B of Appendix Figure IV.50). In sum, our results suggest that Jim Crow was an extractive institution that benefited the wealthiest white families at the cost of Black families while shielding poor white families from most economic harm.

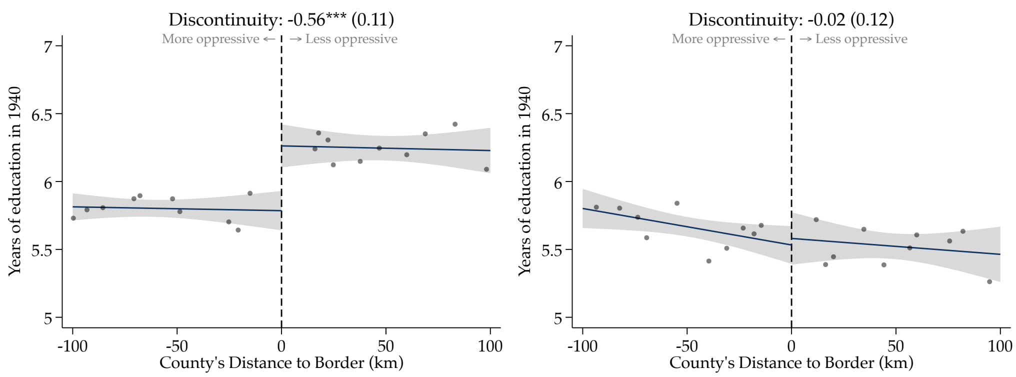
The end of slavery led to a drastic change in the geography of racially oppressive institutions in the US. State governments took the leading role in instituting Jim Crow regimes to limit the economic progress of newly freed enslaved families. Our results show that state institutions became a crucial determinant of how likely a Black family was to experience severe forms of oppression over the next century, shaping Black families' long-run economic progress. In the next section, we provide further evidence that our border discontinuity design isolates the Jim Crow effect without being confounded by other factors.

7.3 Validation of the Border Discontinuity Design

To validate our border discontinuity design, we pool all borders, rather than estimating discontinuities for each border separately. The pooled regression equation closely follows equation (II.9). We equally divide our sample into two types of borders: "high-contrast borders" between states that strongly differ in their Jim Crow intensity (more than the median border difference in the HRR index); and "low-contrast borders" between states that differ less in their Jim Crow intensity (less than the median border difference).

FIGURE II.8: Pooled Regression Discontinuity Estimates

(A) High-contrast borders (B) Low-contrast borders

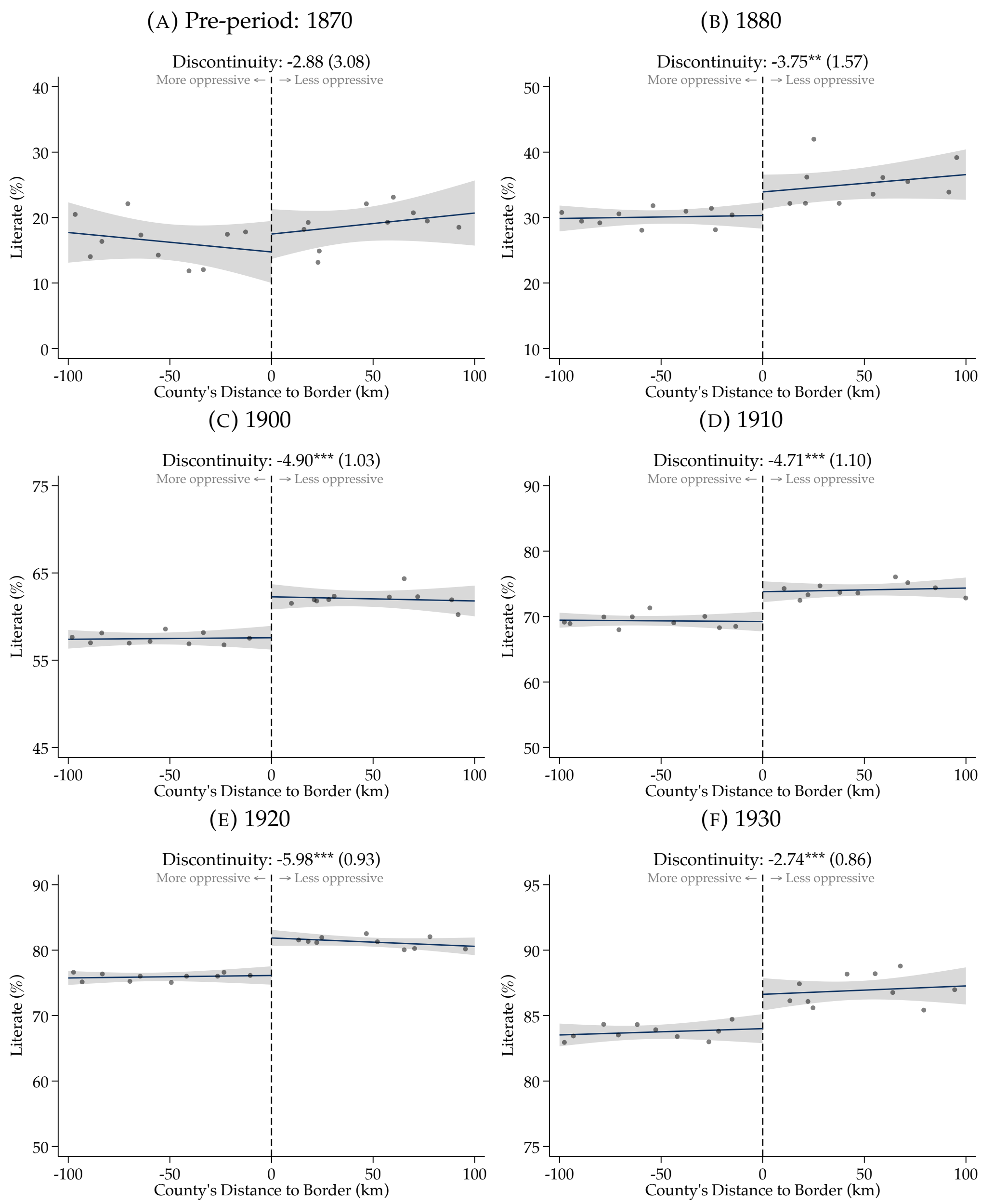


Notes: This figure shows the RD estimates in 1940 years of education for Black families freed across state borders with different Jim Crow intensity in 1865. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Panel A shows "high-contrast borders" where Jim Crow intensity differs more than across the median border (above 0.71 HRR index points, with differences averaging 1.30 HRR index points); panel B shows "low-contrast borders" where it differs less than the median (below 0.71 HRR index points, with differences averaging 0.32 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix 2.2 for details on the sample and data.

Consistent with our main estimates, sharp educational differences only arise for Black families freed across borders where institutions differ substantially (see Figure II.8).[21] Being freed on the more oppressive side of such a high-contrast border sharply reduced the years of education in 1940 by 0.6 years—10 percent of the average among Black men.

[21] Appendix Figure IV.51 shows the pooled RD estimate for all borders—both high- and low-contrast.

FIGURE II.9: Regression Discontinuities in Literacy (High-Contrast Borders)



Notes: This figure shows the RD estimate in literacy for Black families freed across state borders with different Jim Crow intensity in 1865. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). The sample is restricted to high-contrast borders (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix 2.2 for details on the sample and data.

We confirm that differences across high-contrast borders only arise after the onset of Jim Crow (see Figure II.9). Before Jim Crow, there were no differences in literacy among families freed in states that would become more oppressive during Jim Crow.[22] In 1880, the literacy rates of families began to differ. By 1900, Black families attained almost five percentage points lower literacy rates in more oppressive states. These differences grow

[22] Appendix Figure IV.52 shows RD estimates in literacy rates over time, separately by border.

over time in absolute terms but even more so in relative terms. By 1930, while almost 90 percent of all Southern Black men were literate, families freed in more oppressive states were still 4.6 percentage points less likely to be able to read and write.

We also confirm that before Jim Crow, location characteristics evolved smoothly across state borders. In 1860, none of a large array of observable characteristics differed discontinuously across state borders in the South: the number of enslaved people relative to a county's overall population, the share of its Black population, the share of plantation crops (cotton, sugar, tobacco, and rice) of total agricultural output, total agricultural output per capita, cotton output per capita, farm values, white wealth inequality, migration costs to the North, population density, incomes, or the age of enslaved people (see Appendix Figure IV.53). Our validation exercises focus on high-contrast borders where differences in Black economic progress emerged, but the results generalize to low-contrast borders.

We further present evidence that Jim Crow institutions varied sharply across state borders. We find significant gaps in key outcomes directly targeted by Jim Crow across state borders with differing Jim Crow intensities (see Appendix Figures IV.54, IV.55, and IV.36). Specifically, counties in states with more severe Jim Crow regimes have sharply lower voter participation, Black school attendance, Black teacher education, and Black teacher wages, plausibly reflecting the direct impact of suffrage restrictions and reduced school funding instituted in those states. Importantly, neither voter participation nor Black school attendance differ sharply across borders before the Jim Crow era (the other outcomes are not observed pre-Jim Crow). We also find that the number of lynchings between 1883 and 1941 does not vary sharply across borders, supporting the assumption that border differences in economic progress capture the effect of state institutions (see Appendix Figure IV.56).

Our results are also robust to using alternative measures for the intensity of states' Jim Crow regimes. We consider both the Jim Crow index and a state's number of Jim Crow laws (see Appendix Figure IV.21).

Last, we show that our results are robust to different cutoffs for the distance between a county's centroid and a state border between 100 and 350 kilometers (see Appendix Figure IV.22). The pooled RD estimates across high-contrast borders (as shown in panel A of Figure II.8) for those cutoffs all range between -0.61 and -0.46 and are all highly significant. Our baseline bandwidth is 100 kilometers in pooled estimations—close to the mean squared error optimum—and 350 kilometers when separately estimating discontinuities by state pair to reduce the impact of smaller sample sizes.

The results from our regression discontinuity design also strongly support our key identifying assumption—that the birthplace of an enslaved person is orthogonal to their innate "ability." Specifically, we find that the differences in the causal effects of states sharply and fully arise at state borders. Therefore, the main potential threat of selection bias remains the selection of enslaved people into states sharply around borders. However, any plausible selection into the destination of forced migration was based on the crop cultivated in an area that, as we confirm, transcends state borders (along with many other characteristics of border areas). Therefore, the selection of enslaved people into location is implausible to affect our results. In addition, we directly rule out selection based on observable characteristics, showing that the characteristics of enslaved people, such as their age during or their literacy immediately after slavery, do not differ across borders.

In sum, our evidence suggests that states' Jim Crow regimes played a critical role in shaping the South's detrimental effect on Black economic progress. The estimates are a lower bound for Jim Crow's importance because all Southern states adopted Jim Crow regimes. Our estimates only isolate the *additional* effect of more oppressive institutions rather than their aggregate effects.

8 The Mechanism of Limited Access to Education

Leading scholars have pointed out the importance of Jim Crow in limiting Black families' long-run human capital accumulation. Booker T. Washington writes that "few people [have an] idea of the intensive desire which [Black people] showed for education. It was a whole race trying to go to school" (Washington, 1907). However, Black people's desire for education was met with resistance. "[Black Americans'] attempts at education provoked the most intense and bitter hostilities as evincing a desire to render themselves equal to the whites" (Freedmen's Commission Report cited in Du Bois, 1935, p. 645). Robert Higgs argues that governments were the leading force of this resistance:

> *"Most damaging of all [racial discrimination after slavery] was the discriminatory behavior of the southern state and local governments. By providing only scant resources for black education, public school boards helped to perpetuate illiteracy [...], and they thereby set in motion a variety of adverse effects."* (Higgs, 1989, p. 25)

We use our newly built database on laws and their content to explore the relative importance of different domains that Jim Crow regimes affected. We document that the most significant number of laws pertained to education, accounting for one-third of all Jim Crow laws passed across the South until 1950 (see Appendix Figure IV.37).[23]

Jim Crow laws on education established the provision of resources for new schools or colleges for white Americans only. They also required the racial segregation of existing schools or local school boards to comprise only white people. Even school books were regulated, stipulating that once a Black or white child had used a book, children of the other race were not allowed to use the same book. Those laws likely created drastic differences in the educational resources available to Black and white children. Indeed, we find a robust negative correlation between a state's number of education-specific Jim Crow laws and the quality of Black schools ($\rho = -0.70$).

Our analysis of Black teacher wages confirms that disparities in school quality are pronounced right at states' borders, underlining the critical role of institutional factors

[23] A category's number of Jim Crow laws is not a conclusive measure of its importance; suffrage laws are a prime example. Suffrage laws are low in number, but their effects are massive (see e.g., Naidu, 2012). Laws in other categories are likely a downstream outcome of Black voter disenfranchisement (Engerman and Sokoloff, 2011). Therefore, while the number of Jim Crow laws on education is extensive, only through further analysis can one conclude that they were a crucial part of states' Jim Crow regimes.

in shaping the quality of Black schools (see Appendix Figure IV.36 and Margo, 1982, 1990b,a; Naidu, 2012; Card et al., 2022). We also explore the importance of education-specific Jim Crow regimes for Black economic progress by repeating our regression discontinuity design based on the number of education-specific Jim Crow laws and the quality of Black schools (Card and Krueger, 1992; Carruthers and Wanamaker, 2017). Both measures capture the sharp differences in Black economic progress across Jim Crow regimes (see Appendix Figure IV.57). These findings are consistent with Card and Krueger (1996) and Card et al. (2022) who show that state institutions induced critical differences in school quality and educational outcomes among Black children, "helping to explain the persistence of the human capital gap between Blacks and whites."

9 Conclusion

This paper provides new evidence on the long-run impact of racially oppressive institutions, finding that Black Americans' economic status today depends strongly on their ancestors' exposure to those institutions. First, we document that Black families enslaved until the Civil War continue to have considerably lower education, income, and wealth today. Second, we show that this persistence is mostly driven by post-slavery oppression under Jim Crow. We discuss Black Americans' limited access to education as a critical mechanism.

We put forward a new framework for slavery's legacy to incorporate systemic discrimination of the formerly Enslaved and their descendants under Jim Crow. The institution of slavery determined *where* a Black family was freed from slavery. We show that the state where a family was freed determined the Jim Crow regime they likely faced over the subsequent decades. While Jim Crow compressed the economic status of Black Americans *within* states, differences in Jim Crow intensity led to pronounced disparities *across* states, thereby placing descendants of those enslaved until the Civil War at a disproportionate disadvantage. After 1940, the main reason descendants of families enslaved until the Civil War have lower economic status is their concentration in the states that adopted the most strict Jim Crow regimes starting in 1877. Systemic discrimination—the higher exposure to ongoing discrimination *because of past discrimination* (Bohren et al., 2022)—is thus a central aspect of slavery's persisting legacy.

Despite the end of Jim Crow, today's geography of Black economic progress has similarities with that of the past. States that impeded Black economic progress post-slavery also limit intergenerational mobility for low-income children today (see Appendix Figure IV.58 and Berger, 2018). However, different from the Jim Crow era, those differences do not arise sharply across state borders. Future research should investigate why places' capacity to generate upward mobility has persisted despite drastic institutional change. Part of the answer may lie in anti-Black resentment, which remains high in places with historical prevalence of slavery and Jim Crow (Acharya et al., 2018).

Our findings have important implications for policies that aim to reduce the disadvantage faced by descendants of the Enslaved. First, our results highlight the importance of *within-race* disparities that race-specific policies may not address. College affirmative action is a prime example. Massey et al. (2007) show that the more selective a college, the less likely Black students are to descend from the Enslaved. For example, while only 13 percent of 18- to 19-year-old Black Americans have an immigration background, 41 percent of Black Ivy League students do. Affirmative action increases racial diversity on campuses but may be less effective in alleviating disadvantages faced by descendants of the Enslaved.

Second, there has been renewed interest in the specific policy of reparations, i.e., wealth transfers to descendants of the Enslaved (e.g., Darity, 2008; Craemer et al., 2020; Boerma and Karabarbounis, 2021; Albuquerque and Ifergane, 2023). We argue that any assessment of the legacy of slavery should incorporate both *when* and *where* a family was freed—i.e., how long they were enslaved and how intensively they were exposed to Jim Crow after slavery. Our empirical evidence suggests that Black families today are impacted drastically by when and where their ancestors were freed. While some argue that reparations should only be received by those who can prove their ancestors were enslaved, our results suggest that post-slavery institutions also harmed Black Americans who descended from the Free—a group that may find it harder to prove their ancestors had been enslaved decades before the Civil War. We must stress again that we only quantify the *additional* disadvantage faced by those whose ancestors were enslaved until 1865 and concentrated in the Lower South compared to those who gained freedom earlier, mainly in the Upper South and North. Many free Black Americans had been enslaved in earlier periods, and all Black Americans faced discrimination regardless of their specific family history.

This paper has limitations that future work may be able to overcome. First, we limit our analysis to men because automated census-linking methods are unavailable or have poor coverage for women. Women have historically tended to change their surnames upon marriage, making it impossible for conventional methods to link them across census records (Althoff et al., 2024). Second, we emphasize the significance of educational Jim Crow institutions as a crucial mechanism; however, institutions related to other aspects may have further impeded Black economic advancement. Although several of these institutions have been thoroughly investigated (e.g., restrictions on Black suffrage—see Naidu, 2012), numerous others remain relatively unexplored (e.g., constraints on interracial marriage). Third, while this paper quantifies the impact of Jim Crow, future work should explore the political economy underlying the rise of states' different institutional regimes.

Bibliography

AARONSON, D. AND B. MAZUMDER (2011): "The Impact of Rosenwald Schools on Black Achievement," *Journal of Political Economy*, 119, 821–888.

ABRAMITZKY, R., L. BOUSTAN, AND M. RASHID (2020): "Census Linking Project: Version 1.0," dataset: `https://censuslinkingproject.org`.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2021): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

ABRAMITZKY, R., J. CONWAY, R. MILL, AND L. STEIN (2023): "The Gendered Impacts of Perceived Skin Tone: Evidence from African-American Siblings in 1870–1940," Working paper.

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution*," *Quarterly Journal of Economics*, 117, 1231–1294.

ACEMOGLU, D. AND J. A. ROBINSON (2008): "Persistence of Power, Elites, and Institutions," *American Economic Review*, 98, 267–93.

ACHARYA, A., M. BLACKWELL, AND M. SEN (2018): *Deep Roots: How Slavery Still Shapes Southern Politics*.

AGER, P., L. BOUSTAN, AND K. ERIKSSON (2021): "The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War," *American Economic Review*, 111, 3767–3794.

AKBAR, P. A., S. LI, A. SHERTZER, AND R. P. WALSH (2020): "Racial Segregation in Housing Markets and the Erosion of Black Wealth," Tech. rep.

ALBUQUERQUE, D. AND T. IFERGANE (2023): "The Racial Wealth Gap: the Role of Entrepreneurship," .

ALEXANDER, M. (2010): *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*.

ALTHOFF, L. (2021): "Two Steps Forward, One Step Back: Racial Income Gaps among Women since 1950," Working paper.

ALTHOFF, L., H. BROOKES GRAY, AND H. REICHARDT (2024): "The Missing Link(s): Women and Intergenerational Mobility," Working paper.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly harmless econometrics*, Princeton university press.

ASHENFELTER, O. AND A. KRUEGER (1994): "Estimates of the Economic Return to Schooling from a New Sample of Twins," *The American Economic Review*, 84, 1157–1173.

BAKER, R. S. (2022): "The Historical Racial Regime and Racial Inequality in Poverty in the American South," *American Journal of Sociology*, 127.

BAYER, P. AND K. K. CHARLES (2018): "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *Quarterly Journal of Economics*, 133, 1459–1501.

BERGER, T. (2018): "Places of persistence: Slavery and the geography of intergenerational mobility in the United States," *Demography*, 55, 1547–1565.

BERLIN, I. (1974): *Slaves without masters: The free Negro in the antebellum South*, Oxford University Press.

——— (1976): "The Structure of the Free Negro Caste in the Antebellum United States," *Journal of Social History*, 9, 297–318.

BLACKMON, D. A. (2009): *Slavery by another name: The re-enslavement of Black Americans from the Civil War to World War II*, Anchor.

BOERMA, J. AND L. KARABARBOUNIS (2021): "Reparations and Persistent Racial Wealth Gaps," .

BOHREN, J. A., P. HULL, AND A. IMAS (2022): "Systemic Discrimination: Theory and Measurement," Working Paper 29820, National Bureau of Economic Research.

BONILLA-SILVA, E. (2015): "The Structure of Racism in Color-Blind, "Post-Racial" America," *American Behavioral Scientist*, 59, 1358–1376.

BOUSTAN, L. (2016): *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, National Bureau of Economic Research Publications, Princeton University Press.

CARD, D., C. DOMNISORU, AND L. TAYLOR (2022): "The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility," *Journal of Labor Economics*, 40, S1–S493.

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

——— (1996): "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina," *The Journal of Economic Perspectives*, 10, 31–50.

CARRINGTON, W. J., E. DETRAGIACHE, AND T. VISHWANATH (1996): "Migration with Endogenous Moving Costs," *American Economic Review*, 86, 909–930.

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

CARSON, S. A. (2008): "The effect of geography and vitamin D on African American stature in the nineteenth century: evidence from prison records," *The Journal of Economic History*, 68, 812–831.

CASCIO, E. U. AND E. G. LEWIS (2022): "Legal Activism, State Policy, and Racial Inequality in Teacher Salaries and Educational Attainment in the Mid-Century American South," Working Paper 30631, National Bureau of Economic Research.

CHETTY, R. AND N. HENDREN (2018): "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 133, 1163–1228.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment," *American Economic Review*, 106, 855–902.

CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 129, 1553–1623.

CLARK, G. (2014): *The Son Also Rises: Surnames and the History of Social Mobility*, Princeton University Press.

COHEN, W. (1991): *At Freedom's Edge: Black Mobility and the Southern White Quest for Racial Control, 1861–1915*.

COLLINS, W. J. (1997): "When the Tide Turned: Immigration and the Delay of the Great Black Migration," *The Journal of Economic History*, 57, 607–632.

CRAEMER, T., T. SMITH, B. HARRISON, T. LOGAN, W. BELLAMY, AND J. WILLIAM DARITY (2020): "Wealth Implications of Slavery and Racial Discrimination for African American Descendants of the Enslaved," *The Review of Black Political Economy*, 47, 218–254.

DARITY, W. (2008): "Forty Acres and a Mule in the 21st Century," *Social Science Quarterly*, 89, 656–664.

DARITY, W., D. HAMILTON, AND K. ZAW (2016): "Race, Wealth and Incarceration: Results from the National Longitudinal Survey of Youth," *Race and Social Problems*, 8, 103–115.

DELL, M. (2010): "The Persistent Effects of Peru's Mining "Mita"," *Econometrica*, 78, 1863–1903.

DELL, M. AND B. A. OLKEN (2019): "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java," *Review of Economic Studies*, 87, 164–203.

DERENONCOURT, E. (2022): "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, 112, 369–408.

DERENONCOURT, E., C. H. KIM, M. KUHN, AND M. SCHULARICK (2022): "The wealth of two nations: The U.S. racial wealth gap, 1860–2020," Tech. rep.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

DU BOIS, W. E. B. (1935): *Black Reconstruction in America: An Essay Toward a History of the Part Which Black Folk Played in the Attempt to Reconstruct Democracy in America, 1860–1880*.

ENGERMAN, S. L. AND K. L. SOKOLOFF (2011): *Economic Development in the Americas since 1500: Endowments and Institutions*, Cambridge University Press.

FOGEL, R. AND S. ENGERMAN (1974): *Time on the Cross. The Economics of American Negro Slavery*.

FONER, E. (1963): *Forever Free: The Story of Emancipation and Reconstruction*, Knopf Doubleday Publishing Group.

——— (2014): *Reconstruction: America's Unfinished Revolution, 1863-1877*, Harper Collins.

FRAZIER, E. F. (1949): *The Negro in the United States.*, The Macmillan Company.

FRIEDEN, J., R. S. GROSSMAN, AND D. LOWERY (2023): "Was Freedom Road a Dead End? Political and socio-economic effects of Reconstruction in the American South," Working paper.

GALENSON, D. W. (1981): "White Servitude and the Growth of Black Slavery in Colonial America," *The Journal of Economic History*, 41, 39–47.

GOLDSTEIN, J. R., M. ALEXANDER, C. BREEN, A. M. GONZÁLEZ, F. MENARES, M. OSBORNE, M. SNYDER, AND U. YILDIRIM (2021): "CenSoc Mortality File: Version 2.0. Berkeley: University of California, 2021." dataset: `https://censoc.berkeley.edu/data/`.

HIGGS, R. (1989): "Black progress and the persistence of racial economic inequalities, 1865-1940," in *The Question of Discrimination: Racial inequality in the US labor market*, Middletown, CT: Wesleyan University Press, 9–31.

KOUSSER, J. M. (1974): *The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-Party South, 1880-1910*, Yale University Press.

LEEUWEN, M. H. D. V. AND I. MAAS (2011): *HISCLASS: A Historical International Social Class Scheme*, Universitaire Pers Leuven.

LOGAN, T. D. (2020): "Do Black Politicians Matter? Evidence from Reconstruction," *Journal of Economic History*, 80, 1–37.

MARGO, R. A. (1982): "Disfranchisement, School Finance, and the Economics of Segregated Schools in the United States South, 1890-1910," Ph.D. dissertation.

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

——— (1990b): "Race and Schooling in the South: A Review of the Evidence," in *Race and Schooling in the South, 1880-1950: An Economic History*, University of Chicago Press, 6–32.

——— (2016): "Obama, Katrina, and the Persistence of Racial Inequality," *The Journal of Economic History*, 76, 301—341.

MASSEY, D. S., M. MOONEY, K. C. TORRES, AND C. Z. CHARLES (2007): "Black Immigrants and Black Natives Attending Selective Colleges and Universities in the United States," *American Journal of Education*, 113, 243–271.

MELLO, S. (2023): "Fines and Financial Wellbeing," Working paper.

MURRAY, P. (1950): *States' Laws on Race and Color*.

NAIDU, S. (2010): "Recruitment Restrictions and Labor Markets: Evidence from the Postbellum U.S. South," *Journal of Labor Economics*, 28.

——— (2012): "Suffrage, Schooling, and Sorting in the Post-Bellum U.S. South," Working Paper 18129, National Bureau of Economic Research.

OLIVETTI, C. AND M. D. PASERMAN (2015): "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940," *American Economic Review*, 105, 2695–2724.

PERMAN, M. (2001): *Struggle for Mastery: Disfranchisement in the South, 1888-1908*, University of North Carolina Press.

PHILLIPS, U. B. (1918): *American Negro Slavery-A Survey of the Supply, Employment and Control of Negro Labor as Determined by the Plantation Regime*, Louisiana State University Press.

PRITCHETT, J. B. (2001): "Quantitative Estimates of the United States Interregional Slave Trade, 1820-1860," *The Journal of Economic History*, 61.

——— (2019): "Demographic Consequences of the Interregional Slave Trade," .

RANSOM, R. L. AND R. SUTCH (2001): *One Kind of Freedom: The Economic Consequences of Emancipation*, Cambridge University Press.

ROBACK, J. (1984): "Southern Labor Law in the Jim Crow Era: Exploitative or Competitive?" *University of Chicago Law Review*, 51, 1161–1192.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "IPUMS USA: Version 10.0," dataset: `https://doi.org/10.18128/D010.V10.0`.

SACERDOTE, B. (2005): "Slavery and the intergenerational transmission of human capital," *Review of Economics and Statistics*, 87, 217–234.

SCHAMA, S. (2006): *Rough crossings: Britain, the slaves and the American revolution*, Random House.

SMALLWOOD, S. E. (2008): *Saltwater slavery: A middle passage from Africa to American diaspora*, Harvard University Press.

SOWELL, T. (1978): "Essays and data on American ethnic groups," .

STECKEL, R. H. (1979): "Slave height profiles from coastwise manifests," *Explorations in Economic History*, 16, 363.

STECKEL, R. H. AND N. ZIEBARTH (2013): "A Troublesome Statistic: Traders and Coastal Shipments in the Westward Movement of Slaves," *The Journal of Economic History*, 73, 792–809.

——— (2016): "Trader selectivity and measured catch-up growth of American slaves," *The Journal of Economic History*, 76, 109–138.

TADMAN, M. (1977): "Speculators and slaves in the Old South: a study of the American domestic slave trade, 1820-1860," Ph.D. thesis, University of Hull.

——— (1979): "Slave Trading in the Ante-Bellum South: An Estimate of the Extent of the Inter-Regional Slave Trade," *Journal of American Studies*, 13, 195–220.

——— (2008): "6. The Interregional Slave Trade in the History and Myth-Making of the US South," in *The Chattel Principle*, Yale University Press, 117–142.

U.S. SENATE (2019): "Landmark Legislation: The Civil Rights Act of 1964," .

WALTON, H., S. C. PUCKETT, AND D. R. DESKINS (2012): *The African American Electorate: A Statistical History*, Washington, D.C.: CQ Press.

WASHINGTON, B. T. (1907): *Up From Slavery: An Autobiography*.

WESTERN, B. (2006): *Punishment and Inequality in America*, Russell Sage Foundation.

WOLF, E. S. (2006): *Race and Liberty in the New Nation: Emancipation in Virginia from the Revolution to Nat Turner's Rebellion*, LSU Press.

WOODSON, C. G. (1918): *A Century of Negro Migration*.

WOODWARD, C. V. (1955): *The Strange Career of Jim Crow*, Oxford University Press.

WRIGHT, G. (1986): *Old South, New South: Revolutions in the Southern Economy Since the Civil War*, Basic Books.

——— (1997): *Old South, New South: Revolutions in the Southern Economy since the Civil War*, The Louisiana State University Press.

——— (2013): *Sharing the Prize: The Economics of the Civil Rights Revolution in the American South*, Harvard University Press.

III The Missing Link(s): Women and Intergenerational Mobility

Lukas Althoff (Stanford), Harriet Brookes Gray (Yale) & Hugo Reichardt (LSE)

1 Introduction

Studies on the evolution of intergenerational mobility in US history have focused on men, studying the link between fathers' and sons' economic status. This male-centric focus has two main reasons: a lack of intergenerational datasets that include women and the emphasis on income as the primary measure of economic status, which fails to capture mothers' contributions in an era of limited female labor force participation. Other literatures, in contrast, highlight mothers' key role in child development, for example by serving as primary educators before the widespread establishment of schools.

In this paper, we study how both mothers and fathers shaped children's life chances in the US between 1850 and 1940. We find that intergenerational mobility increased from the 19th to the early 20th century when considering a measure of parental background that incorporates human capital alongside income. This finding challenges previous evidence of declining mobility based on income alone. The rise in mobility is driven by the substantial role of mothers' human capital in the early period, which diminished as formal schooling gradually replaced maternal home-education.

By constructing one of the first linked census panels to include women, we trace the parental backgrounds of sons and daughters. We overcome the challenge of linking women's census records despite name changes by leveraging historical administrative data from Social Security Number applications. These applications provide both married and maiden names for applicants' mothers and married female applicants. Using these data, we link the census records of 21 million women along with a similar number of men, resulting in a highly representative panel. We will make this dataset publicly available.

We also develop a novel methodology to account for multiple dimensions of parental background in the intergenerational analysis. To assess the joint importance of mothers and fathers, we propose measuring intergenerational mobility as the share of variation in child outcomes explained by parental background: R^2. Unlike traditional mobility measures, such as the parent-child coefficient, this measure accommodates multiple parental inputs. We show that the R^2 has many desirable properties and—in the special case of using only one parental input—has a one-to-one relationship with the rank-rank coefficient. Another advantage of R^2 is that it can be separated into each parent's predictive power using a statistical decomposition method (Shapley, 1953; Owen, 1977).

Finally, we use cutting-edge statistical techniques to accurately estimate intergenerational mobility despite limitations in historical data. Specifically, we build on a recently developed semi-parametric latent variable method to study rank-rank relationships between parents and children when only binary proxies of the underlying outcomes are observed (Fan et al., 2017). In the historical data, such binary proxies are common; for example, literacy can serve as a proxy for human capital. We extensively validate this method and discuss the assumptions it imposes on the joint distribution of parent and child outcomes.

Our first main finding is that intergenerational mobility increased from the 19th to the early 20th century, challenging previous evidence. Specifically, we find that parents' backgrounds, incorporating human capital alongside income, became less predictive of their children's income over time. The separate importance of parental human capital and income is a central aspect of intergenerational mobility theory (Becker et al., 2018), but prior empirical studies focus on income-to-income transmission alone.

Our second main finding is that maternal human capital is the main driver of increasing intergenerational mobility over time. The predictive power of mothers' human capital initially exceeded fathers', but it gradually declined to make both parents' contributions comparable. Decomposing our R^2 measure, we show that mobility would have decreased had it not been for the diminishing predictive power of mothers' human capital. This finding highlights mothers' key role in intergenerational mobility and shows that previous evidence of declining mobility is due to a focus on paternal factors.[1]

As a potential mechanism for the historically large and declining role of maternal human capital, we explore the shift from home-education to formal schooling. Until around 1900, public schooling was limited in many places and home education was common. Historians have highlighted the pivotal role of parental human capital in child development during this period (Kaestle and Vinovskis, 1978). Mothers, who primarily engaged in home production in this era, were key educators of their children (Dreilinger, 2021). "[T]he middle class mother was advised that she and she alone had the weighty mission of transforming her children into the model citizens of the day" (Margolis, 1984, p. 13). The spread of school access could therefore be a reason why parents' human capital—especially mothers'—became less important and intergenerational mobility increased over time.

We find that, indeed, intergenerational mobility increased with school access and that maternal human capital accounted for this trend. Specifically, mothers' (but not fathers') human capital was more predictive for children whose school access was low due to their race, sex, or place. For example, we find that Black children who lacked equal access to schools during the Jim Crow era relied more on their mother's human capital than white children. Similarly, we find that as school access expanded over time, mothers' predictive power declined. These findings offer an explanation for the importance of maternal human capital in early US history: as the main educators of their time, mothers were key contributors to their children's human capital and, as a consequence, to their broader economic status.

This paper deepens our insights into how mothers shaped Americans' life chances throughout history. Earlier studies focused either on father-child correlations (Olivetti

[1] We validate our panel-based findings on human capital mobility using the cross-section of children aged 13–16 in their parents' household, bypassing the need for record linkage.

and Paserman, 2015; Abramitzky et al., 2021a; Ward, 2023; Craig et al., 2019; Jácome et al., 2021; Buckles et al., 2023b) or the correlation between parents' average status and child outcomes (Chetty et al., 2014b; Card et al., 2022). None of these prior studies assesses mothers' importance in the intergenerational transmission of economic outcomes. Our paper emphasizes mothers' separate role in shaping child outcomes, uncovering that maternal human capital is a stronger predictor than father-based proxies. Espín-Sánchez et al. (2023) develop parametric assumptions under which the role of women in intergenerational mobility can be inferred from the outcomes of male family members. Instead, our methodology overcomes critical measurement issues to estimate women's role in intergenerational mobility directly, allowing us to highlight the mechanisms underlying their impact.

Including mothers in the study of mobility in US history is especially pressing given that evidence from other contexts suggests mothers are key determinants of child outcomes. For Norway, Black et al. (2005) and Abrahamsson et al. (2024) find that education and health interventions have positive intergenerational spillovers to the children of treated mothers but not treated fathers. García and Heckman (2023) show that programs to increase mothers' parenting skills increase intergenerational mobility. Leibowitz (1974) shows that mothers' education is a strong predictor of child human capital whereas fathers' education is not, which they argue is a result of mothers spending more time with their children than fathers.

This paper also expands our knowledge on how women have contributed to the economy throughout US history. Goldin (1977, 1990, 2006) pioneered the effort to study women's contributions as their labor force participation rose mid-20th century (see also Fernández et al., 2004; Olivetti, 2006; Fogli and Veldkamp, 2011; Fernández, 2013). For the era before the rise of female labor force participation, evidence on women's contribution is largely limited to documenting their hours worked in home production (Greenwood et al., 2005; Ramey, 2009; Ngai et al., 2024). While the output of home production is typically hard to measure, we uncover the product of one key aspect: the home-education of children. We find that through their unique role in child development, women made a critical contribution to human capital accumulation in the US economy, even before the rise of female labor force participation.

Lastly, a key contribution of this paper is to construct one of the most extensive and representative panels on intergenerational mobility that includes women, building on the foundations of previous work. Craig et al. (2019) and Bailey et al. (2022) initiated the effort to link women's records by expanding automated record linkage developed for men by Abramitzky et al. (2021b). However, the information they use to do so—historical birth, marriage, and death certificates—are available only for selected states and periods. Buckles et al. (2023b) innovatively use crowd-sourced family trees, leading to vastly larger sample sizes. In contrast to prior work, we leverage historical *administrative* data, allowing for both scale and representativeness.[2]

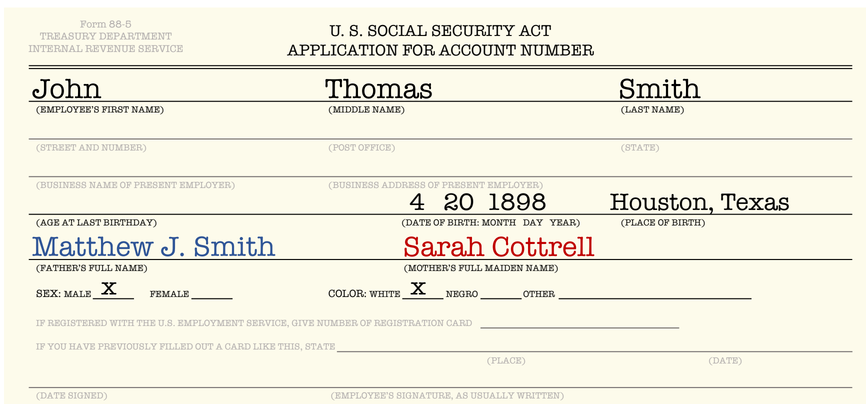
[2] Espín-Sánchez et al. (2023) employ a small subset of the same administrative data.

2 A New Panel that Includes Women (1850–1940)

A main empirical challenge in including women to study the long-run evolution of intergenerational mobility is the lack of suitable panel data. In this section, we describe how we overcome this hurdle by combining census records with historical administrative data that contain the married and maiden names of millions of women. Using these data, we link adult men and women in historical censuses (1850-1940) to their childhood census records. The resulting panel data stands out in its coverage and representativeness, particularly because it includes women.

2.1 Historical Administrative Data (Social Security Administration)

FIGURE III.1: Social Security Application Form


Form 88-5
TREASURY DEPARTMENT
INTERNAL REVENUE SERVICE

U. S. SOCIAL SECURITY ACT
APPLICATION FOR ACCOUNT NUMBER

John (EMPLOYEE'S FIRST NAME) | Thomas (MIDDLE NAME) | Smith (LAST NAME)

(STREET AND NUMBER) | (POST OFFICE) | (STATE)

(BUSINESS NAME OF PRESENT EMPLOYER) | (BUSINESS ADDRESS OF PRESENT EMPLOYER)

(AGE AT LAST BIRTHDAY) | 4 20 1898 (DATE OF BIRTH: MONTH DAY YEAR) | Houston, Texas (PLACE OF BIRTH)

Matthew J. Smith (FATHER'S FULL NAME) | Sarah Cottrell (MOTHER'S FULL MAIDEN NAME)

SEX: MALE X FEMALE ____ COLOR: WHITE X NEGRO ____ OTHER ____

IF REGISTERED WITH THE U.S. EMPLOYMENT SERVICE, GIVE NUMBER OF REGISTRATION CARD ____

IF YOU HAVE PREVIOUSLY FILLED OUT A CARD LIKE THIS, STATE ____ (PLACE) (DATE)

(DATE SIGNED) | (EMPLOYEE'S SIGNATURE, AS USUALLY WRITTEN)

Notes: This figure sketches a filled-in Social Security application form. Besides the applicants' name, address, employer, year and state of birth, and race, the application includes the father's name and the mother's maiden name. We access a digitized version of these data.

The historical administrative data comprise 41 million Social Security Number (SSN) applications, covering the near-universe of applicants. For data privacy reasons, only applicants who died before 2008 are included. The data contain each applicant's name, age, race, place of birth, and the maiden names of their parents (see Figure III.1). Based on these data, we can derive the married and maiden names of millions of women including all applicants' mothers and a smaller group of female applicants who were married at the time of application. We sourced a digitized version of these data from the National Archives and Records Administration (NARA).

Representativeness. Initially, SSN applicants were not representative of the US population, as the SSN system was launched in 1935 to register employed individuals, excluding self-employed and certain other occupations (Puckett, 2009). However, its scope rapidly expanded; for example, Executive Order 9397 in 1943 and the IRS's adoption of SSNs for tax reporting in 1962 increased its coverage to almost 100 percent. Throughout, the share of female applicants has been close to 50 percent (see Appendix Figure IV.70). The representativeness of our sample is further improved by parents who enter our sample irrespective of whether they applied for an SSN.

Coverage. The data has extensive coverage of men and women born in the 1880s or after.

The majority of Americans born in or after 1915 were assigned an SSN and therefore enter our data as applicants—a fact we establish by comparing each cohort's number of births and SSNs (Centers for Disease Control and Prevention, 2023; Social Security Administration, 2023). The share of Americans with an SSN rises from 64 percent for those born in 1915 to 80 percent for those born in 1920, 90 percent for 1935, and close to 100 percent starting with those born in 1950. The inclusion of parents in the SSN application files extend this coverage further back.

2.2 Census Data

We use the full-count census data for all available decades between 1850 and 1940 (Ruggles et al., 2020). These data include each person's full name, state and year of birth, sex, race, marital status, and other information. The data also identify family interrelationships for individuals in the same household. For those who live with their parents or spouses, we therefore also observe parental or spousal information.

2.3 Linking Method

We use a multi-stage linking process to maximize the utility of SSN application data, building on existing methods of automated record linkage (Abramitzky et al., 2021b). This procedure consists of three stages: linking SSN applicants to census records, linking applicants' parents to census records, and tracking census records over time. Appendix 3.5 describes our linking procedure in greater detail.

First stage: Applicant SSN $\leftrightarrow$ census. We start by linking each SSN applicant to their corresponding census record, using a rich set of criteria such as full names of the applicants *and* their parents, year and state of birth, race, and sex. The criteria are then progressively relaxed to the literature standard, which involves only first and last name with spelling variations allowed, state of birth, and year of birth within a 5-year band. A link is established if a unique match is found; if dual matches occur, we discard the observation. For married female applicants, we conduct searches under both maiden and married names; however, if links to a census can be established with both names, we establish no link due to the non-uniqueness of the matches.

Leveraging the combination of both applicants' and their parents' names helps us establish *unique* matches for SSN applicants recorded in the same census household as their parents. Historically, this approach is not only effective for children but also adults in the many existing multi-generational households. During our sample period, 80 to 90 percent of Americans lived in multi-generational households. By the end of our sample period in 1940, 60 percent of 21-year-olds and 20 percent of 30-year-olds lived with at least one parent. Note that while using parental names increases the uniqueness of potential matches of those residing with their parents, we also link adults not observed with their parents.

Second stage: Parent SSN $\leftrightarrow$ census. After linking SSN applicants to their census records, we focus on linking their parents to the census. Since specific birth details for applicants' parents are not available in the SSN applications, we cannot directly link them as we do for applicants. However, if a child's SSN application is successfully matched to a census record, and that census record shows the child residing with their parents, we can link the parents from an SSN application to that specific census household. For parents who are not SSN applicants themselves, we create a synthetic identifier similar to an SSN.

Third stage: Census $\leftrightarrow$ census. Having assigned unique identifiers to millions of individuals in the census records, we can link these records over time irrespective of name changes. We cover all possible pairs of census decades from 1850 to 1940. A person only enters the linked census panel if their SSN application record is linked to at least two different census decades.

In principle, it would be possible to establish additional links across census records by using standard or machine learning methods. These methods would be particularly useful for men and never-married women, where the issue of name changes does not apply. However, we choose not to use these methods for two reasons. First, our dataset's unique value lies in its ability to trace women from childhood to adulthood despite name changes—a feature not replicable by standard linking or machine learning methods. Second, using different methods for different subgroups would compromise the representativeness of our sample, as married women would be linked based on a different set of criteria than other groups.

2.4 Our New Panel

In the first two stages, our process assigns SSNs to 36 million census records—16 million applicants and 20 million parents. Our linking rate is 40 percent for applicants, surpassing the more typical 25 percent of prior studies thanks to our use of more detailed information, notably parent names. In the third stage, we link 112 million census records over time, tracking each of the 36 million individuals through more than three census decade pairs on average.

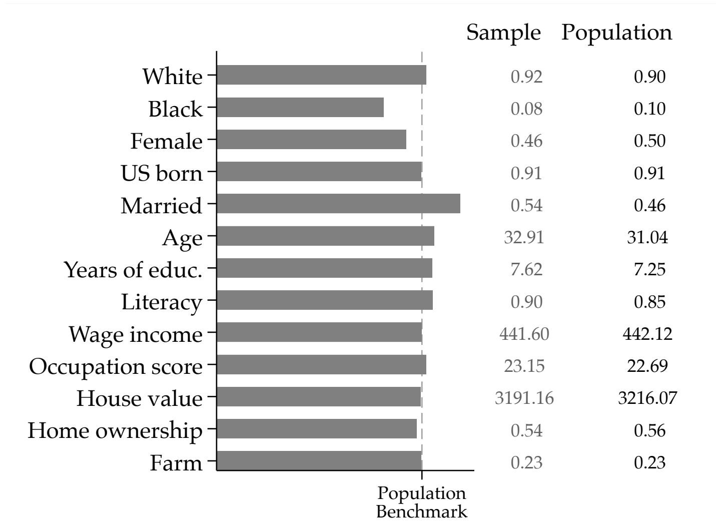
A standout feature of the panel is the inclusion of 12 million women for whom we observe pre- and post-marriage data. The sample sizes are largest for people born between the 1890s and the 1920s, with each birth decade containing 1.5 to 3 million women. These data allows us to overcome critical data limitations to study the role of women in intergenerational mobility throughout US history.

Our panel is highly representative of the overall US population across several metrics, including gender and race (see Figure III.2). Women comprise 46 percent of our linked sample in 1940. The sample mirrors the US-born and foreign-born shares of the population. While Black Americans are slightly underrepresented, our panel exceeds the representativeness of other samples in this dimension as well. Socioeconomic factors like income, home ownership, years of education, and literacy also align well with the broader population. Our sample over-represents married individuals, possibly because we use the names of a person's children or spouse in the linking procedure if they are known to us, improving linking rates for those who have children, a spouse, or both.

We reweight our sample to more closely resemble the US population's characteristics in our empirical analysis.[3] Our reweighted sample is close to perfectly representa-

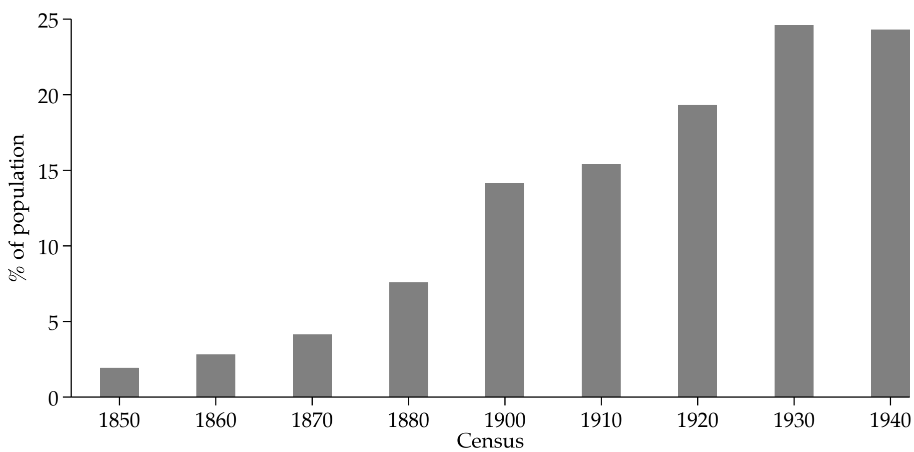
[3] We use a flexible non-parametric method to construct inverse propensity weights (see Appendix 3.6).

FIGURE III.2: Sample Balance Prior to Weighting (1940)



Notes: This figure shows the representativeness of characteristics among individuals in the 1940 census who we successfully assign an SSN compared to the full population in the 1940 census. The sample is exceptionally representative compared to existing panels, most notably with respect to sex and race. Because of the large sample sizes, even economically small differences are statistically significant. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate.

FIGURE III.3: Fraction of US Population Linked in Our New Panel



Notes: This figure shows the fraction of the full population of men and women that we successfully assign a Social Security Number (SSN). This includes parents of SSN applicants who did not apply for an SSN themselves and who we assign synthetic identifiers.

tive of the full population, even in characteristics not directly targeted by the reweighting method. The panel maintains its representative quality even in the earliest census decades (see Appendix Figure IV.71).

Moreover, our panel offers broad coverage. It captures 7–20 percent of the US population from 1910–1940 and 1–5 percent from 1850–1900 (see Figure III.3). This extensive reach makes our sample highly valuable for longitudinal studies.

Compared to existing linked census data, our new panel covers a substantial number of individuals whose records have not previously been linked, while maintaining high agreement rates with existing data for overlapping individuals (see Appendix Figure IV.72). Our panel shares the most data with the novel Census Tree—an innovative, extensive panel that includes women through genealogical data (Buckles et al., 2023a). Agreement rates vary from 80 to nearly 100 percent and are highest with LIFE-M—a panel that leverages vital records in the linking process (Bailey et al., 2022).

2.5 Economic Outcomes

To understand the role of mothers and fathers in shaping child outcomes, we require separate measures of each parent's outcomes. We therefore focus on human capital measures, such as literacy or years of education, reflecting the status of both men and women.

To measure parental background, we additionally consider household-level measures such as income. We incorporate household-level alongside individual-level information only when considering the overall importance of parental background, not when we aim to distinguish mothers' and fathers' separate contributions.

For children, we consider outcomes during both child- and adulthood. During childhood (ages 13–16), we measure literacy (as a proxy for human capital), school attendance, and total years of schooling completed. During adulthood (ages 20–54), we measure literacy, years of education, and occupational income scores.

3 Measuring Intergenerational Mobility with Multiple Inputs

In this section, we propose a statistical model of intergenerational mobility that accounts for the contributions of both fathers' and mothers' human capital to their children's economic outcomes. First, we propose using the R^2 of a regression of child outcomes on multiple parental inputs as a mobility measure that integrates the roles of both parents. Second, we use a simple decomposition method that allows to separate the contributions of mothers and fathers to the overall R^2. Third, we build on a state-of-the-art semi-parametric latent variable method to estimate the R^2 from a rank-rank regression when only binary proxies of underlying outcomes are observed (e.g., literacy as a proxy for human capital).

3.1 A Simple Model of Intergenerational Mobility

We build on standard statistical models of intergenerational mobility where a child's economic outcome is a linear function of parental inputs:

$$\operatorname{rank}(y_i) = \alpha + \boldsymbol{\beta}' \mathbf{rank}\left(\boldsymbol{y}_i^{\text{parental}}\right) + \varepsilon_i, \tag{III.1}$$

where $\operatorname{rank}(y_i)$ is the percentile rank of outcome of i and $\mathbf{rank}\left(\boldsymbol{y}_{\boldsymbol{i}}^{\textbf{parental}}\right)$ is a $k \times 1$ vector of i's ranked parental outcomes. Parental outcomes can include information on mothers, fathers, or both parents.

There are several advantages to the rank-rank approach, which considers mobility in relative positions in the distribution (Chetty et al., 2014a). First, correlations in ranks are not affected by changes in the marginal distribution of outcomes which, given the long time horizon of our study, enhances the interpretability of the coefficients. Second, using ranked outcomes ensures that the marginal distributions of mother's and father's outcomes are identical, so that their relative contributions can be effectively compared.

This statistical model differs from most previous research by allowing for multiple parental inputs—most importantly to explicitly incorporate mothers alongside fathers as contributors to a child's outcomes. While in this paper we focus on human capital and income, the model can be extended to accommodate many different inputs including parents' wealth, grandparents' or other relatives' backgrounds, or neighborhood characteristics.

3.2 R^2 as a Measure of Mobility with Multiple Inputs

We propose using the R^2 of equation (III.1) as an intuitive mobility measure that can account for multiple inputs. It summarizes the joint importance of mothers and fathers:

$$R^2 = \frac{\sum_{i=1}^{N}\left[\widehat{\operatorname{rank}}(y_i) - 50\right]^2}{\sum_{i=1}^{N}\left[\operatorname{rank}(y_i) - 50\right]^2} = \frac{\text{Variance in child outcomes explained by parents}}{\text{Variance in child outcomes}},$$

where $\widehat{\operatorname{rank}}(y_i)$ is the predicted rank of i from equation (III.1) and 50 is the average rank by construction.

We argue that predictability as captured by the R^2 is an intuitive measure of intergenerational mobility. In a perfectly mobile society, child outcomes cannot be predicted by parental background ($R^2 = 0$). In contrast, if child outcomes can be perfectly predicted by parental background ($R^2 = 1$), society is perfect immobile.

The R^2 has a direct relationship with traditional mobility measures—parent-child coefficients or, most commonly, father-son coefficients ($\widehat{\beta}$).[4] In Appendix 3.3.1, we show that in such univariate rank-rank regressions, there is a one-to-one mapping between the parent-child coefficient and our mobility measure: $R^2 = \widehat{\beta}^2$.

The advantage of R^2 is that it can provide an intuitive and easily interpretable measure of mobility even when considering multiple parental inputs. We use this advantage

[4] The parent-child coefficient $\widehat{\beta}$ is the OLS estimate of β: $\operatorname{rank}(y_i) = \alpha + \beta \cdot \operatorname{rank}\left(y_i^{\text{parental}}\right) + \varepsilon_i$.

to include both mothers' and fathers' outcomes, and to include multiple dimensions of parental background. Another advantage is that the R^2 can be decomposed into the contributions of individual inputs, as described in the next section.

3.3 Measuring Individual Inputs' Contribution to R^2

To assess the contribution of individual parent inputs in shaping child outcomes, we decompose the overall R^2 using a statistical method based on Shapley (1953); Owen (1977).

This decomposition method defines the contribution ϕ_j of each set of inputs $\boldsymbol{x}_j \subseteq V$ to the overall R^2:

$$\phi_j = \sum_{T \subseteq V - \{\boldsymbol{x}_j\}} \frac{1}{k!} \left[R^2(T \cup \{\boldsymbol{x}_j\}) - R^2(T) \right],$$

where $R^2(T)$ represents the R^2 of regressing the dependent variable (e.g., rank (y_i)) on a set of variables $T \subseteq V$ (e.g., $V = \left\{ \text{rank}\left(y_i^{\text{mother}}\right), \text{rank}\left(y_i^{\text{father}}\right) \right\}$), and k is the number of variables in V (i.e., $k = |V|$). Intuitively, ϕ_j represents the weighted sum of marginal contributions that a parent makes to the variation in child outcomes explained by different combinations of parental inputs. In Appendix 3.3.2, we describe the decomposition method in more detail and, for the special case of two parental inputs, provide a closed-form expression for ϕ_j in (III.1) in terms of the estimated coefficients and the correlation between the inputs.

The Shapley-Owen decomposition offers several unique advantages, being the only that satisfies three formal conditions defined by Young (1985) and Huettner and Sunder (2011) that can be summarized as follows:

1. *Additivity.* Individual contributions to the R^2 add up to the total R^2.
2. *Equal treatment.* Regressors that are equally predictive receive equal values.
3. *Monotonicity.* More predictive regressors receive larger values.

While the Shapley-Owen decomposition method is popular in the machine learning literature (Lundberg and Lee, 2017; Redell, 2019), it has not been widely used in economics (recent exceptions are Biasi and Ma, 2023; Fourrey, 2023; Redding and Weinstein, 2023).

3.4 Measuring Mobility with Latent Inputs

To estimate rank-rank mobility (R^2) when we only observe binary proxies of the rank variables in equation (III.1), we propose a method based on Fan et al. (2017). Appendix 3.3.3 discusses the method in detail.

Many binary variables can be interpreted as a function of a continuous underlying latent variable that is equal to one if that variable exceeds an unknown threshold and zero otherwise. In our application, we interpret literacy—the only information on human capital in pre-1940 censuses—as such a proxy for human capital.

Under distributional assumptions, we can use the observed binary proxies to identify the parameters and R^2 in equation (III.1). Specifically, we assume that parental and child outcomes in equation (III.1) are drawn from a joint Gaussian copula distribution. That is, we assume that there exists a set of unknown monotonic transformations $f_c, f_{p_1}, \ldots, f_{p_k}$ such that $\left(f_c(y_i), f_{p_1}\left(y_{i,1}^{\text{parental}}\right), \ldots, f_{p_k}\left(y_{i,k}^{\text{parental}}\right)\right)' \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ with $\text{diag}(\boldsymbol{\Sigma}) = \mathbb{1}$.[5] We do not require information on the monotonic transformation themselves. Note that because ranks are themselves monotonic transformations, this assumption implies that not only the outcomes but also their ranks follow the Gaussian copula distribution.

The Gaussian copula distribution is commonly used in the statistics literature due to its flexibility and good performance in practice (e.g. Liu et al., 2009, 2012; Zue and Zou, 2012). It is a family of probability distributions that includes but is not limited to the normal distribution. For instance, since it includes any monotonic transformation of normally distributed random variables, it allows for skewed and multi-modal distributions. Importantly, the Gaussian copula assumption does not impose that the latent variables of interest (e.g., human capital) are themselves normally distributed.

We show that this semi-parametric latent variable method allows us to estimate the rank-rank regression in equation (III.1) even if only binary proxies of the rank variables are observed. Specifically, Fan et al. (2017) show how to estimate $\boldsymbol{\Sigma}$—the correlations between each underlying variable—under such data limitations.[6] $\boldsymbol{\Sigma}$ in turn identifies the pairwise correlations between the ranked variables. We show that any rank-rank regression is identified by the pairwise correlations, and that therefore $\boldsymbol{\Sigma}$ is sufficient to identify equation (III.1) including its R^2. In Appendix 3.3.3, we present an explicit formula for $\widehat{\boldsymbol{\beta}}$ and R^2 as a function of $\widehat{\boldsymbol{\Sigma}}$.

We extensively validate this method and show that it correctly recovers rank-rank mobility by simulation.

First, when observing rank variables to estimate rank-rank mobility directly, we show that our method correctly identifies mobility even after the rank variables are dichotomized arbitrarily. Specifically, we use ranks in educational attainment from the 1940 census and dichotomize this data. We use different cutoffs for children, mothers, and fathers (e.g., 11 years for children, 9 for mothers, 7 for fathers). Our method's mobility estimates by state align well with those derived from the original, undichotomized data (see Panel A, Appendix Figure IV.59). This shows the method's performance in relevant historical data.

Second, we show that the method is robust to cut-offs changing over time, even shifting towards tail ends of the distribution. In our context, an important concern stems from literacy increasing to close to 100 percent over time, changing the information that it contains about a person's human capital rank. To address this concern, we simulate jointly normally distributed data, transform them in ranks, and dichotimize these ranks according to historical literacy rates for each decade from 1870 to 1940. We show that, in contrast to Ordinary Least Squares, our semi-parametric latent variable method yields correct estimates of mobility (R^2) over time, despite changing cut-offs (see Panel B, Appendix Figure IV.59).

[5]Because we allow for any monotonic transformation of the underlying variable, the assumption that the marginal distributions have zero mean and variance equal to 1 is without loss of generality.

[6]The method in Fan et al. (2017) allows for a combination of binary and continuous variables. It can be extended to non-binary ordinal and truncated variables (Dey and Zipunnikov, 2022). Furthermore, they derive statistical properties of the estimator of $\boldsymbol{\Sigma}$, notably $\sqrt{n}$-consistency.

We apply the semi-parametric latent variable method not only to measuring rank-rank mobility in human capital (through literacy), but also to measuring educational rank-rank mobility (through school attendance at a given age). Because we anticipate this method to be useful for future research facing similar data limitations, we developed a Stata command for easy implementation by others.

4 Income Mobility & Parental Human Capital

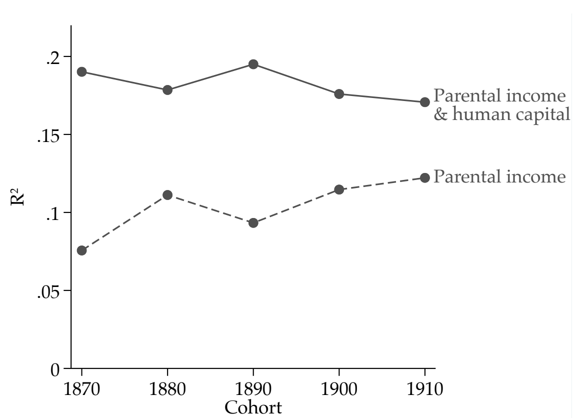
We measure intergenerational mobility as the share of variation in child outcomes that is attributable to parental background. We leverage our new panel that allows us to relate both men's and women's outcomes in adulthood with their parental background measured during childhood. We find that accounting for parental human capital alongside income reveals a trend of rising intergenerational mobility across US history, challenging earlier findings that considered only income. This shift is largely accounted for by the evolving role of maternal human capital—a finding corroborated by historical literature.

4.1 Income Mobility Accounting for Parental Human Capital

Theories of intergenerational mobility indicate that parental human capital, in addition to income, is a critical determinant of children's incomes (Becker et al., 2018). Human capital may not only increase parents' capacity for monetary investments in their children but may also shape their children's human capital directly. However, existing empirical studies focus on parental income and do not take human capital into account.

In addition to the theoretical rationale for including parental human capital, there are significant empirical reasons. The lack of detailed data on economic outcomes in historical US data has forced researchers to rely on occupational income proxies. Factoring in human capital can therefore substantially enhance the measurement of parental background in historical data.

FIGURE III.4: Share of Variation in Income Explained by Parental Background



Notes: This figure shows the share of the variance in a child's household income rank explained by (1) parents' household income ranks and their (latent) human capital ranks (R^2) and (2) parents' household income ranks alone. For parental human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. We use the household head's LIDO occupational income score (Saavedra and Twinam, 2020). Results are based on our new panel and sample weights are applied.

We account for both parental income and human capital by measuring intergenerational mobility as the R^2 in the following version of equation (III.1):

$$\text{rank}\left(inc_i\right) = \alpha + \beta_p \text{rank}\left(inc_i^{\text{parents}}\right) + \beta_m \text{rank}\left(h_i^{\text{mother}}\right) + \beta_f \text{rank}\left(h_i^{\text{father}}\right) + \varepsilon_i, \tag{III.2}$$

where inc is household income and h is (latent) human capital. We measure household income as the household head's LIDO occupational income score. Literacy serves as a binary proxy for latent human capital ranks. We estimate this model using the semi-parametric latent variable method described in section 3.4 and our new representative panel dataset described in section 2.4.

We find that parental human capital accounts for a large share of variation in children's incomes, even conditional on parents' incomes (see Figure III.4). In some periods, the predictive power of parental background doubles after incorporating human capital. Most importantly, the broader measure of parental background that includes both income and human capital suggests that intergenerational mobility in the United States increased over time—challenging the conclusion of declining mobility derived from measures based on income alone (Ferrie, 2005; Long and Ferrie, 2013; Feigenbaum, 2018; Song et al., 2020). We document a similar pattern when using more occupational income scores that are not specific to sex, race, age, or region ("occscore"; see Appendix Figure IV.60).

To understand the reason behind the reversal of the trend in intergenerational mobility, we decompose our mobility measure into multiple components and analyze their individual contributions. Specifically, we decompose R^2 in equation (III.2) into

$$R^2 = \widehat{\beta}_p^2 + \widehat{\beta}_m^2 + \widehat{\beta}_f^2 + 2\left(\widehat{\beta}_p\widehat{\beta}_m\hat{\rho}_{p,m} + \widehat{\beta}_p\widehat{\beta}_f\hat{\rho}_{p,f} + \widehat{\beta}_m\widehat{\beta}_f\hat{\rho}_{m,f}\right) \tag{III.3}$$

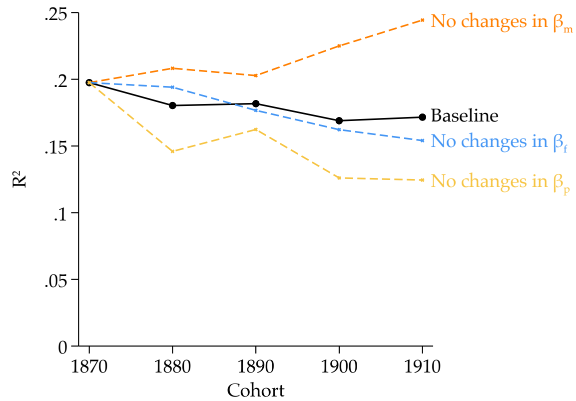
where $\hat{\rho}_{p,m}$, $\hat{\rho}_{p,f}$, and $\hat{\rho}_{m,f}$ are the correlations between parental income and mother's human capital, between parental income and father's human capital, and between mother's and father's human capital.[7] The latter correlation, $\hat{\rho}_{m,f}$, is a measure of assortative mating based on human capital. Using this decomposition, we compute the counterfactual R^2 holding a given parameter constant over time.

Our decomposition shows that the evolving role of maternal human capital ($\widehat{\beta}_m$) is the main reason why intergenerational mobility increased over time (see Figure III.5). Specifically, R^2 would have increased without the changing coefficient of maternal human capital. The importance of father's human capital ($\widehat{\beta}_f$) did not affect mobility significantly. Without changes in the importance of parental income ($\widehat{\beta}_p$) mobility would have increased even further. The rise in $\widehat{\beta}_p$ aligns with decreasing income mobility in previous research. However, we find that the focus of that research on income alone masked important changes in the role of parental background in shaping the outcomes of children (see also Ward, 2023, who documents that accounting for measurement error also reverses the trend).

In contrast to the slope coefficients ($\widehat{\boldsymbol{\beta}}$), none of the correlations between parental inputs ($\widehat{\boldsymbol{\rho}}$)—including assortative mating—had a significant impact on R^2 (see Appendix

[7]For a similar decomposition of R^2 in a rank-rank regression with an arbitrary number of independent variables, see equation (IV.22) in Appendix 3.3.1.2.

FIGURE III.5: The Changing Role of Parental Inputs in Intergenerational Mobility



Notes: This figure shows the role of each parameter on the R^2 in equation (III.2). The baseline represents the observed R^2 shown in Figure III.4. The other three lines represent the counterfactual R^2, had the respective parameter not changed over time, computed using the decomposition in equation (III.3). For parental human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. We use the household head's LIDO occupational income score (Saavedra and Twinam, 2020). Results are based on our new panel and sample weights are applied.

Figure IV.61). For instance, while patters in assortative mating decreased before 1880 and remained constant after (see Appendix Figure IV.62), these changes played a negligible role for intergenerational mobility.

Mobility by group. We show that the predictive power of parental background varies considerably across children of different sex and race (see Appendix Figure IV.63). Sons generally exhibit lower intergenerational mobility compared to daughters, with R^2 around twice as high for sons as for daughters (around 0.3 versus 0.15). White sons are least mobile, with 13 to 19 percent of variation in household incomes linked to parental background. Black sons are more mobile than white sons, followed by White daughters and Black daughters. Black daughters are not only the most mobile group, they are also the only group whose mobility increased over time. It is important to recognize that (1) high within-group mobility does not imply high mobility within the general population and that (2) high mobility does not necessarily equate to high *upward* mobility.

4.2 The Historical Role of Parental Human Capital

Our finding that parental human capital was important—and especially so in the late 19$^{\text{th}}$ century—is consistent with the historical role of parents. Prior to public school access becoming universal in the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries, parental home education was central for children's human capital development. Even children who were enrolled in school in the late 19$^{\text{th}}$ century attended school less than four months a year on average (Dreilinger, 2021).

The specific importance of the mothers' human capital to her children's outcomes also aligns with historical evidence. Women bore most of the responsibility to educate children in the home during the 19$^{\text{th}}$ century—a time marked by women's specialization in home production and a scarcity of public schools. Initially, in the early agrarian phase of US history, both men and women engaged in home-based industries. However, the first industrial revolution (around 1790–1830) ushered in factory work, especially among men, leading home production to be increasingly done by women. Consequently, women became the primary educators of children (Kaestle and Vinovskis, 1978; Margolis, 1984).

Mothers' pivotal role gained recognition from contemporary intellectuals, who advocated for the professionalization of women's role as home-educators. "The mother forms the character of the future man," Catharine Beecher, a famous American educator, wrote (Beecher, 1842). "The mother may, in the unconscious child before her, behold some future Washington or Franklin, and the lessons of knowledge and virtue, with which she is enlightening the infant mind, may gladden and bless many hearts," the Ladies' Magazine wrote (cited in Kuhn, 1947).

During this period, a substantial body of guidance was developed to equip women for this crucial responsibility. Beecher wrote: "Educate a woman, and the interests of a whole family are secured." Some even viewed home education as superior to formal school education. One hour in the "family school" may "do more towards teaching the young what they ought to know, than is now done by our whole array of processes and instruments of instruction" within schools and colleges, William Alcott, another American educator, wrote (cited in Kuhn, 1947).

Motivated by our finding of the importance of maternal human capital for intergenerational mobility and the historical literature, the subsequent analysis studies the specific role of mothers' human capital in shaping their children's outcomes.

5 Mothers & Human Capital Transmission

Motivated by our results in the previous section, we now zero in on the intergenerational transmission of human capital. We find that, mirroring our results on income mobility, human capital mobility increased significantly from the 1850s to 1910s birth cohorts. We decompose the overall predictive power of paternal human capital into the contributions of mothers and fathers. Our findings show that mothers' human capital more strongly predicts child human capital than fathers'. This difference is particularly pronounced for female and Black children.

5.1 Parental Human Capital and Child Outcomes

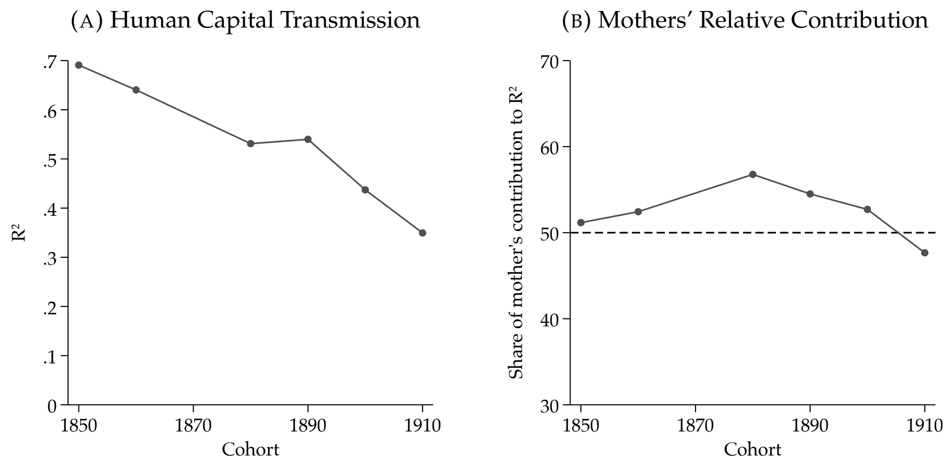
We estimate human capital mobility (R^2) in the following version of equation (III.1):

$$\text{rank}\left(h_i\right) = \delta + \gamma_m \text{rank}\left(h_i^{\text{mother}}\right) + \gamma_f \text{rank}\left(h_i^{\text{father}}\right) + \eta_i, \tag{III.4}$$

where h is (latent) human capital. We estimate this model using the semi-parametric latent variable method described in section 3.4 and use the census cross-section of children in their parents' households. We then use the Shapley-Owen decomposition described in section 3.3 to separate mothers' and fathers' contributions to predicting children's human capital (see Appendix Figure IV.64 for an illustration of the method).

Census cross-sections of children who reside with their parents allow us to study intergenerational mobility in certain outcomes without census linking. Specifically, we use such cross-sections to relate parental background to their children's early life outcomes of literacy and school attendance at ages 13–16. Within this age range, the likelihood of a child living apart from their parents is small, minimizing selection into the sample. Our results based on such census cross-sections provide a valuable benchmark for results derived from our new linked census panel. We also replicate those child-based results for adults using our new panel dataset described in section 2.4.

FIGURE III.6: Transmission of (Latent) Human Capital Ranks Across Cohorts



Notes: Panel A shows the share of the variance in a child's (latent) human capital rank explained by parents' (latent) human capital ranks (R^2) across cohorts. We recover human capital rank-rank transmission using information on literacy and the latent variable method introduced in section 3.4. Panel B shows mothers' relative contribution to the overall R^2 using the Shapley-Owen method. Results are based on the census cross-section of children ages 13–16 in their parents' household.

First, our estimates reveal increasing human capital mobility for American children born from the 1850s to the 1910s (see Panel A of Figure III.6). While parental background accounted for 70 percent of variation in human capital in the earliest cohort, this figure halved to 35 percent for those born in the latest cohort. The largest increases in human capital mobility took place around the end of slavery (1850–1880) and in the era of rapidly rising school attendance (around 1900).

Second, mothers' human capital was more predictive of child human capital than the fathers' (see Panel B of Figure III.6). For cohorts born before 1910, mothers' human capital contributed the majority of the predictive power of child outcomes. Over time, mothers' relative influence on children has diminished and fell below 50 percent for the first time among children born in the 1910s.

Our findings highlight the role of human capital transmission, especially from mothers, in enhancing income mobility over time. Our analysis in section 4 revealed that the declining predictive power of maternal human capital for their child's income led to increased mobility. We show in this section that the diminished predictive power of maternal human capital for income is accounted for by its reduced predictive power for the child's human capital.

We successfully replicate the cross-sectional patterns of human capital mobility using our new panel (see Appendix Figure IV.65). We find that the relative changes in human capital mobility (R^2) match perfectly across both datasets. Similarly, the proportion of human capital transmission attributed to mothers decreases by a similar amount in both datasets. Our panel, while confirming the patterns of *relative changes* over time observed in the cross-section, interestingly shows higher *levels* of human capital mobility. This difference can be explained by two main factors. First, the similarity between parental and child human capital is likely more pronounced in childhood than in adulthood, due to human capital accumulation or depreciation in adult life (intra-generational mobility). Unlike the cross-sectional analysis, our panel includes adult children and accounts for such intra-generational shifts, potentially leading to lower estimates of intergenerational mobility. Second, inaccuracies in automated record linkage might understate the degree of intergenerational persistence through measurement error in parental background.

5.2 Human Capital Mobility by Group

We estimate equation (III.4) separately by race and sex and find that human capital mobility varied significantly for Black and white Americans. The human capital rank of Black children born in the earliest cohort (1850s) was highly predictable by their parents' ($R^2 = 0.7$). However, Black children saw a rapid increase in mobility after slavery ended in 1865 ($R^2 = 0.2$ by 1880). After 1880, Black human capital mobility began to decline again. In contrast, white children's human capital mobility remained low and stable until around 1890 ($R^2 = 0.55$) before it sharply increased around 1900—four decades after the increase in Black mobility had started. The 1910s cohort marked the first time since the Civil War that white children's human capital mobility surpassed Black children's ($R^2 = 0.3$).

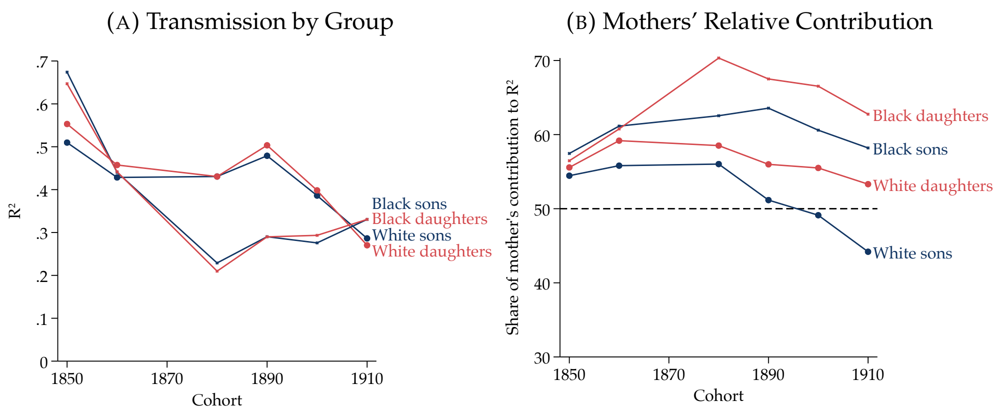
In line with this finding, school access among white children became almost universal in the early 1900s (see Appendix Figure IV.66). In contrast, most Black children—especially those whose ancestors were enslaved and largely denied literacy until 1865—lived in the Jim Crow South with restricted school access, shorter school years, and poor school quality (Card and Krueger, 1992; Althoff and Reichardt, 2023). The denial of equal access to high-quality schooling under Jim Crow may explain why human capital mobility among Black Americans decreased starting around 1880.

The finding that mothers' contributions to their children's human capital are generally larger than fathers' is particularly pronounced among female and Black children (see Panel B of Figure III.7).[8] Mother's large influence on daughters and Black children aligns with the historical lack of access to educational resources for these groups (Kober and Rentner, 2020). For daughters, it could also suggest the presence of gender-specific role model effects (e.g., Bettinger and Long, 2005; Olivetti et al., 2020).

We also estimate a version of equation (III.4) where (latent) human capital ranks are replaced with ranks in formal school attendance completed from the 1940 census. We find that racial differences in educational mobility are larger than those in human capital mobility (see Appendix Figure IV.67). This result underscores the fact that the lack of access to formal schooling was even more persistent across generations among Black families than the racial differences in human capital. In contrast, white Americans, who

[8] Olivetti et al. (2018) find similar gender-specific transmission from paternal and maternal grandparents to their grandsons and granddaughters.

FIGURE III.7: Transmission of (Latent) Human Capital Ranks By Group



Notes: Panel A shows the share of the variance in a child's (latent) human capital rank explained by parents' (latent) human capital ranks (R^2) across cohorts and groups. We recover human capital rank-rank transmission using information on literacy and the latent variable method introduced in section 3.4. Panel B shows mothers' relative contribution to the overall R^2 using the Shapley-Owen method. Results are based on the census cross-section of children ages 13–16 in their parents' household.

had nearly universal access to schools, were able to substitute parental homeschooling with formal schooling, thereby generating even higher mobility than that observed in human capital.

6 The Role of Mothers as Educators

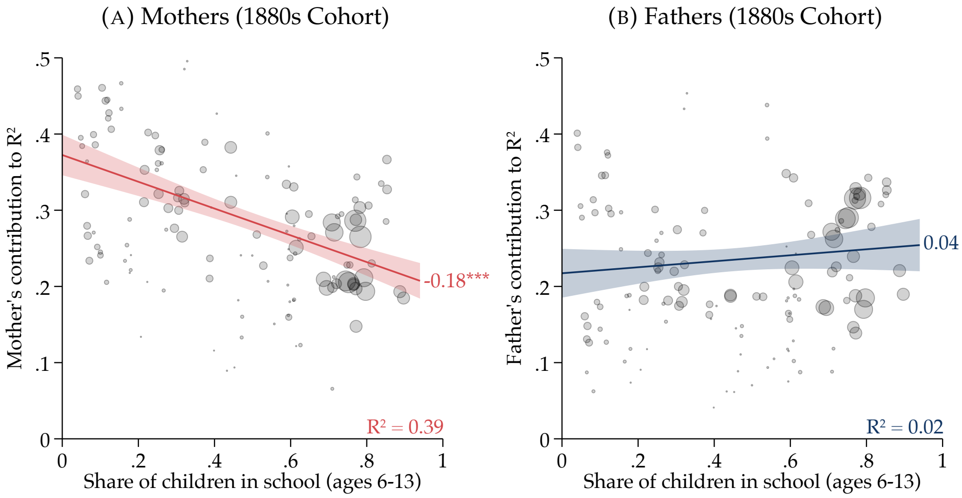
The previous section showed that mothers' human capital is more predictive of their child's human capital than fathers'. This section examines whether mothers' disproportionate importance can be explained by their historical role in home education. We correlate the predictive power of mother's human capital with local school access. Consistent with the role of mothers as home educators, we find that the predictive power of maternal (but not paternal) human capital was substantially greater for groups with limited access to schools.

6.1 Schools and the Rise of Human Capital Mobility

Historians have highlighted mothers' important role in educating their children in the 19th century (Kaestle and Vinovskis, 1978; Margolis, 1984; Dreilinger, 2021). While the spread of school access around 1900 was rapid, it was highly unequal. Specifically, Black children and girls were slower to gain access than white boys. "When public schools did open up to girls, they were sometimes taught a different curriculum from boys and had fewer opportunities for secondary or higher education" (Kober and Rentner, 2020). Similarly, schools for Black children had drastically lower quality than schools for white children (Card and Krueger, 1992; Althoff and Reichardt, 2023).

Consistent with mothers' importance in home schooling, mothers are more predictive of child outcomes in areas with limited school access (see Figure III.8). Maternal human capital explains almost 40 percent of variation in child human capital when school access is minimal, and around 20 percent when school access is universal. Conversely,

FIGURE III.8: Mothers' Human Capital as Substitute for Local Schools



Notes: This figure shows the relationship between local school access and parental contributions to child human capital. We compute the share of the variance in a child's (latent) human capital rank explained by parents' (latent) human capital ranks (R^2) across cohorts and groups. We recover human capital rank-rank transmission using information on literacy and the latent variable method introduced in section 3.4. Panels A and B respectively show mothers' and fathers' contributions to the overall R^2 using the Shapley-Owen method. Each dot represents a group of children born in the 1880s, categorized by race, sex, and state. Sample size weights are applied. School access is determined by the race- and sex-specific share of children aged 6–13 in school.

fathers' contribution was lower and showed no correlation with school access. In fact, the contributions of mothers and fathers were comparable only when school access was universal.

As school access expanded, it diminished the disparities in human capital mobility previously observed among groups with varying levels of school access (see Panel B of Figure IV.68). The reduced influence of parental human capital with improved public school access aligns with Biasi (2023), who shows that equalizing school resources can reduce disparities in intergenerational mobility.

Our analysis reveals a stronger correlation between school access and human capital mobility when refining our measure of school access to reflect children's daily attendance rate. By digitizing data on state-specific school ages, enrollment, attendance, and term lengths from the 1880s Census Statistical Abstracts, we calculate the percentage of children aged 6 to 16 attending school on any given day within each state. This refined measure shows that disparities in school access explain nearly 60 percent of the variation in mothers' contributions to human capital transmission (see Appendix Table IV.30). Conversely, we observe no correlation between fathers' contributions and school access.

In sum, our results suggest that broadening school access in the late 19th and early 20th century contributed to increasing intergenerational mobility. The increase in mobility was driven by a declining role of maternal human capital as schools substituted for home-education. The critical role of schools in increasing intergenerational mobility is consistent with Card et al. (2022) who show that state-level school quality are correlated with higher educational upward mobility in the 1940 census, and with more modern work on the role of education in intergenerational mobility (Chetty et al., 2020; Barrios Fernández et al., 2021; Zheng and Graham, 2022; Black et al., 2023).

7 Conclusion

This paper studies the influence of maternal and paternal background on child outcomes in the US from 1850 to 1940, emphasizing the role of maternal human capital. We construct a representative panel that includes women in early US history, introduce the R^2 mobility measure to accommodate multiple parental inputs, leverage advanced statistical techniques to analyze intergenerational transmission under data constraints, and separate the impact of maternal and paternal inputs. Our findings highlight the significant influence of maternal human capital on children's outcomes, particularly for daughters and Black children. We propose that gaps in school access can explain why the importance of mothers' human capital for child outcomes varies across race, location, and time.

There are several promising avenues for future research. We expanded the parental status measurement to separately encompass maternal and paternal roles. Future research could integrate broader parental background measures like wealth or social norms or consider the role of other relatives including grandparents. Given the importance of the location in which a person grows up—as documented in previous work (e.g., Chetty et al., 2016; Chetty and Hendren, 2018)—future research could also use the R^2 mobility metric to factor in neighborhood quality alongside parental background. Another promising avenue for future work would be to assess changes in maternal transmission of economic outcomes over the 20th century, especially amid rising female labor participation (Goldin, 1977, 1990, 2006; Olivetti, 2014) and single-motherhood (Althoff, 2023).

Lastly, our new panel dataset serves as a foundation for future work on the role of women in shaping US history. Future researchers may find this dataset helpful to reevaluate questions that require panel data but have been studied exclusively for men, as well as to consider new questions that focus specifically on women.

Bibliography

ABRAHAMSSON, S., A. BÜTIKOFER, K. V. LØKEN, AND M. PAGE (2024): "Sources of Generational Persistence in Education and Income," Working paper.

ABRAMITZKY, R., L. BOUSTAN, E. JACOME, AND S. PEREZ (2021a): "Intergenerational Mobility of Immigrants in the United States over Two Centuries," *American Economic Review*, 111, 580–608.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2021b): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

ALTHOFF, L. (2023): "Two Steps Forward, One Step Back: Racial Income Gaps among Women since 1950," Working Paper.

ALTHOFF, L. AND H. REICHARDT (2023): "Jim Crow and Black Economic Progress After Slavery," Working Paper.

BAILEY, M. J., P. Z. LIN, S. MOHAMMED, P. MOHNEN, J. MURRAY, M. ZHANG, AND A. PRETTYMAN (2022): "LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database," dataset: `https://doi.org/10.3886/E155186V2`.

BARRIOS FERNÁNDEZ, A., C. NEILSON, AND S. D. ZIMMERMAN (2021): "Elite universities and the intergenerational transmission of human and social capital," *Available at SSRN 4071712*.

BECKER, G. S., S. D. KOMINERS, K. M. MURPHY, AND J. L. SPENKUCH (2018): "A Theory of Intergenerational Mobility," *Journal of Political Economy*, 126.

BEECHER, C. E. (1842): *Treatise on Domestic Economy*, Boston: T. H. Webb, & Co.

BETTINGER, E. AND B. LONG (2005): "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *American Economic Review*, 95, 152–157.

BIASI, B. (2023): "School Finance Equalization Increases Intergenerational Mobility," *Journal of Labor Economics*, 41, 1–38.

BIASI, B. AND S. MA (2023): "The Education-Innovation Gap," Working Paper 29853, National Bureau of Economic Research.

BLACK, S. E., J. T. DENNING, AND J. ROTHSTEIN (2023): "Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes," *American Economic Journal: Applied Economics*, 15, 26–67.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, 95, 437–449.

BUCKLES, K., A. HAWS, J. PRICE, AND H. WILBERT (2023a): "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Working Paper.

BUCKLES, K., J. PRICE, Z. WARD, AND H. WILBERT (2023b): "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men," Working Paper.

CARD, D., C. DOMNISORU, AND L. TAYLOR (2022): "The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility," *Journal of Labor Economics*, 40, S1–S493.

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

CENTERS FOR DISEASE CONTROL AND PREVENTION (2023): "Live Births, Birth Rates, and Fertility Rates, by Race of Child: United States, 1909-80," dataset: `https://www.cdc.gov/nchs/data/statab/t1x0197.pdf`.

CHETTY, R., J. N. FRIEDMAN, E. SAEZ, N. TURNER, AND D. YAGAN (2020): "Income segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 135, 1567–1633.

CHETTY, R. AND N. HENDREN (2018): "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 133, 1107–1162.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment," *American Economic Review*, 106, 855–902.

CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014a): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 129, 1553–1623.

CHETTY, R., N. HENDREN, P. KLINE, E. SAEZ, AND N. TURNER (2014b): "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review Papers and Proceedings*, 104, 141–147.

CRAIG, J., K. A. ERIKSSON, AND G. T. NIEMESH (2019): "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910," Working Paper.

DEY, D. AND V. ZIPUNNIKOV (2022): "Semiparametric Gaussian Copula Regression modeling for Mixed Data Types (SGCRM)," Working Paper.

DREILINGER, D. (2021): *The Secret History of Home Economics: How Trailblazing Women Harnessed the Power of Home and Changed the Way We Live*, W.W. Norton & Company.

ESPÍN-SÁNCHEZ, J.-A., J. P. FERRIE, AND C. VICKERS (2023): "Women and the Econometrics of Family Trees," Working Paper 31598, National Bureau of Economic Research, Cambridge, MA.

FAN, J., H. LIU, Y. NING, AND H. ZOU (2017): "High dimensional semiparametric latent graphical model for mixed data," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79, 405–421.

FEIGENBAUM, J. J. (2018): "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *The Economic Journal*, 128, F446–F481.

FERNÁNDEZ, R. (2013): "Cultural change as learning: The evolution of female labor force participation over a century," *American Economic Review*, 103, 472–500.

FERNÁNDEZ, R., A. FOGLI, AND C. OLIVETTI (2004): "Mothers and Sons: Preference Formation and Female Labor Force Dynamics*," *The Quarterly Journal of Economics*, 119, 1249–1299.

FERRIE, J. P. (2005): "History lessons: The end of American exceptionalism? Mobility in the United States since 1850," *Journal of Economic Perspectives*, 19, 199–215.

FOGLI, A. AND L. VELDKAMP (2011): "Nature or nurture? Learning and the geography of female labor force participation," *Econometrica*, 79, 1103–1138.

FOURREY, K. (2023): "A Regression-Based Shapley Decomposition for Inequality Measures," *Annals of Economics and Statistics*, 39–62.

GARCÍA, J. L. AND J. J. HECKMAN (2023): "Parenting Promotes Social Mobility Within and Across Generations," *Annual Review of Economics*, 15, 349–388.

GOLDIN, C. (1977): "Female labor force participation: The origin of black and white differences, 1870 and 1880," *Journal of Economic History*, 87–108.

——— (1990): *Understanding the gender gap: An economic history of American women*, Oxford University Press.

——— (2006): "The quiet revolution that transformed women's employment, education, and family," *American economic review*, 96, 1–21.

GREENWOOD, J., A. SESHADRI, AND M. YORUKOGLU (2005): "Engines of Liberation," *The Review of Economic Studies*, 72, 109–133.

HUETTNER, F. AND M. SUNDER (2011): "Decomposing R^2 with the Owen value," Working paper.

JÁCOME, E., I. KUZIEMKO, AND S. NAIDU (2021): "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century," Working Paper 29289, National Bureau of Economic Research.

KAESTLE, C. F. AND M. A. VINOVSKIS (1978): "From Apron Strings to ABCs: Parents, Children, and Schooling in Nineteenth-Century Massachusetts," *American Journal of Sociology*, 84, S39–S80, supplement: Turning Points: Historical and Sociological Essays on the Family.

KOBER, N. AND D. S. RENTNER (2020): "History and Evolution of Public Education in the US," Online report: `https://files.eric.ed.gov/fulltext/ED606970.pdf`.

KUHN, A. L. (1947): *The Mother's Role in Childhood Education: New England Concepts 1830-1860*, Yale University Press.

LEIBOWITZ, A. (1974): "Home Investments in Children," in *Economics of the Family: Marriage, Children, and Human Capital*, ed. by T. W. Schultz, University of Chicago Press, 432–456.

LIU, H., F. HAN, M. YUAN, J. LAFFERTY, AND L. WASSERMAN (2012): "High-dimensional semiparametric Gaussian copula graphical models," *Annals of Statistics*, 40, 2293–2326.

LIU, H., J. LAFFERTY, AND L. WASSERMAN (2009): "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs." *Journal of Machine Learning Research*, 10.

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

LUNDBERG, S. M. AND S.-I. LEE (2017): "A Unified Approach to Interpreting Model Predictions," Working Paper.

MARGOLIS, M. L. (1984): *Mothers and Such: Views of American Women and Why They Changed*, Berkeley and Los Angeles: University of California Press.

NGAI, R., C. OLIVETTI, AND B. PETRONGOLO (2024): "Structural Transformation over 150 years of Women's and Men's Work," *Unpublished Working Paper*.

OLIVETTI, C. (2006): "Changes in Women's Hours of Market Work: The Role of Returns to Experience," *Review of Economic Dynamics*, 9, 557–587.

——— (2014): *The Female Labor Force and Long-Run Development: The American Experience in Comparative Perspective*, University of Chicago Press, 161–197.

OLIVETTI, C. AND M. D. PASERMAN (2015): "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940," *American Economic Review*, 105, 2695–2724.

OLIVETTI, C., M. D. PASERMAN, AND L. SALISBURY (2018): "Three-generation mobility in the United States, 1850–1940: The role of maternal and paternal grandparents," *Explorations in Economic History*, 70, 73–90.

OLIVETTI, C., E. PATACCHINI, AND Y. ZENOU (2020): "Mothers, Peers, and Gender-Role Identity," *Journal of the European Economic Association*, 18, 266–301.

OWEN, G. (1977): "Values of games with a priori unions," in *Essays in Mathematical Economics and Game Theory*, ed. by R. Heim and O. Moeschlin, New York: Springer.

PUCKETT, C. (2009): "The Story of the Social Security Number," *Social Security Bulletin*, 69.

RAMEY, V. A. (2009): "Time Spent in Home Production in the Twentieth-Century United States: New Estimates from Old Data," *The Journal of Economic History*, 69, 1–47.

Redding, S. J. and D. E. Weinstein (2023): "Accounting for Trade Patterns," Working paper.

Redell, N. (2019): "Shapley Decomposition of R-Squared in Machine Learning Models," Working Paper.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2020): "IPUMS USA: Version 10.0," dataset: `https://doi.org/10.18128/D010.V10.0`.

Saavedra, M. and T. Twinam (2020): "A machine learning approach to improving occupational income scores," *Explorations in Economic History*, 75, 101304.

Shapley, L. (1953): "A value for n-person games," in *Contributions to the Theory of Games*, ed. by H. Kuhn and A. Tucker, Princeton University Press, vol. 2.

Social Security Administration (2023): "Number of Social Security card holders born in the U. S. by year of birth and sex," dataset: `https://www.ssa.gov/oact/babynames/numberUSbirths.html`.

Song, X., C. G. Massey, K. A. Rolf, J. P. Ferrie, J. L. Rothbaum, and Y. Xie (2020): "Long-term decline in intergenerational mobility in the United States since the 1850s," *PNAS*, 117, 251–258.

Ward, Z. (2023): "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," *American Economic Review*, 113, 3213–3248.

Young, H. P. (1985): "Monotonic solutions of cooperative games," *International Journal of Game Theory*, 14, 65–72.

Zheng, A. and J. Graham (2022): "Public education inequality and intergenerational mobility," *American Economic Journal: Macroeconomics*, 14, 250–282.

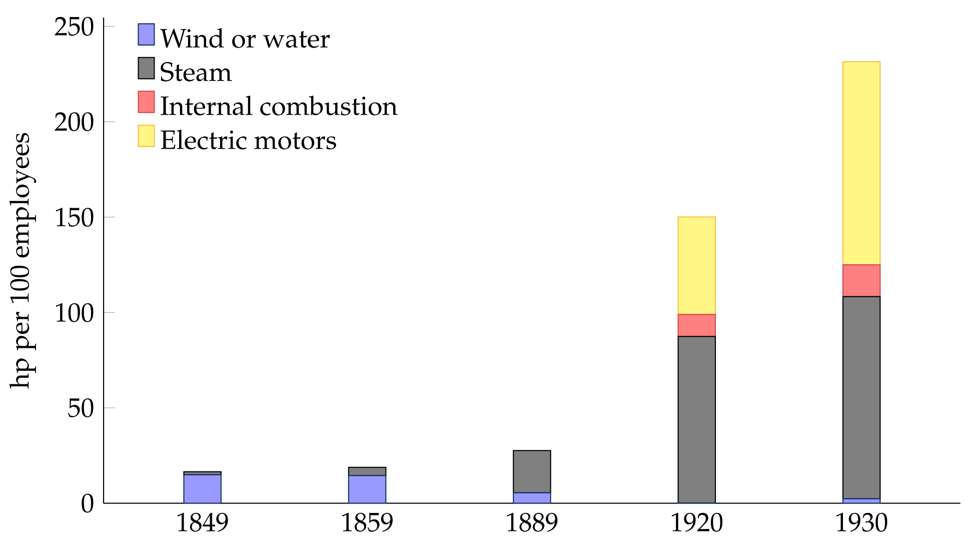
Zue, L. and H. Zou (2012): "Regularized Rank-Based Estimation of High-Dimensional Nonparanormal Graphical Models," *The Annals of Statistics*, 40, 2541–2571.

IV **Appendix**

1 Appendix to "Scale-Biased Technical Change and Inequality"

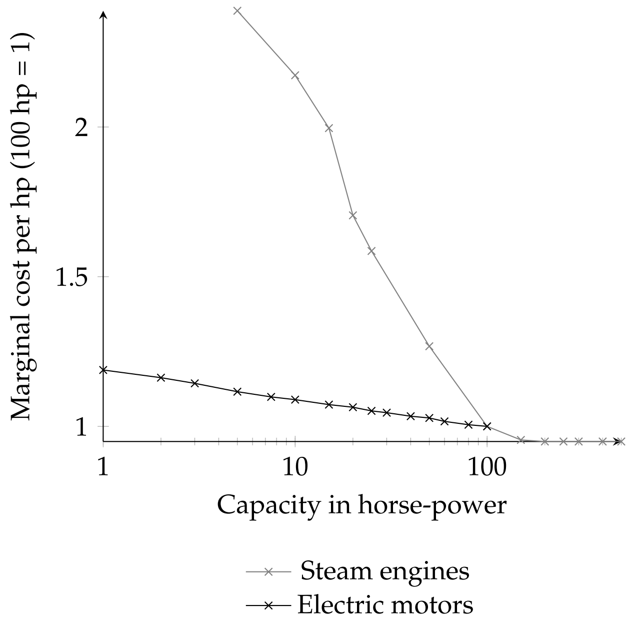
1.1 Figures

FIGURE IV.1: Capacity of primary power by type in horsepower per 100 employees in manufacturing in the Netherlands



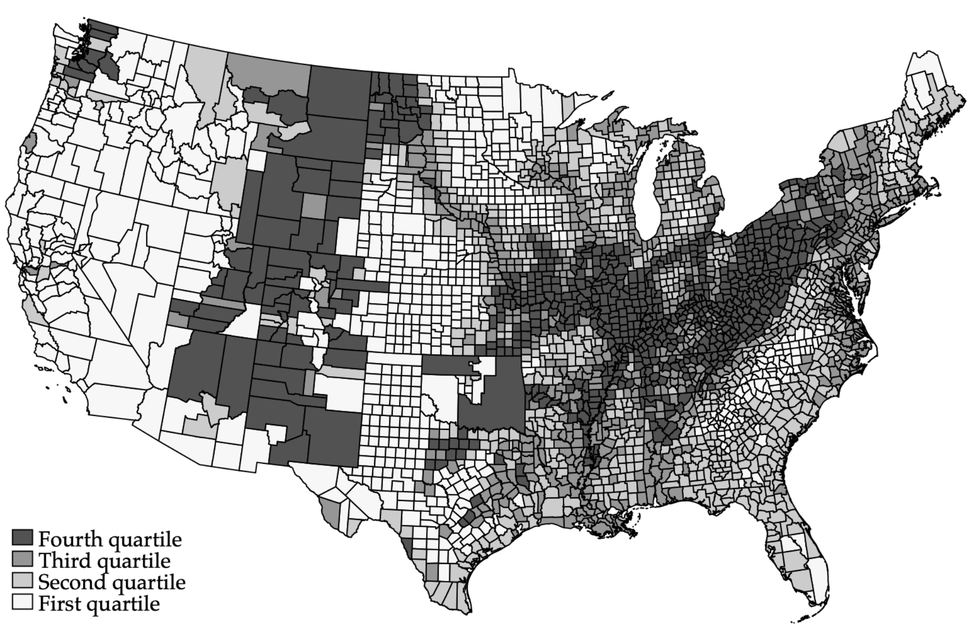
Notes: Electric motors refer to primary electric motors, i.e., electric motors driven by purchased electricity, only. *Sources*: (Blanken and Lintsen, 1981, Table 8) for primary power by type, (Statistics Netherlands, 2001) for employment in manufacturing.

FIGURE IV.2: Marginal cost of steam engines and electric motors of different capacities relative to its 100-horse power equivalent



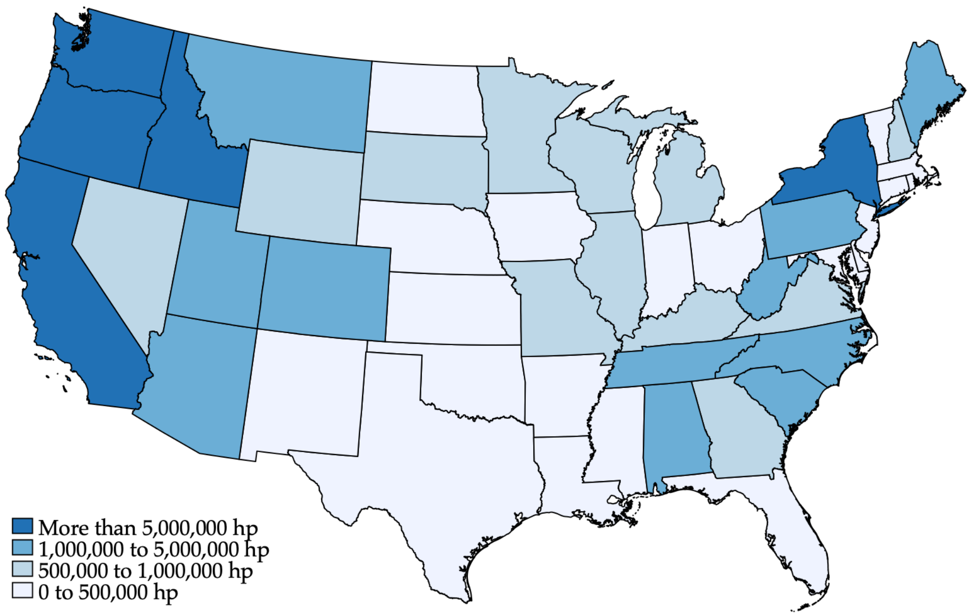
Sources: (Emery, 1883) for coal per horse-power in steam engines; (Bolton, 1926) for full load efficiency of squirrel-cage induction motor.

FIGURE IV.3: Coal access by county



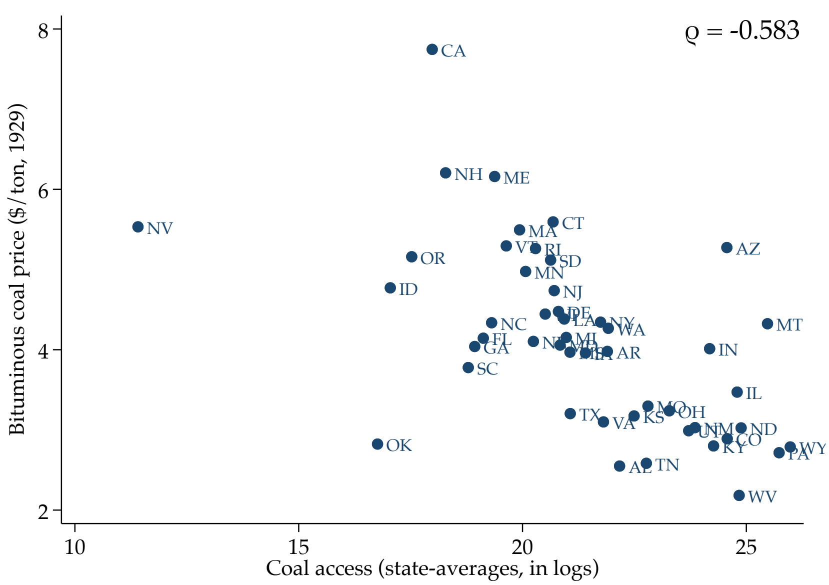
Notes: Coal access is defined in equation (I.16). *Sources:* US Geological Survey, Coal Resources Data System for the coal resources by county. Donaldson and Hornbeck (2016) for transportation costs by county-pair.

FIGURE IV.4: Potential waterpower in horsepower available 50 percent of the time



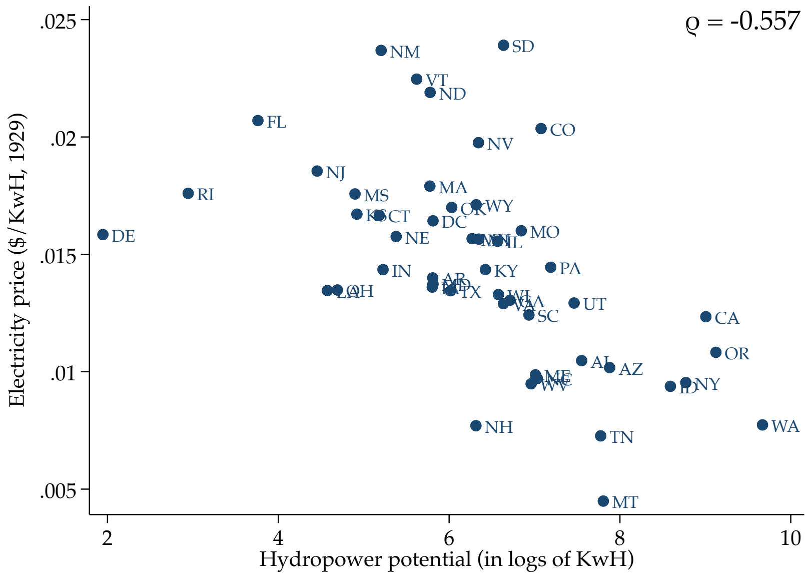
Source: US Geological Survey, (Young, 1964, Table 10).

FIGURE IV.5: Correlation between coal access and coal prices



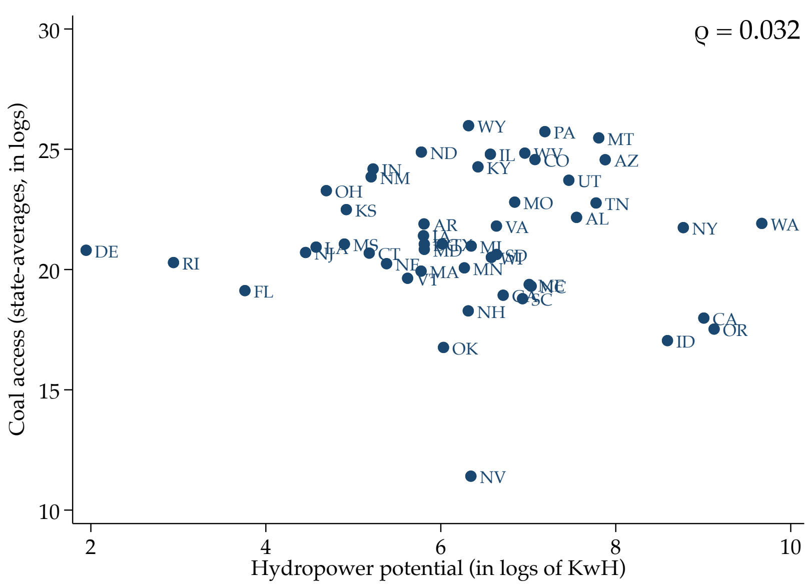
Sources: coal access: National Coal Resources Data System, US Geological Survey and Donaldson and Hornbeck (2016) for transportation costs by county-pair; coal prices: Census of Manufactures, 1929.

FIGURE IV.6: Correlation between hydropower potential and electricity prices



Sources: hydropower potential: US Geological Survey, (Young, 1964, Table 10); electricity prices; Census of Manufactures 1929.

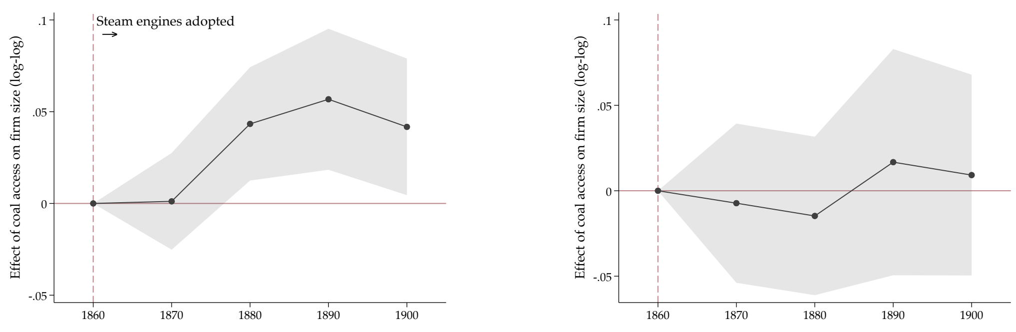
FIGURE IV.7: Correlation between coal access and hydropower potential



Sources: for hydropower potential: US Geological Survey, (Young, 1964, Table 10); for coal access: US Geological Survey, Coal Resources Data System.

FIGURE IV.8: Heterogeneous effects of coal access across industries

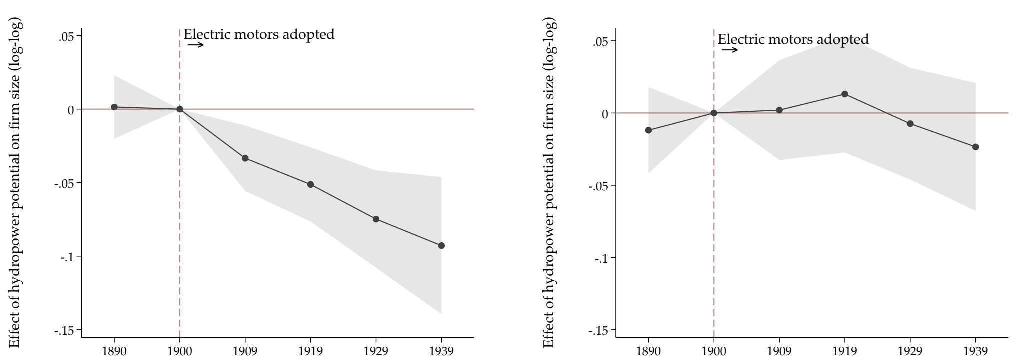
(A) Adopting industries (B) Placebo industries



Notes: This figure shows estimated of the reduced form effects of coal access. Panel (A) shows the effect estimated on a subset of industries that adopt any power nationally in 1890 (measured as being above the 25th percentile in share of establishments reporting the use of power). Panel (B) shows the effect estimated on "placebo" industries, those below the 25th percentile in terms of power use. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the state-level.

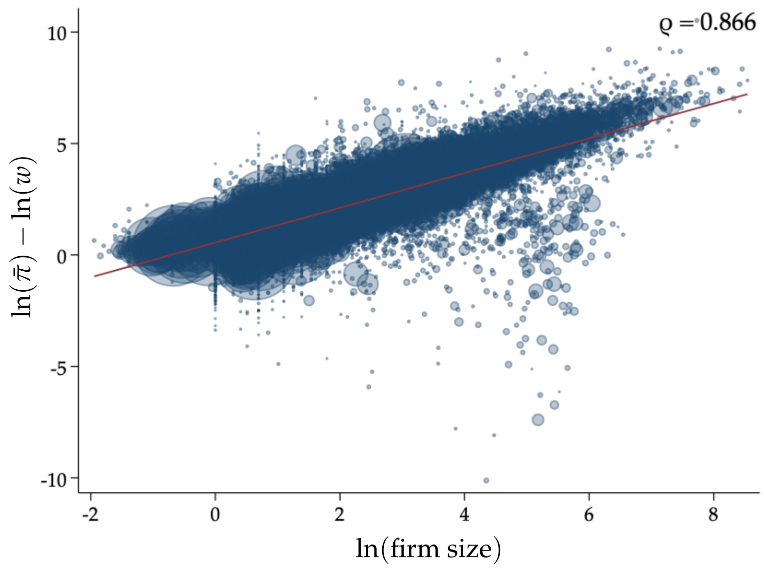
FIGURE IV.9: Heterogeneous effects of hydropower potential across industries

(A) Adopting industries (B) Placebo industries



Notes: This figure shows estimated of the reduced form effects of hydropower potential. Panel (A) shows the effect estimated on a subset of industries that adopt electric motors nationally in 1939 (measured as being above the 25th percentile in share of fuel costs that is electric in 1939). Panel (B) shows the effect estimated on "placebo" industries, those below the 25th percentile in terms of electric motor adoption. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the state-level.

FIGURE IV.10: Free entry condition: correlation between profit-wage ratio and firm size

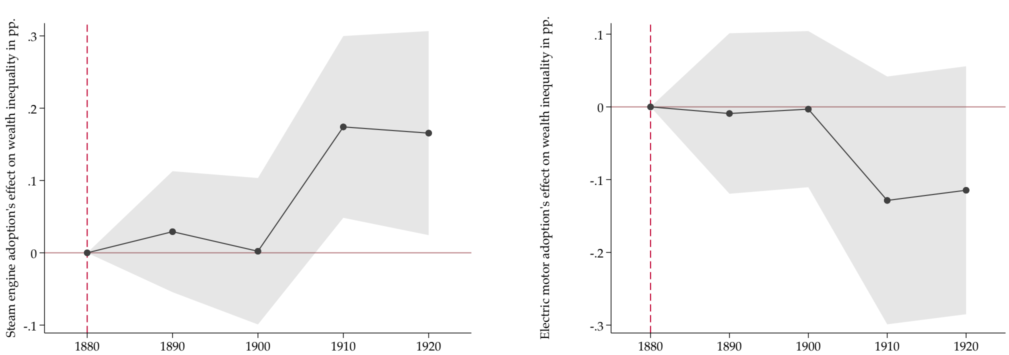


Notes: This figure shows the correlation of the firm size and the ratio between average profits and wages by industry (each in logs). Each dot is an industry-state-year combination. Average profits are approximated by dividing total output minus cost of raw materials and labor costs by the number of establishments. The wage rate is approximated by dividing total wage costs by employment.

FIGURE IV.11: Steam engine adoption increased wealth inequality, electric motors did not

(A) Effect of steam engines

(B) Effect of electric motors

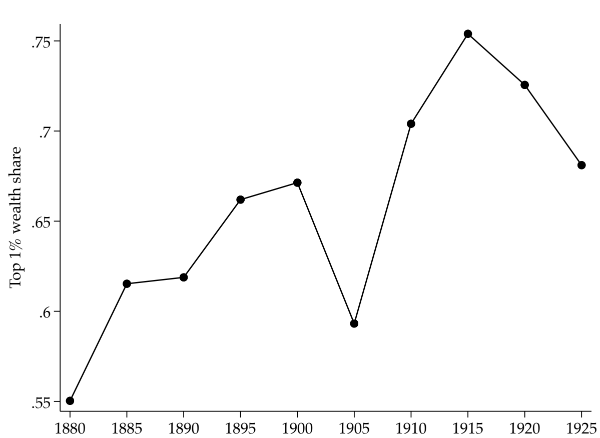


Notes: This figure shows the estimated effects in percentage points of steam engine (in panel A) and electric motor adoption (in panel B) on within-municipality top wealth inequality for each decade relative to 1880. The econometric specifications are detailed in equations (I.25) and (I.26). Observations are weighted by the number of individuals on which the inequality measure is based. Shaded areassent 95% confidence intervals.

FIGURE IV.12: Steam engine adoption relative to electric motor adoption increased wealth inequality.

Notes: This figure shows the estimated effects in percentage points of steam engine adoption on within-municipality top wealth inequality for each decade relative to 1880 relative to electric motor adoption. Observations are weighted by the number of individuals on which the inequality measure is based. Shaded areassent 95% confidence intervals.

FIGURE IV.13: Top 1% wealth share in Enschede, Netherlands



Notes: This figure shows the share of wealth held by the top 1% of decedents aged 20 and over in Enschede between 1879 and 1919. For each year, wealth inequality is computed from the sample of decedents in a 10-year window around it.

1.2 Tables

TABLE IV.1: Little effect of coal access on overall power use (1890)

	Water hp per worker (asinh)			Total hp per worker (asinh)		
Coal access (logs)	-0.030**	-0.028**	-0.037***	-0.001	-0.001	-0.005
	(0.013)	(0.013)	(0.012)	(0.006)	(0.006)	(0.006)
Hydropower potential (logs)		0.017	0.016**		0.002	0.002
		(0.010)	(0.008)		(0.006)	(0.004)
Market access (logs)			X			X
Observations	4237	4237	4237	4237	4237	4237

Notes: This table shows the estimated effect of coal access (in logs) on horsepower of adopted water wheels and total horsepower per employee. Standard errors in parentheses are clustered at the state-level. Industry fixed-effects included. * p< 0.10, ** p< 0.05, *** p< 0.01.

TABLE IV.2: The effect of steam engine and electric motor adoption on the profit-wage ratio

	$\Delta \ln \left(\frac{\text{average profits}_{is}}{\text{wage}_{is}} \right)$					
	1860-1890			1900-1939		
$\text{STEAM}_{is,1890}$	1.134**	1.297**	1.020*			
	(0.529)	(0.533)	(0.512)			
$\text{ELECTRICITY}_{is,1939}$				-0.543**	-0.524**	-0.474*
				(0.250)	(0.250)	(0.254)
$\Delta \ln(\text{population density}_s)$		X	X		X	X
$\Delta \ln(\text{income/wealth p.c.}_s)$			X			X
Observations	1869	1869	1869	1935	1935	1935
Kleibergen-Paap F-stat.	42.8	33.4	24.8	6.6	6.4	5.8

Notes: This table shows the estimated effects of steam engine and electric motor adoption on the change in the log profit-wage ratio in a given state and industry. The explanatory variables are the inverse hyperbolic sine of steam engine horse power in 1890 and megawatt-hour of purchased electricity per worker in 1939. The adoption variables are instrumented with coal access (first three columns) and hydropower potential (last three columns). Observations are weighted by the number of establishments in the base year. Standard errors in parentheses are clustered at the state-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE IV.3: Adoption rates by 2-digit ISIC industry in 1930

ISIC	Name	$STEAM_{1930,i}$	$ELEC_{1930,i}$	Employment
11	Beverages	0.50	0.44	4374
13	Textiles	0.50	0.47	44750
19	Coke and petroleum	0.47	0.42	1129
17	Paper and paper products	0.40	0.57	11000
24	Basic metals	0.35	0.64	6305
23	Other non-metallic mineral products	0.33	0.56	22733
20	Chemicals and chemical products	0.32	0.64	11558
21	Pharmaceuticals	0.29	0.64	1126
22	Rubber and plastics products	0.27	0.71	2540
16	Wood and wood products	0.25	0.40	19081
10	Food products	0.24	0.62	103220
28	Machinery and equipment n.e.c.	0.16	0.82	5313
27	Electrical equipment	0.16	0.84	22380
33	Repair and installation of machinery	0.08	0.89	7030
30	Other transport equipment	0.07	0.87	18723
25	Fabricated metal products	0.07	0.80	34951
15	Leather and related products	0.04	0.40	26855
18	Printing	0.03	0.92	31740
31	Furniture	0.03	0.68	12820
32	Other manufacturing	0.01	0.63	7163
26	Computer and electronic products	0.01	0.32	3748
12	Tobacco products	0.01	0.65	21160
14	Wearing apparel	0.00	0.37	53939

Source: Dutch Census of Companies 1930.

TABLE IV.4: First stage: pre-industrial exposure and technology adoption

	$STEAM_{1930,m}$	$ELECTR_{1930,m}$
$STEAM\_EXP_{1816,m}$	0.535***	
	(0.061)	
$ELECTR\_EXP_{1816,m}$		0.497***
		(0.088)
Constant	0.043*	0.254***
	(0.023)	(0.046)
Observations	835	835

Standard errors in parentheses. Observations are weighted by total manufacturing employment in 1930.

* p< 0.10, ** p< 0.05, *** p< 0.01.

TABLE IV.5: The effect of steam engine and electric motor adoption on the change in income inequality (1946 - 1883)

	ΔINC_INEQUALITY$_{1946,1883}$			
	OLS		IV	
STEAM$_{1930,m}$	0.118**		0.353***	
	(0.052)		(0.120)	
ELECTRICITY$_{1930,m}$		-0.072		-0.876*
		(0.062)		(0.458)
Observations	82	82	78	78
C-D Wald F-stat			24.549	4.895

Notes: Standard errors in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. This table shows the estimated effects of steam engine and electric motor adoption on the change of within-municipality top income inequality between 1946 and 1883. Exposure is computed on the basis of the local industry composition in 1816 and adoption rates by industry in 1930. Observations are weighted by the number of individuals on which the inequality measure in 1946 is based.

1.3 Model appendix

Proof of Proposition 1. I prove Proposition 1 by proving its elements (a) to (c) sequentially. **Proposition 1(a):** Recall that optimal technology adoption implies that the profit gain of adopting a higher fixed, lower marginal, cost relative to a lower fixed, higher marginal cost technology is increasing in productivity ψ. Formally, $\Delta\pi_{jk}(\psi)$ (defined in equation (I.6)) is strictly increasing in ψ if $\kappa_j > \kappa_k$ and $\alpha_j < \alpha_k$. This implies that the least productive entrepreneur uses technology with the highest marginal and lowest fixed cost of all adopted technologies. Also, the least productive entrepreneur has productivity ψ equal to the lowest zero-profit cut-off of all available technologies, $\min_{j \in 1,2,..,J} \bar{\psi}_j$. From equation (I.8), technology t_j is the lowest zero-profit cut-off technology if and only if

$$\alpha_j \kappa_j^{\frac{1}{\sigma-1}} = \min_{k \in \{1,2,..,J\}} \left\{ \alpha_k \kappa_k^{\frac{1}{\sigma-1}} \right\}.$$

The marginal entrepreneur is indifferent between any two technologies t_j and t_k such that $\bar{\psi}_j = \bar{\psi}_k = \bar{\psi}$ as they both give them zero-profit. But since $\Delta\pi_{jk}(\psi)$ in (I.6) is strictly increasing, any entrepreneur with $\psi > \bar{\psi}$ would strictly prefer the technology with higher fixed cost and lower marginal cost. Therefore, out of any technology t_j that minimizes $\bar{\psi}_j$, only the technology with lowest marginal cost is adopted (in the sense of having a strictly positive probability measure of entrepreneurs adopting the technology).

Proposition 1(b): Note that $\Delta\pi_{jk}(\psi) \to \infty$ in (I.6) as $\psi \to \infty$ if and only if $\alpha_j < \alpha_k$. This means that if the marginal cost of a technology is lower than that of any another, there exists a productivity level high enough such that it is profitable to adopt this technology. The assumption that the productivity distribution has semi-infinite support implies that for any $C > 0$, $\Pr(\psi > C) > 0$. Therefore, there always exists a strictly positive share of households that adopt the technology with lowest marginal cost. Note that is true regardless of the fixed cost. Of course, in case there is more than one technology that minimizes marginal cost, the technology with lowest fixed costs amongst those will be adopted. Since no technology can be adopted that is trivially dominated, this must also be the adopted technology with highest fixed cost.

Proposition 1(c): A technology t_j with fixed cost κ_j such that $\kappa_1^* < \kappa_j < \kappa_{J^*}^*$ is adopted if and only if there exists a $\psi > \psi_m$ for which it 1) dominates all technologies with lower fixed costs, 2) dominates all technologies with higher fixed cost, and 3) yields positive profits. Note that condition is 3) is redundant given condition 1) since it can only dominate technology t_1^* if $\psi > \bar{\psi}$ and t_1^* yields positive profits for all $\psi > \bar{\psi}$. Also, recall that technologies in T are arranged in order of increasing fixed costs ($\kappa_1 < .. < \kappa_J$) and thus decreasing marginal costs ($\alpha_1 > .. > \alpha_J$). Therefore, technology t_j is adopted if there exists a $\psi > \psi_m$ such that $\Delta\pi_{jk}(\psi) > 0$ for all $k \in \{1, .., j-1\}$ and $\Delta\pi_{jl}(\psi) > 0$ for all $l \in \{j+1, .., J\}$. Using equation (I.6), this yields the following two restrictions:

$$\frac{Y}{\sigma} (\rho\psi)^{\sigma-1} > \frac{\kappa_j - \kappa_k}{\alpha_j^{1-\sigma} - \alpha_k^{1-\sigma}} \text{ for all } k \in \{1, .., j-1\} \text{ and;} \tag{IV.1a}$$

$$\frac{Y}{\sigma} (\rho\psi)^{\sigma-1} < \frac{\kappa_l - \kappa_j}{\alpha_l^{1-\sigma} - \alpha_j^{1-\sigma}} \text{ for all } l \in \{j+1, .., J\} \tag{IV.1b}$$

Hence, for (IV.1a) and (IV.1b) to hold for some $\psi > \bar{\psi}$, it is necessary and sufficient that the lower bound in (IV.1a) is strictly lower than the upper bound in (IV.1b). Thus, technology j is adopted if and only if for all $k \in \{1, .., j-1\}$ and $l \in \{j+1, .., J\}$

$$\frac{\alpha_l^{1-\sigma} - \alpha_j^{1-\sigma}}{\alpha_j^{1-\sigma} - \alpha_k^{1-\sigma}} < \frac{\kappa_l - \kappa_j}{\kappa_j - \kappa_k}.$$

□

Proposition 2 (Closed-form equilibrium). Suppose that the distribution of productivity ψ is Pareto with shape parameter ξ and a minimum productivity level of $\psi_m > 0$ such that $\xi > 1$ and $\xi > \sigma - 1$. Then, the competitive equilibrium is given in closed-form by

$$L = \frac{\xi}{1+\xi} \tag{IV.2}$$

$$\bar{\psi} = \bar{B}(\xi, \sigma, \psi_m)\left(\bar{\kappa}^{\frac{\sigma-2}{\sigma-1}} \alpha_1^* (\kappa_1^*)^{\frac{1}{\sigma-1}}\right)^{\frac{\sigma-1}{\bar{A}(\xi,\sigma)}} \tag{IV.3}$$

$$Y = \bar{C}(\xi, \sigma, \psi_m)\left(\frac{\bar{\kappa}^{\frac{1}{\xi}}}{\alpha_1^* (\kappa_1^*)^{\frac{1}{\sigma-1}}}\right)^{\frac{\xi(\sigma-1)}{\bar{A}(\xi,\sigma)}} \tag{IV.4}$$

$$w = \rho \frac{1+\xi}{\xi} Y \tag{IV.5}$$

$$\bar{\pi} = \rho \frac{1+\xi}{\xi} \bar{C}(\xi, \sigma, \psi_m) \bar{B}(\xi, \sigma, \psi_m)^{\xi} \psi_m^{-\xi} \bar{\kappa} \tag{IV.6}$$

where $\bar{A}(\xi, \sigma)$, $\bar{B}(\xi, \sigma, \psi_m)$, and $\bar{C}(\xi, \sigma, \psi_m)$ are strictly positive functions of the exogenous (non-technological) parameters ξ, σ, and ψ_m:

$$\bar{A}(\xi, \sigma) \equiv (1+\xi)(\sigma - 1) - \xi$$

$$\bar{B}(\xi, \sigma, \psi_m) \equiv \left(\psi_m^{\xi} (1+\xi)^{\frac{1}{\sigma-1}} \frac{\sigma}{\xi - \sigma + 1}\left(\frac{\xi \psi_m^{\xi}}{\xi - \sigma + 1}\right)^{\frac{1}{1-\sigma}}\right)^{\frac{\sigma-1}{\bar{A}(\xi,\sigma)}}$$

$$\bar{C}(\xi, \sigma, \psi_m) \equiv \bar{B}(\xi, \sigma, \psi_m)^{\frac{-\xi(\sigma-1)}{\bar{A}(\xi,\sigma)}} \frac{\psi_m^{\xi} \sigma}{\xi - \sigma + 1} \frac{\xi}{1+\xi}$$

and $\bar{\kappa}$ is the average fixed cost of all producing entrepreneurs:

$$\bar{\kappa} = \begin{cases} \kappa_1^* & \text{if } J^* = 1 \\ \kappa_1^* + \left(\alpha_1^* (\kappa_1^*)^{\frac{1}{\sigma-1}}\right)^{\xi} \sum_{j=2}^{J^*} \left((\alpha_j^*)^{1-\sigma} - (\alpha_{j-1}^*)^{1-\sigma}\right)^{\frac{\xi}{\sigma-1}} \left(\kappa_j^* - \kappa_{j-1}^*\right)^{\frac{\sigma-1-\xi}{\sigma-1}} & \text{if } J^* > 1. \end{cases}$$

Proof of Proposition 2. We first derive the adopting set Ψ_j^* for each technology $t_j^* \in T^*$. Note that we can restrict ourselves to technologies that are adopted in equilibrium (see Proposition 1), since the adopting set is empty otherwise.

By definition, if T^* is a singleton set, then Ψ_1^* is $[\bar{\psi}, \infty)$. Now suppose $J^* \equiv |T^*| > 1$. From equation (I.6), it follows that an entrepreneur with productivity ψ is indifferent between adopting t_j^* and t_{j+1}^* if and only if $G(\psi, t_{j+1}^*, t_j^*) = 0$. Define $\bar{\psi}_{j,j+1}$ implicitly by

$$G(\psi, t_{j+1}^*, t_j^*) = 0$$

which implies that

$$\bar{\psi}_{j,j+1} = \left(\frac{\kappa_{j+1}^* - \kappa_j^*}{(\alpha_{j+1}^*)^{1-\sigma} - (\alpha_j^*)^{1-\sigma}} \right)^{\frac{1}{\sigma-1}} \left(\frac{\sigma}{Y} \right)^{\frac{1}{\sigma-1}} \frac{w}{\rho} = \bar{\psi} \frac{\left(\frac{\kappa_{j+1}^* - \kappa_j^*}{(\alpha_{j+1}^*)^{1-\sigma} - (\alpha_j^*)^{1-\sigma}} \right)^{\frac{1}{\sigma-1}}}{\alpha_j^* (\kappa_j^*)^{\frac{1}{\sigma-1}}}. \tag{IV.7}$$

Since $G(\psi, t_{j+1}^*, t_j^*)$ is increasing in ψ (see proof of Proposition 1(a)), the more productive entrepreneur chooses the technology that entails higher fixed cost. Specifically, an entrepreneur would choose t_{j+1}^* over t_j^* if and only if $\psi > \bar{\psi}_{j,j+1}$. This means that all entrepreneurs with productivity between $\bar{\psi}$ and $\bar{\psi}_{1,2}$ choose t_1^*, all entrepreneurs with productivity between $\bar{\psi}_{1,2}$ and $\bar{\psi}_{2,3}$ choose t_2^*, and so on and so forth. Formally,

$$\begin{cases} \Psi_j^* = [\bar{\psi}, \bar{\psi}_{j,j+1}] & \text{if } j = 1 \\ \Psi_j^* = [\bar{\psi}_{j-1,j}, \bar{\psi}_{j,j+1}] & \text{if } 1 < j < J^* \\ \Psi_j^* = [\bar{\psi}_{j-1,j}, \infty) & \text{if } j = J^* \end{cases}$$

Combining equation (I.8) (definition of $\bar{\psi}$) and equation (I.11) (labor market clearing) with the Pareto assumption, the probability of being an entrepreneur conditional on entry is

$$1 - F(\bar{\psi}) = \psi_m^{\xi} \bar{\psi}^{-\xi} = \frac{L}{1-L} \frac{\xi - \sigma + 1}{\xi(\sigma - 1)} \frac{w}{\bar{\kappa}} \tag{IV.8}$$

where $\bar{\kappa}$, the average fixed cost across producing entrepreneurs, is

$$\bar{\kappa} = \kappa_1 + \left(\alpha_1^* (\kappa_1^*)^{\frac{1}{\sigma-1}} \right)^{\xi} \sum_{j=2}^{J^*} \left((\alpha_j^*)^{1-\sigma} - (\alpha_{j-1}^*)^{1-\sigma} \right)^{\frac{\xi}{\sigma-1}} \left(\kappa_j^* - \kappa_{j-1}^* \right)^{\frac{\sigma-1-\xi}{\sigma-1}}.$$

Also, labor market clearing in (I.11) combined with the aggregate price equation in (I.13), implies that the labor share is constant and independent of technology:

$$\frac{Lw}{Y} = \rho. \tag{IV.9}$$

Combining the constant labor share with equation (IV.8), shows that the share of output devoted to the fixed costs is constant and independent of technology:

$$\frac{(1-L)\psi_m^{\xi} \bar{\psi}^{-\xi} \bar{\kappa}}{Y} = \frac{\xi - \sigma + 1}{\xi \sigma}. \tag{IV.10}$$

Then, by goods market clearing, the profit share must be constant too:

$$\frac{(1-L)\psi_m^{\xi} \bar{\psi}^{-\xi} \bar{\pi}}{Y} = 1 - \rho - \frac{\xi - \sigma + 1}{\xi \sigma} = \frac{\rho}{\xi}. \tag{IV.11}$$

Together with the free entry condition in equation (I.9) and the labor share in equation (IV.9), the constant profit share implies that the share of entrants is constant and independent of technology too:

$$L = \frac{\xi}{1+\xi}. \tag{IV.12}$$

Lastly, the pricing equation in (I.13) combined with the Pareto distribution yields

$$\left(\frac{w}{\rho}\right)^{\sigma-1} = (1-L)\left(\frac{\xi\psi_m^{\xi}}{\xi-\sigma+1}\right)\bar{\psi}^{\sigma-1-\xi}\frac{\bar{\kappa}}{(\alpha_1^*)^{\sigma-1}\kappa_1^*} \tag{IV.13}$$

Equations (IV.8), (IV.9), (IV.12), (IV.13) together lead to the closed-form solutions for L, $\bar{\psi}$, Y, and w in equations (IV.2), (IV.3), (IV.4), and (IV.5), respectively. Lastly, the solution for $\bar{\pi}$, the average profits, in (IV.6) result from equations (IV.3), (IV.4), and (IV.5) together with the free-entry condition in (I.9). □

Lemma 1. Suppose that the assumptions in Proposition 2 (Pareto distribution) hold and that $\sigma > 2$. Then, if a new technology t_{new} is added to the technology set T and it is adopted in equilibrium, it increases output Y, wages w, and total profits $(1-L)\psi_m^{\xi}\bar{\psi}^{-\xi}\bar{\pi}$.

Proof of Lemma 1. Suppose towards contradiction that output Y does not increase. Since Y and wages w are positively linearly related (equation (IV.5)), the profit function can be rewritten as

$$\pi_j(\psi) = \frac{1}{\sigma}\left(\frac{\xi}{1+\xi}\right)^{\sigma-1} Y^{2-\sigma}\left(\frac{\psi}{\alpha_j}\right)^{\sigma-1} - \kappa_j. \tag{IV.14}$$

Given $\sigma > 2$, if Y does not increase, it means that profits can not go down for any productivity level and for any technology choice. Also, given that the technology is adopted, it must yield strictly higher profits for some entrepreneurs. Therefore, total profits must go up. But by equation (IV.11), profits are a fixed share of output. Hence, the increase in total profits implies that output Y *increases*, a contradiction. Therefore, output must increase in response to a new technology that is adopted. Since output, wages, and total profits are positively and linearly related, wages and total profits must also go up in response to an adopted new technology. □

Proof of Proposition 3. I prove Proposition 3 by proving its elements (a) and (b) sequentially.

Proposition 3(a): If t_{new} is adopted and has the highest fixed cost, it must have lowest marginal cost. By the reasoning in Stage 3 (equation (I.6)), this technology is only adopted by the entrepreneurs above a certain threshold for ψ. The entrepreneurs above this threshold reduce their marginal cost and thus increase their total factor productivity.

Because it becomes the highest fixed cost technology, the average fixed cost among producing entrepreneurs, $\bar{\kappa}$, increases.[1] Since $\bar{\kappa}$ increases, the entry threshold $\bar{\psi}$ increases too (seen from equation (IV.3)). Hence, the technical change would lead some entrepreneurs to no longer produce, i.e., decreasing their total factor productivity to 0. It also means that the thresholds above which an entrepreneur uses technology $j+1$ instead of j increase for each j (see equation (IV.7)): at least some entrepreneurs that do not adopt the new technology "downgrade" their technology, because the increased total

[1] To see this formally, note that output increases by Lemma 1. By equation (IV.4), if output increases while the entry technology, i.e. $\alpha_1^*(\kappa_1^*)^{\frac{1}{\sigma-1}}$ remains unchanged, $\bar{\kappa}$ must increase.

output (see Lemma 1) by those using the new technology reduces their demand. Hence, for all entrepreneurs that do not adopt the new technology, the marginal cost either decreases or remains unchanged. This proves that technical change is large-scale-biased *if* the new technology has higher fixed than any other adopted technology.

Now suppose the new technology does not have highest fixed cost of all adopted technologies. Then, entrepreneurs that previously adopted the technology with lowest marginal cost can not decrease their marginal cost. For any $k > \psi_m$, there exists a subset of entrepreneurs with $\psi > k$ that adopts the technology with lowest marginal cost before and after the technical change. Hence, there does not exist a k such that all entrepreneurs with $\psi > k$ strictly increase total factor productivity, so that the technical change can not be large-scale-biased, which proves that technical change can be large-scale-biased *only if* the new technology has higher fixed than any other adopted technology.

Proposition 3(b): If t_{new} is adopted and has the lowest fixed cost, it must have highest marginal cost. First, the entry threshold $\bar{\psi}$ in equation (IV.3) decreases because both κ_1^* (the fixed cost of the lowest adopted fixed-cost technology) and $\bar{\kappa}$ decrease. Therefore, there exists a range of entrepreneurial productivities $[\bar{\psi}_{new}, \bar{\psi}_{old}]$ such that entrepreneurs within that range exited before the technical change and enter after. Therefore, these entrepreneurs increase their total factor productivity from 0 to a strictly positive value. For any $\psi > \psi_{old}$, none chooses a technology that has lower marginal cost than before the technical change, because the increased total output (see Lemma 1) reduces their demand for any given price. Hence, some entrepreneurs with $\psi > \psi_{old}$ "downgrade" their technology relative to before the technical change and others do not change their adoption choice. This proves that technical change is small-scale-biased *if* the new technology has lower fixed than any other adopted technology.

Now suppose the new technology does not have lowest fixed cost of all adopted technologies. By Lemma 1, output and wages increase as a result of the technical change. Also, output and wages are positively linearly related (equation (IV.5)). Thus, if output goes up while the entry technology remains unchanged, $\bar{\psi}$ must increase by equation (I.8). That is, $\bar{\psi}(T_{new}) > \bar{\psi}(T_{old})$. This means the range of entrepreneurs with $\psi \in (\bar{\psi}(T_{old}), \bar{\psi}(T_{new}))$ see their TFP decrease. Hence, such technical change can not be small-scale-biased, which proves that technical change can be small-scale-biased *only if* the new technology has lower fixed than any other adopted technology. □

Proposition 3A (Scale-biased technical change with obsolescence). Suppose that the assumptions in Proposition 2 (Pareto distribution) hold, that $\sigma > 2$ and that $T^*_{new} = \tilde{T}_{old} \cup \{t_{new}\}$ where $\tilde{T}_{old} \subset T^*_{old}$ (the new technology makes at least of one of the previously adopted technologies obsolete), then

(a) the technical change is large-scale-biased if and only if the conditions a.1 and either a.2 or a.3 are satisfied:

(a.1) $\alpha_{new} < \min_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \alpha_j$

(a.2) $\alpha_{new} \kappa_{new}^{\frac{1}{\sigma-1}} > \min_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \alpha_j \kappa_j^{\frac{1}{\sigma-1}}$

(a.3) $\alpha_{new} \kappa_{new}^{\frac{1}{\sigma-1}} \leq \min_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \alpha_j \kappa_j^{\frac{1}{\sigma-1}}$ and $\alpha_{new} \kappa_{new} > \min_{\{\alpha_j, \kappa_j\} \in T^*_{old}} \left[\alpha_j \kappa_j^{\frac{1}{\sigma-1}} \right] \bar{\kappa}_{old}$

(b) the technical change is small-scale-biased if and only if the conditions b.1, b.2, and b.3 are satisfied:

(b.1) $\alpha_{new} \geq \min_{\{\alpha_j,\kappa_j\}\in T^*_{old}} \alpha_j$

(b.2) $\alpha_{new}\kappa_{new}^{\frac{1}{\sigma-1}} < \min_{\{\alpha_j,\kappa_j\}\in T^*_{old}} \alpha_j\kappa_j^{\frac{1}{\sigma-1}}$

(b.3) $\alpha_{new}\kappa_{new}^{\frac{1}{\sigma-1}}\bar{\kappa}_{new} \leq \min_{\{\alpha_j,\kappa_j\}\in T^*_{old}} \alpha_j\kappa_j^{\frac{1}{\sigma-1}}\bar{\kappa}_{old}$

Proof of Proposition 3A. I prove Proposition 3A by proving its elements (a) and (b) sequentially.

Proposition 3A(a): If the new technology satisfies a.1 it is the technology with lowest marginal cost. Therefore, it is adopted by all entrepreneurs above a certain threshold of productivity. This range of entrepreneurs would see TFP increase. If a.2 is true (besides a.1), it means the new technology does not become the entry technology. From there, the same reasoning as in the proof of Proposition 3(a), proves that the technical change is large-scale-biased. If a.3 is true (besides a.1), the new technology becomes the only technology that is adopted in equilibrium by Proposition 1 and the entry threshold increases by Proposition 2. Therefore, every entrepreneur with productivity above the new entry threshold increases TFP, while those below the threshold lose out. That is, $\bar{\psi}(T_{new}) > \bar{\psi}(T_{old})$. This means the range of entrepreneurs with $\psi \in (\bar{\psi}(T_{old}), \bar{\psi}(T_{new}))$ see their TFP decrease, while those with $\psi > \bar{\psi}(T_{new})$ see their TFP increase. This proves that technical change is large-scale-biased if the conditions a.1 and either a.2 or a.3 are satisfied.

To prove that technical change is large-scale-biased only if the conditions a.1 and either a.2 or a.3 are satisfied, now suppose technical change is large-scale-biased. By definition, the new technology increases TFP for all entrepreneurs above a certain productivity threshold $k > \min\{\bar{\psi}(T_{new}), \bar{\psi}(T_{old})\}$. Therefore, the marginal cost of the new technology must be lower than any previously adopted technology, such that a.1 is satisfied. Also, by definition of large-scale bias, TFP does not increase for all entrepreneurs with $\psi < k$. Therefore, if the new technology becomes the only technology that is adopted in equilibrium (such that a.2 is not satisfied), it must be that the entry threshold increases, hence a.3 is satisfied. This proves that technical change is large-scale-biased only if the conditions a.1 and either a.2 or a.3 are satisfied.

Proposition 3A(b): Suppose conditions b.1, b.2, and b.3 are satisfied. Then, because the new technology does not have the lowest marginal cost (b.1), it is not adopted by the most productive entrepreneurs. Because b.2 is satisfied, it is adopted by the least productive entrepreneurs. Because b.3 is satisfied, it reduces the entry threshold (by Proposition 2). Therefore, it increases TFP for a range of entrepreneurs that did not enter before the technical change. If it increases TFP for some $\psi' > \bar{\psi}_{old}$, it also increases TFP for any $\bar{\psi}_{new} > \psi'' > \psi' > \bar{\psi}_{old}$. This can be seen by realizing that the new technology can only increase TFP for ψ' if it is adopted by ψ', in which case it must also be adopted by any entrepreneur with lower productivity (since it is adopted by the marginal entrepreneur). Also, because it is not adopted by the most productive entrepreneurs, there is a productivity threshold above which the new technology is not adopted and therefore does not increase TFP. This proves that technical change is small-scale-biased if the conditions b.1, b.2, and b.3 are satisfied. This proves that technical change is large-scale-biased only if the conditions b.1, b.2, and b.3 are satisfied.

Now suppose technical change is small-scale-biased. Since there exists a productivity threshold above which the technical change does not increase TFP, its marginal cost must not be lower than the lowest marginal cost of any existing technology (b.1). Also, since there exists a productivity threshold below which it increases TFP, it must be adopted by the marginal entrepreneur (so that b.2 is satisfied by Proposition 1). Lastly, again since there exists a productivity threshold below which it increases TFP, it cannot increase the entry threshold (so that b.3 must be satisfied by Proposition 2). □

Proof of Proposition 4. I prove Proposition 4 by proving its elements (a), (b), and (c) sequentially.

Proposition 4(a): If technical change is large-scale-biased, it increases average fixed cost: $\bar{\kappa}_{new} > \bar{\kappa}_{old}$ (see the proof of Proposition 3(a)). Since it increases the average fixed cost without affecting $\alpha_1^*(\kappa_1^*)^{\frac{1}{\sigma-1}}$, it increases $\bar{\psi}$ by equation (IV.3). The average employment by firm is the number of workers divided by the number of entrepreneurs. The number of workers L is constant in equilibrium by equation (IV.2). The number of entrepreneurs is $(1-L)(1-F(\bar{\psi}))$ which is decreasing in $\bar{\psi}$. Therefore, the average employment size increases in response to large-scale-biased technical change.

If technical change is small-scale-biased the entry threshold $\bar{\psi}$ in equation (IV.3) decreases because both κ_1^* (the fixed cost of the lowest adopted fixed-cost technology) and $\bar{\kappa}$ decrease. Therefore, the average employment size decreases in response to large-scale-biased technical change.

Proposition 4(b): By Proposition Proposition 4(a), if technical change is large-scale biased, it increases $\bar{\psi}$. Thus, by the free-entry condition in equation (I.9), it increases the ratio between average profits of producting entrepreneurs and wages. The opposite is true for small-scale-biased technical change.

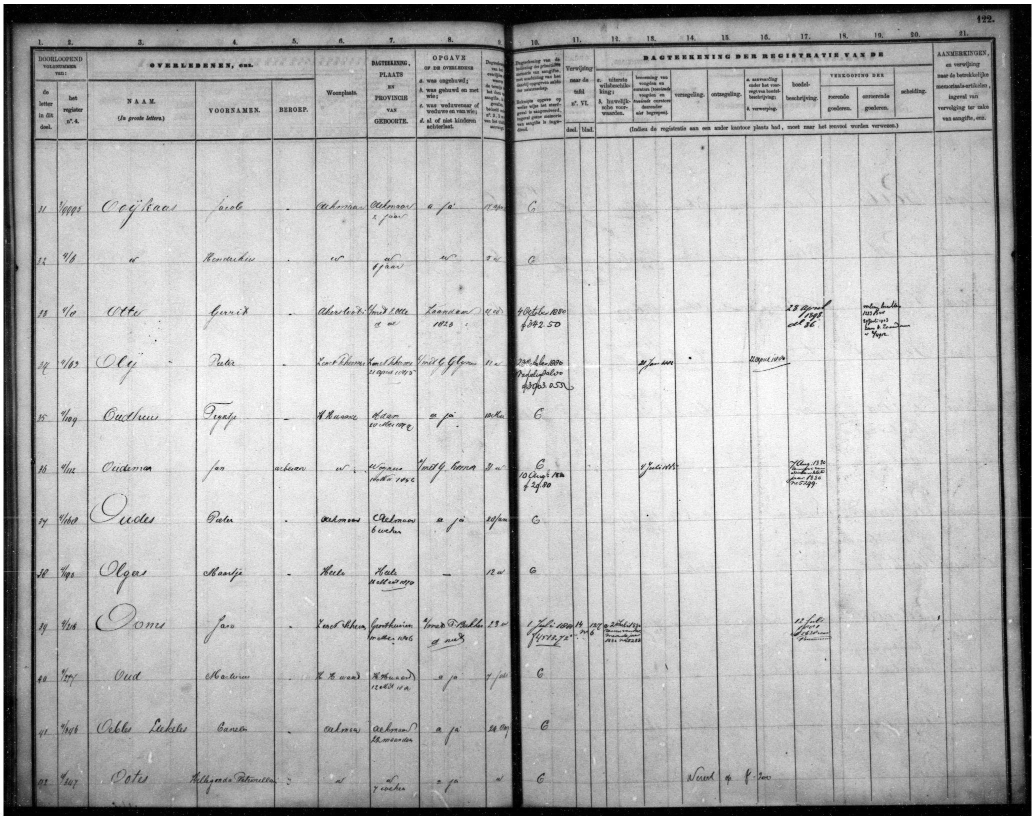
Proposition 4(c): Any entrepreneur that does not adopt the technology, sees a reduction in profits as a result of technical change. This can be seen by noting that equation (IV.14) is decreasing in output Y and output increases when a new technology is added by Lemma 1. If technical change is small-scale-biased, entrepreneurs above a certain productivity threshold do not adopt the technology and their profits must therefore decline. However, total output and wages go up. Hence, there exists a $\bar{k} \in (0, 100)$ such that average income growth of the top k% of incomes is lower than average income growth of the bottom $(100-k)$% of incomes for all $k < \bar{k}$.

If technical change is large-scale-biased, it increases profits of entrepreneurs above a certain productivity threshold, while it decreases profit for those below it. Because the profit share of output is constant (see equation (IV.11)) and wages are a linear function of output (IV.5), total profit growth equates wage growth. Because only adopting entrepreneurs experience a profit increase, while other entrepreneurs' profit decline, their income growth must exceed wage growth. Furthermore, even among adopting entrepreneurs, proportional profit growth is increasing in ψ. Therefore, there exists a $\bar{k} \in (0, 100)$ such that average income growth of the top k% of incomes is higher than average income growth of the bottom $(100-k)$% of incomes for all $k < \bar{k}$. □

1.4 Data appendix

1.4.1 Examples of source files

FIGURE IV.14: Example of a source image of the Dutch inheritance tax files.



Notes: The template form was consistent nationally and over time between 1879 and 1927.

FIGURE IV.15: Example of a source image for the income and wealth distribution by municipality

(A) Income distribution (1946)

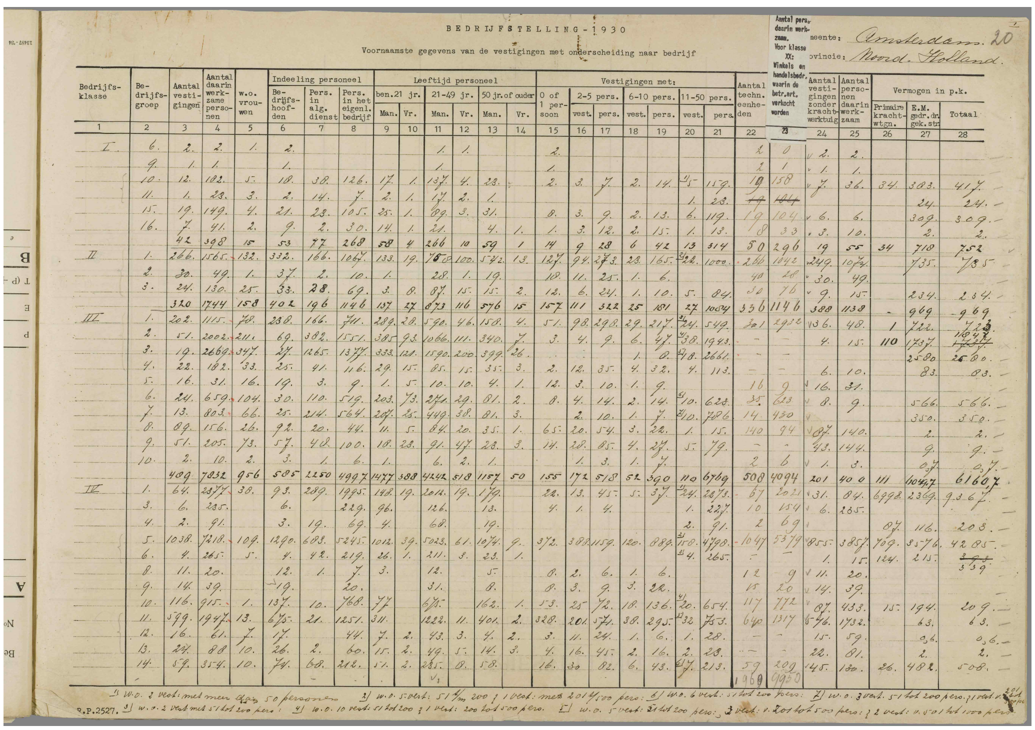
Inkomensklasse (x f. 1 000)	Aagtekerke Z.-43 Aantal	Inkomen
< 1	68	36 377
1 - < 2	100	141 537
2 - < 3	33	72 459
3 - < 4	1) 17	1) 57 864
4 - < 5	18	138 655
5 - < 6	2)	2)
6 - < 7	2)	2)
7 - < 8	2)	2)
8 - < 9	2)	2)
9 - < 10		
10 - < 15	2)	2)
15 - < 20	2)	2)
20 - < 50	2)	2)
50 - < 100		
100 en meer		
Totaal	236.	446 892
Totaal belasting	51 945	
Gem. inkomen:		
per inwoner	679	
per belastingpl.	1 894	

(B) Wealth distribution (1947)

Vermogensklasse (x f. 1 000)	Aagtekerke Z.-43 Aantal	Vermogen x f.1000
- < 10	1)	1)
10 - < 15	1)	1)
15 - < 20	2) 21	2) 248,5
20 - < 30	14	356,5
30 - < 50	10	390,5
50 - < 100	2) 18	2) 2 512,0
100 - < 200	1)	1)
200 - < 300	1)	1)
300 - < 500	1)	1)
500 - < 1 000		
1 000 en meer..		
Totaal	63	3 507,5
Totaal belasting	f. 12 285	
Gem.vermogen:		
per inwoner	" 4 772	
per belastingpl.	" 55 675	

Notes: The first column indicates the income or wealth bracket, the second column indicates the number of individuals in that bracket, and the third column the total bracket income or wealth. The notes 1) and 2) indicate which brackets have been grouped together for privacy reasons. *Source*: (Statistics Netherlands, 1952, 1953).

FIGURE IV.16: Example of a source image of the Census of Companies by municipality in 1930



Notes: The example is for Amsterdam. The data contains the broad and detailed industry classification (columns 1 and 2), the number of establishments and workers by size (columns 15-21), and information on power adoption (columns 24-28). *Source*: (Statistics Netherlands, 2010).

1.4.2 Census of Manufactures industry crosswalks

1.4.3 1860-1900 crosswalks

Industry	Census of Manufactures industries
agricultural implements	agricultural implements; agricultural implements - fanning mills; agricultural implements - grain cradles and scythe snaths; agricultural implements - grain drills; agricultural implements - handles, plough and other; agricultural implements - hoes; agricultural implements - miscellaneous; agricultural implements - mowing and reaping machines; agricultural implements - ploughs, harrows, and cultivators; agricultural implements - rakes; agricultural implements - straw cutters; agricultural implements - threshers, horse-powers, and separators; agricultural implements, ns; mowing-machine knives; scythe rifles; scythes; shovels and spades; shovels, spades, forks, and hoes
agriculture	bee-hives; clover hulling; clover seed cleaning; cotton ginning; fences, patent; flowers; grain threshing; hay and straw, baling; hay pressing; prepared moss; rice cleaning; rice, cleaning and polishing; seeds, garden and flower
artificial limbs and surgical appliances	artificial limbs; shoulder braces; splints; surgical appliances
ashes, pot and pearl	ashes, pot and pearl
awnings and tents	awnings and tents; awnings, tents, and sails
bagging, flax, hemp, and jute	bagging; bagging, flax, hemp, and jute; hemp hose
bags, other than paper	bags; bags, other than paper
bags, paper	bags paper; paper bags
baking and yeast powders	baking and yeast cakes and powders; baking and yeast powders; baking-powders; saleratus
belting and hose	belting and hose leather; leather belting and hose; racking-hose
billiard tables and materials	billiard and bagatelle tables; billiard and bagatelle tables and materials; billiard cues; billiard tables and materials
blacking and other polishes	blacking; blacking and water-proof composition; cleansing and polishing preparations; furniture polish; polishing preparations; stove polish
blacksmithing	blacksmithing; blacksmithing and wheelwrighting; horse-shoes
bleaching, dyeing, and cleaning	bleaching and dyeing; bleaching straw goods; dyeing and bleaching; dyeing and cleaning; dyeing and finishing textiles; straw bonnet bleaching
bolts, nuts, washers, and rivets	bolts, nuts, washers, and rivets; iron and steel, bolts,nuts,washers, and rivets; iron, bolts, nuts, washers, and rivets
bookbinding	bookbinding; bookbinding and blank books; bookbinding and blank-book making
boots and shoes	boot and shoe cut stock; boot and shoe findings; boot and shoe patterns; boot and shoe uppers; boots and shoes; boots and shoes factory product; boots and shoes, custom work and repairing; boots and shoes, including custom work and repairing; shoe and boot tips; shoe findings; shoe strings
boxes, fancy and paper	boxes fancy and paper; boxes, fancy; boxes, paper

brassware and bells	bells; brass; brass and bell founding; brass and copper tubing; brass book clasps and badges; brass castings; brass castings and brass finishing; brass founding and brass ware; brass founding and finishing; brass ornaments; brass wire and wire cloth; brass, rolled; brassware
bread and bakery products	bread and crackers; bread and other bakery products; bread, crackers, and other bakery products
brick, stone, and tile	brick; brick and tile; fire-brick; masonry brick and stone; plastering and stuccowork; sand, washed
bridge building	bridge-building; bridges
bronze	bronze castings; bronze powders
brooms and brushes	broom handles; brooms; brooms and brushes; brooms and wisp-brushes; brush blocks; brush handles and stocks; brushes; mops and dusters
butter, cheese, etc	butter reworking; cheese; cheese and butter urban dairy product; cheese and butter, factory; cheese butter and condensed milk factory product
canning and preserving	fish, cured and packed; fruits and vegetables, canned and preserved; fruits and vegetables, canning and preserving; oysters canning and preserving; pickles, preserves, and sauces; preserves and sauces; provisions
carpentering	carpentering; carpentering and building
carpets	carpets; carpets and rugs other than rag; carpets, other than rag; carpets, rag
carriage and wagon materials	carriage and wagon materials; hubs, spokes, bows, shafts, wheels, and felloes; spokes, hubs, felloes, shafts, and bows; wheelwrighting
carriages and wagons	carriages; carriages and sleds, childrens; carriages and wagons; carriages and wagons, including custom work and repairing; carriages childrens; carriagesmithing; wagons and carts
cases	clock cases and materials; clock-cases; hydrant cases; jewelry and instrument cases; jewelry boxes and cases; sewing machine cases; show cases; stereoscopic cases; watchcases
chemical pigments	blueing; bluing; bone-, ivory-, and lamp-black; bone-black; ivory-black; lampblack; washing blue; white lead; whiting
chemicals, other	acid, pyroligneous; acid, sulphuric; acids, (not specified); barilla; benzoline; calcium lights; celluloid and celluloid goods; chemicals bichromate of potash; chemicals bisulphate of lime; fire clay; fire extinguishers chemical; isinglass; lye, condensed; moulding sand; mucilage and paste; oil - water; potters clay and materials; putty; saltpeter; saltpetre and nitrate of soda; sulphur; taxidermy; water lime; wood preserving
chocolate	chocolate; chocolate and cocoa products
chromos and lithographs	photolithographing and engraving; photolithographing and photoengraving
clocks and watches	clock materials; clocks; watch and clock materials; watch and clock repairing; watch clock and jewelry repairing; watch materials; watches; watches, watch repairing, and materials

clothing, general	belt clasps and slides; belts, childrens; buttons; clothing mens custom work and repairing; clothing, childrens; clothing, mens; clothing, mens, factory product; clothing, mens, factory product, buttonholes; clothing, ns; collars and cuffs, paper; furnishing goods mens; shirts; suspenders
clothing, women's	car fixtures and trimmings; carriage-trimmings; clothing - ladies; clothing, womens; clothing, womens, dressmaking; clothing, womens, factory product; coach lace; coffin trimmings; corsets; dress patterns; fancy articles; fancy articles not elsewhere specified; fruit-jar trimmings; hatters trimmings; hoop-skirts and corsets; lamp trimmings; millinery; millinery and dress making; millinery and lace goods; millinery goods; millinery, custom work; skirt supporters
coffee and spices, roasted and ground	coffee and spice, roasting and grinding; coffee and spices, ground; coffee and spices, roasted and ground; coffee roasting; coffee, essence of
coffins	coffin screws; coffins; coffins and burial cases, trimming and finishing; coffins burial cases and undertakers goods
combs	comb plates; combs; combs, shell and other
confectionery	confectionery
construction, other	building stone, artificial; cement pipe; cisterns; stair building; well curbs
cooperage	cooperage; staves, heading, hoops, and shooks
copper	copper - sheet and bolt; copper smelting; copper work; copper, milled and smelted; copper, rolled; coppersmithing; speaking tubes
cordage and twine	cordage; cordage and twine; cotton braid, thread, lines, twine, and yarn; cotton cordage; cotton thread, twine, and yarn
cork	cork cutting; corks
cotton compressing	cotton batting and wadding; cotton compressing; cotton pressing
cotton goods	cotton bags; cotton coverlets; cotton flannel carding; cotton goods; cotton goods, (not specified); cotton lamp wick; cotton mosquito netting; cotton small wares; cotton table-cloths; cotton-ties
cutlery, edge tools, and axes	cutlery; cutlery and edge tools; cutlery and edge-tools, (not specified); edge tools and axes
decorative work, other	artificial feathers and flowers; bath tubs; bead work; china and glass decorating; china decorating; embroidery; feathers, cleaned, dressed, and dyed; kaolin and ground earths; kaolin and other earth grinding; ornaments - terra cotta; pearl goods; pencils and pens, gold; pens, gold; pipes - clay; pipes - meerschaums; porcelain ware; spelter; stuffed birds; teeth, porcelain; terra-cotta ware; veneers
dentistry	dentistry; dentistry, mechanical; dentists materials
drugs, chemicals, and medicines	chemicals; drug grinding; druggists preparations not including prescriptions; drugs and chemicals; drugs, ground; magnesia; manganese; medicines, extracts, and drugs; nitro-glycerine; patent medicines and compounds; zinc, oxide of
dyestuffs and extracts	bark - ground; bark - sumac, and sumac prepared; dye stuffs and extracts; dye woods and dye stuffs; gum and gum cleaning; hemlock-bark, extract; liquor coloring

electrical, telegraph, and telephone apparatus	electrical apparatus and supplies; telegraph and telephone apparatus
emery	corundum; emery; emery wheels; emery, reduced and ground
enameled goods	enameled goods; enameling; enameling and enameled goods; enamelling
engines and railroad cars	car brakes; car wheels; cars and general shop construction and repairs by steam-railroad companies; cars and general shop construction and repairs by street railroad companies; cars steam railroad not including operations of railroad companies; cars street-railroad not including operations of railroad companies; cars, omnibuses, and repairing; cars, railroad, street, and repairs; fire engines; locomotive engines and repairing; machinery, fire-engines
engraving	carving; engravers materials; engraving; engraving and die-sinking; engraving and stencil-cutting; engraving steel including plate printing; engraving, calico; engraving, steel; engraving, wood; gilding; watch engraving
envelopes	envelopes; envelopes and cards, embossed
explosives and fireworks	explosives; explosives and fireworks; fireworks; high explosives
fertilizers	fertilizers
files	files
fisheries	fisheries
fishing supplies	fish hooks; fishing lines, nets, and tackle; hunting and fishing tackle; nets; nets and seines; nets, fish, and seines
flags and banners	flags and banners; regalia and society banners and emblems; regalias, banners, and flags
flax, dressed	flax dressing; flax, dressed
flour and grist mills	flour and meal; flouring and grist mill products
food products, other	barley, pearl; bone boiling; cocoa; cordials and sirups; dippers, cocoa-nut; fish canning and preserving; flavoring extracts; food preparations; food preparations, animal; food preparations, macaroni and vermicelli; food preparations, vegetable; ginseng; hemp dressing; hominy; macaroni and vermicelli; milk, condensed; mustard; mustard, ground; oleomargarine; rice flour; sumac, ground
fuel, charcoal and coke	charcoal; charcoal, pulverized; coke
fuel, gas	gas; gas illuminating and heating; gas, illuminating
fuel, kerosene and camphene	camphene and burning fluid; coal-oil, rectified; oil - coal; oil - kerosene
fuel, other	fuel, artificial; granular fuel; oil, illuminating, not including petroleum refining
furniture	beds, spring; furniture; furniture factory product; furniture, (not specified); furniture, cabinet, school, and other; furniture, cabinetmaking, repairing and upholstering; furniture, chairs; furniture, iron bedsteads; furniture, refrigerators; house-furnishing goods, not elsewhere classified; housefurnishing goods; mattresses and beds; mattresses and spring beds; medicine chests; money drawers; printers chases, furniture, and rollers; refrigerators; refrigerators and water-coolers
furs	fur goods; furs; furs, dressed

glass	aquariums; artificial eyes; bottle moulds; bottling; glass; glass cutting staining and ornamenting; glass sand; glass ware; glass, cut; glass, cut, stained, and ornamented; glass, plate; glass, stained; glass, window; looking-glasses; mineral water apparatus; mirrors; optical goods; soda-water apparatus; spectacles and eye-glasses
gloves and mittens	gloves and mittens
glue	glue
gold and silver leaf and foil	gold and silver leaf and foil; gold, leaf and foil
gold and silver refining	gold and silver assaying and refining; gold and silver reducing and refining not from the ore; gold and silver, reduced and refined
grease, hides, and tallow	grease; grease and tallow; hides and tallow; lard, refined
gun- and lock-smithing	ammunition; bank locks; fire bomb-lances; fire-arms; gun locks and materials; gunsmithing; keys, metallic; lock and gun smithing; locksmithing and bellhanging; percussion-caps; powder flasks and percussion caps
gunpowder	gunpowder
hair-work	hair jewelry; hairwork; wigs and hair work
hardware	hardware; hardware saddlery
hats and caps	cap fronts; fur hats; hat and cap materials; hat materials; hat-bodies; hat-tips; hats and caps; hats and caps not including fur hats and wool hats; hats and caps, not including wool hats; wool hats
hones and whetstones	hones and whetstones; whetstones
hooks and eyes	hooks and eyes
hosiery and knit goods	hand knit goods; hosiery; hosiery and knit goods
ice	ice; ice, artificial; ice, manufactured
ink	ink; ink, printing; ink, writing
instruments, professional and scientific	globes, terrestial and celestial; instruments; instruments professional and scientific
iron and steel products, other	anchors and chains; axles; candle moulds; carpet-sweepers; cheese presses and vats; chimney flues; eave troughs; grates and fenders; handspikes; hydrants; iron anchors and cable-chains; iron and steel, doors and shutters; iron doors and shutters; iron, castings, stoves, heaters, and hollow ware; ironwork, architectural and ornamental; metallic caps and lables; plugs and wedges; plumbers materials; sad-irons; sash, metal; sieve hoops; stair rods; tinned iron ware; torpedoes; truss hoops; vats; wheelbarrows; whitesmithing
iron and steel, forged and wrought	fire-escapes; hinges, wrought and cast; iron - forged, rolled, and wrought; iron and steel forgings; iron and steel pipe wrought; iron forgings; iron pipe, wrought; iron, forged and rolled; iron, railing, wrought; steel, forged
iron and steel, general	iron - cast; iron and steel; iron, castings, (not specified); steel, (not specified); steel, and manufactures of; steel, cast
iron and steel, other	galvanizing; iron, blooms; steel, bessemer
iron and steel, pig	iron, pig; iron, pigs
ivory and bone work	ivory and bone work; ivory-work; turning, ivory and bone

japanned ware	japanned ware; japanning
jewelry	jewelry; jewelry, (not specified)
kindling wood	kindling wood
lapidary work	lapidaries work; lapidary work
lasts	lasts; lasts and boot trees
lead	lead bar pipe and sheet; lead, bar and sheet; lead, bar, pipe, sheet, and shot; lead, manufactures of; lead, pipe; lead, shot; plumbago, black and silver lead
leather	leather; leather board; leather morocco; leather patent and enameled; leather patent and enamelled leather; leather skin dressing; leather tanned, curried, and finished; leather, curried; leather, dressed skins; leather, morocco, tanned and curried; leather, tanned; leather, tanned and curried
leather products, other	leather goods; razor-strops; watch guards
lightning rods	lightning-rods
lime and cement	cement; lime; lime and cement
linen and linen goods	belting and hose, linen; flax and linen goods; linen goods; thread, linen
liquors and beverages, other	alcohol; cider; cider refined; liquors - bottled; liquors - cordials; malt kilns
liquors, distilled	liquors, distilled
liquors, malt	liquors malt; small beer
liquors, rectified	liquors - rectified
liquors, vinous	liquors - wine; liquors vinous
lithographing	chromos and lithographs; lithographing; lithographing and engraving; lithography
looking-glass and picture frames	looking-glass and picture frames
lumber	lumber and other mill products from logs or bolts; lumber and timber products; lumber, ns; lumber, planed; lumber, planing mill products , including sash , doors, and blinds; lumber, sawed; timber cutting and timber hewed; timber products, not manufactured at mill
machinery, iron and steel	anvils and vices; automaton pressmen; bellows; bookbinders machinery; coffee, roasters; cotton gins; crucibles; electro-magnetic machines; foundery and machine-shop products; foundry and machine shop products; furnaces, ranges, registers, and ventilators; gas and oil stoves; gas stoves; gas works, portable; gas-retorts; hoisting apparatus and machines; machinery - hay and cotton presses; machinery - paper; machinery - rice machines; machinery - shingle machines; machinery - silk; machinery - stamp machines; machinery - steam-engines, and c; machinery - turbine water-wheels; machinery - wood working; machinery, railroad repairing; machinery, steam engines and boilers; metal spinning; newspaper directing machines; oil-tanks; paint mills; pipe tongs; portable forges; printing and lithographic presses; registers cash; registers, car-fare; seal and copying presses; steering apparatus; sugar evaporators; watchmakers lathes; windmills
machinery, other	foundery supplies; foundry supplies; machinery, (not specified); shoe peg machines; vanes, weather; windlasses

machinery, wooden	machinery - cotton and woollen; machinery - ribbon looms; machinery, cotton and woolen; washing machines and clothes dryers; washing machines and clothes wringers
malt	malt
marble and stone work	mantels slate marble and marbleized; marble and stone work; marble and stone work, (not specified); marble and stone work, monuments and tombstones; monuments and tombstones
matches	matches
mats and matting	mats and matting; mats and rugs
military goods	military goods
milled quartz	quartz, milled
millstones	millstones; millstones and mill furnishing
millwrighting	millwrighting
mineral and soda waters	mineral and soda waters; mineral water
mining, coal	coal - anthracite; coal - bituminous; coal, ns
mining, gold and silver	gold mining; silver mining
mining, iron	iron ore
mining, lead	lead mining and smelting; lead, pig
mining, other	asphaltum work; chrome mining; clay mining; copper mining; nickel ore; zinc ore
musical instruments	musical instrument materials; musical instruments - melodeons; musical instruments - miscellaneous; musical instruments - piano-fortes; musical instruments and materials not specified; musical instruments organs; musical instruments, nec; musical instruments, organs and materials; musical instruments, pianos and materials; piano-forte stools
nails and spikes	horse-shoe nails; iron and steel nails and spikes cut and wrought including wire nails; iron, nails and spikes, cut and wrought; nails, cut, wrought, and spikes
non-metal minerals, other	foundry facings; glaziers diamonds; graphite; graphite and graphite refining; grindstones; oil-stones; paving and paving materials; paving materials; scythe stones; soap-stone
oilcloth	clothing - oil; oil and enamelled cloth; oil floor cloth; oil-cloth, silk; oilcloth, enameled; oilcloth, floor
oils	oil - cocoa-nut; oil - cotton-seed; oil - fish, whale and other; oil - lard; oil - neatsfoot; oil - rosin; oil cotton-seed and cake; oil, animal; oil, castor; oil, essential; oil, fish; oil, linseed; oil, not elsewhere specified; oil, resin; oil, vegetable, (not specified); oil, vegetable, castor; oil, vegetable, cotton-seed; oil, vegetable, essential; oil, vegetable, linseed; oils - essential; pitch, brewers and burgundy
oils, lubricating	axle grease; oil, lubricating; oils - chemical
other metal products	babbitt metal and solder; brass and copper, rolled; brass and german silver, rolled; candlesticks; copper and brass ware; electroplating; metal, repaired and white; stamped ware; tin foil
painting and paperhanging	painting; painting and paperhanging; painting house sign etc; paperhanging; paperhangings
paints	paints; paints, (not specified); paints, lead and zinc; zinc paint

paper	paper; paper and wood pulp; paper goods not elsewhere specified; paper, (not specified); paper, printing; paper, writing
paper, other	card boards; card cutting; card cutting and designing; cardboard; cards - enameled; cards - hand; cards - playing; cards, other than playing; ornaments - paper; paper clay; paper patterns; paper ruling; paper shades; paper staining; paper, wrapping; postal cards; valentines
patterns and models	models and patterns; patterns and models
perfumery and cosmetics	perfumery and cosmetics; perfumery and fancy soaps
photography	cameras; photographic apparatus; photographic materials; photographing; photographing materials; photographs; photography
pipes	pipe, wooden; pipes, tobacco
plumbing, heating, and lighting	drain and sewer pipe; drain tile; drain-pipe; electric light and power; electric lights; gas and lamp fixtures; gas fixtures, lamps, and chandeliers; gas machines and meters; gasometers; gasometers and tanks; heating apparatus; lamp fixtures; lamps; lamps and lanterns; lamps and reflectors; metal cocks and faucets; meters, gas; meters, water; plumbers supplies; plumbing and gas and steam fitting; plumbing and gasfitting; steam and gas fittings and valves; steam and water gauges; steam fittings and heating apparatus; steam heaters and heating apparatus
pocket-books	pocket-books, porte-monnaies, and wallets; pocketbooks
printing and publishing	printing and publishing; printing and publishing, (not specified); printing and publishing, book and job; printing and publishing, music; printing and publishing, newspaper; printing and publishing, newspapers and periodicals; printing materials; printing, job
printing and publishing, other	block letters; charts, hydrographic; map mounting and coloring; maps; maps and atlases; music printing; printers fixtures; show cards; signs; stencils and brands
pumps	pumps; pumps and hydraulic rams; pumps not including steam pumps
quarrying	barytes; grindstones and grindstone quarrying; ochre; slate quarrying
roofing and plastering	coal-tar; ornaments - plaster; plaster, and manufactures of; plaster, ground; plastering; roofing; roofing and roofing materials; roofing materials; shingles and lath; shingles, split; stucco and stucco work
rubber and elastic goods	belting and hose, rubber; boots and shoes rubber; gutta-percha goods; india-rubber and elastic goods; india-rubber goods; rubber and elastic goods; rubber, vulcanized; safety-fuse
saddlery and harness	saddlery and harness; saddlery and harness materials
safes, doors, and vaults	safes - cheese; safes - fire-proof; safes - provision; safes and vaults; safes, doors, and vaults, (fire-proof)
salt	salt; salt ground
sand and emery paper and cloth	sand and emery paper and cloth; sand-paper
sash, doors, and blinds	curtain fixtures; sash, doors, and blinds; venetian blinds; window blinds and shades; window shades; wooden door knobs
saws	saws

scales and balances	scales and balances
screws	jack-screws; screws; screws machine; screws wood
sewing machines	needle-threaders; needles; needles and pins; pins; sewing birds; sewing machine needles; sewing machine repairing; sewing machine shuttles; sewing machines and attachments; sewing-machine fixtures; sewing-machines
ship and boat building	blocks and spars; blocks, pumps, and spars; boats; iron steamships; iron, ship building and marine engines; mast hoops and hanks; masts and spars; oakum; oars; rigging; sails; ship and boat building; ship and boat building wooden; ship building, repairing, and ship materials; shipbuilding; ship-building iron and steel
shoddy	shoddy
silk and silk goods	silk and fancy goods, fringes, and trimmings; silk and silk goods; silk goods, (not specified); silk, sewing and twist
silverware	plated and britannia ware; plated ware; silver, manufactures of; silver-plated and britannia ware; silversmithing; silverware
slaughter and meat packing	butchering; meat, cured and packed, (not specified); meat, packed, beef; meat, packed, pork; sausage; slaughtering and meat packing; slaughtering and meat packing, wholesale; slaughtering wholesale not including meat packing
smelting and refining, other	copper smelting and refining; lead smelting and refining; nickel and cobalt; quicksilver; quicksilver, smelted; smelting and refining; smelting and refining, not from the ore
soap and candles	candles - adamantine; candles - wax; candles, adamantine and wax; soap and candles; wax work
springs	springs steel car and carriage; springs, car, carriage, locomotive, and other; steel, springs
stationery and school supplies	artists materials; chalk and crayons; chalk, prepared; pencils, indelible; pencils, lead; pens fountain and stylographic; pens, steel; school apparatus; stationery; stationery goods; stationery goods, not elsewhere classified
stereotyping and electrotyping	stereotyping and electrotyping
stone- and earthen-ware	clay and pottery products; pottery and stone ware; pottery terra-cotta and fire-clay products; stone and earthen ware
straw goods	straw goods
sugar, glucose, and starch	arrow-root; glucose; molasses, refined; sirups, other than sorghum; sorghum sirup; starch; sugar and molasses; sugar and molasses beet; sugar and molasses refining; sugar and molasses, refined; sugar refining
tar and turpentine	tar; tar and turpentine; turpentine - crude; turpentine - distilled; turpentine and rosin
textile products, other	calico printing; car linings; carpet cleaning; cloth finishing; cloth sponging and refinishing; clothing, horse; costumes; filter bags; fly nets; hair-cloth; hammocks; horse-covers; labels and tags; laundry work; life-preservers; mixed textiles; printing cotton and woolen goods; quilts; satinet printing; tags; tapes and binding; trusses, bandages, and supporters; weaving, (not specified); webbing; wool cleaning and pulling
tin, copper, and sheet-iron ware	tin and terne plate; tin, copper, and sheet-iron ware; tinsmithing coppersmithing and sheet-iron working; tinware, copperware, and sheet-iron ware

tobacco	cigars; tobacco and cigars; tobacco and snuff; tobacco chewing smoking and snuff; tobacco cigars and cigarettes; tobacco stemming; tobacco, chewing and smoking, and snuff; tobacco, cigars; tobacco, stemming and rehandling
tools	blacksmiths tools; bookbinders tools; brick machinery and tools; carpenters tools; confectioners tools; coopers tools; curriers tools; hatters tools; jewelers dies, tools, and machinery; machinists tools; shoemakers tools; stencil tools; stone-cutters tools; tinners tools and machines; tools; tools not elsewhere specified
toys, games, and sporting goods	base-ball goods; croquet sets; sporting goods; toy books and games; toys; toys and games; toys, tin
trunks, carpet bags, and valises	trunk and carpet bag frames; trunks and valises; trunks, carpet bags, and valises; trunks, seamens chests; trunks, valises and satchels
type founding	metal type; type and type and stereotype founding; type founding
umbrellas, whips, and canes	umbrella furniture; umbrellas and canes; whips; whips and canes; whips, whip-lashes, sockets, and canes
upholstery	curled hair; curtains; husks, prepared; sponges; upholstering; upholstering materials; upholstery; upholstery materials
varnish	varnish
vault lights	vault lights; vault lights and ventilators
vinegar	vinegar; vinegar and cider
willow ware, baskets, and rattan	baskets; baskets, and rattan and willow ware; baskets, rattan and willow ware; whalebone and ratan; whalebone and rattan; whalebone and rattan, prepared; willow furniture and willow ware; willow ware and rustic ornaments
wire	wire; wire cloth; wire rope; wire work - sieves and bird cages; wire, insulated; wired steel; wirework; wirework including wire rope and cable
wood products, other	carpets, wood; churns; cigar molds; drain pipe, wooden; dumb waiters; engravers blocks and wood; fans; hand stamps; handles; handles, wooden; hat and bonnet blocks; pulp goods; pulp,wood; rules ivory and wood; shoe-pegs; sugar moulds; type, wooden; veneering; water-closets; wood cutting; wood pulp; wood work, miscellaneous; wood, brackets, moldings and scrolls; wooden clothes frames; wooden screws
wood, turned and carved	turning, scroll sawing, and moulding; wood, turned and carved
wooden boxes	box shooks; boxes - packing; boxes - sugar; boxes - tobacco; boxes, cheese; boxes, cigar; boxes, ns; boxes, wooden packing
wooden ware	wooden ware; woodenware, not elsewhere specified
wool-carding and cloth-dressing	wool-carding and cloth-dressing
woolen goods	wool pulling; wool scouring; woolen goods; woollen goods; woollen yarn
worsted goods	worsted goods
yarn and cloth, other	felt goods; felting; jute and jute goods
zinc	zinc; zinc smelting and refining; zinc, (statuary and building ornaments); zinc, smelted and rolled

1.4.4 1890-1939 crosswalks

Industry	Census of Manufactures industries
agricultural implements	agricultural implements; agricultural machinery (except tractors); tractors
aircraft and parts	aircraft and parts; aircraft and parts, including aircraft engines; airplanes, seaplanes, and airships, and parts
artificial flowers and feathers and plumes	artificial and preserved flowers and plants; artificial feathers and flowers; artificial flowers; artificial flowers and feathers and plumes; feathers and plumes; feathers, plumes, and artificial flowers; feathers, plumes, and manufactures thereof
artists materials	artists materials
automobiles including bodies and parts	automobile bodies and parts; automobile trailers (for attachment to passenger cars); automobiles; automobiles including bodies and parts; motor vehicles, motor-vehicle bodies, parts and accessories; motor vehicles, not including motorcycles; motor-vehicle bodies and motor-vehicle parts
awnings tents and sails	awnings, tents, and sails; awnings, tents, sails, and canvas covers
axle grease	axle-grease; lubricating greases; lubricating oils and greases, not made in petroleum refineries
bags other than paper	bagging, flax, hemp, and jute; bags, other than paper; bags, other than paper, not including bags made in textile mills; bags, other than paper, not made in textile mills; textile bags—not made in textile mills
bags paper	bags, paper; bags, paper, exclusive of those made in paper mills; bags, paper, not including bags made in paper mills; paper bags, except those made in paper mills
baking and yeast powders	baking and yeast powders; baking powder, yeast, and other leavening compounds; baking powders and yeast; baking powders, yeast, and other leavening compounds; baking-powders
baskets and rattan and willowware	baskets and rattan and willow ware; baskets and rattan and willow ware, not including furniture; baskets for fruits and vegetables; rattan and willowware (except furniture) and baskets other than vegetable and fruit baskets
belting and hose	belting and hose woven and rubber; belting and hose, leather; belting and hose, linen; belting and hose, other than rubber; belting and hose, rubber; belting and hose, woven, other than rubber; belting, leather; belting, other than leather and rubber, not made in textile mills; belts (apparel), regardless of material; industrial leather belting and packing leather
beverages	beverages; liquors, malt; liquors, malt, including cereal beverages; malt liquors; mineral and soda waters; nonalcoholic beverages
bicycles motorcycles and parts	bicycles and tricycles; bicycles motorcycles and parts; motorcycles, bicycles, and parts

billiard tables and materials	billiard and pool tables, bowling alleys, and accessories; billiard tables and accessories; billiard tables and materials; billiard tables, bowling alleys, and accessories
blacking and cleansing and polishing preparations	blacking; blacking and cleansing and polishing preparations; blacking, stains, and dressings; cleaning and polishing preparations; cleaning and polishing preparations, blackings, and dressings; cleansing and polishing preparations; cleansing preparations
bluing	bluing
bone ivory and lamp black	bone and carbon black; bone black, carbon black, and lampblack; bone, carbon, and lamp black; bone-, ivory-, and lampblack
boots and shoes including cut stock and findings	boot and shoe cut stock; boot and shoe cut stock and findings; boot and shoe cut stock, not made in boot and shoe factories; boot and shoe findings; boot and shoe findings, not made in boot and shoe factories; boot and shoe uppers; boots and shoes; boots and shoes custom work and repairing; boots and shoes factory product; boots and shoes including cut stock and findings; boots and shoes, not including rubber boots and shoes; boots and shoes, other than rubber; footwear (except rubber)
boots and shoes rubber	boots and shoes rubber; rubber boots and shoes (including rubber-soled footwear with fabric uppers)
boxes cigar	boxes, cigar; boxes, cigar, wooden; cigar boxes wooden, part wooden
boxes fancy and paper	boxes, fancy and paper; boxes, paper and other, not elsewhere specified; boxes, paper, not elsewhere classified; boxes, paper, shipping containers; boxes, set-up paper boxes; boxes, set-up paper boxes and cartons; paperboard containers and boxes not elsewhere classified
bread and other bakery products	biscuit, crackers, and pretzels; bread and other bakery products; bread and other bakery products (except biscuit, crackers, and pretzels)
brick and tile pottery terracotta and fire clay products	brick and hollow structural tile; brick and tile; brick and tile, terra-cotta , and fire clay products; clay and pottery products; clay products (except pottery) not elsewhere classified; clay products (other than pottery) and non-clay refractories; clay refractories, including refractory cement (clay); floor and wall tile (except quarry tile); nonclay refractories; roofing tile; sand-lime brick; sand-lime brick, block and tile; sewer pipe and kindred products; terra cotta
brooms and brushes	brooms; brooms and brushes; brooms, from broom corn; brushes; brushes, other than rubber
butter cheese and condensed milk	butter; butter cheese and condensed milk; butter reworking; cheese; cheese and butter urban dairy product; cheese butter and condensed milk factory product; condensed and evaporated milk; condensed milk; creamery butter
buttons	buttons

canning and preserving	canned and dried fruits and vegetables (including canned soups); canned fish, crustacea, and mollusks; canning and preserving; canning and preserving fish, crabs, shrimps, oysters, and clams; canning and preserving fruits and vegetables; canning and preserving fruits and vegetables pickles, jellies, preserves, and sauces; canning and preserving, fish; canning and preserving, fruits; canning and preserving, oysters; canning and preserving, vegetables; canning and preserving, vegetables and dried fruits; cured fish; fish canning and preserving; fruits and vegetables, canning and preserving; oysters canning and preserving; pickled fruits and vegetables and vegetable sauces and seasonings; pickles, preserves, and sauces; preserves, jams, jellies, and fruit butters; quick-frozen foods; salad dressings
card cutting and designing	card cutting and designing
carpets and rugs other than rag	carpet yarn, woolen and worsted; carpets and rugs other than rag; carpets and rugs, wool; carpets and rugs, wool, other than rag; carpets, wood
carpets rag	carpets and rugs, rag; carpets, rag
carriages and sleds childrens	carriages and sleds, childrens; childrens vehicles
carriages and wagons and materials	carriage and wagon materials; carriages and wagons; carriages and wagons and materials; carriages and wagons, including custom work and repairing; carriages and wagons, including repairs; carriages and wagons, repair work only; carriages, wagon, sleigh, and sled materials; carriages, wagons, sleighs, and sleds; transportation equipment, nec
cars and general shop construction by railroad companies	cars and general construction and repairs, electric-railroad repair shops; cars and general construction and repairs, steam railroad repair shops; cars and general shop construction and repairs by electric-railroad companies; cars and general shop construction and repairs by steam-railroad companies; cars and general shop construction and repairs by street-railroad companies
cash registers and calculating machines	cash registers and calculating machines; cash registers, and adding, calculating, and card-tabulating machines; registers cash; registers, car fare
chemicals	chemicals; chemicals, not elsewhere classified; coal-tar products; coal-tar products, crude and intermediate; hardwood distillation and charcoal manufacture; rayon and allied products; sulphuric, nitric, and mixed acids; wood distillation; wood distillation and charcoal manufacture; wood distillation not including turpentine and rosin; wood naval stores
china decorating	china decorating; china decorating, not including that done in potteries; china firing and decorating (for the trade); china firing and decorating, not done in potteries
chocolate and cocoa products	chocolate and cocoa products; chocolate and cocoa products, not including confectionery
clocks and watches including cases and materials	clocks; clocks and watches including cases and materials; clocks, clock movements, time-recording devices, and time stamps; clocks, watches, and materials and parts (except watchcases); watch and clock materials; watch and clock materials and parts, except watchcases; watch and clock materials, except watchcases; watch cases; watch materials, except watchcases; watch, clock and jewelry repairing; watches

clothing mens including shirts	childrens and infants wear not elsewhere classified—made in inside factories or by jobbers engaging contractors; childrens dresses—made in contract factories; childrens dresses—made in inside factories or by jobbers engaging contractors; clothing (except work clothing), mens, youths, and boys, not elsewhere classified; clothing mens factory product buttonholes; clothing mens including shirts; clothing, mens; clothing, mens, buttonholes; clothing, mens, factory product; clothing,mens, custom work and repairing; coats, suits, and skirts (except fur coats)—made in inside factories or by jobbers engaging contractors; mens and boys shirts (except work shirts), collars, and nightwear—made in contract factories; mens and boys shirts (except work shirts), collars, and nightwear—made in inside factories or by jobbers engaging contractors; mens and boys suits, coats, and overcoats (except work clothing)—made in contract factories; mens and boys suits, coats, and overcoats (except work clothing)—made in inside factories or by jobbers engaging contractors; mens and boys underwear—made in inside factories or by jobbers engaging contractors; mens neckwear—made in contract factories; mens neckwear—made in inside factories or by jobbers engaging contractors; raincoats and other waterproof garments (except oiled cotton); robes, lounging garments, and dressing gowns; shirts; trousers (semidress), wash suits, and washable service apparel; womens and misses blouses and waists—made in contract factories; womens and misses blouses and waists—made in inside factories or by jobbers engaging contractors
clothing womens	childrens and infants wear not elsewhere classified—made in contract factories; clothing womens dressmaking; clothing, womens; clothing, womens, factory product; clothing, womens, not elsewhere classified; clothing, work (including sheep-lined and blanket-lined work coats but not including shirts), mens; womens and misses clothing, not elsewhere classified—made in inside factories or by jobbers engaging contractors; womens and misses dresses (except house dresses)—made in contract factories; womens and misses dresses (except house dresses)—made in inside factories or by jobbers engaging contractors; womens, childrens, and infants underwear and nightwear of cotton and flannelette woven fabrics; womens, childrens, and infants underwear and nightwear of knitted fabrics; womens, childrens, and infants underwear and nightwear of silk and rayon woven fabrics
cloth sponging and refinishing	cloth sponging and miscellaneous special finishing; cloth, sponging and refinishing
coffee and spice roasting and grinding	coffee and spice, roasting and grinding; peanuts grading roasting cleaning and shelling; peanuts, walnuts, and other nuts, processed or shelled
coffins burial cases and undertakers goods	caskets, coffins, burial cases, and other morticians goods; coffins, burial cases, and undertakers goods
coke	beehive coke; coke; coke, not including gas-house coke; oven coke and coke-oven byproducts
confectionary and ice cream	ice cream; ice cream and ices
confectionery and ice cream	candy and other confectionery products; chewing gum; confectionery; confectionery and ice cream

copper tin and sheet iron products	aluminum manufactures; aluminum products (including rolling and drawing and extruding), not elsewhere classified; aluminum ware, kitchen, hospital, and household (except electrical appliances); copper tin and sheet-iron products; copper, tin, and sheet-iron work; copper, tin, and sheet-iron work, including galvanized-iron work, not elsewhere classified; enameled goods; enameling; enameling and enameled goods; enameling and japanning; enameling, japanning, and lacquering; sheet-metal work not specifically classified; stamped and enameled ware, not elsewhere specified; stamped and pressed metal products (except automobile stampings); stamped ware; stamped ware, enameled ware, and metal stamping, enameling, japanning, and lacquering; stamped ware, not elsewhere specified; tin and terne plate; tin cans and other tinware not elsewhere classified; tin plate and terneplate; tinsmithing coppersmithing and sheet-iron working; tinware, not elsewhere specified
cordage and twine	linen goods
cordage and twine and jute and linen goods	cordage and twine; cordage and twine and jute and linen goods; jute and jute goods; jute goods; jute goods (except felt)
cork cutting	cork cutting; cork products
corsets	corsets; corsets and allied garments
cotton goods including cotton smallwares	carpets, rugs, and mats made from such materials as paper fiber, glass, jute, flax, sisal, cotton, cocoa fiber, and rags; cotton broad woven goods; cotton goods; cotton goods including cotton small wares; cotton lace; cotton narrow fabrics; cotton small wares; cotton thread
crucibles	crucibles
cutlery and tools not specified	cutlery (except aluminum, silver, and plated cutlery) and edge tools; cutlery (not including silver and plated cutlery) and edge tools; cutlery and edge tools; cutlery and tools not elsewhere specified; tools, not elsewhere specified
dentists materials	dental equipment and supplies; dental goods; dental goods and equipment; dentists materials
drug grinding	drug grinding
dyeing and finishing textiles	cotton yarn; dyeing and finishing cotton, rayon, silk, and linen textiles; dyeing and finishing textiles; dyeing and finishing textiles, exclusive of that done in textile mills
dyestuffs and extracts	dye stuffs and extracts; dyestuffs and extracts—natural; tanning materials, natural dyestuffs, mordants and assistants, and sizes; tanning materials, natural dyestuffs, mordants, assistants, and sizes
electrical machinery apparatus and supplies	automotive electrical equipment; batteries, storage and primary (dry and wet); beauty-shop and barber-shop equipment; carbon products for the electrical industry, and manufactures of carbon or artificial graphite; communication equipment; electric lamps; electrical apparatus and supplies; electrical appliances; electrical machinery, apparatus, and supplies; electrical measuring instruments; electrical products not elsewhere classified; generating, distribution, and industrial apparatus, and apparatus for incorporation in manufactured products, not elsewhere classified; insulated wire and cable; radios, radio tubes, and phonographs; wiring devices and supplies; x-ray and therapeutic apparatus and electronic tubes
electroplating	electroplating; electroplating, plating, and polishing

emery and other abrasive wheels	emery and other abrasive wheels; emery wheels; emery wheels and other abrasive and polishing appliances
enameling and japanning	japanning
engravers materials	engravers materials
engraving and die sinking	engraving (other than steel, copperplate, or wood), chasing, etching, and diesinking; engraving and die-sinking; engraving on metal (except for printing purposes)
engraving wood	engraving, wood
explosives	explosives; gunpowder; high explosives
fancy articles not specified	combs; combs and hairpins, except those made from metal or rubber; combs and hairpins, not made from metal or rubber; fancy and miscellaneous articles, not elsewhere classified; fancy articles not elsewhere specified; fancy articles, not elsewhere-specified; ivory and bone work; ivory, shell, and bone work, not including buttons, combs, or hairpins; ivory, shell, and bone work, not including combs and hairpins; signs and advertising novelties; signs, advertising displays, and advertising novelties
fertilizers	fertilizers
files	files
firearms and ammunition	ammunition; ammunition and related products; fire-arms; firearms and ammunition
fire extinguishers chemical	fire extinguishers, chemical
fireworks	fireworks
flags banners regalia society badges and emblems	flags and banners; flags banners regalia society badges and emblems; flags banners regalia society banners and emblems; regalia and society banners and emblems; regalia, and society badges and emblems; regalia, badges, and emblems
flavoring extracts and flavoring sirups	cordials and flavoring sirups; cordials and sirups; flavoring extracts; flavoring extracts and flavoring sirups; flavoring extracts and flavoring sirups, not elsewhere classified
flour mill and grist mill products	flour and other grain-mill products; flour-mill and gristmill products; flouring and grist mill products
food preparations	blended and prepared flour made from purchased flour; cereal preparations; feeds, prepared, for animals and fowls; food preparations; food preparations, not elsewhere specified; macaroni, spaghetti, vermicelli, and noodles; prepared feeds (including mineral) for animals and fowls; special dairy products

foundry and machine shop products	automobile repairing; bells; blowers exhaust and ventilating fans; bridges; cars and trucks, industrial; cast-iron pipe; cast-iron pipe and fittings; cold-rolled steel sheets and strip and cold-finished steel bars made in plants not operated in connection with hot-rolling mills; commercial laundry, dry-cleaning, and pressing machinery; construction and similar machinery (except mining and oil-field machinery and tools); elevators, escalators, and conveyors; enameled-iron sanitary ware and other plumbers supplies (not including pipe and vitreous and semivitreous china sanitary ware); engines, steam, gas, and water; engines, turbines, tractors, and water wheels; food-products machinery; foundry and machine shop products; foundry and machine-shop products, not elsewhere classified; gas and oil stoves; gas machines; gas machines and gas and water meters; gas machines and meters; gas machines, gas meters, and water and other liquid meters; gas stoves; gray-iron and semisteel castings; hardware; hardware not elsewhere classified; hardware saddlery; heating and cooking apparatus, except electric, not elsewhere classified; industrial machinery, not elsewhere classified; internal-combustion engines; iron and steel, cast-iron pipe; iron and steel, tempering and welding; iron and steel, welding; ironwork architectural and ornamental; lightning-rods; machine tools; machine-shop products, not elsewhere classified; machine-shop repairs; machine-tool accessories and small metal working tools, not elsewhere classified; machine-tool and other metalworking-machinery accessories, metal-cutting and shaping tools, and machinists precision tools; malleable-iron castings; mechanical power-transmission equipment; metalworking machinery and equipment, not elsewhere classified; mining machinery and equipment; oil burners, domestic and industrial; oil-field machinery and tools; paper-mill, pulp-mill, and paper-products machinery; plumbers supplies; plumbers supplies, not elsewhere specified; plumbers supplies, not including pipe or vitreous-china sanitary ware; power boilers and associated products; printing-trades machinery and equipment; pumping equipment and air compressors; pumps (hand and power) and pumping equipment; pumps not including steam pumps; pumps, not including power pumps; pumps, steam; pumps, steam and other power; special-industry machinery, nec; steam and hot-water heating apparatus (including hot-water furnaces); steam engines, turbines, and water wheels; steam fittings and heating apparatus; steam fittings and steam and hot-water heating apparatus; steam fittings, regardless of material; steel barrels, drums, and tanks; steel barrels, drums, and tanks, portable; steel barrels, kegs, and drums; stokers, mechanical, domestic and industrial; stoves and furnaces including gas and oil stoves; stoves and hot-air furnaces; stoves and ranges (other than electric) and warm-air furnaces; stoves, gas and oil; stoves, ranges, water heaters, and hot-air furnaces (except electric); structural and ornamental iron and steel work, not made in plants operated in connection with rolling mills; structural ironwork, not made in steel works or rolling mills; textile machinery; textile machinery and parts; vending, amusement, and other coin-operated machines; woodworking machinery
foundry supplies	foundry supplies
fur goods	fur coats and other fur garments, accessories, and trimmings; fur goods

furnishing goods mens	collars and cuffs, mens; furnishing goods mens; furnishing goods, mens, not elsewhere classified; gloves and mittens, cloth or cloth and leather combined, made from purchased fabrics; gloves and mittens, cloth, not including gloves made in textile mills; mens and boys underwear—made in contract factories; suspenders, garters, and elastic woven goods; suspenders, garters, and other elastic woven goods, made from purchased webbing; suspenders, garters, and other goods made from purchased elastic material; work gloves and mittens cloth, cloth and leather combined
furniture and refrigerators	furniture; furniture and refrigerators; furniture cabinetmaking repairing and upholstering; furniture factory product; furniture, chairs; furniture, except rattan and willow; furniture, including store and office fixtures; furniture, store and office fixtures; furniture, wood, other than rattan and willow; household furniture, except upholstered; laboratory, hospital, and other professional furniture; office furniture; partitions, shelving, cabinet work, and office and store fixtures; public-building furniture; refrigerators; refrigerators and refrigerator cabinets, exclusive of mechanical refrigerating equipment; refrigerators, domestic (mechanical and absorption), refrigeration machinery and equipment, and complete air-conditioning units; refrigerators, mechanical; upholstered household furniture
furs dressed	furs, dressed; furs, dressed and dyed
galvanizing and other coating processes	galvanizing; galvanizing and other coating, not done in plants operated in connection with rolling mills; galvanizing and other coating—carried on in plants not operated in connection with rolling mills
gas and electric fixtures and lamps and reflectors	gas and electric fixtures; gas and electric fixtures and lamps and reflectors; gas and electric fixtures lamps, lanterns, and reflectors; gas and lamp fixtures; lamps; lamps and reflectors; lighting fixtures
gas illuminating and heating	gas illuminating and heating; gas, manufactured, illuminating and heating
glass	flat glass; glass; glass containers; tableware, pressed or blown glass, and glassware not elsewhere classified
glass cutting staining and ornamenting	glass products (except mirrors) made from purchased glass; glass, cutting, staining, and ornamenting
gloves and mittens leather	gloves and mittens; gloves and mittens leather; leather gloves and mittens
glucose and starch	corn sirup, corn sugar, corn oil, and starch; glucose; glucose and starch; starch
glue and gelatin	glue; glue and gelatin; glue, not elsewhere specified
gold and silver leaf and foil	gold and silver leaf and foil; gold, leaf and foil
gold silver and platinum reducing and refining not from the ore	gold and silver reducing and refining not from the ore; gold, silver, and platinum, reducing and refining, not from the ore; secondary smelting and refining, gold, silver, and platinum
graphite and graphite refining	graphite; graphite and graphite refining; graphite, ground and refined
grease and tallow	grease and tallow; grease and tallow (except lubricating greases); grease and tallow, not including lubricating greases
grindstones	grindstones
hairwork	hair work

handstamps and stencils and brands	hand stamps; hand stamps and stencils and brands; hand stamps, stencils, and brands; stencils and brands
hat and cap materials	hat and cap materials; hat and cap materials trimmings, etc; hat and cap materials, mens
hats and caps not including wool hats	finishing of mens and boys hats of fur-felt, wool-felt, and straw; fur hats; hat bodies and hats, fur-felt; hat bodies and hats, wool-felt; hats and caps not including fur hats and wool hats; hats and caps, except felt and straw, mens; hats and caps, not including wool hats; hats and caps, other than felt, straw, and wool; hats, fur-felt; hats, straw; hats, straw, mens; hatters fur; mens and boys hats and caps (except felt and straw)
hones and whetstones	hones and whetstones
hosiery and knit goods	hand knit goods; hosiery and knit goods; hosiery—full-fashioned; hosiery—seamless; knit goods; knitted cloth; knitted gloves; knitted outerwear (except knit gloves)—contract factories; knitted outerwear (except knit gloves)—regular factories or jobbers engaging contractors; knitted underwear
housefurnishing goods not specified	curtains, draperies, and bedspreads—contract factories; curtains, draperies, and bedspreads—made in regular factories or by jobbers engaging contractors; house-furnishing goods, not elsewhere classified; housefurnishings (except curtains, draperies, and bedspreads)
ice manufactured	ice manufactured; ice, artificial
ink printing	ink, printing; printing ink
ink writing	ink, writing; writing ink
instruments professional and scientific	instruments, professional and scientific
iron and steel blast furnaces steel works and rolling mills	blast-furnace products; ferroalloys; iron and steel; iron and steel blast furnaces; iron and steel, steel works and rolling mills; steel castings; steel works and rolling mills
iron and steel bolts nuts washers and rivets	bolts, nuts, washers, and rivets, not made in plants operated in connection with rolling mills; bolts, nuts, washers, and rivets—made in plants not operated in connection with rolling mills; iron and steel bolts nuts washers and rivets; iron and steel bolts nuts washers and rivets not made in steel works or rolling mills; iron and steel, bolts, nuts, washers, and rivets, not made in rolling mills
iron and steel doors and shutters	doors, shutters, and window sash and frames, metal; doors, window sash, frames, molding, and trim (made of metal); iron and steel doors and shutters
iron and steel forgings	forgings, iron and steel, not made in plants operated in connection with rolling mills; forgings, iron and steel—made in plants not operated in connection with rolling mills; iron and steel, forgings; iron and steel, forgings, not made in steel works or rolling mills
iron and steel nails and spikes cut and wrought including wire nails	iron and steel nails and spikes cut and wrought including wire nails; iron and steel, nails and spikes, cut and wrought, including wire nails, not made in steel works or rolling mills; nails, spikes, etc, not made in wire mills or in plants operated in connection with rolling mills
iron and steel pipe wrought	iron and steel pipe wrought; iron and steel, wrought pipe; wrought pipe, welded and heavy riveted, not made in plants operated in connection with rolling mills; wrought pipes, welded and heavy riveted—made in plants not operated in connection with rolling mills

jewelry	costume jewelry and costume novelties (jewelry other than fine jewelry); jewelers findings and materials; jewelry; jewelry (precious metals)
jewelry and instrument cases	jewelry and instrument cases; jewelry cases and instrument cases
labels and tags	labels and tags
lapidary work	lapidary work
lasts	lasts; lasts and related products
leather goods	bellows; clothing, leather and sheep-lined; leather goods; leather goods, nec; pocket-books; pocketbooks, purses, and cardcases; saddlery and harness; saddlery, harness, and whips; small leather goods; trunks and valises; womens pocketbooks, handbags, and purses
leather tanned curried and finished	leather morocco; leather tanned, curried, and finished—contract factories; leather tanned, curried, and finished—regular factories or jobbers engaging contractors; leather, dressed skins; leather, patent and enameled; leather, tanned and curried; leather, tanned, curried, and finished
lime and cement	cement; lime; lime and cement
liqours distilled	alcohol, ethyl, and distilled liquors; liquors distilled; liquors, distilled, grain alcohol; liquors, distilled, grain alcohol and rum; liquors, rectified or blended
liquors vinous	liquors vinous; wines
looking glass and picture frames	looking-glass and picture frames; mirror and picture frames; mirror frames and picture frames
lumber and timber products	boxes, wooden packing, except cigar boxes; boxes, wooden, except cigar boxes; boxes, wooden, packing; logging camps and logging contractors (not operating sawmills); lumber and other mill products from logs or bolts; lumber and timber products; lumber and timber products, not elsewhere classified; lumber, planing mill products , including sash , doors, and blinds; lumber, planing-mill products, not including planing mills connected with sawmills; planing mills not operated in conjunction with sawmills; planing-mill products (including general mill-work), not made in planing mills connected with saw mills; plywood mills; sawmills, veneer mills, and cooperage-stock mills, including those combined with logging camps and with planing mills; timber products, not manufactured at mill; venetian blinds; window and door screens; window and door screens and weather strip; window and door screens and weather strips; wooden boxes, except cigar boxes
malt	malt
marble and stone work	artificial stone; artificial stone products; concrete products; marble and stone work; marble, granite, slate, and other stone products; monuments and tombstones
masonry brick and stone	masonry, brick and stone
matches	matches
mattresses and spring beds	mattresses and bed springs, not elsewhere classified; mattresses and bedsprings; mattresses and spring beds; mattresses and spring beds not elsewhere specified

millinery and lace goods	childrens coats—made in contract factories; childrens coats—made in inside factories or by jobbers engaging contractors; coats, suits, and skirts (except fur coats)—made in contract factories; embroideries; embroideries schiffli-machine products; embroideries, other than schiffli-machine products—contract factories; embroideries, other than schiffli-machine products—made in regular factories or by jobbers engaging contractors; handkerchiefs; handkerchiefs—made in contract factories; handkerchiefs—made in inside factories or by jobbers engaging contractors; house dresses, uniforms, and aprons—made in contract factories; house dresses, uniforms, and aprons—made in inside factories or by jobbers engaging contractors; lace goods; millinery; millinery and lace goods; millinery and lace goods, not elsewhere specified; trimmings (not made in textile mills) and stamped art goods for embroidering; trimmings (not made in textile mills), stamped art goods, and art needlework—contract factories; trimmings (not made in textile mills), stamped art goods, and art needlework—made in regular factories or by jobbers engaging contractors; womens and misses clothing, not elsewhere classified—made in contract factories; womens neckwear, scarfs, etc
minerals and earths ground	kaolin and ground earths; kaolin and other earth grinding; minerals and earths, ground or otherwise treated
mirrors	mirrors; mirrors and other glass products made of purchased glass; mirrors, framed and unframed; mirrors, framed and unframed, not elsewhere specified
models and patterns not including paper patterns	models and patterns; models and patterns (except paper patterns); models and patterns not including paper patterns
mucilage and paste	mucilage and paste; mucilage, paste, and other adhesives, except glue and rubber cement; mucilage, paste, and other adhesives, not elsewhere specified
musical instruments pianos and organs and materials	musical instrument parts and materials piano and organ; musical instruments and materials not specified; musical instruments and parts and materials, not elsewhere classified; musical instruments pianos and organs and materials; musical instruments, organs; musical instruments, organs and materials; musical instruments, piano and organ materials; musical instruments, pianos; musical instruments, pianos, and materials; organs; piano and organ parts and materials; pianos
needles pins and hooks and eyes	hooks and eyes; needles and pins; needles, pins, and hooks and eyes; needles, pins, hooks and eyes, and slide and snap fasteners; needles, pins, hooks and eyes, and snap fasteners
nonferrous metal alloys and products not including aluminum products	alloying and rolling and drawing of nonferrous metals, except aluminum; babbitt metal and solder; brass; brass and bronze products; brass and copper, rolled; brass castings and brass finishing; brass, bronze, and copper products; brassware; lead, bar, pipe, and sheet; nonferrous-metal alloys and products, not including aluminum products; nonferrous-metal foundries (except aluminum); nonferrous-metal products not elsewhere classified
oilcloth and linoleum	linoleum, asphalted-felt-base, and other hard-surface floor coverings, not elsewhere classified; oilcloth and linoleum; oilcloth and linoleum floor; oilcloth floor; oilcloth, enameled
oil cottonseed and cake	cottonseed oil, cake, meal, and linters; oil and cake, cottonseed; oil cotton-seed and cake; oil, cake, and meal, cottonseed
oil essential	essential oils; oils - essential

oil linseed	linseed oil, cake, and meal; oil - linseed; oil, cake, and meal, linseed
oleomargarine	oleomargarine; oleomargarine and other butter substitutes; oleomargarine, not made in meat-packing establishments
optical goods	ophthalmic goods lenses and fittings; optical goods; optical instruments and lenses
paints and varnishes	colors and pigments; paint and varnish; paints; paints and varnishes; paints, varnishes, and lacquers; varnish; varnishes
paper and wood pulp	paper; paper and paperboard mills; paper and wood pulp; pulp (wood and other fiber); pulp mills; pulp,wood
paper goods not specified	coated and glazed paper; converted paper products not elsewhere classified; envelopes; paper goods, not elsewhere classified
patent medicines and compounds and druggists preparations	druggists preparations; druggists preparations, not including prescriptions; drugs and medicines (including drug grinding); insecticides, fungicides, and related industrial and household chemical compounds; patent and proprietary medicines; patent medicines and compounds; patent medicines and compounds and druggists preparations; patent or proprietary medicines and compounds; perfumery and cosmetics; perfumes, cosmetics, and other toilet preparations
paving materials	paving and paving materials; paving blocks and paving mixtures asphalt, creosoted wood, and composition; paving materials; paving materials asphalt, tar, crushed slag, and mixtures
pencils	pencils (except mechanical) and crayons; pencils lead; pencils, lead (including mechanical); pens, mechanical pencils, and pen points
pens fountain stylographic and gold	pens fountain and stylographic; pens fountain stylographic and gold; pens gold; pens, fountain and stylographic pen points, gold, steel, and brass
petroleum refining	petroleum refining
phonographs and graphophones	phonographs; phonographs and graphophones
photo engraving	gravure, rotogravure, and rotary photogravure (including preparation of plates); photo-engraving, not done in printing establishments; photoengraving; photoengraving, not done in printing establishments (including preparation of plates); photolithographing and engraving; photolithographing and photoengraving
photographic apparatus and materials	photographic apparatus; photographic apparatus and materials; photographic apparatus and materials and projection equipment (except lenses); photographic materials
pipes tobacco	pipes tobacco; tobacco pipes and cigarette holders
plumbing and gas and steam fitting	plumbing and gas and steam fitting; plumbing and gasfitting
pottery terracotta and fire clay products	hotel china; porcelain electrical supplies; pottery; pottery products, nec; pottery terra-cotta and fire-clay products; pottery, earthen and stone ware; pottery, including porcelain ware; vitreous-china plumbing fixtures; vitreous-enameled products, including kitchen, household, and hospital utensils; whiteware

printing and publishing	bookbinding and blankbook making; bookbinding and related industries; books printing without publishing; books publishing without printing; books, publishing and printing; engraving (steel, copperplate, and wood) plate printing; engraving, steel and copper plate, including plate printing; engraving, steel and copper plate, including pre-printing; engraving, steel and copperplate, and plate printing; engraving, steel, including plate printing; general commercial (job) printing; greeting cards (except hand-painted); lithographing; lithographing and engraving; lithographing and photo-lithographing (including preparation of stones or plates and dry transfers); machine and hand typesetting (including advertisement typesetting); newspapers publishing and printing; newspapers publishing without printing; paper patterns; periodicals publishing and printing; periodicals publishing without printing; printing and publishing; printing and publishing book and job; printing and publishing music; printing and publishing newspapers and periodicals; printing and publishing, book and job job printing; printing and publishing, job printing; printing and publishing, newspaper and periodical; printing,tip
pulp goods	fabricated plastic products, not elsewhere classified; pulp goods; pulp goods (pressed, molded)
railroad cars	cars steam-railroad not including operations of railroad companies; cars street railroad not including operations of railroad companies; cars, electric and steam railroad, not built in railroad repair shops
rice cleaning and polishing	rice cleaning and polishing
roofing materials	roofing and roofing materials; roofing materials; roofing, built-up and roll asphalt shingles roof coating (except paint); roofing, built-up and roll asphalt shingles roof coatings other than paint
rubber goods not specified	rubber and elastic goods; rubber goods (other than rubber boots and shoes) and rubber tires and inner tubes; rubber goods not elsewhere specified; rubber goods other than tires, inner tubes, and boots and shoes; rubber products not elsewhere classified; rubber tires and inner tubes; rubber, tires, tubes, and rubber goods, not elsewhere specified; tires and inner tubes
safes and vaults	safes and vaults
salt	salt
sand and emery paper and cloth	sand and emery paper and cloth; sandpaper, emery paper, and other abrasive paper and cloth
saws	saws
scales and balances	scales and balances
screw machine products and wood screws	screw-machine products and wood screws; screws wood; screws, machine
sewing machines cases and attachments	sewing machine cases; sewing machines and attachments; sewing machines cases and attachments; sewing machines, domestic and industrial
shipbuilding	boat building and boat repairing; ship and boat building wooden; ship and boat building, steel and wooden, including repair work; shipbuilding; shipbuilding and ship repairing; shipbuilding including boat building; shipbuilding iron and steel; shipbuilding, steel; shipbuilding, steel, new vessels; shipbuilding, steel, new vessels and repair work; shipbuilding, steel, new vessels and small boats; shipbuilding, wooden, including boat building

silk and silk goods including throwsters	rayon broad woven goods—contract factories; rayon broad woven goods—regular factories or jobbers engaging contractors; rayon narrow fabrics; rayon throwing and spinning—contract factories; rayon yarn and thread, spun or thrown—regular factories or jobbers engaging contractors; silk and rayon manufactures; silk and silk goods; silk and silk goods including throwsters; silk broad woven goods—contract factories; silk broad woven goods—regular factories or jobbers engaging contractors; silk goods; silk goods, including throwsters; silk narrow fabrics; silk throwing and spinning—contract factories; silk yarn and thread, spun or thrown—regular factories or jobbers engaging contractors
silverware and platedware	plated and britannia ware; plated ware; silversmithing; silversmithing and silverware; silverware; silverware and plated ware
slaughtering and meat packing	custom slaughtering, wholesale; meat packing, wholesale; sausage; sausage casings—not made in meat-packing establishments; sausage, meat puddings, headcheese, etc, and sausage casings, not made in meat-packing establishments; sausage, not made in slaughtering and meat-packing establishments; sausages, prepared meats, and other meat products—not made in meat-packing establishments; slaughtering and meat packing; slaughtering and meat packing, wholesale; slaughtering wholesale not including meat packing
smelting and refining copper	copper smelting and refining; smelting and refining copper
smelting and refining lead	lead smelting and refining; smelting and refining, lead
smelting and refining not from the ore	secondary smelting and refining of nonferrous metals, not elsewhere classified; smelting and refining; smelting and refining not from the ore; smelting and refining, metals other than gold, silver, or platinum, not from the ore
smelting and refining zinc	smelting and refining, zinc; zinc smelting and refining
soap and candles	candles; soap; soap and candles; soap and glycerin
soda water apparatus	soda fountains, beer dispensing equipment, and related products; soda-water apparatus
sporting and athletic goods	sporting and athletic goods; sporting and athletic goods not elsewhere classified; sporting and athletic goods, not including firearms or ammunition; sporting goods
springs steel car and carriage	springs, steel (except wire)—made in plants not operated in connection with rolling mills; springs, steel, car and carriage; springs, steel, car and carriage, not made in steel works or rolling mills; springs, steel, except wire, not made in plants operated in connection with rolling mills
stationery goods not specified	stationery goods not elsewhere specified
steam packing	steam and other packing pipe and boiler covering; steam and other packing, pipe and boiler covering, and gaskets, not elsewhere classified; steam packing
stereotyping and electrotyping	electrotyping and stereotyping, not done in printing establishments; stereotyping and electrotyping; stereotyping and electrotyping, not done in printing establishments
sugar and molasses beet	beet sugar; sugar and molasses, beet; sugar, beet
sugar and molasses not including beet	cane sugar—except refineries; cane-sugar refining; sugar and molasses; sugar and molasses refining; sugar refining, cane; sugar, cane; sugar, cane, not including products of refineries; sugar, refining, not including beet sugar

surgical appliances and artificial limbs	artificial limbs; surgical and medical instruments; surgical and orthopedic appliances, including artificial limbs; surgical appliances; surgical appliances and artificial limbs; surgical supplies and equipment not elsewhere classified orthopedic appliances
tobacco manufactures	cigarettes; cigars; cigars and cigarettes; tobacco chewing and smoking, and snuff; tobacco manufactures; tobacco stemming and rehandling; tobacco, chewing, smoking and snuff; tobacco, cigars; tobacco, cigars and cigarettes; tobacco, smoking; tobacco, smoking, and snuff
tools not including edge tools machine tools files or saws	tools (except edge tools, machine tools, files, and saws); tools, not including edge tools, machine tools, files, or saws
toys and games	games and toys (except dolls and childrens vehicles); toys (not including childrens wheel goods or sleds), games, and playground equipment; toys and games
trunks suitcases and bags	luggage; suitcases, brief cases, bags, trunks, and other luggage; trunks, suitcases, and bags
turpentine and rosin	gum naval stores (processing but not gathering or warehousing); tar and turpentine; turpentine and rosin
type founding and printing materials	printing materials; printing materials, not including type or ink; type founding; type founding and printing materials
typewriters and supplies	typewriters and parts; typewriters and supplies
umbrellas and canes	umbrellas and canes; umbrellas, parasols, and canes
upholstering materials	haircloth; upholstering materials; upholstering materials, excelsior; upholstering materials, not elsewhere classified; upholstery materials
vinegar and cider	vinegar; vinegar and cider
wallpaper	paper hangings; wall paper, not made in paper mills; wallpaper
washing machines and clothes wringers	laundry equipment, domestic; washing machines and clothes wringers; washing machines, wringers, driers, and ironing machines, for household use
waste	cotton waste; waste; waste, cotton
whips	whips
windmills	windmills; windmills and windmill towers
window shades and fixtures	window shades; window shades and fixtures
wire	wire; wire drawn from purchased rods; wire, drawn from purchased bars or rods
wirework not specified	wirework; wirework including wire rope and cable; wirework, nec
wood preserving	wood preserving
wood turned and shaped and other wooden goods not specified	cooperage; cooperage and wooden goods not elsewhere specified; wood products, nec; wood, turned and carved; wood, turned and shaped and other wooden goods, not elsewhere classified; wooden goods, not elsewhere specified

woolen worsted and felt goods and wool hats	dyeing and finishing woolen and worsted; felt goods; felt goods, wool, hair, and jute (except woven felts and hat bodies and hats); felt goods, wool, hair, or jute; hats, wool-felt; wool hats; wool pulling; wool scouring; wool shoddy; woolen and worsted goods; woolen and worsted manufactures—contract factories; woolen and worsted manufactures—regular factories or jobbers engaging contractors; woolen goods; woolen worsted and felt goods and wool hats; worsted goods

1.4.5 Income distribution by Dutch municipality

1883

The main source of the data reports the income distribution of 79 municipalities. I added data on the income distribution for 8 large municipalities with an income tax. The data for each additional cities derives from the same source as the other 79 municipalities. Table IV.8 documents the relevant year that the income distribution was measured and the source of the data.

TABLE IV.8: Sources of income distribution data for 8 additional cities

City	Year	Archive	Source
Breda	1881	Stadsarchief Breda	Municipal year report ("Gemeenteverslag") 1880
Delft	1893	Stadsarchief Delft	Municipal year report ("Gemeenteverslag") 1893
Eindhoven	1885	RHC Eindhoven	Original assessment lists, archive number 10246.925
Enschede	1880	Stadsarchief Enschede	Original assessment lists, archive number 1.1226
Hilversum	1880	Archive Prof. Van Zanden	Original assessment lists
Nijmegen	1880	Regionaal Archief Nijmegen	Overview by income class, archive number 2.14167
Utrecht	1888	Utrechts Archief	Municipal year report ("Gemeenteverslag") 1900
Vlissingen	1883	Zeeuws Archief	Original assessment lists, available here.

1.4.6 Matching the inheritance tax records to the civil registry

I first download all deaths recorded between 1879 and 1927 in the civil registry databases from four regional archives, each covering the near-universe of deaths in their province: Brabants Historich Informatie Centrum (Noord-Brabant), Collectie Overijssel (Overijssel), Gelders Archief (Gelderland), Noord-Hollands Archief (Noord-Holland). These datasets contain high quality hand-collected information on each deaths. While the type of information that was digitized varies somewhat by archive, each archive has digitized the name(s) of the decedent and their parents, the date of death, the sex, and the place of death. In all cases except Noord-Brabant, the age at death was also collected. Amsterdam is the only place in the regions covered for which digitized records of the civil death registry are not available. To maximize the amount of information available for each person that appears in the death records, I also link the civil death records to the civil marriage and birth records.

The inheritance tax records were ordered by place and date of death. Furthermore, all decedents on the same inheritance tax table share the same first letter of the surname.

For instance, Figure IV.14 shows a page for individuals with last names starting with the letter ”O”. I use this to narrow down the possible matches in the civil registry data for each person in the inheritance tax data. In record linking terminology, I use the relevant image set and the first letter of the surname as *blocking variables* for the linking between the inheritance tax records and the civil registry data. This generates for each individual in the inheritance tax records, a set of possible matches from the civil registry.

From the set of available matches, I choose the most appropriate match (if any) by using a heuristic multi-stage matching algorithm. The algorithm takes into account information on the name, date of death, and date of birth.

1.5 Details on steam engine and electric motor costs

In this section, I explain in detail the sources, assumptions and computations underlying the average and marginal cost curves of steam engines and electric motors shown in Figures I.4 and IV.2. The underlying data for steam engines, taken directly from (Emery, 1883), are displayed in Table IV.9. The data for electric motors, from (Bolton, 1926), are displayed in Table IV.10. I take these to be a full description of the costs.

TABLE IV.9: Cost parameters (in $, 1874) of steam engines of different capacities

	Purchase costs		Yearly operating costs ($)				
HP	Price ($)	Life (yrs)	Engineer	Firemen	Oil, etc.	Repairs	Coal
5	645	30	540.75		61.80	40.17	226.64
10	988	30	540.75		77.25	49.44	412.44
15	1487	30	618.00		83.43	52.53	568.33
20	1981	30	618.00		92.70	67.98	647.14
25	2441	30	695.25		101.90	83.43	752.41
50	5331	30	618.00	432.60	111.24	135.96	1202.82
100	9207	30	695.25	463.50	123.60	237.93	1898.28
150	13046	30	772.50	463.50	145.23	309.00	2718.00
200	16785	30	772.50	463.50	169.95	383.16	3603.86
250	20426	30	849.75	463.50	200.85	454.23	4504.68
300	23899	30	927.00	463.50	247.20	525.30	5406.08
400	29958	30	927.00	695.25	293.55	679.80	7207.72
500	36220	30	927.00	927.00	355.35	886.83	9009.94

Source: (Emery, 1883, p. 430).

TABLE IV.10: Cost (in £, 1925) of electric motors (squirrel-cage induction motors) of different capacities

		Purchase costs		Electricity input	
HP	Efficiency	Price (£)	Life (yrs)	kWh	£
1	0.770	12.90	15	2304	15.83
2	0.787	14.50	16	4608	31.66
3	0.800	16.20	17	6913	47.49
5	0.820	22.20	18	11521	79.15
7.5	0.833	26.80	18	17282	118.72
10	0.840	31.50	19	23042	158.30
15	0.853	39.25	19	34563	237.45
20	0.860	46.20	20	46084	316.60
25	0.870	52.80	20	57605	395.75
30	0.875	58.80	20	69126	474.90
40	0.885	69.90	20	92169	633.20
50	0.890	81.25	20	115211	791.50
60	0.900	92.00	20	138253	949.80
80	0.910	110.50	20	184337	1266.40
100	0.915	132.20	20	230421	1582.99

Notes: The price of electricity per kWh in 1925 was £0.00687 (Hannah, 1979). *Source of all other data*: (Bolton, 1926, p. 344).

Both the coal and electricity input costs are based on the assumption that the engine/motor is run at capacity 309 days per year, 10 days per hour. For steam engines, coal input data comes directly from (Emery, 1883). For electric motors, I computed the cost using electricity prices. For example, running a 1 horsepower electric motor at full capacity for 309×10 hours requires 3090 horsepower-hour, which corresponds to $0.7457 \times 3090 \approx 2304$ kWh. The price of electricity per kWh in the UK in 1925 was £0.00687.

From the data in Tables IV.9 and IV.10, I compute the annualized cost of purchase and renewal using the sinking fund formula:

$$\text{Annualized purchase cost} = \text{Price} \times \frac{r}{(1+r)^{\text{Life}} - 1}. \tag{IV.15}$$

I set the interest rate r equal to 0.05. Then, for example, the annualized cost of renewal of a 5 horsepower steam engines every 30 years becomes \$9.71. In other words, with an interest rate of 5 percent, a deposit of \$9.71 each year would yield \$645 every 30 years. From there, the total annual costs per horsepower per year are calculated as the sum of the annualized purchase costs and the yearly operating costs. Figure I.4 illustrates the data on cost per horsepower per year tabulated in Table IV.11.

TABLE IV.11: Total and per horsepower annualized cost of purchase, renewal, maintenance and operation (including and excluding of fuel) of a steam engine and electric motor of different sizes at capacity for 309 days, 10 days per hour.

	Steam engines (in 1874 \$)				Electric motors (in 1925 £)			
	Excl. fuel		Incl. fuel		Excl. fuel		Incl. fuel	
HP	Total	Per HP	Total	Per HP	Total	Per HP	Total	Per HP
1					0.60	0.78	16	21
2					0.61	0.39	32	21
3					0.63	0.26	48	20
5	652	130	879	176	0.79	0.19	80	19
7.5					0.95	0.15	120	19
10	682	68	1095	109	1.03	0.12	159	19
15	776	52	1345	90	1.29	0.10	239	19
20	808	40	1456	73	1.40	0.08	318	18
25	917	37	1670	67	1.60	0.07	397	18
30					1.78	0.07	477	18
40					2.11	0.06	635	18
50	1378	28	2581	52	2.46	0.06	794	18
60					2.78	0.05	953	18
80					3.34	0.05	1270	17
100	1659	17	3557	36	4.00	0.04	1587	17
150	1887	13	4605	31				
200	2042	10	5646	28				
250	2276	9	6780	27				
300	2523	8	7929	26				
400	3047	8	10254	26				
500	3641	7	12651	25				

Notes: To compute the cost per horsepower per year for electric motors, an efficiency loss relative to capacity that varies across sizes is taken into account (see Table IV.10).

Bibliography

BLANKEN, I. AND H. LINTSEN (1981): "Mechanische kracht in de industrialisatie van Nederland (1850-1950)," .

BOLTON, D. (1926): "The variation of efficiency with size, and the economic choice of electrical machinery," *Journal of the Institution of Electrical Engineers*, 64, 337–348.

DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 131, 799–858.

EMERY, C. E. (1883): "The Cost of Steam Power," *Transactions of the American Society of Civil Engineers*, 12, 425–431.

HANNAH, L. (1979): *Electricity before nationalisation*, Springer.

STATISTICS NETHERLANDS (1952): "Gegevens per gemeente betreffende de inkomensverdeling 1946," The Hague, Netherlands.

——— (1953): "Gegevens per gemeente betreffende de vermogensverdeling 1947," The Hague, Netherlands.

——— (2001): "Tweehonderd jaar statistiek in tijdreeksen 1800–1999," The Hague, Netherlands.

——— (2010): "Bedrijfstelling 1930," .

YOUNG, L. L. (1964): *Summary of Developed and Potential Waterpower of the United States and Other Countries of the World, 1955-62*, vol. 483, US Government Printing Office.

2 Appendix to "Jim Crow and Black Economic Progress After Slavery"

2.1 Robustness Checks

2.1.1 Adjusting Estimates for Misclassification Bias

Potential misclassification of ancestors' enslavement status could bias our estimates of the Free-Enslaved gap towards zero. It is valuable to distinguish two kinds of misclassification: false negatives, which refer to individuals incorrectly classified as formerly Enslaved despite having free paternal ancestry (due to imperfect linking rates); and false positives, which refer to individuals incorrectly classified as Free when their paternal ancestry was enslaved until the Civil War (due to incorrect links to the 1850 or 1860 census).

To mitigate misclassification bias, we use an instrumental variable (IV) approach designed to correct for both false negatives and false positives. We use our surname-based measure as an instrument for the linking-based measure. The resulting IV estimates offer an unbiased assessment of the Free-Enslaved gap, contingent upon the measurement errors in the linking-based measure being uncorrelated with the surname-based measure (Ashenfelter and Krueger, 1994; Angrist and Pischke, 2008). This assumption is plausible given that our surname-based measure is independent of census-linking methods.

The IV results suggest that measurement error reduces our initial estimates of the Free-Enslaved gap by an average of 9 percent across various outcomes (see Appendix Table IV.16). For instance, the education gap, as estimated via the IV approach, is 1.67 years—a 5 percent increase compared to the OLS estimate of 1.59 years.

We also separately address potential bias from false negatives, which is more likely to be significant due to the conservative nature of our linking approach that makes false positives unlikely. The linking criteria require both uniqueness within and matches across two census waves, based on several attributes including name, year and state of birth, sex, and race. Our methodology may incorrectly categorize many Black families as descendants of the Enslaved, particularly if they originated in slave states with a significant pre-Civil War free Black population. For instance, in Maryland, approximately 50 percent of Black Americans were free before the Civil War according to the 1860 census. In our sample, 70 percent of Black Americans with ancestors from Maryland are classified as descendants of the Enslaved in 1940—20 points more than expected.

We adjust our estimates for bias that may arise from this type of misclassification. We use that our original estimates are a weighted average of the (unknown) unbiased estimate and the non-causal estimate for free Black Americans:

$$\widehat{\beta}_{\text{original}} = \frac{\text{Enslaved}_{s,\text{links}}}{\text{Enslaved}_{s,1860}} \cdot \widehat{\beta}_{\text{unbiased}} + \left(1 - \frac{\text{Enslaved}_{s,\text{links}}}{\text{Enslaved}_{s,1860}}\right) \cdot \widehat{\beta}_{\text{free}}, \tag{IV.16}$$

where $\text{Enslaved}_{s,\text{links}}$ is the share of Black Americans who descend from the Enslaved of state s according to our classification in 1940, $\text{Enslaved}_{s,1860}$ is the true share of Black Americans who descend from the Enslaved of state s according to the 1860 census, and

$\widehat{\beta}_{\text{free}}$ is the non-causal estimate for outcomes of those with ancestors from state s.

We find that adjusting for the gap between the actual proportion of free Black individuals before the Civil War and our smaller classified share has a small impact on our Free-Enslaved gap estimates. Appendix Figure IV.24 shows that the share of Black Americans who descend from the Enslaved only deviates from our classification for three small slave states. Accordingly, adjusting our original estimates of the causal effect of each state barely affects our estimates. Even when excluding states with a high pre-Civil War free Black population, our gap estimate remains largely unchanged (see Appendix Figure IV.25).

2.1.2 Adjusting Estimates for Intermarriage

We distinguish between two estimands in our analysis: 1) the Free-Enslaved gap based on paternal enslavement ancestry, and 2) the variation in economic status of a Black individual based on the *share* of their maternal and paternal ancestors who were Free vs. Enslaved.

The Free-Enslaved gap accurately captures the former estimand, i.e., differences between Black Americans whose male ancestry line goes back to people enslaved until the Civil War vs. Black Americans whose male ancestry line goes back to people free before the Civil War.

The second estimand is more difficult to quantify and depends on the frequency of Free-Enslaved intermarriages. Some individuals who we identify as descending from the Free or Enslaved via their paternal ancestry line may descend from the opposite group via other ancestry lines. However, our estimates of the Free-Enslaved gap can be informative about this second estimand depending on intermarriage levels.

Estimating intermarriage directly is not feasible without census links for women. As an approximation, we use a person's state of birth as a proxy for their enslavement status. Using this proxy, we estimate that intermarriage was relatively rare. Specifically, the probability of a Black person's mother being born in a slave state, given that their father was also born in a slave state, is between 98 and 100 percent throughout 1870 to 1940. Conversely, for fathers born in free states, the probability that the mother was also from a free state ranges between 64 and 86 percent (while free Black Americans in free states only account for 5 percent of the Black population).

This analysis has two limitations. First, some intermarriages between ancestor regions may actually be marriages within, not across, Free-Enslaved status. For example, we show that free Black Americans in the South have a far higher likelihood to migrate North before 1940 than descendants of the enslaved. Thus, many marriages between Southern-born and Northern-born Black Americans may be Free-Free marriages, not Free-Enslaved intermarriage as classified by the birthplace proxy. Our approximation could therefore *overstate* the actual frequency of intermarriages. Second, Free-Enslaved intermarriages may also occur within region of origin, not just across those regions. Our approximation could therefore *understate* the actual frequency of intermarriages. However, the small geographic overlap between the two groups makes such intermarriage within locations less likely to be quantitatively important.

While data challenges limit our ability to provide conclusive quantitative evidence of Free-Enslaved intermarriages, historical accounts support the notion that such intermarriages were relatively rare, even within location. After the Civil War, Black Americans free before the Civil War maintained a distinct social and cultural identity, often isolating themselves from the majority of people enslaved until the Civil War:

> *"After the Civil War, the free mulatto class continued to hold itself aloof from the masses of freedmen. In Louisiana, the hostility of some members of this class to the newly emancipated blacks was so great that they opposed giving political rights to the freedmen. [...] Even in their religious affiliations, the descendants of the free mulattoes held aloof from the Negro masses. [...] The descendants of the free mulattoes became, after the Civil War, the core of a small upper class which undertook to maintain the American pattern of family life and conventional sex mores. In some small communities in the South, a single family with this social and cultural background would live in complete isolation rather than associate with the masses of Negroes"* (Frazier et al., 1957)

In conclusion, the limited available evidence suggests that intermarriages across Free-Enslaved status were relatively uncommon, primarily due to geographic and socioeconomic divides. While the Free-Enslaved gap we estimate based on paternal ancestry provides important insights, we acknowledge that in later generations, quantifying the exact share of ancestors enslaved until the Civil War poses empirical challenges.

Formally, in addition to the Free-Enslaved gap, estimated via $y_i = \alpha + \beta \cdot s_i + \varepsilon_i$, we may also be interested in $y_i = a + b \cdot share_i + e_i$, where $share_i$ is the share of i's ancestors who were slave until the Civil War. For our estimate of the Free-Enslaved gap, we have

$$\hat{\beta} \xrightarrow{p} \mathbb{E}[y|s=1] - \mathbb{E}[y|s=0] = b \cdot (\mathbb{E}[share_i|s=1] - \mathbb{E}[share_i|s=0]) \,. \tag{IV.17}$$

In the following sections, we use this expression to derive the attenuation bias that makes the Free-Enslaved gap a lower bound for the group differences between families with high vs. low shares of ancestors enslaved.

2.1.2.1 First generation after slavery For the first generation of descendants, we know that

$$\begin{aligned}
\mathbb{E}[share_{i,1}|s=1] &= 1 \cdot \mathbb{P}(share_{i,1}=1|s_i=1) + 0.5 \cdot \mathbb{P}(share_{i,1}=0.5|s_i=1) + 0 \\
&= 1 \cdot \mathbb{P}(\text{mother slave}|\text{father slave}) + 0.5 \cdot \mathbb{P}(\text{mother free}|\text{father slave}) \\
\mathbb{E}[share_{i,1}|s=0] &= 1 \cdot \mathbb{P}(share_{i,1}=1|s_i=0) + 0.5 \cdot \mathbb{P}(share_{i,1}=0.5|s_i=0) + 0 \\
&= 0.5 \cdot \mathbb{P}(\text{mother slave}|\text{father free})
\end{aligned}$$

Therefore, we have

$$\hat{\beta} \xrightarrow{p} b_1 \cdot [0.5 + 0.5 \cdot \mathbb{P}(\text{mother slave}|\text{father slave}) - 0.5 \cdot \mathbb{P}(\text{mother slave}|\text{father free})] \,.$$

If there was no intermarriage, we would have $\hat{\beta} \xrightarrow{p} b_1$.[2] If marriage between formerly enslaved families and free Black families were random—in the sense that free and enslaved fathers have an equal probability of marrying an enslaved mother—we would have $\hat{\beta} \xrightarrow{p} 0.5 \cdot b_1$.[3] Given that it is implausible that free Black men were more likely than formerly enslaved Black men to marry formerly enslaved women, it seems reasonable that $b_1 \in [\hat{\beta}, 2 \cdot \hat{\beta}]$.

We empirically assess this bias by analyzing the likelihood that a Black person descends from one parent born in a slave state and another parent born in a free state for 20-40 year old Americans in the 1910 census (whose parents were likely born towards the end of slavery). We are not able to quantify intermarriage between the formerly Enslaved and Free within state of origin because we do not have information on women's enslavement status beyond her birthplace.

We estimate that in 1910,

$$\begin{aligned} \widehat{\mathbb{P}}(\text{mother slave}|\text{father slave}) &= 0.99 \\ \widehat{\mathbb{P}}(\text{mother slave}|\text{father free}) &= 0.20, \end{aligned}$$

suggesting that the gap between individuals whose grandparents are either all formerly Enslaved or all Free could be 1.1 times as large as the Free-Enslaved gap.

2.1.2.2 Second generation after slavery If there was no intermarriage, we would have $\hat{\beta} \xrightarrow{p} b_2$. If marriage between formerly enslaved families and free Black families were random we would have $\hat{\beta} \xrightarrow{p} 0.25 \cdot b_2$. Thus, $b_2 \in [\hat{\beta}, 4 \cdot \hat{\beta}]$. The details of the derivation are available upon request.

We empirically assess this bias by analyzing the likelihood of having parents born in slave or free states for married couples between 20 and 40 years old in the 1910 census (whose parents were likely born towards the end of slavery). Our estimates suggest that the gap between individuals whose grandparents are either all formerly Enslaved or all Free could be 1.5 times as large as the Free-Enslaved gap.

2.1.2.3 $\mathbf{n^{th}}$ generation after slavery Generally, if there was no intermarriage, we would have $\hat{\beta} \xrightarrow{p} b_n$. If marriage between formerly enslaved families and free Black families were random we would have $\hat{\beta} \xrightarrow{p} 2^{-n} \cdot b_n$. Thus, $b_n \in [\hat{\beta}, 2^n \cdot \hat{\beta}]$.

Our geographic ancestry analysis from 1880 to 1940 indicates little intermarriage between slave and non-slave states even in the latest decades of our sample period. Specifically, the probability of a Black person's mother being born in a slave state, given that their father was also born in a slave state, is between 98 and 100 percent throughout this period. Conversely, for fathers born in free states, the probability that the mother was also from a free state ranges between 64 and 86 percent (while free Black Americans in free states only account for 5 percent of the Black population).

[2] Without intermarriage: $\mathbb{P}(\text{mother slave}|\text{father slave}) = 1$ and $\mathbb{P}(\text{mother slave}|\text{father free}) = 0$.

[3] With random intermarriage: $\mathbb{P}(\text{mother slave}|\text{father free}) = \mathbb{P}(\text{mother slave}|\text{father slave}) = \mathbb{P}(\text{mother slave})$.

2.1.3 Placebo Exercises

In two types of placebo exercises, we test our method of quantifying the Free-Enslaved gap. First, we estimate the placebo Free-Enslaved gap for white Americans. White families who cannot be linked to the 1850 or 1860 censuses are classified as (placebo) descendants of the Enslaved. The (placebo) Free-Enslaved gaps for white Americans are economically insignificant, especially in comparison to the actual Free-Enslaved gaps estimated on the Black population (see Appendix Figure IV.23). This also holds for a wider range of variables observed in 1940 (see Appendix Table IV.14). Note that this exercise may not yield pure placebo estimates because white families immigrating after 1860 may be different from those who immigrated earlier.

Second, we estimate the Free-Enslaved gap on the Black population using 1875 as the (placebo) end of slavery. Appendix Table IV.15 shows that this placebo Free-Enslaved gap is economically negligible. This finding is consistent with Figure II.2 which shows that there are no gaps between Black Americans who can be linked back to 1880 (but not 1870 or earlier) and those who can be linked back to 1870 or earlier.

2.1.4 The Direct Effect of Locations After Accounting for Migration

Our estimates of how being freed in a given location affected the economic progress of Black families reflects both the effect of the original location and the expected effects of future locations conditional on the 1870 location. Under a mild assumption, we can recover the treatment effect of each destination location.

Assumption 2 (No direct long-run effect of enslavement location)**.** The pre-1865 effect of enslavement location ℓ ceases to directly affect a family's descendants by 1940. That is,

$$\rho\gamma_c^0 = 0$$

where ρ is the intergenerational elasticity from 1865 to 1940 and γ_ℓ^0 is the effect that location ℓ had on Black families who lived there.

This assumption is plausible for two reasons. First, the vast majority of enslaved people were freed from slavery with little to no measured physical or human capital with little variation across locations. Second, plausible values for ρ are likely small given the high intergenerational mobility of Black Americans following the end of slavery and the amount of time that elapsed until 1940.

Under this assumption, we can recover a state's treatment effect from the originally estimated intent-to-treat (ITT) using standard instrumental variable methods in settings with multiple treatments under imperfect compliance—each treatment being a potential state of birth and non-compliance arising through migration. As described in Section 6.1, the ITT effect of location ℓ, η_ℓ, is the average of all potential future locations' treatment effects, $\gamma_{\ell'}^1$, weighted by the probability of migrating from ℓ to ℓ'. We invert the migration probability matrix to recover the effect of living in each state until 1940.

We find that the original *ITT effect* of living in a state after 1865, estimated as the causal effect of being born into slavery in that state, is almost identical to the *treatment effect* of living in the state after 1865 (see Appendix Figure IV.27). In essence, this finding results from high "compliance rates" due to limited geographic mobility in the Deep South before 1940.

2.1.5 Empirical Bayes Shrinkage

When estimating place effects with many geographic units (counties), a common problem is that some estimates may be noisy. While these estimates are unbiased, they are on average further from the truth—in a total squared error sense—than optimal (Efron, 2010). Shrinkage techniques address this problem.

Empirical Bayes methods have become a popular means to shrink noisy estimates (e.g., Angrist et al., 2017; Chetty and Hendren, 2018). The method is motivated by the fact that under the assumption of place effects resulting from a common (unknown) distribution, the optimal point estimator has the form of a Bayesian posterior mean (Armstrong et al., 2021). One does not need to make any assumptions on the specific distribution that the place effects result from.

We apply an empirical Bayes shrinkage to our baseline county effects. We provide two forms of shrinkage estimates. The first set does not use covariates, shrinking the baseline estimates toward a common mean. The second set includes covariates, shrinking the baseline estimates toward the place effect predicted by the covariates.

Figure IV.28 shows the place effects before and after shrinkage. While the negative effects are concentrated in the Lower South before *and* after, the shrunk estimates are more spatially correlated. Figure IV.29 shows the correlation of causal place effects on Black economic progress with the same places' (non-causal) effects on the outcomes of white and free Black Americans. Before and after shrinkage, there is no correlation between the effects for descendants of the Enslaved and white Americans, but a strong positive correlation between those for descendants of the Enslaved and the Free.

2.1.6 Assessing Linking Bias

Any study that uses automated linking methods faces the problem that individuals who can be linked across decades may not represent the overall population. For example, families with a high socioeconomic status may choose more unique names for their children, making it easier to create a unique match across census records. A socioeconomic gap between two sub-populations is only biased if the linking procedure differentially selects them into the sample. Table IV.12 shows that, if anything, the linking procedure biases the Free-Enslaved gap toward zero.

In addition, a family's socioeconomic status may affect not only *whether* they can be linked across decades but also *over how many decades* they can be linked. For example, children who grow up with single mothers can typically not be linked to their grandparents because women cannot be linked due to name changes at marriage. Our classification algorithm identifies descendants of the Free mainly through whether they can be linked back to 1850 or 1860, which could lead to an almost mechanically higher socioeconomic status. We addressed this concern in Section 3.4 (see Figure II.2).

One may be also concerned that the outcomes of Black men in the 1940 census depend on whether they can be linked to ancestors in the 1850 to 1880 censuses. However, Table IV.17 alleviates those concerns by showing that our linked sample of Black prime-age men is comparable to the general population of Black prime-age men. We present means both with and without conditioning on having US-born parents, the former excluding recent immigrants to maximize comparability to our linked sample. The observable characteristics of our linked sample closely align with these populations, with the exception of slightly higher labor force participation in our sample (91.7%) compared to the population's average (88.8%–90.6%).

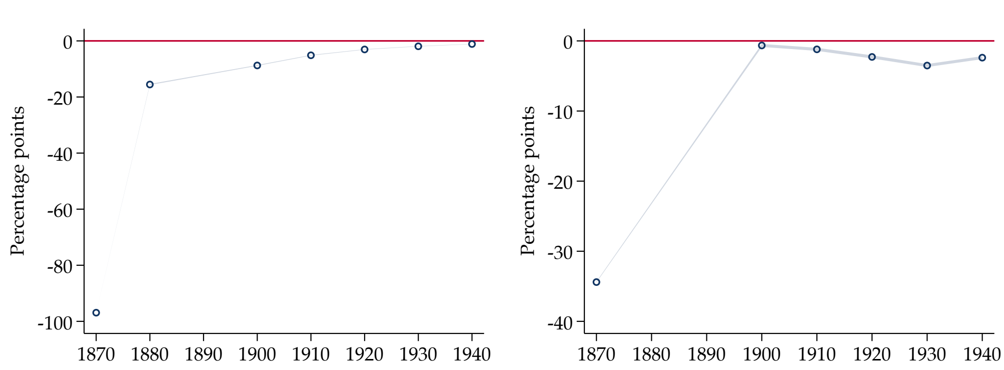
Last, one may be concerned that the effect of place in 1870 on outcomes in 1940 may be biased by differences in linking rates across those locations. In particular, areas with large Black populations may have lower linking rates because the linking relies on the *uniqueness* of a person's identifying characteristics. Lower linking rates may imply that only individuals with particularly rare names—and therefore potentially different socioeconomic statuses—are selected into the sample. Appendix Figure IV.26 addresses this concern by showing counties' average likelihood of a resident in 1870 being linkable to the 1940 census. Linking rates are similar across the country except for the most sparsely populated counties in the North (which do not contribute to our causal analysis).

2.1.7 Figures

FIGURE IV.17: Benchmark for Speed of Convergence—White Americans Whose Ancestors Did vs. Did Not Have Any Physical or Human Capital

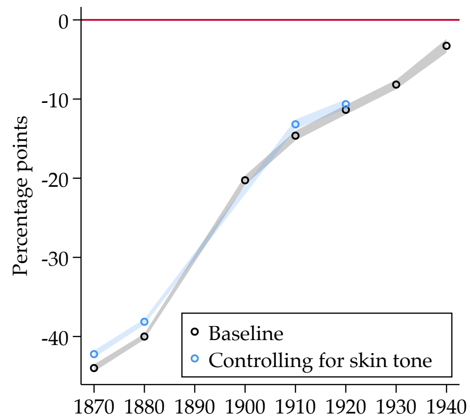
(A) Literacy

(B) Homeownership



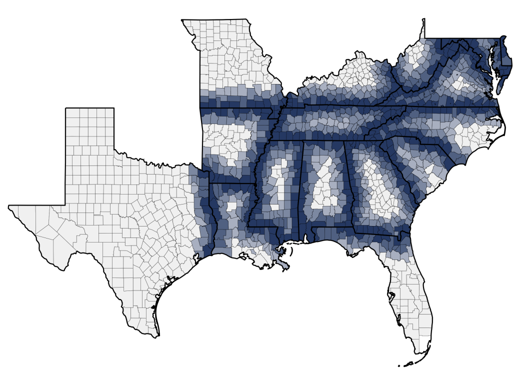
Notes: This figure shows the gaps in literacy and homeownership among white prime-age (20-54) male descendants of ancestors with vs. without any physical or human capital in 1870. Physical capital is measured in terms of real and personal property; Human capital is measured in terms of literacy. The comparison yields a benchmark for the convergence of large economic gaps from 1870 to 1940. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. All estimates control for a quadratic function in age and include 95 percent confidence bands that are clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.18: Free-Enslaved Gap in Literacy Conditional on "Mulatto"-Status



Notes: This figure shows the Free-Enslaved gap in literacy before and after including a dummy for whether a person is classified as "Mulatto" (instead of "Black") in the census. This classification does not exist in the 1900 census or any census after 1920. The sample includes both the South and North of the US. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. The sample includes only Black prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix 2.2 for details on the sample and data.

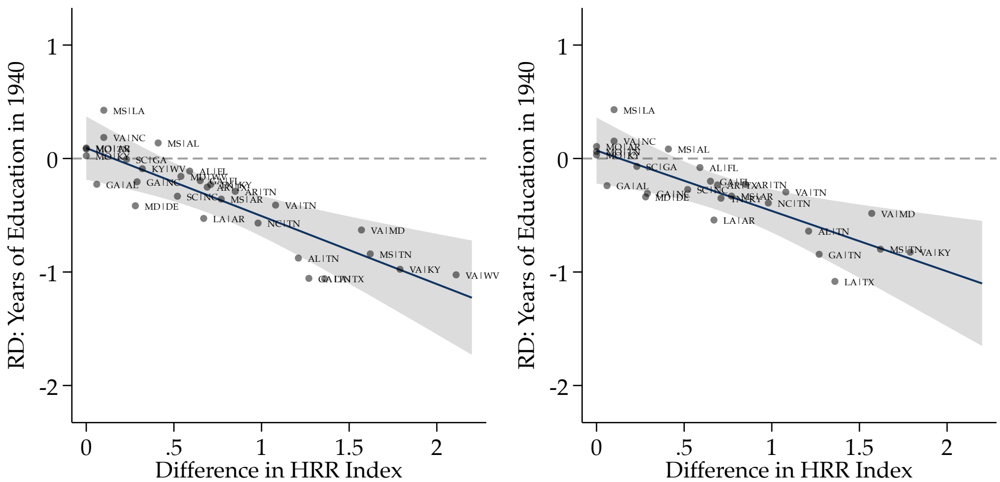
FIGURE IV.19: Southern Counties' Distance to State Borders



Notes: This map shows each county's distance to the closest state border within the South. Darker shades correspond to closer proximity to a border. Distances are measured from a county's centroid to the border. In our main analysis, we limit our analysis to counties within 100 kilometers (62 miles) of any border but show that our results are robust to other cutoffs.

FIGURE IV.20: RD Estimates Using Different Sets of Control Variables

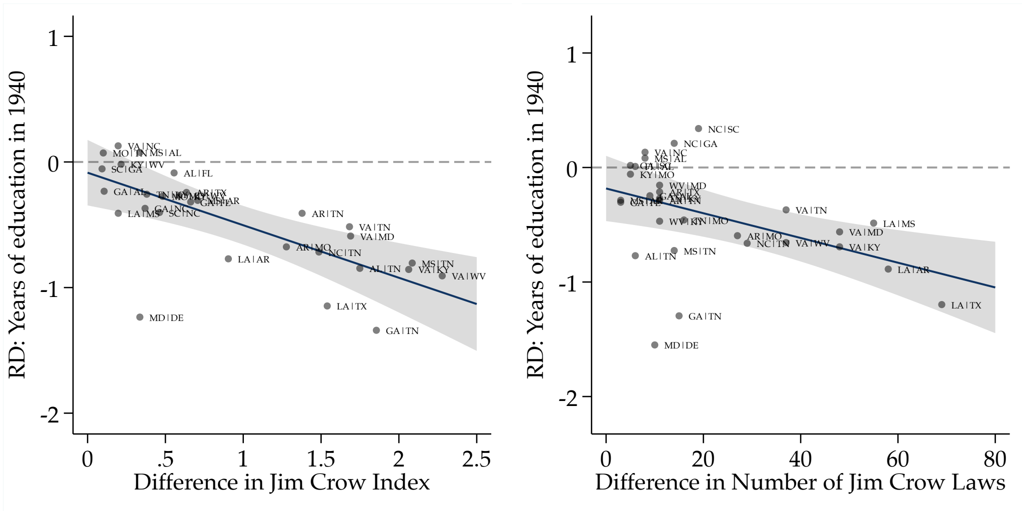
(A) Controls: demographics (B) Controls: demographics, crops, economic



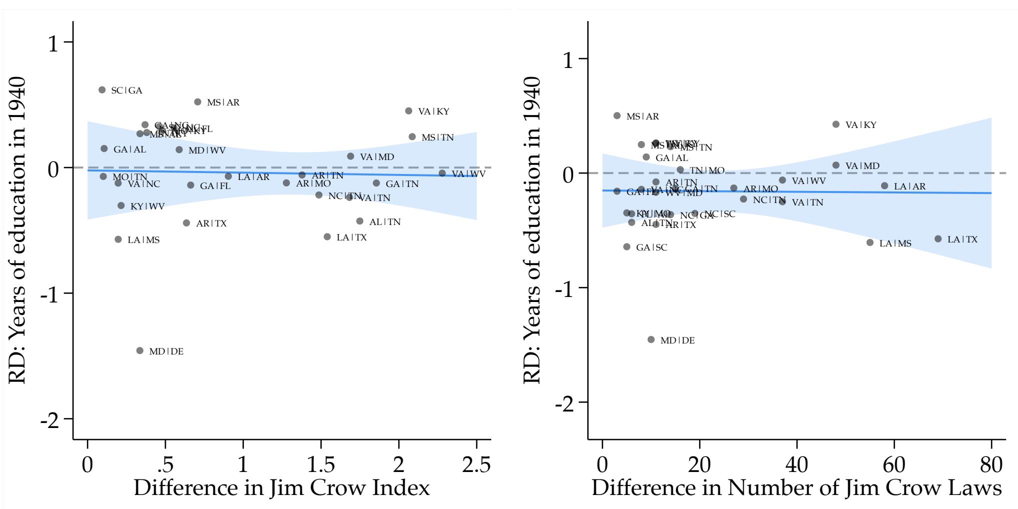
Notes: This figure shows each separate RD estimate in 1940 years of education for Black families freed across state borders with different Jim Crow intensity in 1865 after controlling for different sets of county-level variables in 1860. Panel A includes controls for the fraction Black; the fraction free among Black persons; and the age and sex of enslaved persons. Panel B includes controls for the farm share; wealth; population density; share Black; migration cost to the North; per-capita tobacco, cotton, and cane sugar output; farm values; and share slaveholders. Each label shows the more oppressive before the less oppressive state. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.21: RD Estimates Using Alternative Jim Crow Intensity Measures

(A) Black Americans



(B) White Americans

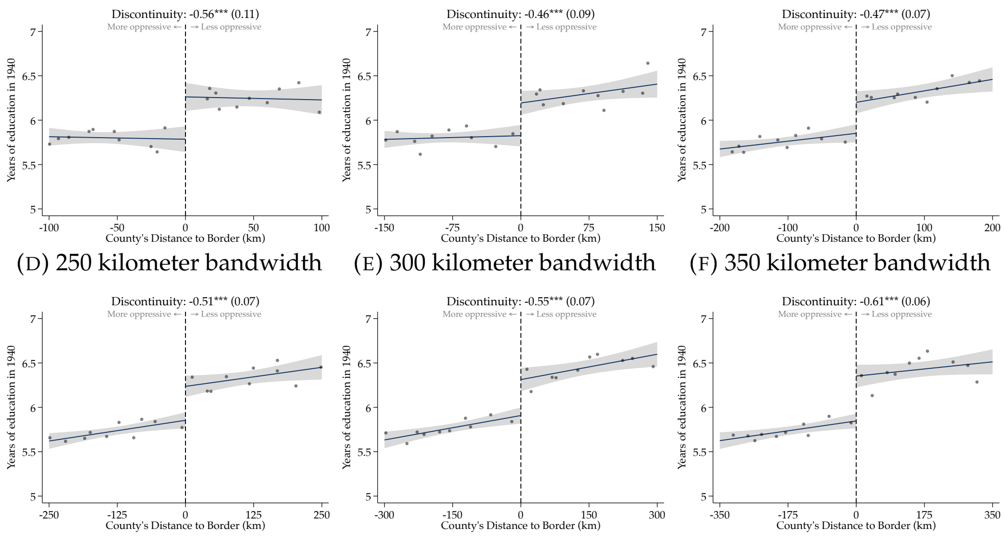


Notes: Panel A of this figure shows each separate RD estimate in 1940 years of education for Black families whose ancestors were freed on different sides of state borders in 1865. Panel B shows the same for white families depending on where their ancestors lived in 1870. Each label shows the more oppressive before the less oppressive state. Negative estimates reflect lower education in more oppressive states. Lines show the best linear fit between RD estimates and the differences in Jim Crow intensity, weighted by the inverse of the estimates' standard error. Shaded areas represent robust 95 percent confidence bands. For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.22: Different Bandwidths for Pooled RD Estimates

(A) 100 kilometer bandwidth (B) 150 kilometer bandwidth (C) 200 kilometer bandwidth

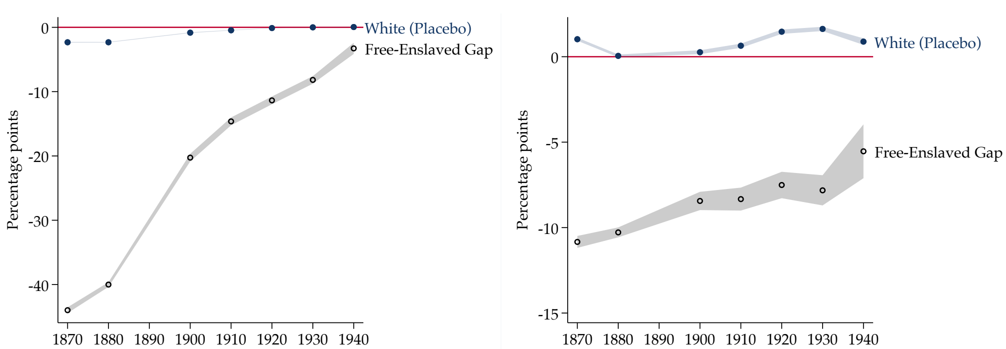
(D) 250 kilometer bandwidth (E) 300 kilometer bandwidth (F) 350 kilometer bandwidth



Notes: This figure shows the RD estimate in 1940 years of education for Black families freed across state borders with different Jim Crow intensity in 1865. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). The analysis is limited to "high-contrast borders" where Jim Crow intensity differs more than across the median border (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). Panels (A) to (D) show 100, 150, 200, 250, 300, and 350 kilometer bandwidths respectively. The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.23: Free-Enslaved Gap (1870–1940) vs. Placebo for White Americans

(A) Literacy (B) Occupational skill

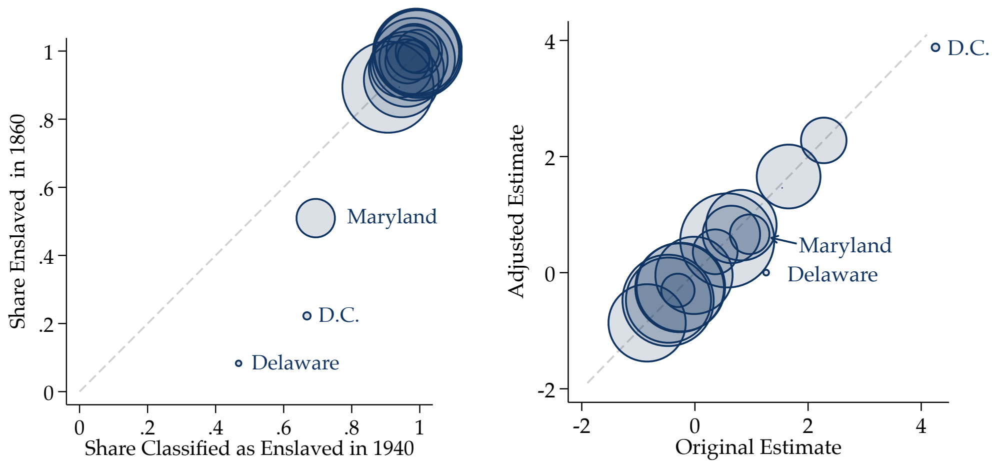


Notes: This figure shows the true and placebo gaps in literacy ratesand occupation skill levels among prime-age (20-54) male descendants of enslaved vs. free Black Americans in each census decade. The placebo applies the exact same procedure to the sample of white Americans. The comparison shows that some linking bias may affect results in early periods, but all of it vanishes over time. The sample includes both the South and North of the US. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. We assign "skilled" to occupations classified as "medium skilled workers" or above by the HISCLASS scheme (Leeuwen and Maas, 2011); and "unskilled" to others. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. All estimates control for a quadratic function in age and include 95 percent confidence bands that are clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.24: Misclassification and Bias

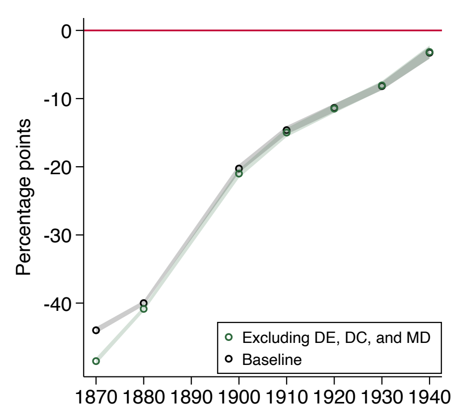
(A) Rate of misclassification

(B) Minimal bias due to misclassification



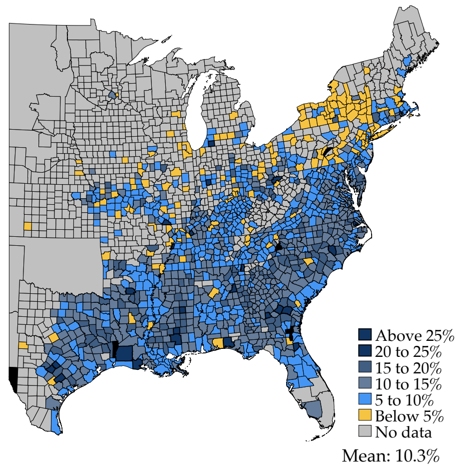
Notes: This figure assesses on misclassification of the Free-Enslaved status and the impact misclassification has on our estimates. Panel A shows the extent of misclassification as descendants of the Enslaved or the Free among Black Americans in 1940 with ancestors born in a given state before 1870. Panel B shows our causal estimates of living in each state before and after adjusting for misclassification bias. The sample includes the South of the US. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.25: Free-Enslaved Gap in Literacy (1870–1940)



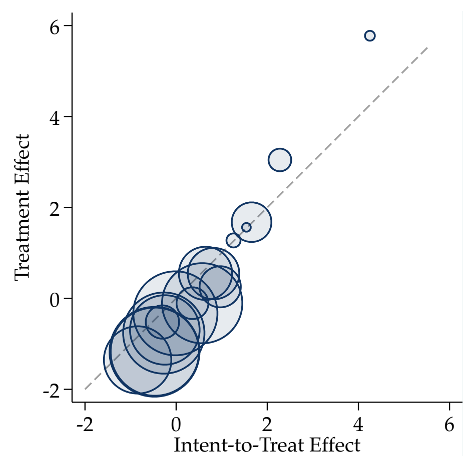
Notes: This figure shows the gaps in literacy among prime-age (20-54) male descendants of enslaved vs. free Black Americans in each census decade before and after excluding Delaware, DC, and Maryland. The sample includes both the South and North of the US. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.26: Linking Rates by County from 1870 to 1940



Notes: This figure shows the average linking rate for Black prime-age (20–54) men in 1870 to 1940. Only counties with a Black population of at least 50 prime-age men in 1870 are included.

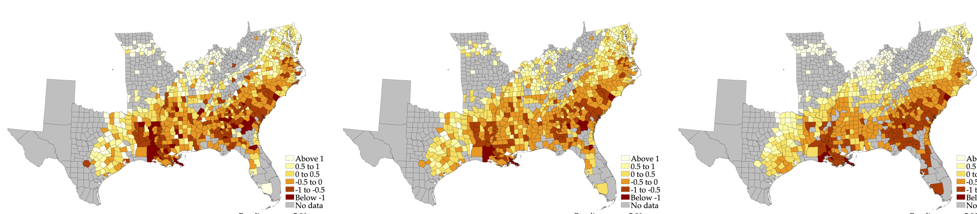
FIGURE IV.27: ITT Effect and Treatment Effect of Living in Each Southern State (1870–1940) on Years of Education in 1940



Notes: This figure compares our original (ITT) estimates of how being freed in a given state affected a Black family's economic progress to the direct treatment effect that living in that state had. The estimates are in years of education in 1940. See Data Appendix 2.2 for details on the sample and data.

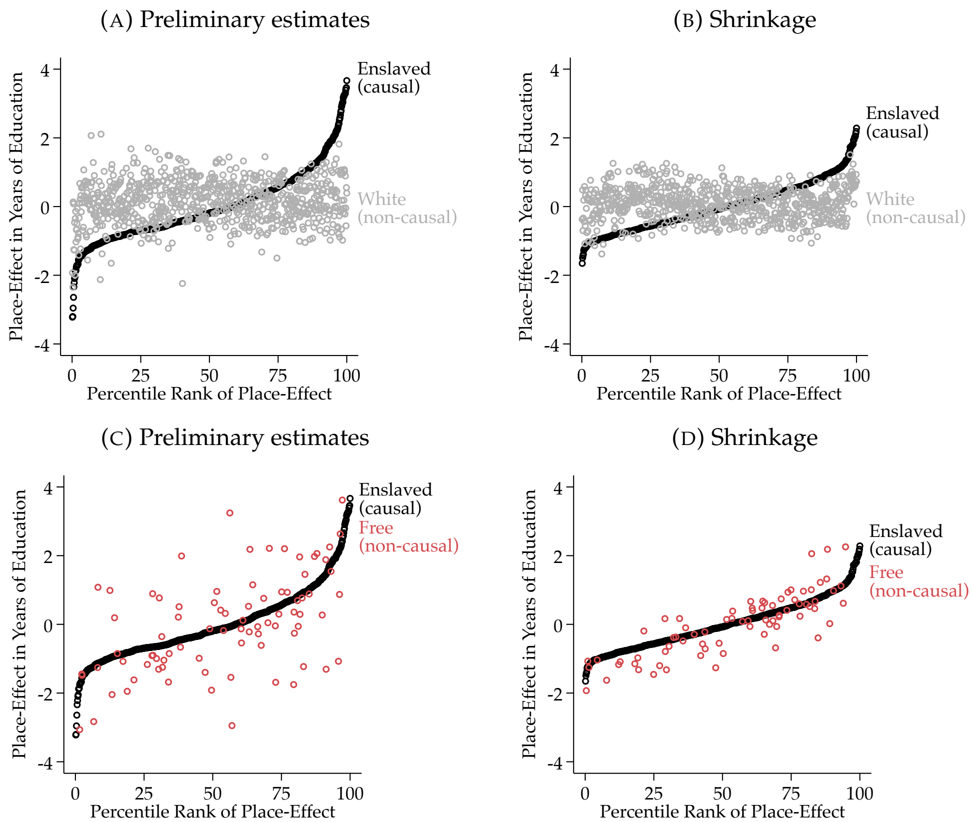
FIGURE IV.28: Causal Place Effects on 1940 Years of Education

(A) Preliminary Estimates (B) Shrinkage (No Covariates) (C) Shrinkage (Covariates)



Notes: This figure shows the 1870 ancestor county fixed effect (FE) estimates on 1940 years of education for descendants of the Enslaved. Panel A shows the preliminary estimates. Panel B shows the estimates after shrinking them to their common mean. Panel C shows the estimates after shrinking them to the regression line based on various covariates. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.29: Place Effects Across Groups Before and After Shrinkage



Notes: This figure compares the 1870 ancestor county fixed effect estimates on years of education in 1940 for descendants of the Enslaved (causal) with those of white Americans and descendants of free Black Americans (non-causal). Panels (A) and (C) show the estimates before shrinkage, Panels (B) and (D) show the shrinkage estimates. The shrinkage does not preserve a county's original rank. County-fixed effects based on ten observations or fewer are discarded. See Data Appendix 2.2 for details on the sample and data.

2.1.8 Tables

TABLE IV.12: Assessing Linking Bias

	Free (1860)			Enslaved (1870)		
	Linked	Population	Δ	Linked	Population	Δ
Literacy (%)	65.1	66.8	-3%	20.4	20.4	0%
Occupation Score	6.0	6.1	-1%	3.7	3.8	-1%
Real property ($)	1,217	1,230	-1%	1,400	1,270	10%
Personal property ($)	312	316	-1%	312	293	6%
Lives in North (%)	45.1	52.1	-13%	7.8	8.2	-4%
Lives on Farm (%)	21.2	18.2	17%	23.8	23.2	3%
Observations	20,994	79,374		190,676	726,667	

Notes: This table shows that there is little selection into the linked sample. If anything, the linked sample is negatively selected for the Free and positively selected for the formerly Enslaved, attenuating the Free-Enslaved gap toward zero. The left panel compares the Free who can be linked to any future decade to the entire 1860 population (which only contains free Black Americans). The right panel compares our linked sample to the 1870 population (89 percent of whom were enslaved until 1865).

TABLE IV.13: Free-Enslaved Gap Based on the Distribution of Surnames (1940)

	Education (Years) Mean: 5.70		Wage Income (USD) Mean: 588.60		Homeownership (%) Mean: 21.53		House Value (USD) Mean: 1,616.81	
P(Ancestor Enslaved until Civil War)	**-1.25*****	**-1.40*****	**-88.36*****	**-113.15*****	**-1.95****	**-2.31****	**-1,098.68*****	**-1,194.53*****
	(0.07)	(0.09)	(21.22)	(25.50)	(0.87)	(1.05)	(237.09)	(282.83)
Name-measure	Exact	NYSIIS	Exact	NYSIIS	Exact	NYSIIS	Exact	NYSIIS
Controls (age, age^2)	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.03	0.03	0.01	0.01	0.01	0.01	0.00	0.00
Observations	2,598,739		2,842,572		2,618,795		556,422	

Notes: This table repeats Table II.1 showing the gap in years of education, total income, homeownership, and house value among prime-age (20-54) male descendants of enslaved vs. free Black Americans in 1940. Without record linkage, we cannot assure that all Black families in the sample were present in the US during both slavery and Jim Crow. However, we weight observations in the 1940 census to hold the distribution of surnames constant at its 1870 level. The sample includes both the South and North of the US. The sample includes the entire universe of prime-age Black men, not just those linkable. The coefficients can be interpreted as a 100 percentage point increase in the likelihood of descending from the Enslaved based on their (exact) surname. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Robust standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.14: Placebo Free-Enslaved Gap (1940) for White Americans

	Education (Years) Mean: 9.76	Wage Income (USD) Mean: 892.68	Home Ownership (%) Mean: 49.74	House Value (USD) Mean: 3,284.56
Placebo	**-0.17*****	**-1.68**	**0.09**	**12.17**
	(0.00)	(1.04)	(0.05)	(9.63)
Baseline Free-Enslaved gap	-1.59***	-145.92***	-7.24***	-694.69***
Controls (age, age^2)	Y	Y	Y	Y
Adjusted R^2	0.03	0.06	0.01	0.00
Observations	5,015,270	4,770,969	5,012,884	2,425,204
Ancestor Free	*3,158,604*	*3,001,138*	*3,155,980*	*1,536,909*

Notes: This table shows the placebo gaps in years of education, total income, homeownership, and house value among prime-age (20-54) male white Americans in 1940. The placebo applies our linking-based method to measure a person's (placebo) Free-Enslaved status. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.15: Placebo Free-Enslaved Gap (1940)

	Education (Years) Mean: 5.99	Wage Income (USD) Mean: 380.61	Home Ownership (%) Mean: 29.21	House Value (USD) Mean: 1,368.20
Placebo	**0.04***	**-6.84*****	**-0.01**	**-76.89****
	(0.02)	(2.44)	(0.26)	(30.66)
Baseline Free-Enslaved gap	-1.59***	-145.92***	-7.24***	-694.69***
Controls (age, age^2)	Y	Y	Y	Y
Adjusted R^2	0.03	0.04	0.01	0.00
Observations	162,387	153,368	163,195	46,574
Ancestor Free	*75,583*	*71,474*	*76,048*	*21,873*

Notes: This table shows the placebo gaps in years of education, total income, homeownership, and house value among prime-age (20-54) male Black Americans in 1940. The placebo uses 1875 as the (placebo) year of Emancipation, applying our linking-based method to measure a person's Free-Enslaved status. The sample includes both the South and North of the US. Only observations that can be linked to the 1870 or 1880 census are included. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.16: Free-Enslaved Gap (1940): IV Design to Reduce Measurement Error in Enslavement Status

	Education (Years) Mean: 6.08	Wage Income (USD) Mean: 390.18	Home Ownership (%) Mean: 29.71	House Value (USD) Mean: 1,422.37
IV: Ancestor Enslaved	**-1.67*****	**-170.12*****	**-9.69*****	**-554.68*****
until Civil War	(0.15)	(17.69)	(1.89)	(149.68)
OLS: Ancestor Enslaved	-1.59***	-145.92***	-7.24***	-694.69***
Controls (age, age^2)	Y	Y	Y	Y
F-Statistic (weak id.)	2,077.22	1,998.63	2,049.38	994.86
Adjusted R^2	0.05	0.05	0.01	0.01
Observations	158,032	149,252	158,787	45,311
Ancestor Free	*9,078*	*8,551*	*9,070*	*3,227*

Notes: This table shows instrumental variable (IV) estimates of the gap in years of education, wage income, homeownership, and house value (conditional on ownership) among prime-age (20–54) male descendants of enslaved vs. free Black Americans in 1940. We use our surname-based measure of a Free-Enslaved status as an instrument for our linking-based measure. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.2 Data Appendix

2.2.1 Individual-Level Outcome Variables

Our main outcome variables can be categorized as (proxies of) income, education, or wealth. Most individual-level data draw on census records provided through IPUMS (Ruggles et al., 2020). We use additional individual-level data from a major US credit bureau to extend our results to 2023.

Income

TABLE IV.17: Sample Balance of 1940 Sample Linked to Ancestors 1850–1880

	Linked Sample	Population	
	Black prime-age men linked to ancestors 1850–80	Black prime-age men with US-born parents	Black prime-age men
Literacy (%)	91.5	92.5	89.9
Years of education	6.0	6.4	5.7
LFP (%)	91.7	88.8	90.6
Wage income ($)	381.2	296.3	399.7
Occupation Score	4.9	4.6	4.9
Homeownership (%)	29.3	31.4	21.8
House value ($)	1,372.0	1,288.4	1,632.2
Urban (%)	47.0	44.4	53.7
Lives in North (%)	22.3	20.6	25.5
Lives on Farm (%)	36.1	39.7	29.2
Observations	168,138	327,393	3,000,331

Notes: This table compares our sample of Black prime-age (20–54) men linked to ancestors in 1850, 1860, 1870, and/or 1880 to the overall population of Black prime-age men in the census. The first population column conditions on having US-born parents according to the 1940 census; the second column includes all Black prime-age men. Note that in the 1940 census, parents' birthplace was a "sample-line" feature, available only for a random subset of the population.

- **Occupational income scores, 1850–1940 (census).** Because the census does not include any continuous measure of income before 1940, researchers have instead relied on occupational income scores. The most popular version, "occscore," reflects the median total income of a person in that occupation in 1950.

- **Lido income scores, 1850–1940 (Saavedra and Twinam, 2020).** Occupational income scores do not contain any age-, sex-, or race-specific information. The recent literature has used regression and machine learning techniques to improve on the traditional occupational income score (e.g., Saavedra and Twinam, 2020; Abramitzky et al., 2021a). We use the Lido score constructed by Saavedra and Twinam (2020). The authors constructed it using machine learning techniques using 1950 and 2000 census data to validate their results against occscore in the 1915 Iowa census. According to Abramitzky et al. (2021a), the Lido score has a correlation of 0.99 with their own measure.

- **Occupational skill, 1850–1940 (Leeuwen and Maas, 2011).** We use HISCLASS, a classification to compare occupations based on the skill they typically required. The classification ranges from "higher managers" to "unskilled farm workers." We coarsen this classification by assigning "skilled" to every occupation classified as "medium skilled workers" or above and "unskilled" to everyone else.

- **Wage income, 1940 (census).** We use wage income for 1940, the only year it is available for in our sample period.

- **Predicted total income, 2019–2023 (credit bureau).** Measures a household's gross total compensation for the most recent year reported. This measure is estimated based on proprietary data and prediction models. For more details, see Appendix 2.2.3.

- **Predicted disposable income, 2019–2023 (credit bureau).** Measures a household's income available to spend, invest, or save after accounting for fixed expenses. This measure is estimated based on proprietary data and prediction models. For more details, see Appendix 2.2.3.

- **Hourly job, 2019–2023 (credit bureau).** Measures whether a person is employed as an hourly or salary worker.

Education

- **Literacy, 1850–1940 (census).** We use literacy for all years. In 1940, literacy becomes unavailable, and instead the census starts to include educational attainment. We proxy for literacy by having completed at least the second grade. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate.

- **Years of education, 1940 (census).** We impute years of education from the highest educational level attained ("educd").

- **High school, 1940 (census).** We impute whether a person holds a high school degree based on whether they completed at least 12 years of schooling ("educd").

- **College, 1940 (census).** We impute whether a person holds a college degree based on whether they completed at least 16 years of schooling ("educd").

- **Graduate, 1940 (census).** We impute whether a person holds a graduate degree based on whether they completed at least 17 years of schooling ("educd").

Wealth

- **Personal property, 1860–1870 (census).** Measures "the contemporary dollar value of all stocks, bonds, mortgages, notes, livestock, plate, jewels, and furniture" as reported to the census. It is not clear whether zeros indicate missing values or true zero personal property, and therefore we replace zeros with "missing."

- **Real property, 1850–1870 (census).** Measures "the contemporary dollar value of any real estate owned by the respondent" as reported to the census. It is not clear whether zeros indicate missing values or true zero personal property, and therefore we replace zeros with "missing."

- **Homeownership, 1850–1940 (census).** Measures whether the individual rents or owns their home. For 1900 to 1940, the census reports homeownership directly. For 1850 to 1870, we follow Collins and Margo (2011) in imputing homeownership status using information on wealth, where every household with positive real property is classified as owner-occupied. Collins and Margo (2011) exempt households who live in multi-family homes from this classification but the information necessary to follow them in doing so is not included in the full-count version of the census we use. However, creating homeownership proxies using their and our method yields a correlation of 0.9733 in the 1 percent sample.

- **House value, 1930–1940 (census).** Measures the house value conditional on owning the house.

- **Credit score, 2019–2023 (credit bureau).** The VantageScore® 3.0 measures a person's credit health. The score takes into account a rich set of indicators on a person's financial situation. It ranges from 300 to 850. Scores above 700 are typically considered "good" and scores below 550 "very poor."

2.2.2 Neighborhood-Level Outcome Variables

While we cannot link our data to censuses after 1940, we can link the 1940 census to administrative mortality records from 1988 and 2007 using the CenSoc-Numident file (Goldstein et al., 2021). Importantly, the mortality records contain the nine-digit ZIP codes of residence at the time of death. We link these codes to statistical census geographic areas, i.e., census tracts, block groups, and blocks (see Section 2.2.7 for more detail on the procedure). Census tracts contain between 1,200 and 8,000 people and are designed to be "relatively homogeneous units with respect to population characteristics, economic status, and living conditions" (Census Bureau, 2017). Block groups (between 600 and 3,000 people) and blocks are subdivisions of a census tract.

We assigned to each decedent various economic characteristics based on these statistical areas at the time of death. Since the sample is about evenly split between deaths before 2000 and deaths after 2000, we used the aggregated census data for the year 2000 from the NHGIS database. For variables from other sources, we selected the data to refer to a period as close to 2000 as availability allowed.

One potential concern with this data may be that many people live in retirement homes, possibly making the neighborhood a less precise proxy of a person's economic status. To assess this potential issue, we compare the density of deaths with a ZIP code's population density and find that the two are highly correlated ($\rho = 0.91$). Our results are robust to excluding ZIP codes that have far higher rates of deaths than predicted by their population density.

Income

- **Income, 2000 (NHGIS).** The median household income by race of householder. Available by ZCTA, census tracts, and block groups.

Wealth

- **House value, 2000 (NHGIS).** The median value of owner-occupied housing units by race of householder. Available by ZCTA and census tracts.

- **Homeownership, 2000 (NHGIS).** The share of occupied housing units that is occupied by the owner (relative to a renter) by race. Available by ZCTA, census tracts, block groups, and blocks.

Education

- **High school degree, 2000 (NHGIS).** The share of the population over 25 years old by race and sex who hold a high school degree. Available by ZCTA, census tracts, and block groups.

- **College degree, 2000 (NHGIS).** The share of the population over 25 years old by race and sex who hold a college degree. Available by ZCTA, census tracts, and block groups.

Demographics

- **Age at death, 1988–2007 (BUNMD, Goldstein et al., 2021).** The median age at death by race and sex. Available by five-digit ZIP code, census tracts, block groups, and block.

- **Percentage Black, 2000 (NHGIS)**. The share of the population that is Black. Available by ZCTA, census tracts, block groups, and blocks.

2.2.3 Credit Bureau Sample

We analyze data from a major US credit bureau, which includes comprehensive monthly credit reports for individuals from January 2010 to the present. These reports, updated on the final Tuesday of each month, contain information from various sources, such as financial institutions, debt collection agencies, and public records, along with proprietary data. Our focus is on the March 2023 snapshot.

Our sample is restricted to Black prime-age (20-54) men. The credit bureau uses a predictive method to determine race, based on 1) a person's first and last name and 2) their detailed neighborhood (nine-digit ZIP code). Names are analyzed both in terms of their frequency across racial groups as well as for prefixes and suffixes that may contain information about the ethnic origin of a person. A person's neighborhood of residence allows the credit bureau to leverage information on the racial composition of the neighborhood.

This method, given the detailed geographic information it leverages, is far more accurate than common proxies that rely solely on surnames. Using a separate dataset—our Social Security mortality records—we find that surnames capture 22 percent of the variation in whether a person is Black or not; nine-digit ZIP codes capture 76 percent; and both combined capture 90 percent.

The bureau combined our probabilistic surname-based classification of Free-Enslaved status of Black individuals with their credit reports, subsequently anonymizing the data. We access these anonymized individual-level credit reports for around 550,000 Black prime-age men whose names were successfully merged to our Free-Enslaved classification via a secure server, allowing real-time estimation of the Free-Enslaved gap in employment and credit. Based on our continuous surname-based measure of ancestors' enslavement status, the average likelihood of descending from free Black Americans across our credit bureau sample is 9.5 percent—close to the share of Black Americans recorded as free in the 1860 census: 11 percent.

The credit bureau predicts individual income using a comprehensive set of demographic, financial, and property data aggregated from various sources, including banks and insurance providers. Because this income prediction relies on models and data proprietary to the credit bureau, our ability to validate the predictions are limited. However, recent work using similar credit bureau data validate the accuracy of these predictions using payroll records (Mello, 2023). The credit bureau's income prediction model consists of two main components. First, predicted salary is based on the credit bureau's proprietary database of payroll records. Second, predicted financial income, which includes income from investments, businesses, and retirement, is estimated using various data from the credit bureau and its partners. The credit bureau's internal validation exercises show that predicted incomes are predictive of individuals' consumption patterns, such as purchasing a luxury car. Moreover, the distribution of predicted incomes aligns with the income distribution documented by the census.

2.2.4 Jim Crow Database

We build a rich dataset on states' Jim Crow regimes by combining newly collected information on Jim Crow laws and existing data on states' institutions and outcomes directly affected by those institutions, including voter participation and educational resources.

2.2.4.1 Jim Crow Index As an alternative to the Historical Racial Regime (HRR) index to measure the intensity of each state's Jim Crow regime, we introduce a composite metric—the "Jim Crow index." This index is constructed using principal component analysis and encompasses multiple factors, each serving as a proxy for specific aspects of anti-Black institutions. Our index builds on the HRR index from Baker (2022) but focuses on institutional factors and the Jim Crow era specifically.

Our new Jim Crow index is based on five factors. The first factor is the anti-Black share of race-related laws a state passed until 1950. For this measure, we collected new information on laws that mention race or color and classify those laws as to whether they are anti-Black discriminatory or not (see next section). The second factor is a state's number of disenfranchisement devices (i.e., literary tests, poll tax, grandfather clause, and white primary; Walton et al., 2012; Baker, 2022). The third factor is a state's share of congressional delegates that signed the Southern Manifesto (Baker, 2022). The fourth factor is the racial gap in states' school year lengths—i.e., the legislative term length of Black schools relative to that of white schools (Card and Krueger, 1992). The fifth and final factor is the year in which a state introduced legislation for minimum teacher pay—legislation central to narrowing the large wage penalty historically suffered by Black teachers (Card et al., 2022; Cascio and Lewis, 2022).

Appendix Table IV.20 presents each state's Jim Crow index alongside the corresponding input variables. The Deep South—Mississippi, Louisiana, Georgia, South Carolina, and Alabama—emerge as the most oppressive according to our index. Notably, Louisiana ranks in the top quartile of most oppressive states across all measures. In contrast, the border states—Delaware, West Virginia, Kentucky, Maryland, and Missouri—are categorized as the least oppressive.

We consider a variety of alternative measures for states' Jim Crow intensity. Figure

IV.33 shows the correlations between different proxies of Jim Crow intensity (discussed in the following two sections). While these measures are very different in nature and capture both de jure and de facto aspects of Jim Crow, they are correlated and using them, we consistently arrive at the same conclusions. Key outcomes directly affected by Jim Crow institutions are also highly correlated with our Jim Crow index: overall votes cast per adult male between 1900 and 1940 ($\rho = -0.89$, not available by race) and our causal effects on long-run economic progress of Black families ($\rho = -0.93$).

2.2.4.2 New Database on Jim Crow Laws We collect information from 800 Jim Crow laws from four sources, covering both race-specific and "race-blind" Jim Crow laws. We first digitize a comprehensive collection of laws that refer to race and color by state in 1950 Murray (1950). We categorize the laws as discriminatory, anti-discriminatory, or neutral. We restrict our sample to discriminatory laws and further categorize the domain they pertain to, such as education, suffrage, or employment. Our remaining sources add Jim Crow laws that made no explicit mention of race. We collect laws that limited geographic mobility and regulated employment arrangements from Roback (1984) and Cohen (1991). We further collect laws that restricted suffrage from Walton et al. (2012). Appendix Figure IV.39 shows the number of total Jim Crow laws passed by each state until 1950. Appendix Figure IV.40 shows the distribution over years in which Southern governments passed laws of different types.

2.2.4.3 Other Data on Jim Crow Regimes **Historical Racial Regime (HRR) Index.** As our main measure of a state's Jim Crow intensity, we use the HRR index (Baker, 2022). This index "measures different manifestations of the US racial regime across different historical periods—slavery and Jim Crow—and is based on state-level institutions including slavery, sharecropping, disfranchisement, and segregation." It is a principal component of four factors: a state's share of the population enslaved in 1860, its number of disenfranchisement devices, the share of sharecroppers who were Black in 1930, and the share of Congressional delegates who signed the Southern Manifesto.

Votes cast per adult male. As a second alternative measure for the intensity of Jim Crow regimes, we compute a county's aggregate votes cast per adult male in decennial presidential elections in the South from 1900 to 1940 (ICPSR, 1999; Bernini et al., 2023). We divide the total number of votes cast in each election by a county's total population (see panel A of Appendix Figure IV.36). Data on the number of votes cast by race are not available. Panel A of Appendix Figure IV.55 shows border discontinuities in votes cast per adult male.

Black school quality index. Last, as a third alternative measure for the intensity of Jim Crow regimes, we construct an aggregate measure of Black school quality in the South (Card and Krueger, 1992). We extract a principal component from three measures of Black school quality by state prior to 1940: student-teacher ratios, term lengths, and teacher wages. We also use individual-level data on Black teachers' wages from the 1940 census to assess whether or not Black school quality differed sharply across state borders (see panel B of Appendix Figure IV.36). Appendix Figure IV.54 shows border discontinuity estimates in Black teachers' education and wages.

2.2.5 Identifying Descendants of the Free and Enslaved

2.2.5.1 Main Method: Linking Historical Census Records Figure IV.41 illustrates our new method to identify descendants of the Free and descendants of the Enslaved in census records between 1870 and 1940. It mainly relies on census-linking methods (Abramitzky et al., 2021b) but also uses information on place and year of birth.

The method consists of three steps. First, we identify the Free themselves before identifying their descendants. In 1850 and 1860, the enslaved population was excluded from the individual-level censuses. By definition, every Black American included in the census was therefore free before 1865. We link the 1850 and 1860 censuses forward to all census decades between 1870 and 1940 and then classify every Black American who can be linked to 1850 or 1860 as free.

In addition to linking, we use information on place and year of birth in our classification algorithm. All Northern states had begun banning or restricting slavery by 1804—some of them decades earlier. Any Black person born in those states was either free upon birth or would be emancipated by a certain age (typically in their 20s). While the latter case opens up the possibility of a Northern-born Black person being sold into slavery in other states before their emancipation, this possibility was ruled out by law.

In Appendix Table IV.21, we compare the de jure to the de facto status of slavery in the North. As a de facto measure, we show the number of slaves in the state in absolute numbers and as a fraction of the state's Black population. Based on this evidence, we classify any Black American born outside of the slave states after 1804 and before 1865 as Free. In addition, we use the state-specific years in which slavery was abolished or restricted in non-slave states to go even further back in time.

Second, we identify the *descendants* of the Free by using information on the relationship between individuals within census households. Specifically, we classify Black people with a free Black American *ancestor* as being descendants of the Free. Any person without a free ancestor is classified as a descendant of the Enslaved. In 1940, the final year of our sample, we identify 9,400 descendants of the Free and 155,800 descendants of the Enslaved. Because we can only link men, the descendant classification is determined exclusively through the male ancestry line.

2.2.5.2 Alternative Method of Free-Enslaved Classification: Distribution of Surnames While our main method provides a high-accuracy classification of descendants of the Free and Enslaved, accuracy comes at the cost of reduced sample sizes due to imperfect linking rates across the decades. To use the full census sample of Black Americans after 1870, rather than a linked sub-sample thereof, we develop an additional strategy for identifying descendants of the Free and Enslaved based on surnames. Figure IV.31 shows that the name-based measures are highly correlated with the Free-Enslaved status based on our preferred measure, though they are attenuated as expected.

Our alternative classification algorithm uses changes in the distribution of surnames from 1850–1860 to 1870–1880. Before 1865, the census only included free Black Americans—after, it also included the formerly Enslaved and their descendants. Census pooling (1850 and 1860; 1870 and 1880) reduces the impact of imperfect coverage in any given decade.

We compute the relative frequency of each surname before and after 1865. We then create a measure of how likely a person is to descend from the Free by dividing their surname's relative frequency before 1865 by its relative frequency after 1865. For example, the surname Du Bois appears with relatively high frequency in the 1850 and 1860 censuses, while Freedman does not appear at all. After the four million formerly enslaved individuals entered the census sample in 1870 and 1880, the name Du Bois is far less (one-tenth) frequent, whereas a substantial number of individuals entered the sample with the surname Freedman for the first time. These changes suggest that anyone named Du Bois after 1865 likely descends from the Free, whereas anyone named Freedman likely descends from the Enslaved. Note that not all names give us a good idea of whether a person descends from the Enslaved or not. Some names very common among Black Americans before 1865, such as Johnson, Brown, or Smith, remain very common after 1865. Other names such as Washington did exist among Black Americans before 1865 but became more common after many newly freed enslaved people chose this name in honor of the country's first president.

Formally, using the example of the surname Du Bois, we estimate the name-specific likelihood of descending from free Black Americans defined as

$$\begin{aligned} P(Free_{it} = 1 | Name_i = \#DuBois_t) &= \frac{P(Free_{it} = 1, Name_{it} = \#DuBois_t)}{P(Name_{it} = \#DuBois_t)} \\ &= \frac{P(Free_{i,1860} = 1, Name_{i,1860} = \#DuBois_t)}{P(Name_{i,1870} = \#DuBois_t)} \\ &= \frac{P(Name_{i,1860} = \#DuBois_t)}{P(Name_{i,1870} = \#DuBois_t)}, \end{aligned}$$

where the second equation follows from assuming that a surname conveys a constant probability of descending from free Black Americans. The last equation follows from the fact that the 1860 census only contained free Black Americans. This equation can be approximated by

$$\widehat{P}(Free_{it} = 1 | Name_{it} = \#DuBois_t) = \frac{\#(\#DuBois_t)_{1860} / BlackPop_{1860}}{\#(\#DuBois_t)_{1870} / BlackPop_{1870}},$$

where $\#DuBois_t$ is the number of individuals with the surname Du Bois in a given year and $BlackPop_t$ is the population of all Black Americans (free and enslaved). Before 1865, we compute the population by adding up the census sample size (the Free) and the number of the Enslaved (Berlin, 1974). We truncate our estimated probability by 0 and 1. Names that only appear pre-1865 but not post-1865 are assigned probability 1; those that only appear post-1865 are assigned probability 0. Appendix Table IV.18 shows a Black person's probability of descending from ancestors who were enslaved until 1865, given their surname.

To allow for misspellings, we also compute this measure based on the phonetics of surnames. Specifically, we transform surnames using the New York State Identification and Intelligence System (NYSIIS) phonetic code. For example, the surnames "Browne" and "Brown" both become "Bran." For placebo exercises, we also compute the above measure as a pseudo-probability of being free for white Americans as well as for 1875 as a time placebo for Emancipation.

2.2.6 County Characteristics

We compile a dataset on county characteristics combining data from the IPUMS National Historical Geographic Information System (NHGIS, Manson et al., 2021), the census (Ruggles et al., 2020), and various other sources.

- **Age of enslaved people, 1860 (NHGIS).** Enslaved people's average age within a county.
- **Agricultural output, 1860 (NHGIS).** County's value of total agricultural output in USD per capita.
- **Share of Black population, 1860 (NHGIS).** Share of county's 1860 population that is Black.
- **Distance to the North, East (NHGIS).** County's distance to the North and the East is proxied by its centroid's latitude and longitude.
- **Farm share, 1870 (NHGIS).** Fraction of county's population living on a farm in 1870.
- **Farm value, 1860 (NHGIS).** County's value of farms in USD.
- **Free share, 1860 (NHGIS).** Percentage of county's 1860 Black population that is free.
- **Intergenerational mobility, 1996–2012 (Chetty and Hendren, 2018).** Causal effect of a county on the expected rank in the national income distribution conditional on one's parents' income ranking at the 25th percentile during childhood.
- **Intergenerational mobility, 1994–2015 (Chetty et al., 2020).** Non-causal effect of a commuting zone on the expected rank in the national income distribution conditional on one's parents' income ranking at the 25th percentile during childhood. We use estimates specific to Black Americans.
- **Lynchings, 1883–1941 (Seguin and Rigby, 2019).** Number of lynchings that occurred in a county between 1883 and 1941.
- **Migration cost North, 1870 (Donaldson and Hornbeck, 2016).** Transportation cost through land and water ways from a given county to the Northern cities that were the main destinations of the Great Migration: Chicago, Detroit, Pittsburgh, and New York. The migration cost estimates are based on the 1870 railroad network.
- **Occupational income, 1860 (census).** County's average occupational income score among prime age (20-54) men.
- **Plantation crop share, 1860 (NHGIS).** County's value of cotton, tobacco, sugar, and rice output as a share of the total value of agricultural output.
- **Population density, 1870 (NHGIS).** County's 1870 population per square kilometer area.

- **School, 1870 (NHGIS).** Fraction of county's Black children (ages 6–16) attending school in 1870.

- **Slaves per capita, 1860 (NHGIS)**. Average number of enslaved people per capita.

- **Tobacco, cotton, rice, and sugar, 1860 (NHGIS).** Value of a county's tobacco, cotton, rice, or sugar output in USD per capita in 1860.

- **Top-1% wealth share, 1860 (census).** County's top-1% share of white Americans' wealth, including real property and personal property. To compute the top-1% share, we restrict the sample to white prime-age men (20-54).

- **Votes cast per adult male, 1860–1940 (ICPSR, 1999; Bernini et al., 2023).** Number of votes cast in decennial Presidential elections from 1860 to 1940 as a share of the total population eligible based on sex and age (men aged 21 or older).

- **Wealth Gini index, 1860 (census).** County's Gini index of white Americans' wealth, including real property and personal property. To compute the Gini index, we restrict the sample to white prime-age men (20-54).

2.2.7 Nine-Digit ZIP to Census 2000 Crosswalks

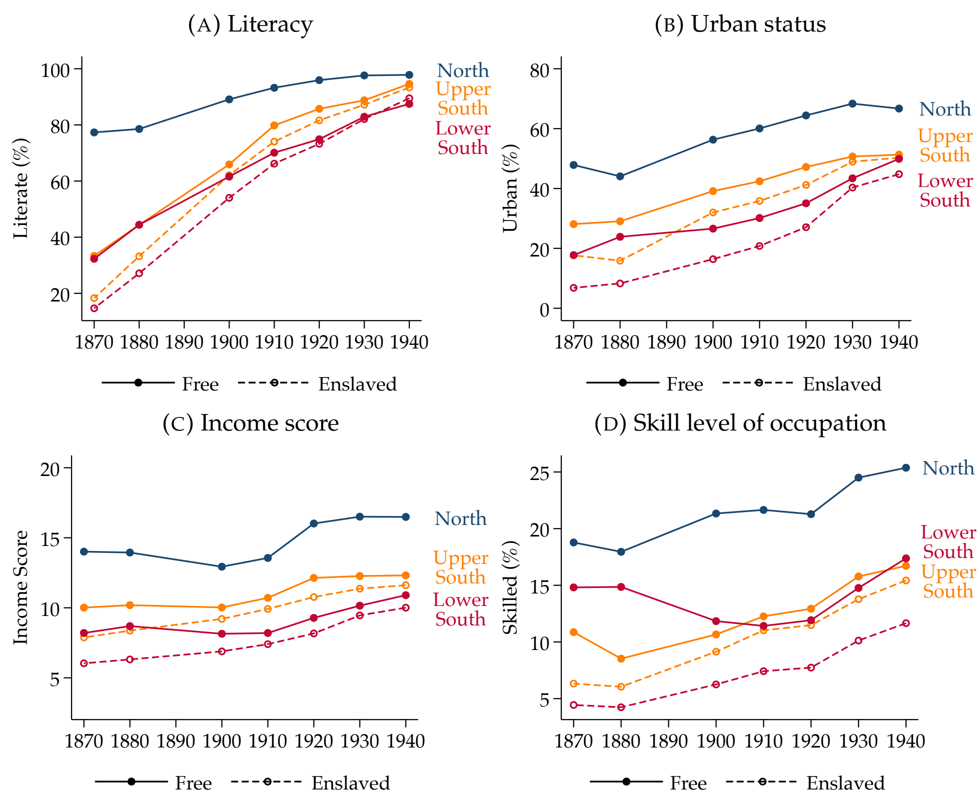
The administrative mortality records contain nine-digit ZIP codes ("ZIP9") of the place of residence at the time of death. We use the Census Bureau's TIGER/Line ASCII files from 1994 to 2006 to link ZIP9s to 2000 census statistical areas (i.e., census blocks, block groups, and census tracts). A ZIP9 comprises a range of addresses, usually a side or segment of a street.

In most cases, a ZIP9 maps into a unique block (and hence maps into a unique block group and census tract). For instance, in 2000, 81 percent of ZIP9s were matched to a unique block. For block groups and census tracts, 96 percent and 97 percent of the ZIP9 matches were unique, respectively. In cases where a ZIP9 occurs in more than one statistical area, we assign the area that has the largest number of matches with the relevant ZIP9. This yields a one-to-one mapping of ZIP9s to blocks. However, not all ZIP9s in the mortality records occur in the TIGER/Line files. To improve the coverage, we sort the data by ZIP9 for each version and interpolate the census statistical areas in case the next non-missing census area is exactly equal to the previous non-missing area (using that the ZIP9s are ordered geographically).

Using this procedure, we link around 84 percent of the decedents with ZIP9s to a census tract, 82 percent to a block group, and 77 percent to a block. For decedents for which we can find the census area corresponding to their ZIP9 both before and after their death, the agreement rate between the different versions is high (98 percent for census tracts, 96 percent for block groups, and 88 percent for blocks).

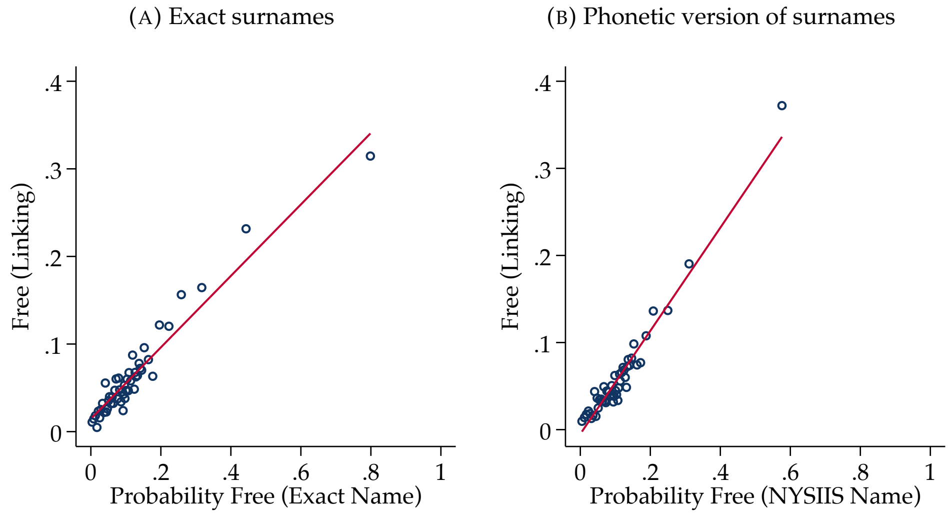
2.2.8 Figures

FIGURE IV.30: Socioeconomic Characteristics of Family by Region of Origin (1870–1940)



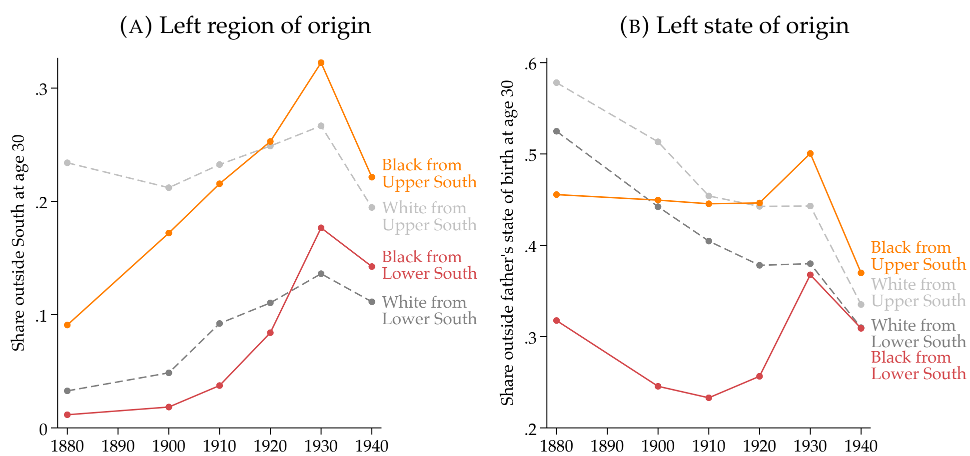
Notes: This figure shows the averages of characteristics in the cross-section of prime-age male descendants of the Free and Enslaved by their ancestor's region (family's residence pre-1880). Incomes Score uses the Lido score developed by Saavedra and Twinam (2020). In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. We assign "skilled" to occupations classified as "medium skilled workers" or above by the HISCLASS scheme (Leeuwen and Maas, 2011); and "unskilled" to others. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.31: Comparing Name-Based and Linking-Based Measures



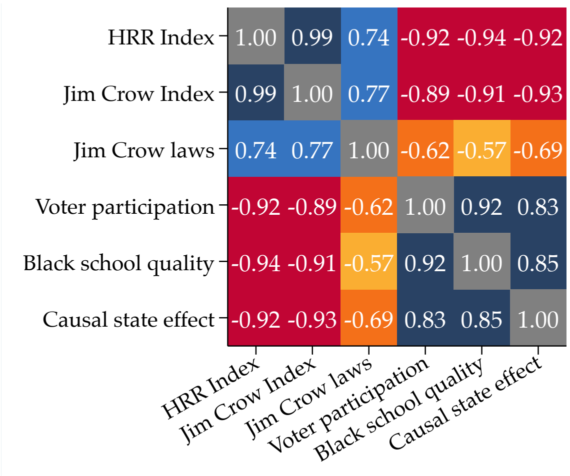
Notes: This figure compares the probabilistic measures of descending from free Black Americans with our preferred measure based mainly on census linking. This binned scatter plot shows that among Black prime-age men in the 1940 census, the fraction of people classified as Free closely coincides with the predicted probability based on the people's surnames.

FIGURE IV.32: Long-Term Migration Rates across Regions and States by Race



Notes: This figure shows the fraction of Black and white individuals aged 30 who have migrated from their father's birth region (panel A) or father's birth state (panel B) in each census year. The data is derived from the 1850–1940 censuses, focusing on the Southern-born fathers' states of birth, and does not require census linking.

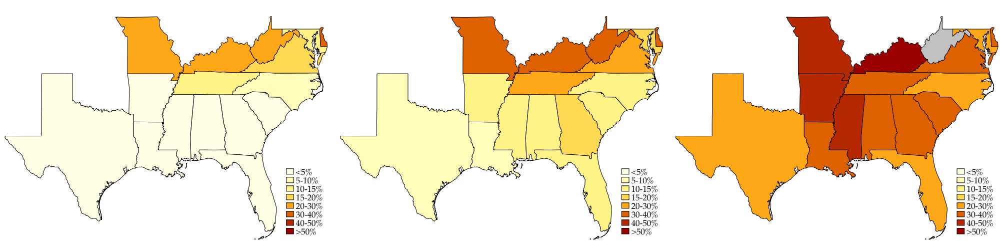
FIGURE IV.33: Correlations Between Proxies of Jim Crow Intensity



Notes: This figure shows the correlation between a state's Historical Racial Regime (HRR) index (Baker, 2022), Jim Crow index, number of Jim Crow laws, votes cast per adult male (ICPSR, 1999; Bernini et al., 2023), quality of Black schools (Card and Krueger, 1992), and causal 1870-ancestor state effects on Black Americans' 1940 years of education as shown in panel A of Appendix Figure IV.44.

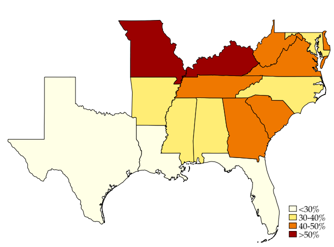
FIGURE IV.34: Black Families Leaving the Slave States by 1870 State of Origin

(A) By 1920 (B) By 1940 (C) By 2000



Notes: This figure shows the cumulative fraction of Black families who live outside the slave states, by the state their 1870 ancestor was born. The figure highlights that the first wave of the Great Migration from 1910 to 1940 was mainly an Upper Southern phenomenon (see Panels A and B). Black families with roots to the Lower South only caught up with those rates of migration to the North after 1940 (see panel C).

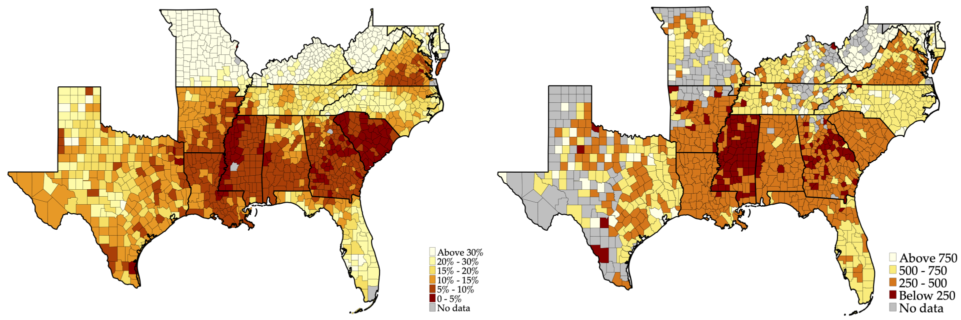
FIGURE IV.35: Black Families Leaving their 1870 State of Origin by 1940



Notes: This figure shows the fraction of Black families who in 1940 live outside the state in which their ancestors were enslaved. As the state of enslavement, we use the state of birth of formerly enslaved ancestors in the 1870 census.

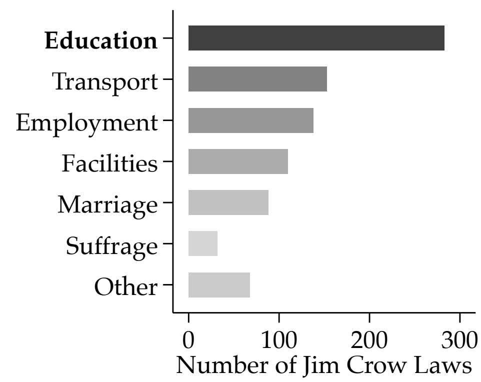
FIGURE IV.36: Outcomes Directly Targeted by Jim Crow Differ Sharply Across States

(A) Votes cast per adult male (1900–1940) (B) Black teachers' median wages (1940)



Notes: Panel A of this figure shows the average fraction of each county's population that cast a vote in decennial Presidential elections between 1900 and 1940. Panel B of this figure shows the median annual wage income of Black teachers in the 1940 census for each Southern county. Results for the Black-white ratio in teachers' median annual wage income are very similar and available upon request. Appendix Figure IV.54 shows border discontinuity estimates in both outcomes.

FIGURE IV.37: Jim Crow Laws by Type



Notes: This figure shows the number of Jim Crow laws across Southern states that pertain to each category. See Data Appendix 2.2 for details on the data.

FIGURE IV.38: County Population of Enslaved and Free (1790)

(A) Enslaved (B) Free

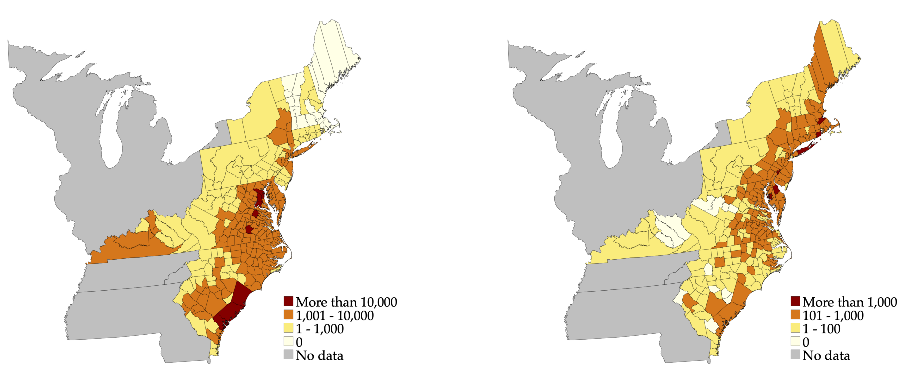
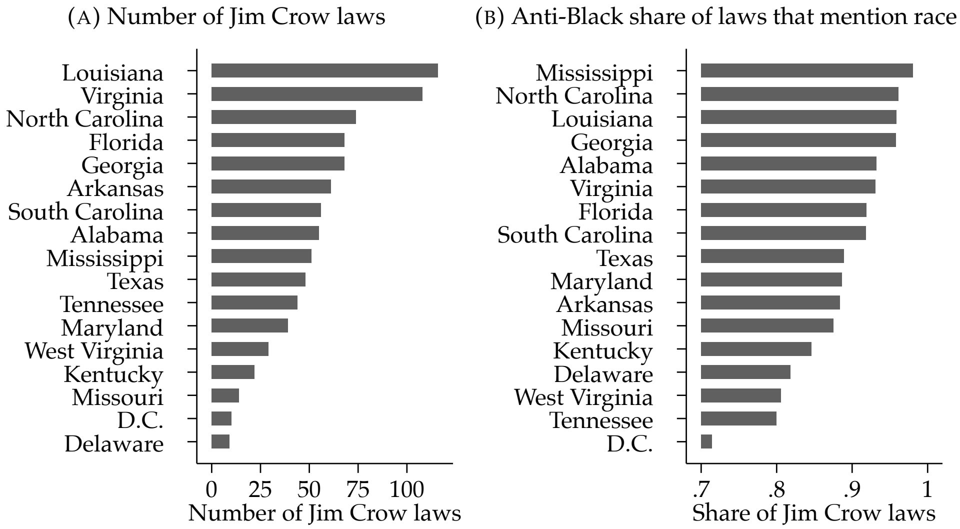


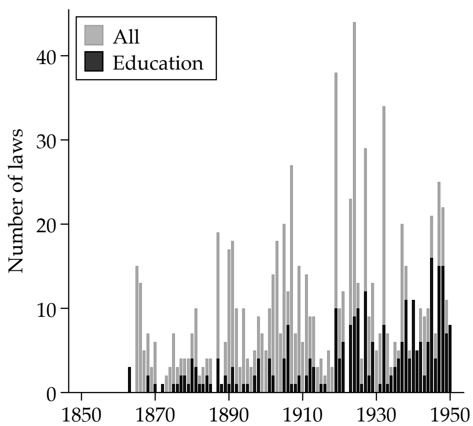
FIGURE IV.39: Jim Crow laws by State



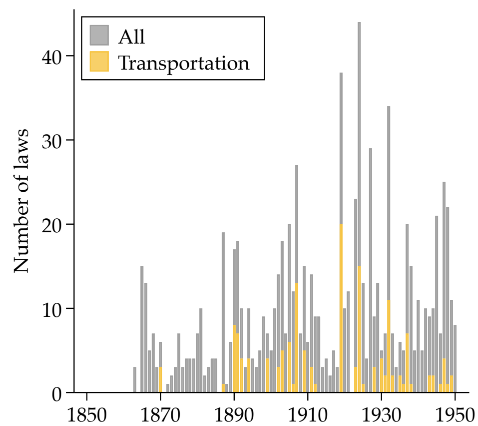
Notes: Panel A of this figure shows the cumulative number of Jim Crow laws passed by state until 1950. Panel B shows the anti-Black discriminatory share of all race-specific laws a state passed until 1950. We categorized each law as discriminatory (Jim Crow) or not based on its content and context provided by other sources.

FIGURE IV.40: Annual Jim Crow Laws Passed Across the South by Type

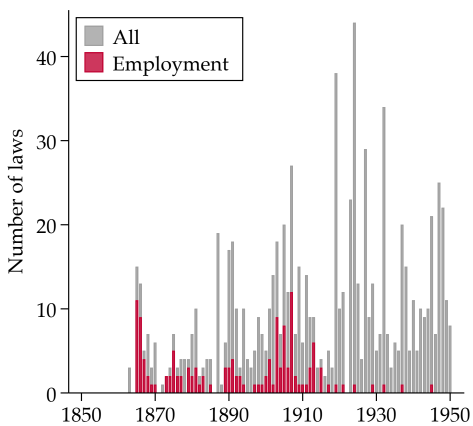
(A) Education



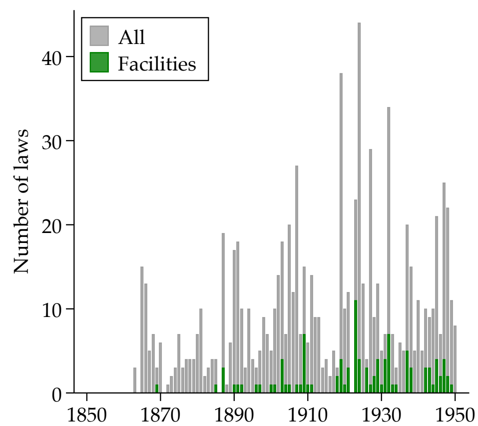
(B) Public transport



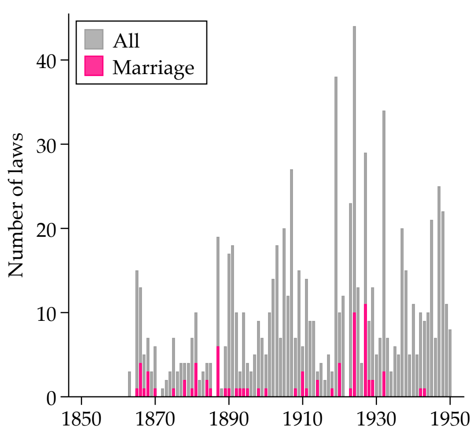
(C) Employment & geographic mobility



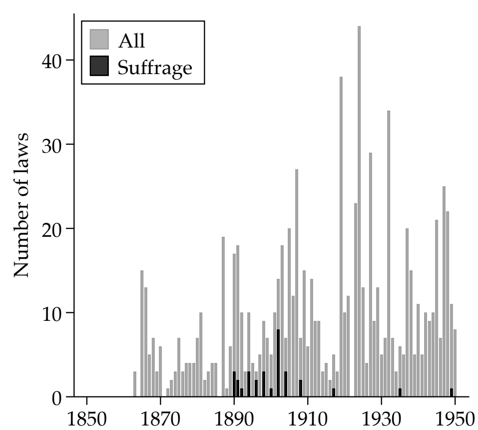
(D) Facilities



(E) Racial intermarriage

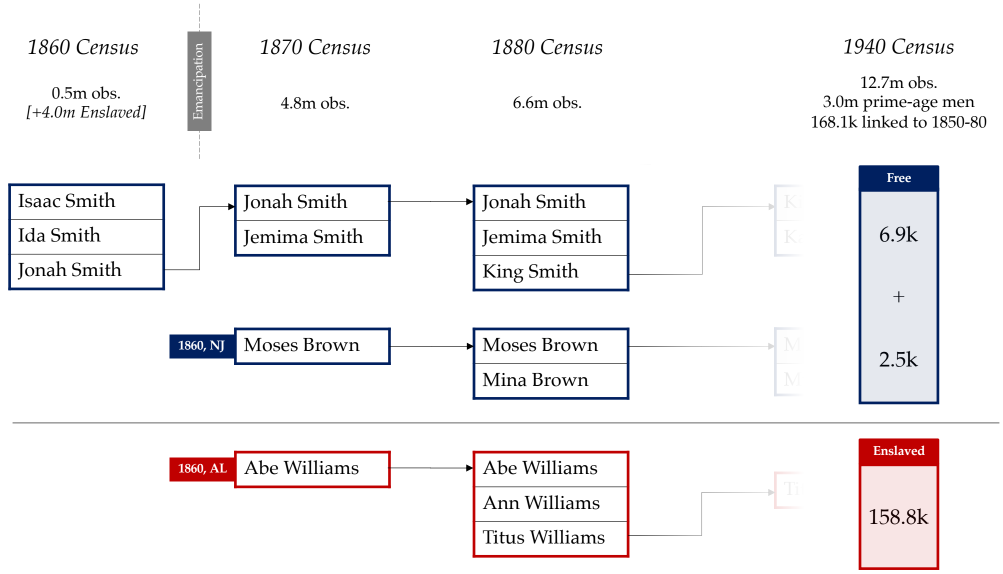


(F) Suffrage



Notes: This figure shows the number of Jim Crow laws passed by type across all Southern states and years.

FIGURE IV.41: Illustration of Our Free-Enslaved Classification Algorithm



Notes: This figure illustrates our new method to identify descendants of the Free and Enslaved in census records 1870–1940. The names are chosen are arbitrary examples and do not reflect real data. Jonah Smith is identified as a descendant of the Free because he can be linked back to the 1860 census; Moses Brown because he was born in a state (New Jersey) that had abolished slavery by the time of his birth (1860). Abe Williams does not fall into either category and is therefore classified as formerly enslaved or a descendant of the Enslaved. The Free-Enslaved status is assigned to descendants based on their male ancestor. In 1940, the final year of our sample, we identify 9,400 descendants of the Free (6,800 through direct linking to 1850–1860 and 2,600 through their ancestor's birthplace) and 155,800 descendants of the Enslaved. While not comprehensively illustrated here, we do link across all adjacent and non-adjacent census records of 1850, 1860, 1870, 1880, 1900, 1910, 1920, 1930, and 1940.

2.2.9 Tables

TABLE IV.18: Selected Surnames and Enslavement Status

Surname	Likelihood Enslaved
Wanamaker	0%
Du Bois	1%
Cumberland	2%
Dewitt	6%
Radcliffe	10%
McCollins	16%
Dupas	21%
Freemann	28%
Butcher	44%
Freeman	66%
Tubman	70%
Baptiste	85%
Jackson	86%
Broom	87%
Douglass	87%
Johnson	87%
Smith	89%
Carter	90%
Robinson	90%
Hamilton	91%
King	91%
Morrison	91%
Williams	91%
Hughes	92%
Jefferson*	92%
Marshall	92%
Baldwin	94%
Jordan	94%
Lincoln	95%
Knowles	96%
Washington*	96%
Cooks*	97%
Broadnax*	99%
Boykins*	100%
Doyley*	100%
Gadson*	100%
Freedman	100%
Merriweather*	100%
Rockingham*	100%

Notes: This table shows estimates of the probability of descending from enslaved Black Americans by surname (conditional on being Black). Some of the examples (marked by *) are mentioned by Clark (2014), who lists a number of surnames that "sound classically English" but tend to be predominantly Black today, suggesting that they were likely "adopted in the slavery era from masters whose own families died out or left few descendants." Consistent with that idea, our estimates suggest that Black people with those surnames are almost certain to descend from ancestors who were enslaved until the Civil War.

TABLE IV.19: Family Tree's Linking Rates

	Individual		Family
	Adjacent only	Incl. non-adjacent	
1870 to 1900	12.8%	25.9%	**27.6%**
1870 to 1910	3.5%	19.4%	**24.8%**
1870 to 1920	1.1%	12.3%	**26.0%**
1870 to 1930	0.3%	6.2%	**14.2%**
1870 to 1940	0.1%	3.1%	**9.8%**

Notes: This table shows the linking rates for Black men from 1870 to each decade from 1900 to 1940. The first column shows the linking rate when conditioning on finding a person in each adjacent decade (e.g., 1870 to 1900 would require a person to be linked from 1870 to 1880 and from 1880 to 1900). The second column shows the linking rate when allowing for intermediate decades to be skipped (e.g., 1870 to 1900 would require a person to be linked *either* from 1870 to 1880 and from 1880 to 1900 *or* from 1870 to 1900 directly). The third column shows the linking rate when linking either the individual or their ancestors or descendants in the same household (again, allowing intermediate decades to be skipped).

TABLE IV.20: The Jim Crow Index

State	**Jim Crow Index**	Share of laws discriminatory	Disenfranchisement devices	Southern Manifesto	Black-white ratio in term length	Minimum teacher pay introduced
Louisiana	**1.33**	96%	4	100%	0.77	1948
Mississippi	**1.14**	98%	3	100%	0.78	1924
South Carolina	**1.00**	92%	3	100%	0.76	1945
Georgia	**0.91**	96%	4	100%	0.91	1937
Alabama	**0.80**	93%	4	100%	0.89	1927
Virginia	**0.73**	93%	4	100%	0.95	1946
North Carolina	**0.54**	96%	4	71%	0.96	1919
Arkansas	**0.43**	88%	2	100%	0.88	1957
Florida	**0.24**	92%	2	80%	0.96	1955
Texas	**-0.21**	89%	2	21%	0.93	1949
Missouri	**-0.85**	88%	0	0%	1.05	1985
Tennessee	**-0.95**	80%	1	36%	0.99	1925
Maryland	**-0.96**	89%	0	0%	0.96	1904
Delaware	**-1.29**	82%	0	0%	1.00	1919
Kentucky	**-1.33**	85%	0	0%	1.05	1912
West Virginia	**-1.54**	81%	0	0%	1.00	1882

Notes: This table shows each states' Jim Crow index, ordered from most to least oppressive. The Jim Crow index is a principal component extracted from five factors, as shown in the remaining columns. The top-quartile (most oppressive) is highlighted in red; the bottom-quartile (least oppressive) in blue.

TABLE IV.21: Abolition of Slavery in the North

Year	State	*De Jure* Abolition of Slavery	*De Facto* Number of Slaves	
			Year	Total
1777	Vermont	Slavery was banned immediately upon founding of Vermont (Constitution of Vermont, 1777).	1790 1800 1810 1820 1830 1840 1850	0[4] 0 0 0 0 0 0
1780	Pennsylvania	Law of gradual emancipation passed in 1780 (Pennsylvania General Assembly, 1780). Black Americans born to enslaved mothers after 1780 would be freed at age 28. Slavery was ended in 1847.	1790 1800 1810	3,737 (36%) 1,706 (10%) 795 (3%)

Table IV.21: Abolition of Slavery in the North

Year	State	*De Jure* Abolition of Slavery	*De Facto* Number of Slaves: Year	*De Facto* Number of Slaves: Total
			1820 1830 1840 1850	211 (1%) 403 (1%) 64 (0%) 0
1781	Maine Massachusetts	Slavery was abolished by Supreme Judicial Court rulings in three related court cases, collectively known as the "Quock Walker case" (Cushing, 1961; Zilversmit, 1968). Slavery was ruled incompatible with the new state constitution of 1780.	1790 1800 1810 1820 1830 1840 1850	0 0 0 0 3 (0%) 0 0
1783	New Hampshire	Similar to Massachusetts, New Hampshire's constitution essentially abolished slavery by stating "all men are born equal and independent" (Constitution of the State of New Hampshire, 1783). However, it is not clear whether court rulings indeed interpreted the constitution as being at odds with slavery or not.	1790 1800 1810 1820 1830 1840 1850	158 (20%) 8 (1%) 0 0 3 (0%) 1 (0%) 0
1784	Rhode Island	Law for gradual emancipation passed in 1784 (General Assembly of Rhode Island, 1784). Black Americans born to enslaved mothers after 1784 would be freed at age 18 (women) or 21 (men).	1790 1800 1810 1820 1830 1840 1850	952 (22%) 381 (10%) 108 (3%) 48 (1%) 17 (0%) 5 (0%) 0
1784	Connecticut	Law for gradual emancipation passed in 1784 (Connecticut General Assembly, 1784). Black Americans born to enslaved mothers after 1784 would be freed at age 25. This age was lowered to 21 in 1797. Slavery was abolished in 1848.	1790 1800 1810 1820 1830 1840 1850	2,759 (50%) 951 (15%) 310 (5%) 97 (1%) 25 (0%) 17 (0%) 0
1787	Ohio Indiana Illinois Michigan Wisconsin Minnesota	The Confederation Congress's Northwest Ordinance of 1787 both banned and enforced slavery (Confederation Congress, 1787). A clause allowed Northerners to capture and enslave runaway slaves. Slavery was abolished by Ohio in 1802, Indiana in 1816, and Illinois in 1818.	1790 1800 1810 1820 1830 1840 1850	– 135 (21%) 429 (28%) 1,106 (40%) 788 (5%) 348 (1%) 0
1799	New York	Law for gradual emancipation passed in 1799 (New York State Legislature, 1799). Black Americans born to enslaved mothers after 1799 would be freed at age 25 (women) or 28 (men). In 1817, state decided to free all slaves born before 1799 (but not their children) in 1827 (New York State Legislature, 1817).	1790 1800 1810 1820 1830 1840 1850	21,324 (82%) 20,343 (66%) 15,017 (37%) 10,088 (26%) 75 (0%) 4 (0%) 0
1804	New Jersey	Law for gradual emancipation passed in 1804 (New Jersey State Legislature, 1804). While not freeing living slaves, Black Americans born to enslaved mothers after 1804 would be freed at age 21 (women) or 25 (men).[5]	1790 1800 1810 1820 1830 1840 1850	11,423 (81%) 12,422 (74%) 10,851 (58%) 7,557 (38%) 2,254 (11%) 674 (3%) 236 (1%)

Notes: This table provides a timeline for the abolition of slavery in the North. The first column indicates the year which we choose as the states' final year of slavery. We classify any Black American born in the state after this cutoff as free. The third column shows the laws that abolished slavery. In many cases, slavery was not abolished outright, but rather it was restricted in ways that would imply a person is free before 1865 in all likelihood. The final column shows the actual number of slaves who reside in the state and the percentage of the state's Black population being enslaved in parentheses. The number of slaves is taken from aggregate counts in census records (1790–1850).

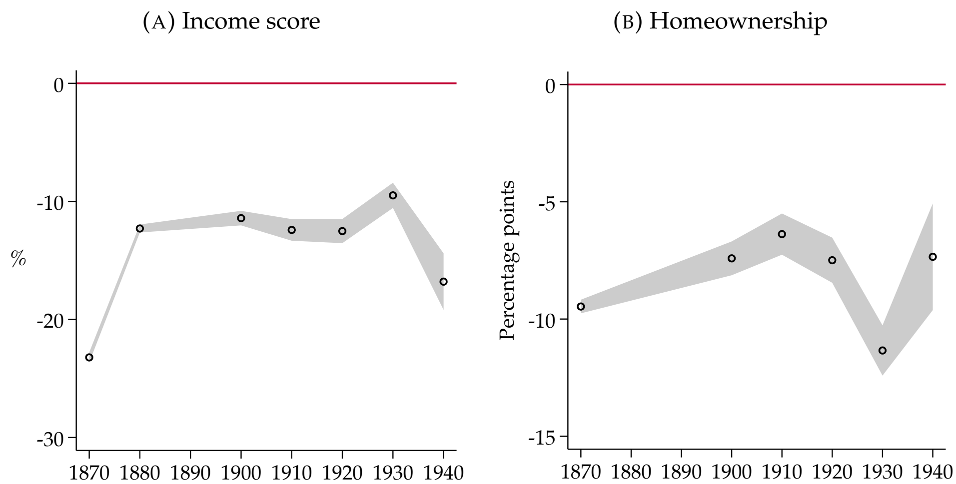
[5] While the 1790 census states that 16 slaves were in Vermont that year, this is likely an error.

[5] There is some evidence that after 1804, some Black Americans were sold to slave states before they reached the age to be emancipated (Armstead et al., 2016, p.104).

2.3 Additional Results

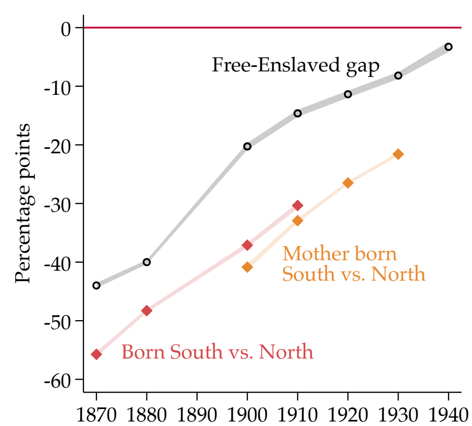
2.3.1 Figures

FIGURE IV.42: Free-Enslaved Gap (1870–1940)



Notes: This figure shows the gaps in income (occupational income score) and homeownership among prime-age (20-54) male descendants of enslaved vs. free Black Americans in each census decade. The sample includes both the South and North of the US. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

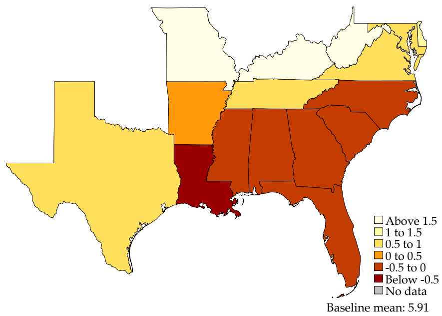
FIGURE IV.43: Free-Enslaved and Southern-Northern Born Gap in Literacy (1870–1940)



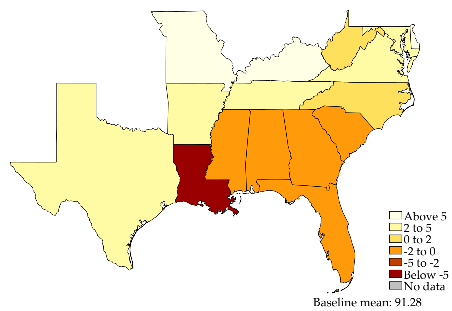
Notes: This figure shows the gaps in literacy among prime-age (20-54) male descendants of free and enslaved Black Americans, as well as those born in the North and South, over each census decade. The gap between Southern and Northern-born individuals is estimated using full census data (not requiring record linkage) that include birthplaces or maternal birthplaces. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. For the Free-Enslaved gap, we restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.44: Long-Run Effect of Ancestor's State of Emancipation on Outcomes

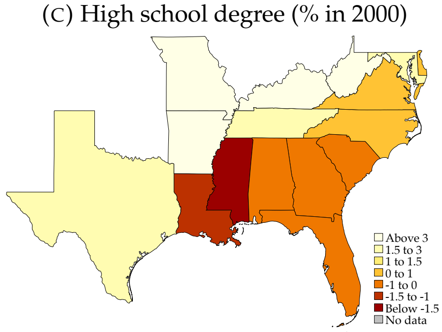
(A) Years of education (1940)



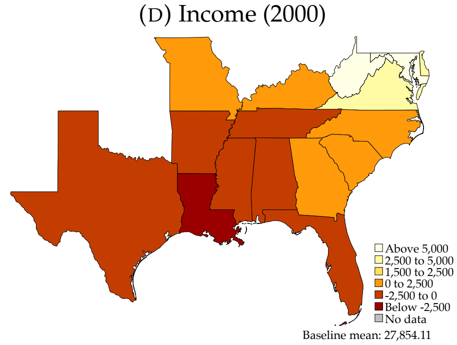
(B) Literacy (% in 1940)



(C) High school degree (% in 2000)



(D) Income (2000)

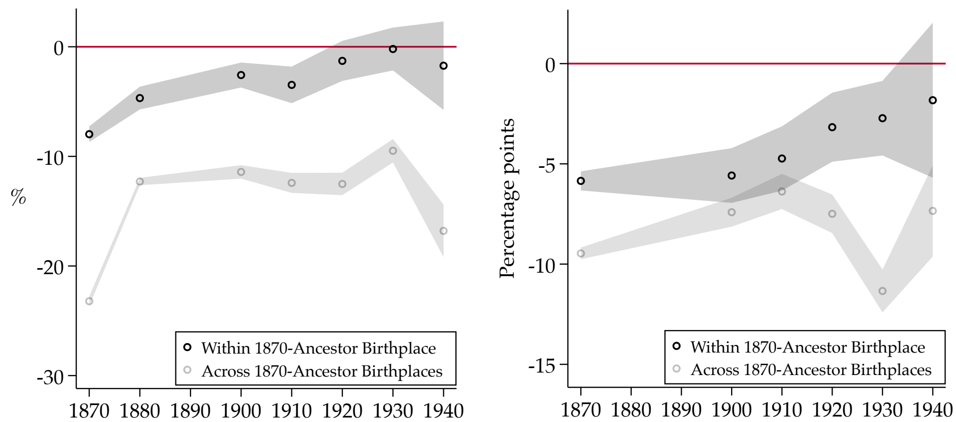


Notes: This figure shows the 1870 ancestor state of birth fixed effect estimates on years of education and literacy rates in 1940, neighborhood-level high school completion rates in 2000, and neighborhood-level income in 2000. A state's FE is the deviation from the population-weighted average across all states (baseline mean) after controlling for a quadratic function of age. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. The sample includes Black prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.45: Free-Enslaved Gap Conditional on Ancestor State (1870–1940)

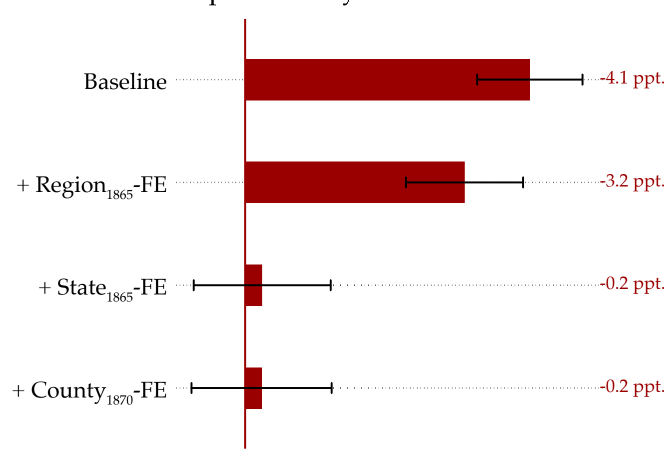
(A) Income score

(B) Homeownership



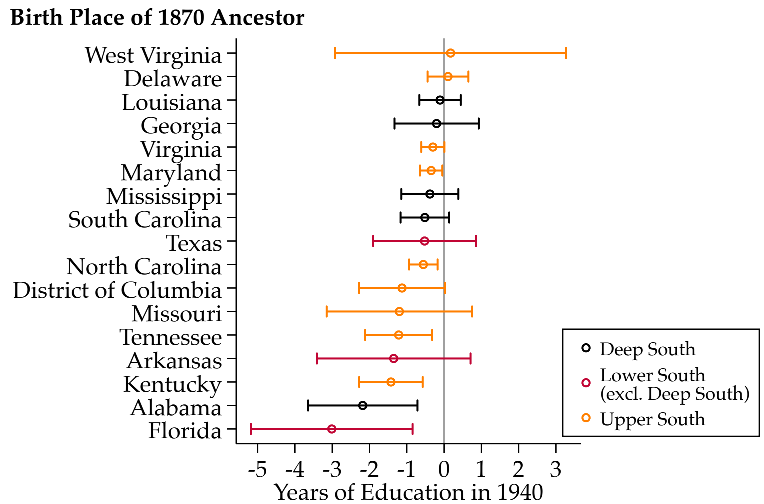
Notes: This figure shows the gaps in income (occupational income score) and homeownership among prime-age (20-54) male descendants of enslaved vs. free Black Americans in each census decade before (light) and after (dark) including fixed effects for 1870 ancestor state of birth. The sample includes both the South and North of the US. We restrict the sample to observations linked to ancestors in 1850, 1860, 1870, or 1880. We control for a quadratic function in age and include 95 percent confidence bands clustered at the family level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.46: Free-Enslaved Gap in Literacy Conditional on Ancestor Location (1940)



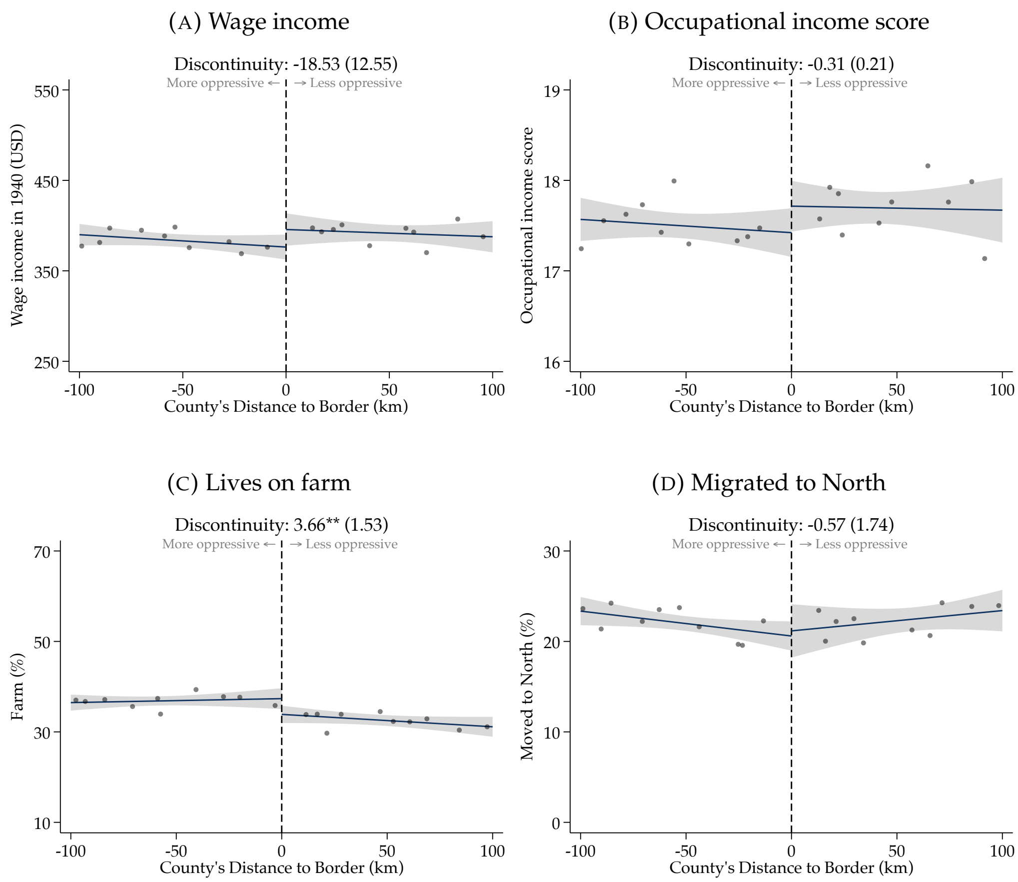
Notes: This figure shows the 1940 Free-Enslaved gap in literacy before and after including different levels of origin location fixed effects. We successively add fixed effects for the region (South or North) and state a family's 1870 ancestor were born, and the county in which their 1870 ancestors lived. The sample includes only Black prime-age (20–54) men whose ancestors can be located in 1870. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.47: Free-Enslaved Gap in 1940 Years of Education by 1870 Ancestor Birthplace



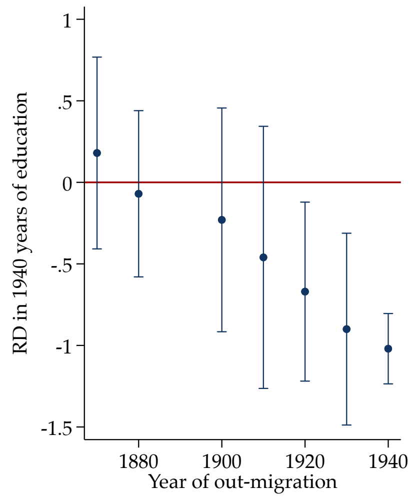
Notes: This figure shows the gaps between descendants of Free and Enslaved in 1940 years of education by 1870 ancestor state of birth. The comparison is made between prime-age (20-54 years) male descendants in each census decade. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included, minimizing bias due to the fact that the Free by definition have a link to 1850 or 1860. Both panels control for age and include 95 percent confidence bands that are clustered at the family level.

FIGURE IV.48: Border Discontinuities in Additional 1940 Outcomes



Notes: This figure shows the RD estimate in additional 1940 outcomes for Black families freed across state borders with different Jim Crow intensity in 1865. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). The sample is restricted to "high-contrast borders" where Jim Crow intensity differs more than across the median border (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level. See Data Appendix 2.2 for details on the sample and data.

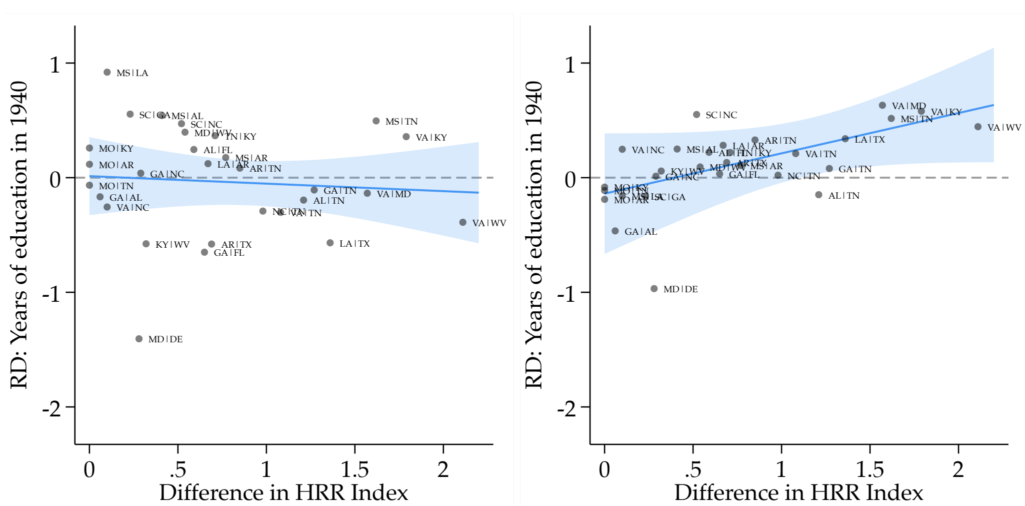
FIGURE IV.49: RD Estimates by Year of Outmigration from Ancestor State



Notes: This figure shows RD estimates in 1940 years of education for Black families whose ancestors were freed on different sides of state borders in 1865 and stayed there for different amounts of time. Each estimate shows the pooled RD estimate for families who stayed in the state where their ancestors were freed from slavery until a given year (x-axis). Jim Crow intensity is measured via the Historical Racial Regime index (Baker, 2022). Negative estimates reflect lower education in the more oppressive state. Bars represent 95 percent confidence intervals. See Data Appendix 2.2 for details on the sample and data.

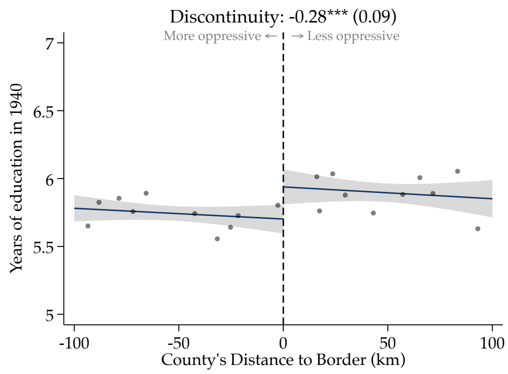
FIGURE IV.50: RD Estimates for Poor and Wealthy White Americans

(A) Poor (B) Wealthy



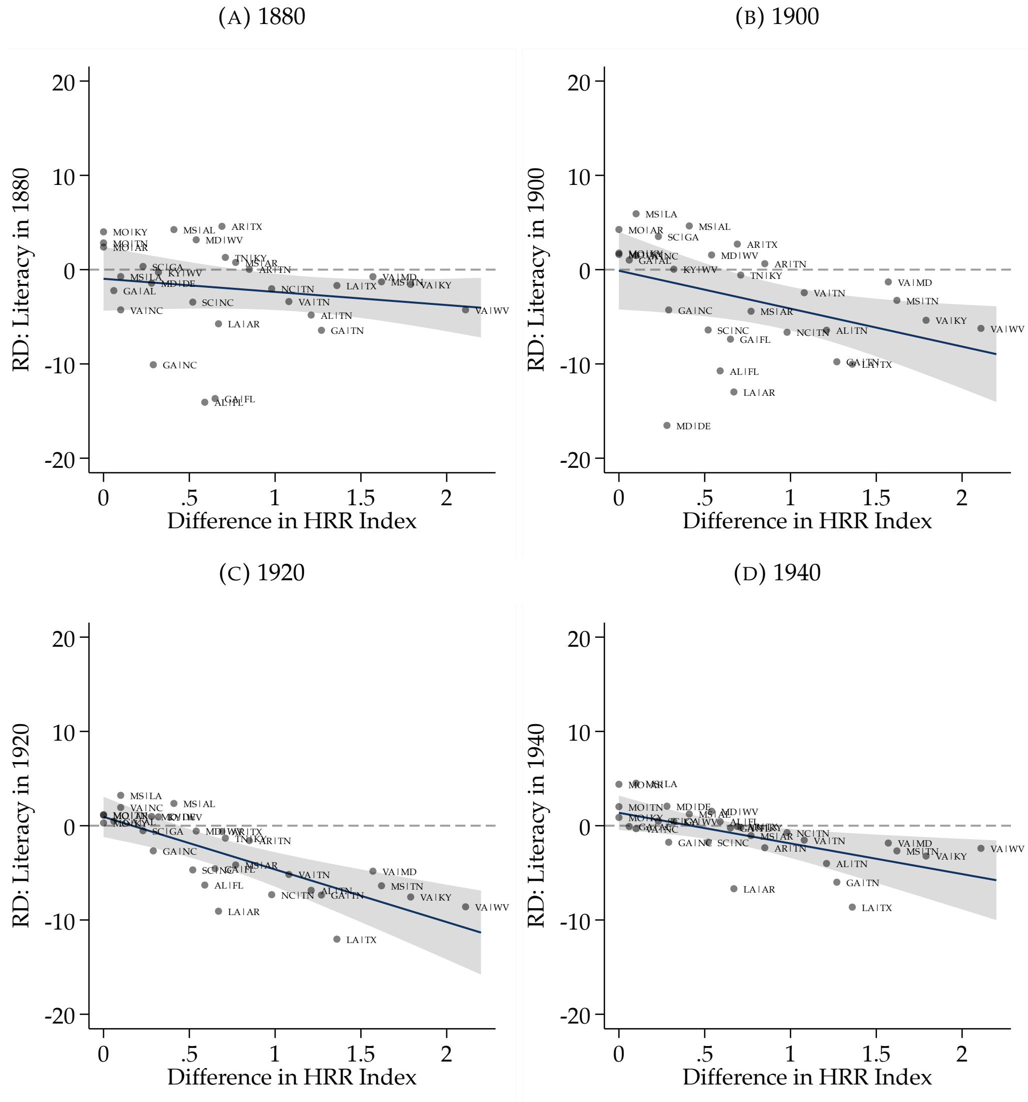
Notes: This figure shows each separate RD estimate in 1940 years of education for white families who had no physical or human capital in 1870, i.e., illiterate and zero wealth (panel A) or were in the top decile in terms of real property in 1870 (panel B). Each label shows the more oppressive before the less oppressive state. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Negative estimates reflect lower education in the more oppressive state. Lines show the best linear fit between RD estimates and the differences in Jim Crow intensity, weighted by the inverse of each estimate's standard error. Shaded areas represent robust 95 percent confidence bands. For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.51: RD Estimates Pooling High- and Low-Contrast Borders



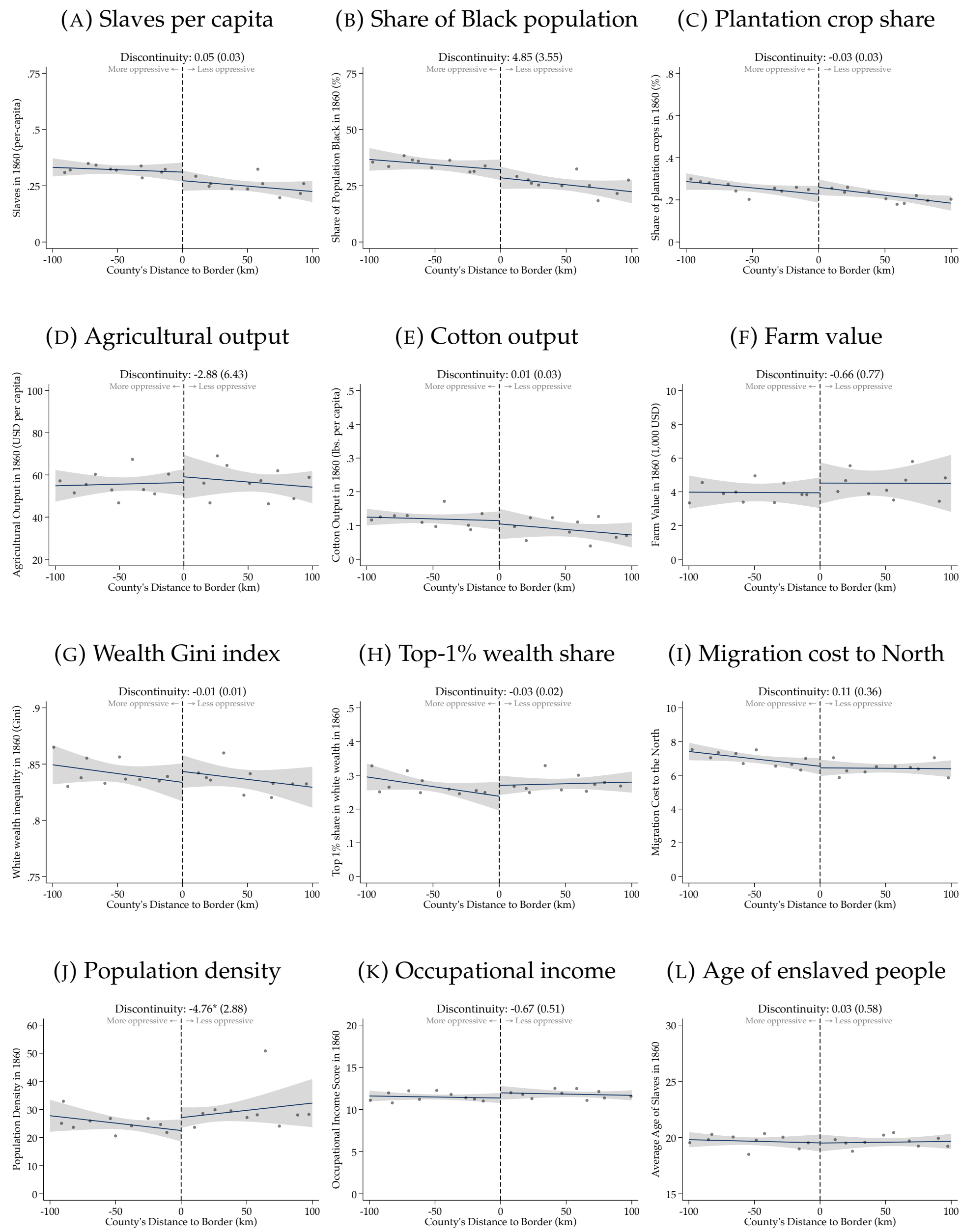
Notes: This figure shows the RD estimate in 1940 years of education for Black families freed across state borders with different Jim Crow intensity in 1865. The left half of the figure represents more oppressive states; the right half less oppressive states. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands clustered at the 1870 county level.

FIGURE IV.52: RD Estimates in Literacy over Time



Notes: This Figure shows each separate RD estimate in literacy in 1880, 1900, 1920, and 1940 for Black families whose ancestors were freed on different sides of state borders in 1865. Each label shows the more oppressive before the less oppressive state. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Negative estimates reflect lower literacy in the more oppressive state. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. Lines show the best linear fit between RD estimates and the differences in Jim Crow intensity, weighted by the inverse of the estimates' standard error. Shaded areas represent robust 95 percent confidence bands. For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data.

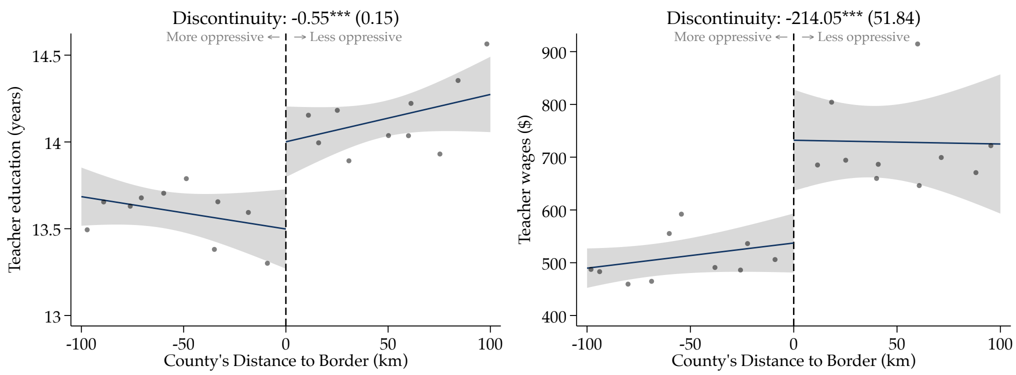
FIGURE IV.53: No Border Discontinuities in 1860 Location Characteristics



Notes: This figure shows the RD estimate in counties' characteristics in 1860 across state borders with different Jim Crow intensities in 1865. Average income is calculated based on occupational income scores. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). The sample is restricted to high-contrast borders (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit weighted by county population. Shaded areas represent 95 percent confidence bands clustered at the county level. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.54: Border Discontinuities in Black Teacher Education and Wages

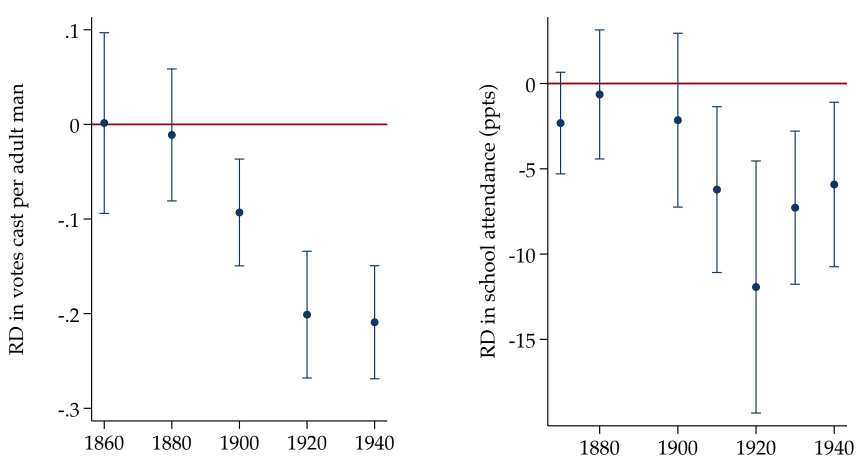
(A) Black teacher education (1940) (B) Black teacher wages (1940)



Notes: This figure shows the RD estimates for counties' Black teacher education (years of education attained) in 1940 and counties' Black teacher wages in 1940. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). The sample is restricted to "high-contrast borders" where Jim Crow intensity differs more than across the median border (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Each dot is the average across a decile of the border population. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands. See Data Appendix 2.2 for details on the sample and data.

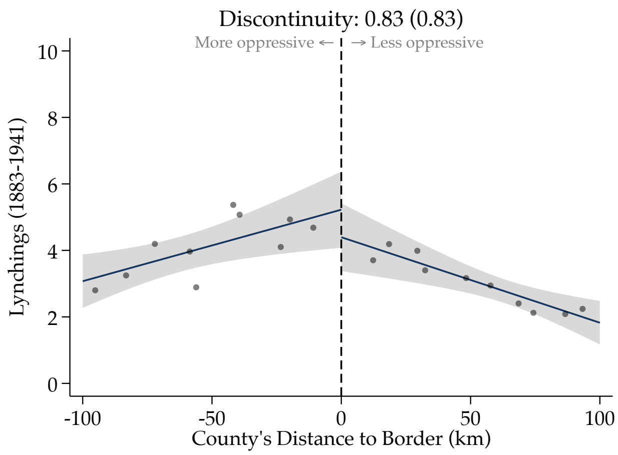
FIGURE IV.55: Border Discontinuities Over Time

(A) Votes cast in presidential elections (B) Black children's school attendance



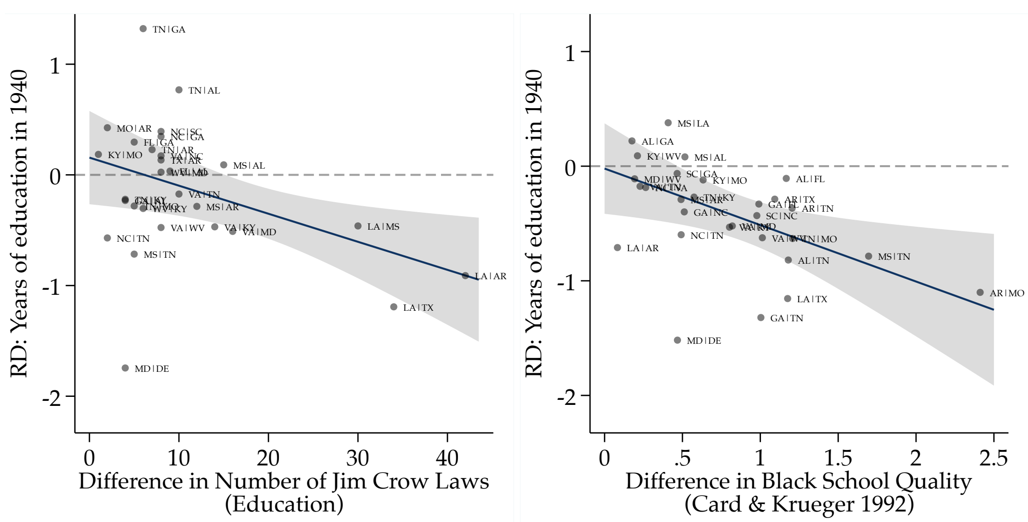
Notes: This figure shows the RD estimates for counties' number of votes cast per adult male in decennial Presidential elections from 1860 to 1940 as a share of the total population eligible based on sex and age (men aged 21 or older); and Black children's school attendance from 1870 to 1940. The sample is limited to "high-contrast borders" (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). Each estimate is the difference between outcomes in the more oppressive compared to the less oppressive state. Vertical bars represent 95 percent robust confidence bands. See Data Appendix 2.2 for details on the sample and data.

FIGURE IV.56: No Border Discontinuities in Lynchings between 1883 and 1941



Notes: This figure shows the RD estimate in counties' number of lynchings of Black Americans between 1883 and 1941. The sample is restricted to high-contrast borders (above 0.71 HRR index points, with differences averaging 1.30 HRR index points). The left half of each panel represents more oppressive states; the right half less oppressive states. Jim Crow intensity is measured via the Historical Racial Regime (HRR) index (Baker, 2022). Each dot is the average across a decile of counties. Lines show the best linear fit. Shaded areas represent 95 percent confidence bands. See Data Appendix 2.2 for details on the sample and data.

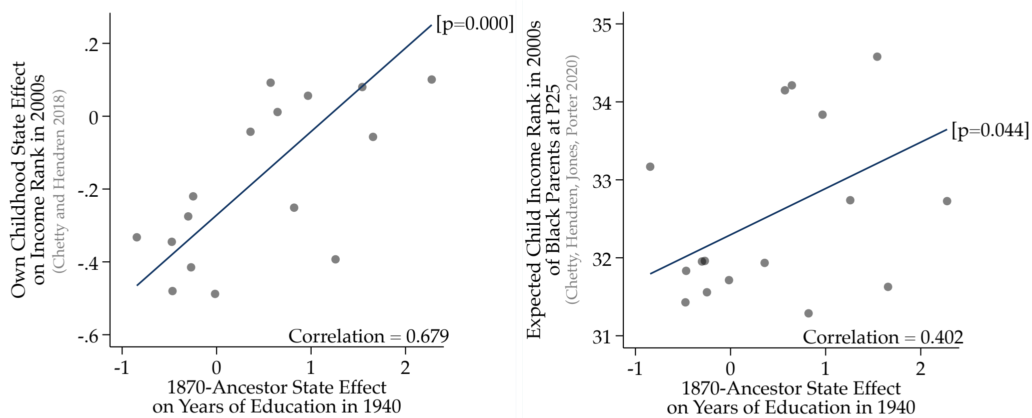
FIGURE IV.57: Regression Discontinuity Estimates and Education under Jim Crow



Notes: This figure shows each separate RD estimate in 1940 years of education for Black families whose ancestors were freed on different sides of state borders in 1865. Each label shows the more oppressive before the less oppressive state. Negative estimates reflect lower education in the more oppressive state. Lines show the best linear fit, weighted by the inverse of each estimate's standard error. Shaded areas represent robust 95 percent confidence bands. For point estimates, we use a 350km bandwidth and empirical Bayesian shrinkage as described in Appendix 2.1.5. See Data Appendix 2.2 for details on the sample and data. Our results are robust to using an alternative measure of school quality from Carruthers and Wanamaker (2017) instead of Card and Krueger (1992).

FIGURE IV.58: Persistence of a State's Capacity to Generate Upward Mobility

(A) Causal Estimates Across Races (B) Non-Causal Estimates for Black Americans



Notes: This figure is a binned scatter plot relating a state's causal effect on Black economic progress from 1865 to 1940 (as shown in panel A of Appendix Figure IV.44) to (A) the state's causal effect on intergenerational mobility in recent decades (as estimated by Chetty and Hendren, 2018) and (B) the state's non-causal estimate of expected child income rank among Black parents (as estimated by Chetty et al., 2020). The modern estimates reflect a child's mean percentile rank in the national household income distribution at age 26 conditional on growing up with parents at the 25th percentile. See Data Appendix 2.2 for details on the sample and data.

2.3.2 Tables

TABLE IV.22: Free-Enslaved Gap (1940) in Different Income Measures

	OCCSCORE (1950-$) Mean: 1,604.09	LIDO Score (1950-$) Mean: 1,161.69	Wage Income (1940-$) Mean: 381.20	Total Income (1940-$) Mean: 793.47	Song et al. Score Mean: 43.42
Ancestor Enslaved until Civil War	**-148.39*****	**-279.00*****	**-145.92*****	**-204.29*****	**-9.29*****
	(10.86)	(8.59)	(6.13)	(10.29)	(0.39)
Controls (age, age^2)	Y	Y	Y	Y	Y
Adjusted R^2	0.04	0.04	0.05	0.09	0.01
Observations	168,138	142,743	154,463	146,871	168,138
Ancestor Free	*9,325*	*7,517*	*8,551*	*8,100*	*9,325*

Notes: This table shows the Free-Enslaved gap in income across different measures: Occupational income score (OCCSCORE), a refined occupational income score (LIDO from Saavedra and Twinam, 2020), wage income, total predicted income, and the Song et al. (2020) score. We compute the Song et al. (2020) score by computing the average literacy rate by occupation and birth decade and converting this measure into ranks. The sample includes both the South and North of the US. All estimates are for Black prime-age men in 1940. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.23: Free-Enslaved Gap (1940) in Different Education Measures

	Literacy (%) Mean: 91.49	Education (Years) Mean: 5.99	High School (%) Mean: 9.28	College (%) Mean: 1.70	Graduate (%) Mean: 0.46
Ancestor Enslaved until Civil War	**-4.25*****	**-1.59*****	**-7.86*****	**-1.86*****	**-0.74*****
	(0.26)	(0.05)	(0.45)	(0.21)	(0.12)
Controls (age, age^2)	Y	Y	Y	Y	Y
Adjusted R^2	0.01	0.04	0.01	0.00	0.00
Observations	163,549	163,549	163,549	163,549	163,549
Ancestor Free	*9,078*	*9,078*	*9,078*	*9,078*	*9,078*

Notes: This table shows the Free-Enslaved gap in education across different measures: Literacy, years of education, and the probability of holding a high school, college, or graduate degree. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. The sample includes both the South and North of the US. All estimates are for Black prime-age men in 1940. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.24: Free-Enslaved Gap using Mortality Records (1988–2007)

	HS Degree (%) Mean: 68.85	College Degree (%) Mean: 12.31	Income (USD) Mean: 29,875.58	House Value (USD) Mean: 87,921.78
Ancestor Enslaved until Civil War	**-3.02*****	**-2.45*****	**-4,795.93*****	**-15,755.30*****
	(0.51)	(0.55)	(636.79)	(2,462.82)
Level of outcome	Tract×Race×Sex	Tract×Race×Sex	Tract×Race	Tract×Race
Controls (age, age^2)	Y	Y	Y	Y
Adjusted R^2	0.01	0.00	0.01	0.00
Observations	26,765	26,765	26,803	25,787
Ancestor Free	*1,713*	*1,713*	*1,715*	*1,634*

Notes: This table shows the Free-Enslaved gap in 2000 neighborhood-level outcomes: high school and college degrees, median incomes, and median house values (conditional on ownership). A neighborhood is a census tract. Each person is assigned the value of the census tract in which they last lived according to administrative mortality records. The sample includes both the South and North of the US. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.25: Free-Enslaved Gap (1940) between and within Ancestor's Birthplace

	Education (Years) Mean: 5.91		Wage Income (USD) Mean: 388.01		Home Ownership (%) Mean: 29.48		House Value (USD) Mean: 1,412.17	
Ancestor Enslaved until Civil War	-1.49***	**-0.41*****	-137.00***	**-20.22****	-6.76***	**-1.61**	-574.06***	**8.40**
	(0.07)	(0.08)	(8.51)	(9.84)	(0.86)	(1.04)	(90.08)	(115.61)
1870 State of Birth-FE	N	Y	N	Y	N	Y	N	Y
Controls (age, age^2)	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.04	0.08	0.04	0.07	0.01	0.03	0.01	0.03
Observations	75,583	75,583	71,474	71,474	76,048	76,048	21,873	21,873
Ancestor Free	*4,617*	*4,617*	*4,371*	*4,371*	*4,640*	*4,640*	*1,624*	*1,624*

Notes: This table shows the gap in years of education, total income, homeownership rate, and house value among prime-age (20-54) male descendants of enslaved vs. free Black Americans in 1940. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Columns 1, 3, 5, and 7 repeat Table II.1 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and Enslaved. Figure II.5 and Appendix Figure IV.45 show the evolution of the conditional Free-Enslaved gap over time. See Data Appendix 2.2 for details. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.26: Free-Enslaved Gap Between and Within Ancestor's Birthplace using Mortality Records (1988–2007)

	HS Degree (%) Mean: 69.20		College Degree (%) Mean: 12.32		Income (USD) Mean: 30,143.90		House Value (USD) Mean: 88,830.12	
Ancestor Enslaved until Civil War	-2.57***	**-0.89**	-2.07***	**-0.29**	-5,032.50***	**-1,014.92**	-13,391.02***	**-780.04**
	(0.74)	(0.82)	(0.78)	(0.78)	(921.89)	(1,005.32)	(3,498.95)	(3,829.19)
Level	Tract×Race×Sex		Tract×Race×Sex		Tract×Race		Tract×Race	
1870 State of Birth-FE	N	Y	N	Y	N	Y	N	Y
Controls (age, age^2)	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.00	0.01	0.00	0.01	0.01	0.03	0.00	0.02
Observations	11,931	11,931	11,931	11,931	11,932	11,932	11,500	11,500
Ancestor Free	*863*	*863*	*863*	*863*	*861*	*861*	*830*	*830*

Notes: This table shows the Free-Enslaved gap at the neighborhood-level in the fraction of people who hold a high school degree, the fraction of people who hold a college degree, the median income earned, and the median house value in 2000. The sample includes both the South and North of the US. Columns 1, 3, 5, and 7 repeat Table IV.24 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership and therefore exclude zeros. Each person is assigned the respective value of the census block in which they lived at the time of death. Sample means are computed for the combined sample of the Free and Enslaved. See Data Appendix 2.2 for details on the sample and data. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE IV.27: Free-Enslaved Gap (1940) for Free Without Physical or Human Capital in 1860

	Education (Years) Mean: 5.83		Wage Income (USD) Mean: 381.64		Homeownership (%) Mean: 29.08		House Value (USD) Mean: 1,380.43	
Ancestor Enslaved until Civil War	**-1.00******	**-0.12**	**-90.43******	**26.85**	**-6.16******	**-1.42**	**-343.74****	**440.28****
	(0.15)	(0.15)	(21.13)	(21.44)	(1.95)	(2.00)	(159.58)	(184.15)
1870 State of Birth-FE	N	Y	N	Y	N	Y	N	Y
Controls (age, age^2)	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.03	0.07	0.04	0.07	0.01	0.02	0.00	0.03
Observations	71,574	71,574	67,672	67,672	72,013	72,013	20,455	20,455
Ancestor Free	*608*	*608*	*569*	*569*	*605*	*605*	*206*	*206*

Notes: This table shows the gap in years of education, total income, homeownership rate, and house value among prime-age (20-54) male descendants of a subset of the enslaved vs. free Black Americans in 1940. Among the Free, we only include those whose ancestors had no measurable physical capital (real and personal property) or human capital (literacy) in 1860. The sample includes both the South and North of the US. Only observations that can be linked to the 1850, 1860, 1870, or 1880 census are included. Columns 1, 3, 5, and 7 repeat Table II.1 but hold the sample constant to the other columns. Columns 2, 4, 6, and 8 add fixed effects for 1870 ancestor state of birth. House values are measured conditional on ownership. Sample means are computed for the combined sample of the Free and Enslaved. Appendix Figure IV.45 shows the evolution of the conditional Free-Enslaved gap over time. See Data Appendix 2.2 for details. Standard errors are clustered at the family level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.4 Model Appendix

2.4.1 Importance of Geography in Perpetuating Free-Enslaved Gap

We can decompose the average treatment effect (ATE) of descending from ancestors enslaved until the Civil War defined in equation (II.4) into the sum of 1) the intergenerational effect conditional on location and "ability" ($-\rho\delta$), 2) the geographic effect of the ancestor's enslavement location (*geographic endowment effect*), and 3) the effect of opportunities to migrate to more favorable locations (*location choice effect*). Formally, we decompose the ATE into

$$\text{ATE} = -\rho\delta + \theta + \kappa$$

where θ is the *geographic endowment effect* and κ is the *location choice effect*, and

$$\begin{aligned}
\theta \equiv \int \sum_{\ell \in \mathcal{L}} & \left(\Pr\left(\ell_{(i,0)} = \ell \mid s_i = 1\right) - \Pr\left(\ell_{(i,0)} = \ell \mid s_i = 0, \alpha_{i,0}\right)\right) \times \\
& \left(\rho\gamma_\ell^0 + \mathbb{E}\left[\gamma^1_{\ell(i,1)} \mid s_i = 1, a_{i,0}, \ell_{(i,0)} = \ell\right]\right) dF(\alpha_{i,0}) \\
\kappa \equiv \int \sum_{\ell \in \mathcal{L}} & \Pr\left(\ell_{(i,0)} = \ell \mid s_i = 0, \alpha_{i,0}\right) \times \\
& \left(\mathbb{E}\left[\gamma^1_{\ell(i,1)} \mid s_i = 1, a_{i,0}, \ell_{(i,0)} = \ell\right] - \mathbb{E}\left[\gamma^1_{\ell(i,1)} \mid s_i = 0, a_{i,0}, \ell_{(i,0)} = \ell\right]\right) dF(\alpha_{i,0}).
\end{aligned}$$

We imposed Assumption 1: location is independent of ability for the enslaved population.

We argue that the geographic disadvantage that the Enslaved population faced relative to the Free *within the South* provides a lower bound (in absolute terms) for the *geographic endowment effect* (θ). In the North, descendants of the Free tended to face more favorable conditions after slavery than those in the South. A large part of the *geographic endowment effect* therefore likely results from the fact that around half of the Free population lived in the North before 1865—an effect that we ignore to provide a lower bound. Formally, we assume that the *geographic endowment effect*$\theta \leq Z$ with Z defined as

$$Z \equiv \sum_{\ell \in \mathcal{L}} \left(\Pr\left(\ell_{(i,0)} = \ell \mid s_i = 1\right) - \Pr\left(\ell_{(i,0)} = \ell \mid s_i = 0, \ell \in S\right)\right)\left(\eta_\ell - \eta_{\ell'}\right),$$

where $S \subset \mathcal{L}$ denotes all states in the South, $\ell' \in S$ is an arbitrary reference state in the South, and $\eta_\ell - \eta_{\ell'}$ as defined in equation (II.7) is the intent-to-treat effect of having a formerly enslaved ancestor born in state ℓ (relative to state ℓ'). We estimate Z using the state effects estimated in regression equation (II.6). Specifically, we estimate Z via

$$\hat{Z} = \sum_{\ell \in \mathcal{L}} \left(\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\ell_{(i,0)} = l \mid s_i = 1\right) - \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\ell_{(i,0)} = l \mid s_i = 0, l \in S\right)\right)\left(\hat{\eta}_l - \hat{\eta}_l'\right)$$

where $\mathbb{1}(\cdot)$ is the indicator function and $\hat{\eta}_l - \hat{\eta}_l'$ are the state fixed effects obtained in (II.6).

We find that the estimated upper bound of Z is around two-thirds of the Free-Enslaved gap. We also argued that Z is plausibly a lower bound of the geographic endowment effect. Under the additional assumption that $-\rho\delta$ and κ are both negative,[6] this implies that 1) at least two-thirds of the Free-Enslaved gap is *causal*, i.e. did not arise from selection into freedom, and 2) that the difference in the initial geographic distribution induced by slavery was the most important channel underlying this causal effect.

TABLE IV.28: Decomposition of the Free-Enslaved Gap in 1940

	Free-Enslaved gap & ancestor location			Geography's effect as % of gap		
	National	Within South	Within state	Less conservative	Conservative	Lower bound
Literacy (%)	-4.2	-3.2	-0.4	138%	90%	67%
Years of education	-1.6	-1.2	-0.4	113%	75%	50%

Notes: This table decomposes the 1940 Free-Enslaved gaps in literacy and years of education. We successively add fixed effects for the region (South or North) and state a family's 1870 ancestor were born, and the county in which their 1870 ancestors lived. Columns 4 and 5 show the fraction of the national Free-Enslaved gap (column 1) that can be accounted for by state variation (column 3), respectively including (less conservative) or excluding (conservative) extrapolated effects for the North. The extrapolation predicts causal state effects for the North based on the relationship between causal state effects among Enslaved in the South and non-causal state effects among Free in the South. Column 6 shows the result of our formal decomposition. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate. The sample includes only Black prime-age (20–54) men whose ancestors can be located in 1870. See Data Appendix 2.2 for details on the sample and data.

We further estimate how the Enslaved would have progressed had they been geographically distributed as they Free within *the South and the North*. To do so, we extrapolate Northern states' effects. We cannot estimate those effects directly because we lack plausibly exogenous variation in location assignment there. Our extrapolation predicts Northern state effects based on 1) Northern non-causal state effects among the Free and 2) the relationship between Southern causal state effects among the Enslaved and non-causal state effects among the Free. This exercise shows that the Free-Enslaved gap would have closed entirely by 1940 (see Appendix Table IV.28). Overall, our results show that group differences in initial location were the primary driver of the persistent Free-Enslaved gap.

2.4.2 Direct Evidence on Selection into Freedom Before the Civil War

Combining (II.2), (II.3), and (II.4), the observed Free-Enslaved gap is equal to

$$\mathbb{E}[y_{i,1} \mid s_i = 1] - \mathbb{E}[y_{i,1} \mid s_i = 0] = ATE - B, \tag{IV.18}$$

where the (negative of) the selection bias B, arising from 1) potential selection into being free, 2) potential selection into location by (descendants of) the Free, and 3) potential

[6] Intuitively, this assumption imposes that 1) being enslaved longer did not benefit descendants ($-\rho\delta < 0$) and 2) migration opportunities were not better from enslaved people's locations than from free Black Americans' locations ($\kappa < 0$).

selection into location by (descendants of) the Enslaved, is equal to:

$$
\begin{aligned}
B = & \underbrace{\mathbb{E}\left[(\lambda+\rho)\,\alpha_{i,0} \mid s_i=0\right] - \mathbb{E}\left[(\lambda+\rho)\,\alpha_{i,0} \mid s_i=1\right]}_{\text{Potential selection into being free}} + \\
& \underbrace{\left(\mathbb{E}[\rho\gamma^0_{\ell(i,0)} + \gamma^1_{\ell(i,1)} \mid s_i=0] - \int \mathbb{E}[\rho\gamma^0_{\ell(i,0)} + \gamma^1_{\ell(i,1)} \mid s_i=0, \alpha_{i,0}]dF(\alpha_{i,0})\right)}_{\text{Potential selection into location by (descendants of) the Free}} - \\
& \underbrace{\left(\mathbb{E}\left[\rho\left(\gamma^0_{\ell(i,0)} - \delta\right) + \gamma^1_{\ell(i,1)} \mid s_i=1\right] - \int \mathbb{E}\left[\rho\left(\gamma^0_{\ell(i,0)} - \delta\right) + \gamma^1_{\ell(i,1)} \mid s_i=1, \alpha_{i,0}\right] dF(\alpha_{i,0})\right)}_{\text{Potential selection into location by (descendants of) the Enslaved}}.
\end{aligned}
$$

If being free before the Civil War was a matter of pure chance, the differences between the Free and the Enslaved have a causal interpretation. A priori, this assumption is strong. However, the plausibility of the assumption depends crucially on the conditions under which freedom was attained.

There were five main channels into freedom between the Revolutionary War (1775–1783) and the abolition of slavery in 1865: 1) by emancipation through abolition of slavery in the North in the late 18th and early 19th century, 2) by manumission through one's master, 3) by manumission through self-purchase, 4) by manumission through purchase by a third party, or 5) by running away. A person born to a free mother inherited their mother's freedom. In rare occasions, enslaved people were unintentionally freed by accompanying their masters on a trip to a free state. Setting foot on free soil freed enslaved people by law and some sued to enforce their rights (see, e.g., Rose, 2009).

In 1860, around half of the free population was born in the North, which we argue is a reasonable approximation of the share of the free families freed through general emancipation in the North. Within the remaining half, it is hard to estimate the share of people freed "legally" and those who ran away.

Dittmar and Naidu (2012) use runaway slave advertisements placed in Southern newspapers between 1840 and 1860 and suggest that such advertisements were placed for around 8,000 runaway slaves throughout the final two decades of slavery. However, the authors also point out that "it is clear that among the many absconders only a small fraction remained at large for a lengthy period." The odds of a successful escape were especially small in the Lower South. This is corroborated by the fact that in a Pennsylvania census of Free Black Americans, only 2 out of 314 people who were not born free indicated that they attained freedom through escape.[7] It is therefore safe to conclude that the vast majority of those who became free in the South did so through manumission (as opposed to escape).

Since slavery had been de facto abolished in the North by 1850 (see Table IV.21), the enslaved people there were freed non-selectively. That is, as long as one is willing to assume that those enslaved in the North were not inherently different from those enslaved in the (Upper) South around 1800, those in the North were freed independently of any observed or unobserved characteristics. In the South, the degree of selection into manumission varied largely across time and locations. Around the 1780s, the early years

[7]Pennsylvania Abolition Society and Society of Friends Manuscript Census Schedules, 1838. Available in machine-readable form through `https://doi.org/10.3886/ICPSR03805.v1`.

after the Revolutionary War, there was a stream of manumissions motivated by morality or religion. In later antebellum years, manumission turned into an instrument to uphold slavery (Berlin, 1974). It did not, in most cases, arise from anti-slavery sentiments. On the contrary, many owners manumitted their slaves as a reward for loyalty and by doing so "reinforced rather than challenged the values, assumptions, and discipline of slavery" (Wolf, 2006, p. 44).

One could imagine that the practice of manumission induced a degree of selection into being free. Indeed, some quantitative evidence on the presence of selection into manumission exists. Cole (2005) finds that in Louisiana, manumitted people were 62.5 percent female (43.6 percent in the enslaved population) and much more likely to be "Mulatto" (38.5 percent) than the slave population (5.8 percent). This is consistent with the observation that manumission in the Lower South was reserved for "illicit offspring, special favorites, or least productive slaves" (Berlin, 1974). Bodenhorn (2011), too, finds evidence of preferential manumission for people of mixed race in Virginia. Similarly, Berlin (1974) argues that skilled slaves had a larger chance of accumulating enough wealth to be manumitted through self-purchase. Little is known about selection into being manumitted through purchase by other people (usually other free Black people). Runaways, however, "as a group, had always been more skilled, sophisticated, and aggressive than the mass of slaves" (Berlin, 1974, p. 160). Table IV.29 summarizes the discussion.

TABLE IV.29: Relative prevalence of and selectivity in different roads to freedom

	%	Degree of selection
Emancipation in North	≈ 50	None
Manumission by master	30–40	Varied across time and locations
Manumission by self-purchase	5–10	Potentially high
Manumission by a third buyer	5–10	Unknown
Escape	< 5	Potentially high

Notes: This table indicates a rough breakdown of the relative probability of attaining freedom in various ways. The percentage emancipated in the North is estimated by the fraction of free Black people born in the North in the 1860 census. The fraction that escaped is a conservative upper bound given the observations mentioned in the text. The remaining probability is attributed to manumissions. The distribution within manumissions is derived from (Bodenhorn, 2011): 10-20 percent through self-purchase, 10-20 percent through a third buyer, and the remaining 60-80 percent by the master.

Bibliography

ABRAMITZKY, R., L. BOUSTAN, E. JACOME, AND S. PEREZ (2021a): "Intergenerational Mobility of Immigrants in the United States over Two Centuries," *American Economic Review*, 111, 580–608.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2021b): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

ANGRIST, J. D., P. D. HULL, P. A. PATHAK, AND C. R. WALTERS (2017): "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132, 871–919.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly harmless econometrics*, Princeton university press.

ARMSTEAD, S., B. SUTTER, P. WALKER, AND C. WIESNER (2016): ""And I Poor Slave Yet": The Precarity of Black Life in New Brunswick, 1766–1835," in *Scarlet and Black: Slavery and Dispossession in Rutgers History*, ed. by M. J. Fuentes and D. G. White, Rutgers University Press, 91–122.

ARMSTRONG, T. B., M. KOLESŔ, AND PLAGBORG-MØLLER (2021): "Robust Empirical Bayes Confidence Intervals," Tech. rep.

ASHENFELTER, O. AND A. KRUEGER (1994): "Estimates of the Economic Return to Schooling from a New Sample of Twins," *The American Economic Review*, 84, 1157–1173.

BAKER, R. S. (2022): "The Historical Racial Regime and Racial Inequality in Poverty in the American South," *American Journal of Sociology*, 127.

BERLIN, I. (1974): *Slaves without masters: The free Negro in the antebellum South*, Oxford University Press.

BERNINI, A., G. FACCHINI, AND C. TESTA (2023): "Race, Representation, and Local Governments in the US South: The Effect of the Voting Rights Act," *Journal of Political Economy*, 131, 994–1056.

BODENHORN, H. (2011): "Manumission in nineteenth-century Virginia," *Cliometrica*, 5, 145–164.

CARD, D., C. DOMNISORU, AND L. TAYLOR (2022): "The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility," *Journal of Labor Economics*, 40, S1–S493.

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

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

CASCIO, E. U. AND E. G. LEWIS (2022): "Legal Activism, State Policy, and Racial Inequality in Teacher Salaries and Educational Attainment in the Mid-Century American South," Working Paper 30631, National Bureau of Economic Research.

CENSUS BUREAU, U. (2017): "Census tract," Available at https://web.archive.org/web/20170513191843/https://factfinder.census.gov/help/en/census_tract.htm. Accessed on 14 November 2021.

CHETTY, R. AND N. HENDREN (2018): "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 133, 1163–1228.

CHETTY, R., N. HENDREN, M. R. JONES, AND S. R. PORTER (2020): "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *The Quarterly Journal of Economics*, 135, 711–783.

CLARK, G. (2014): *The Son Also Rises: Surnames and the History of Social Mobility*, Princeton University Press.

COHEN, W. (1991): *At Freedom's Edge: Black Mobility and the Southern White Quest for Racial Control, 1861–1915*.

COLE, S. (2005): "Capitalism and freedom: Manumissions and the slave market in Louisiana, 1725–1820," *Journal of Economic History*, 65, 1008–1027.

COLLINS, W. J. AND R. A. MARGO (2011): "Race and Home Ownership from the Civil War to the Present," *American Economic Review: Papers and Proceedings*, 101, 355–359.

CONFEDERATION CONGRESS (1787): "An Ordinance for the Government of the Territory of the United States North West of the River Ohio," Library of Congress: https://lccn.loc.gov/90898154.

CONNECTICUT GENERAL ASSEMBLY (1784): "An Act Concerning indian, Molatto and Negro Servants and Slaves," Western CT State University: https://archives.library.wcsu.edu/omeka/items/show/2625.

CONSTITUTION OF THE STATE OF NEW HAMPSHIRE (1783): https://www.nh.gov/glance/constitution.htm.

CONSTITUTION OF VERMONT (1777): https://avalon.law.yale.edu/18th_century/vt01.asp.

CUSHING, J. D. (1961): "The Cushing Court and the Abolition of Slavery in Massachusetts: More Notes on the "Quock Walker Case"," *The American Journal of Legal History*, 5, 118–144.

DITTMAR, J. AND S. NAIDU (2012): "Contested Property: Fugitive Slaves in the Antebellum U.S. South," Tech. rep.

DONALDSON, D. AND R. HORNBECK (2016): " Railroads and American Economic Growth: A "Market Access" Approach *," *The Quarterly Journal of Economics*, 131, 799–858.

EFRON, B. (2010): *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Cambridge University Press.

FRAZIER, E. F. ET AL. (1957): *Black bourgeoisie*, Free Press.

GENERAL ASSEMBLY OF RHODE ISLAND (1784): "Gradual Emancipation Act," Library of Congress: https://lccn.loc.gov/90898154.

GOLDSTEIN, J. R., M. ALEXANDER, C. BREEN, A. M. GONZÁLEZ, F. MENARES, M. OSBORNE, M. SNYDER, AND U. YILDIRIM (2021): "CenSoc Mortality File: Version 2.0. Berkeley: University of California, 2021." dataset: https://censoc.berkeley.edu/data/.

ICPSR (1999): "United States Historical Election Returns, 1824-1968," Inter-university Consortium for Political and Social Research.

LEEUWEN, M. H. D. V. AND I. MAAS (2011): *HISCLASS: A Historical International Social Class Scheme*, Universitaire Pers Leuven.

MANSON, S., J. SCHROEDER, D. V. RIPER, T. KUGLER, AND S. RUGGLES (2021): "IPUMS National Historical Geographic Information System: Version 16.0," dataset: http://doi.org/10.18128/D050.V16.0.

MELLO, S. (2023): "Fines and Financial Wellbeing," Working paper.

MURRAY, P. (1950): *States' Laws on Race and Color*.

NEW JERSEY STATE LEGISLATURE (1804): "An act for the gradual abolition of slavery," https://www.loc.gov/item/rbpe.0990100b/.

NEW YORK STATE LEGISLATURE (1799): "An act for the gradual abolition of slavery," New York laws of 1799, Chapter 62, p. 1: https://webarchive.loc.gov/all/20101110015236/http://iarchives.nysed.gov/dmsBlue/viewImageData.jsp?id=177879.

——— (1817): "An act relative to slaves and servants," New York State Archives: http://www.archives.nysed.gov/education/act-relative-slaves-and-servants-1817.

PENNSYLVANIA GENERAL ASSEMBLY (1780): "The Gradual Abolition Act," https://www.mountvernon.org/library/digitalhistory/digital-encyclopedia/article/gradual-abolition-act-of-1780/.

ROBACK, J. (1984): "Southern Labor Law in the Jim Crow Era: Exploitative or Competitive?" *University of Chicago Law Review*, 51, 1161–1192.

ROSE, B. Z. (2009): *Mother of Freedom: Mum Bett and the Roots of Abolition*, TreeLine Press.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "IPUMS USA: Version 10.0," dataset: https://doi.org/10.18128/D010.V10.0.

SAAVEDRA, M. AND T. TWINAM (2020): "A machine learning approach to improving occupational income scores," *Explorations in Economic History*, 75, 101304.

SEGUIN, C. AND D. RIGBY (2019): "National crimes: A new national data set of lynchings in the United States, 1883 to 1941," *Socius*, 5, 2378023119841780.

SONG, X., C. G. MASSEY, K. A. ROLF, J. P. FERRIE, J. L. ROTHBAUM, AND Y. XIE (2020): "Long-term decline in intergenerational mobility in the United States since the 1850s," *Proceedings of the National Academy of Sciences*, 117, 251–258.

WALTON, H., S. C. PUCKETT, AND D. R. DESKINS (2012): *The African American Electorate: A Statistical History*, Washington, D.C.: CQ Press.

WOLF, E. S. (2006): *Race and Liberty in the New Nation: Emancipation in Virginia from the Revolution to Nat Turner's Rebellion*, LSU Press.

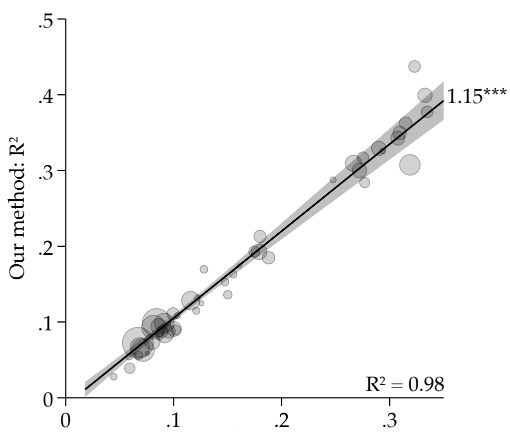
ZILVERSMIT, A. (1968): "Quok Walker, Mumbet, and the Abolition of Slavery in Massachusetts," *The William and Mary Quarterly*, 25, 614–624.

3 Appendix to "The Missing Link(s): Women and Intergenerational Mobility"

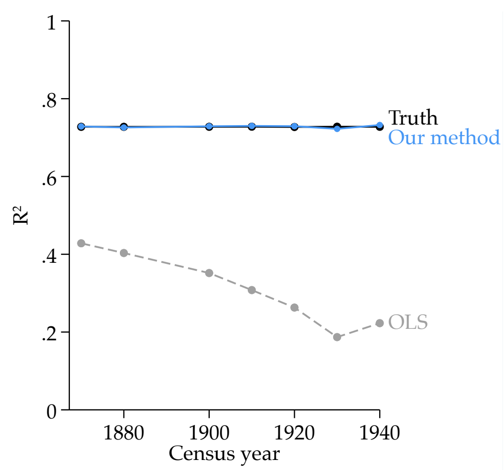
3.1 Figures

FIGURE IV.59: Validation of the Semi-parametric Latent Variable Method

(A) Education ranks vs. dummies (1940 census)

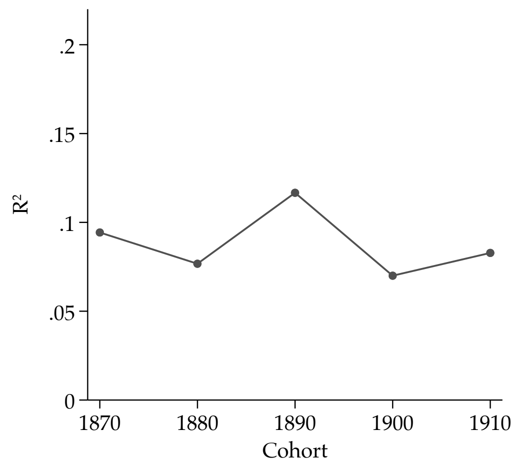


(B) Literacy dummies over time (simulation)



Notes: This figure demonstrates the effectiveness of our semi-parametric latent variable method in identifying rank-rank relationships from binary proxies. Panel A contrasts the R^2 values from rank-rank regressions using actual and binarized educational data from the 1940 census. We binarize the data by arbitrarily categorizing individuals based on their educational attainment: more than 11 years for children, 9 for mothers, and 7 for fathers. Each dot represents a US state, weighted by sample size and focusing on children aged 13—21 living with parents. Panel B illustrates a simulation where literacy serves as a binary proxy for human capital. We simulate human capital ranks, convert them into literacy dummies based on historical literacy rates, and compare the R^2 values from regressions using these dummies. The "Truth" line represents the R^2 from a human capital rank-rank regression, "Our method" from our latent variable method using literacy dummies, and "OLS" from a standard OLS regression with the same literacy dummies. In the 1940 census, instead of literacy, we observe the highest year of school or degree completed. We classify individuals who have completed at least two grades of school as literate; others we classify as illiterate.

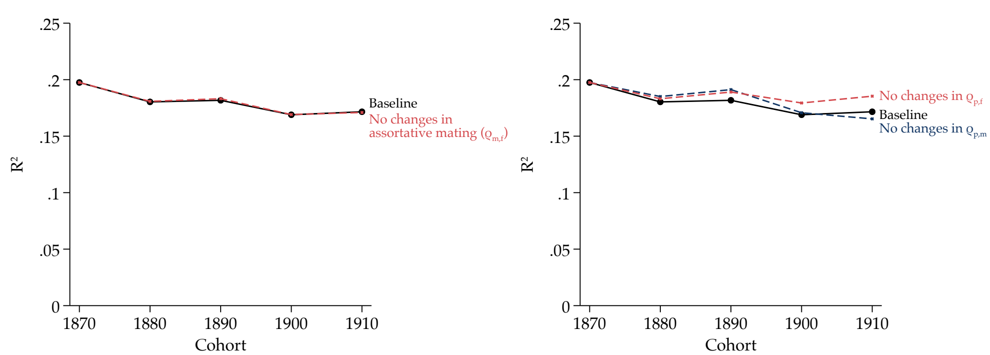
FIGURE IV.60: Mobility Estimates Based on "occscores"



Notes: This figure shows the share of the variance in a child's household income rank explained by (1) parents' household income ranks and their (latent) human capital ranks (R^2) and (2) parents' household income ranks alone. For parental human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. We use the household head's occupational income score ("occscore"). Results are based on our new panel and sample weights are applied.

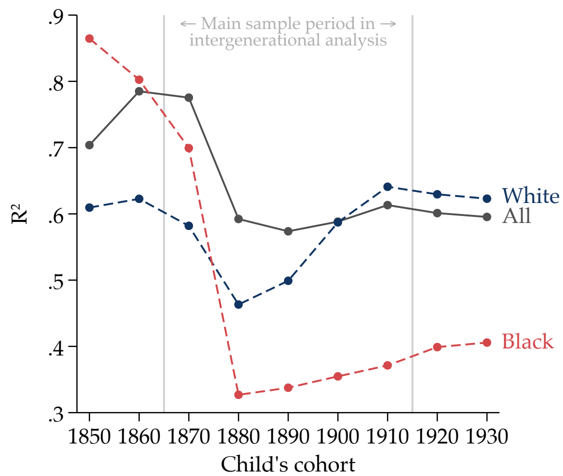
FIGURE IV.61: Mobility and the Impact of Evolving Parental Input Correlations

(A) Assortative mating (B) Correlations in human capital and income



Notes: This figure shows the role of each parameter on the R^2 in equation (III.2). The baseline represents the observed R^2 shown in Figure III.4. The other three lines represent the counterfactual R^2, had the respective parameter not changed over time, computed using the decomposition in equation (III.3). For parental human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. We use the household head's LIDO occupational income score (Saavedra and Twinam, 2020). Results are based on our new panel and sample weights are applied.

FIGURE IV.62: Assortative Mating Estimates by Group

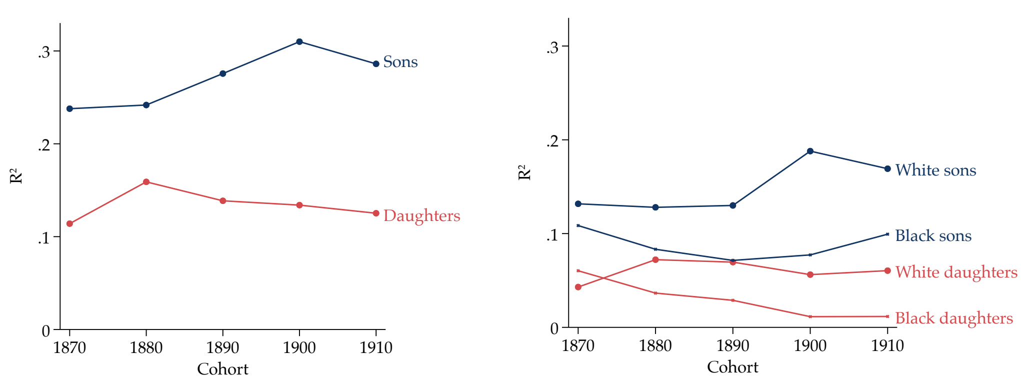


Notes: This Figure shows the share of the variance in a person's (latent) human capital rank explained by their spouse's (latent) human capital rank (R^2) across their child's cohort. For human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. Results are based on the full census cross-section of two-parent households with children aged 1 to 16. Note that as we show in Appendix 3.3.1, in this univariate rank-rank model, $R^2 = \beta^2 = \rho^2_{x,y}$, allowing researchers to directly compare our estimates of assortative mating to (the square of) conventional rank-rank correlations.

FIGURE IV.63: Within-Group Mobility Estimates

(A) Mobility by sex

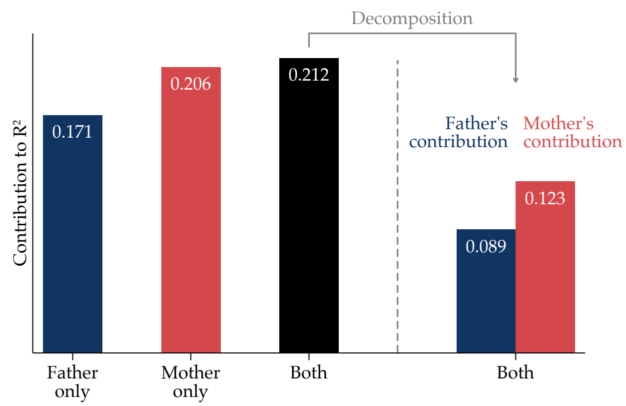
(B) Mobility by sex and race



Notes: This Figure shows the share of the variance in a child's household income rank explained by parents' household income ranks and their (latent) human capital ranks (R^2) across cohorts and groups. For parental human capital ranks, we use information on parental literacy and the latent variable method introduced in section 3.4. We use the household head's LIDO occupational income score (Saavedra and Twinam, 2020). Results are based on our new panel and sample weights are applied.

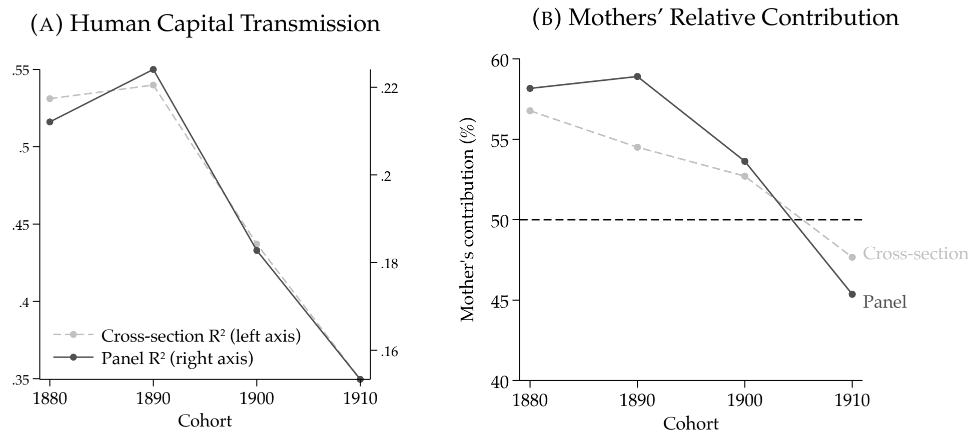
FIGURE IV.64: Illustrating our Decomposition Method

Intergenerational Transmission of Human Capital



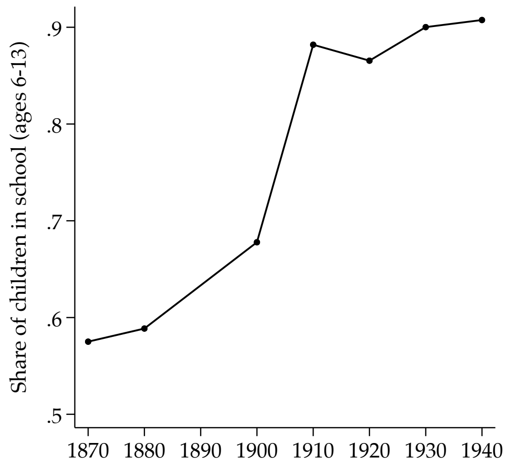
Notes: This figure shows the share of the variance in a child's (latent) human capital rank explained by parents' (latent) human capital ranks (R^2). We recover human capital rank-rank transmission using information on literacy and the latent variable method introduced in section 3.4. We decompose the overall R^2 using the Shapley-Owen method to quantify each parent's contribution. Results are based on our new panel, specifically children born in the 1880s; sample weights are applied.

FIGURE IV.65: Panel-Based Estimates of Human Capital Mobility Across Cohorts



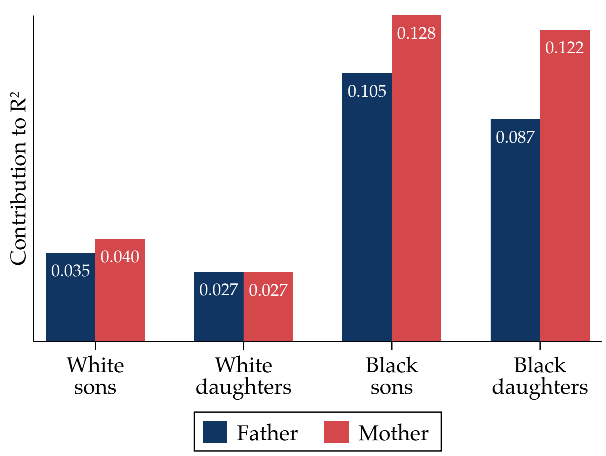
Notes: This figure compares our baseline results of human capital transmission from the cross-section of children who live with their parents to estimates based on our new panel. Panel A shows the share of the variance in a child's (latent) human capital rank explained by parents' (latent) human capital ranks (R^2) across cohorts. We recover human capital rank-rank transmission using information on literacy and the latent variable method introduced in section 3.4. Panel B shows mothers' relative contribution to the overall R^2 using the Shapley-Owen method. Cross-sectional results are based on the census cross-section of children ages 13–16 in their parents' household; panel results are based on individuals of any age.

FIGURE IV.66: Increasing Access to Schools



Notes: This figure shows the share of children aged 6–13 who attend school across time.

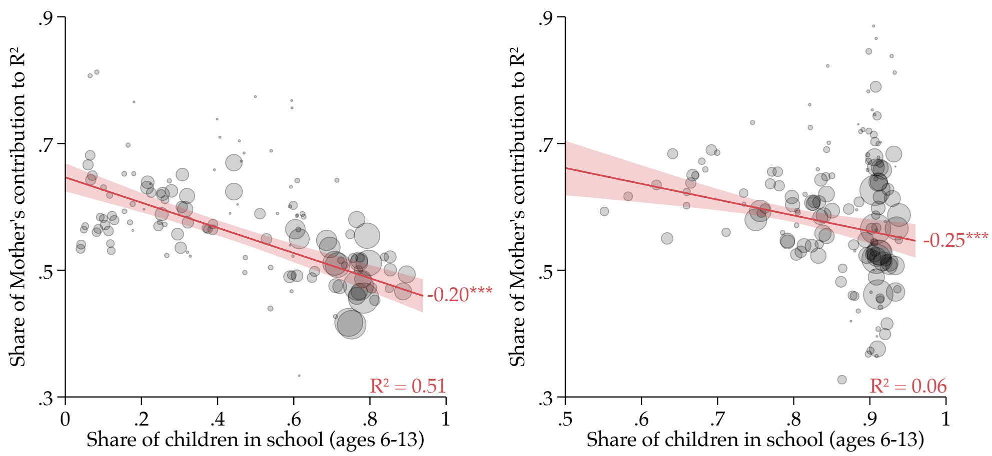
FIGURE IV.67: Intergenerational Transmission of Formal Schooling (1920s cohort)



Notes: This figure shows the share of the variance in a child's years of education rank explained by parents' years of education ranks (R^2). The figure focuses on the 1920s cohort (children aged 13–16 in the 1940 census—the only historical census that records years of education). We decompose the overall R^2 using the Shapley-Owen method to quantify each parent's contribution. Results are based on the census cross-section of children in their parents' household.

FIGURE IV.68: Mothers' Human Capital as Substitute for Local Schools

(A) 1880s Cohort (B) 1920s Cohort



Notes: This figure shows the relationship between local school access and mothers' *relative* contributions to child human capital (as a share of total variation explained). Literacy is used as the measure for rank-based transmission of human capital (section 3.4). Each dot represents a group of children born in the 1880s or 1920s, categorized by race, sex, and state. Sample size weights are applied. School access is determined by the race- and sex-specific share of children aged 6–13 in school. Results are based on the census cross-section of children ages 13–16 in their parents' household.

3.2 Tables

TABLE IV.30: Mothers & Schools—Robustness to Measures of School Access

	ϕ_{Mother}	ϕ_{Father}	$\frac{\phi_{\text{Mother}}}{R^2}$	ϕ_{Mother}	ϕ_{Father}	$\frac{\phi_{\text{Mother}}}{R^2}$
Baseline measure of school access	-0.18***	0.04	-0.20***			
	(0.03)	(0.05)	(0.03)			
Refined measure of school access				-0.47***	0.15	-0.58***
(accounts for attendance, term lengths, etc.)				(0.08)	(0.11)	(0.10)
R^2	0.39	0.02	0.51	0.37	0.04	0.57
Observations	133	133	133	128	128	128

Notes: This table shows the relationship between local school access and parents' contributions to child human capital. Columns 1–3 (baseline) contain the results from Figure III.8 and Panel A of Appendix Figure IV.68. For this baseline, school access is determined by the race- and sex-specific share of children aged 6–13 in school according to the 1880 census. Columns 4–6 show that these results are even stronger when we use an alternative measure of school access. For this measure, we newly digitized data on state-specific school ages, enrollment, attendance, and term lengths from the Census Statistical Abstracts. From these data, we compute the average likelihood of attending school on any given day in the year between ages 6–16, specific to each state. These data are incomplete for Arkansas and Wyoming, leading to slightly lower sample sizes.

3.3 Methods Appendix

3.3.1 Relation Between R^2 and Coefficients

3.3.1.1 One input In a linear regression with a single explanatory variable, $y_i = \alpha + \beta x_i + \varepsilon_i$, the coefficient β and the R^2 are defined as follows:

$$\widehat{\beta} = \text{cor}(x, y) \cdot \sqrt{\frac{\text{Var}(y)}{\text{Var}(x)}} \tag{IV.19}$$

$$R^2 = \text{cor}(x,y)^2 = \widehat{\beta}^2 \cdot \frac{\text{Var}(x)}{\text{Var}(y)}, \tag{IV.20}$$

where $\text{cor}(x,y)$ is the correlation between y and x and $\text{Var}(y)$ is the variance of y_i.

Rank-rank coefficients. Rank-rank coefficients are a popular measure of mobility. By construction, quantile-ranked outcomes share the same distribution. Therefore, if both y and x are outcomes in quantile-ranks, we have $\text{Var}(y) = \text{Var}(x)$ so that $R^2 = \widehat{\beta}^2$.

Intergenerational elasticity coefficients. Intergenerational elasticities are another common measure of mobility. Such elasticities are estimated in a regression of $\log(y)$ and $\log(x)$ where y and x are a child and a parent's outcome, respectively. Such an elasticity is equal to $\sqrt{R^2}$ if and only if $\text{Var}(\log(y)) = \text{Var}(\log(x))$. A sufficient condition for these variances to equate is that the marginal distribution of children's outcomes are a shifted version of that of the parents, i.e. $y \sim bx$ for some $b > 0$.

3.3.1.2 Multiple inputs

In a multivariate linear regression, $y_i = \alpha + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \varepsilon_i$, the R^2 depends on the parameters $\beta_1, \ldots, \beta_k$ and the variance-covariance matrix of the explanatory variables. That is,

$$R^2 = \frac{Var\left(\sum_{j=1}^{k} \widehat{\beta}_j x_{i,j}\right)}{Var(y)} = \frac{\sum_{j=1}^{k} \widehat{\beta}_j^2 Var(x_j) + 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} \widehat{\beta}_j \widehat{\beta}_l Cov\left(x_j, x_l\right)}{Var(y)}. \tag{IV.21}$$

Rank-rank coefficients. Again, using that quantile-ranked outcomes share the same distribution by construction—i.e., $\text{Var}(y) = \text{Var}(x_j)\ \ \forall j = 1, \ldots, k$—we obtain

$$R^2 = \sum_{j=1}^{k} \widehat{\beta}_j^2 + 2\sum_{j=1}^{k-1}\sum_{j=i+1}^{k} \widehat{\beta}_j \widehat{\beta}_l \widehat{\rho}_{j,l} \tag{IV.22}$$

where $\widehat{\rho}_{j,l}$ is the correlation between x_j and x_l.

3.3.2 Shapley-Owen Decomposition of the R^2

The Shapley-Owen decomposition of R^2 (Shapley, 1953; Owen, 1977) provides a way to quantify the contribution of each independent variable to a model. The method was introduced in cooperative game theory as a method for fairly distributing gains to players. It has been used more recently as a way to interpret black-box model predictions in machine learning (Redell, 2019; Lundberg and Lee, 2017), as well as in some economics research on inequality (Azevedo et al., 2012; Fourrey, 2023).

For a given set of k vectors of regressors $V = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k\}$, we create sub-models for each possible permutation of vectors of regressors.

The marginal contribution of each vector of regressor $\boldsymbol{x}_j \in V$ is:

$$\Delta_j = \sum_{T \subseteq V - \{\boldsymbol{x}_j\}} \left[R^2(T \cup \{\boldsymbol{x}_j\}) - R^2(T)\right]$$

where $R^2(T)$ represents the R^2 of regressing the dependent variable on a set of variables $T \subseteq V$ (e.g., $V = \{y_i^{\text{mother}}, y_i^{\text{father}}\}$). The marginal contribution gives us the sum of the contributions that the vector of regressors $\boldsymbol{x}_j$ makes to the R^2 of each sub-model. Then, the Shapley-value ϕ_j for the vector of regressors $\boldsymbol{x}_j$ is obtained by normalizing each marginal contribution so that they sum to the total R-squared:

$$\phi_j = \frac{\Delta_j}{k!}, \tag{IV.23}$$

where k is the number of vectors of regressors in V (i.e., $k = |V|$). Each ϕ_j then corresponds to the goodness-of-fit of a given vector of regressor, and they sum up to equal the model's total R^2. Using this method, perfect statistical substitutes will receive the same Shapley value.

3.3.2.1 Example with two inputs Table IV.31 shows an example for the Shapley-Owen decomposition of the R^2 for the case of two parental inputs, omitting their interaction. We add variables at every column, leading up to the full two-parent model containing the outcomes of both fathers and mothers. Note that the individual parental contributions (i.e., Shapley values) sum up to the total R^2 of 0.25 in the two-parent model. In this case, mothers account for 64 percent of the variation in child outcomes explained by parental background.

TABLE IV.31: Example of Shapley-Owen Decomposition

Empty Model		One-Parent Model		Two-Parent Model		Marginal Contribution (Δ_j)	
Regressors	R^2	Regressors	R^2	Regressors	R^2	Father	Mother
∅	0.0	Father	0.08	Father, Mother	0.25	$0.08 - 0 = 0.08$	$0.25 - 0.08 = 0.17$
∅	0.0	Mother	0.15	Father, Mother	0.25	$0.25 - 0.15 = 0.10$	$0.15–0 = 0.15$
				Shapley Value (ϕ_j)		$\frac{0.08+0.1}{2!} = 0.09$	$\frac{0.17+0.15}{2!} = 0.16$

3.3.2.2 Unpacking the Shapley-value with two inputs To better understand what the Shapley-value for each parental input comprises, we express it as a function of regression coefficients, variances, and covariances in the two-input case. Let ϕ_1 be one parent's Shapley value—i.e., the contribution that the parent's input makes to the overall R^2 when regressing child outcomes on both parents' inputs. Applying equation (IV.23), we have

$$\phi_1 = \frac{1}{2}\Big(R^2(\{x_1, x_2\}) - R^2(\{x_2\}) + R^2(\{x_1\}) - R^2(\{\emptyset\})\Big).$$

Further, using equation (IV.21), we have

$$\phi_1 = \frac{1}{2}\left(\left[\widehat{\beta}_1^2 + \widehat{\beta}_{1,univ}^2\right]\frac{Var(x_1)}{Var(y)} + \left[\widehat{\beta}_2^2 + \widehat{\beta}_{2,univ}^2\right]\frac{Var(x_2)}{Var(y)} + 2\widehat{\beta}_1\widehat{\beta}_2\frac{Cov(x_1, x_2)}{Var(y)}\right),$$

where $\widehat{\beta}_{1,univ}^2$ is the coefficient on the mother's input in a univariate regression and $\widehat{\beta}_1^2$ the coefficient on the mother's input in the multivariate regression including the father's input. Using the omitted variable bias formula, $\widehat{\beta}_{1,univ}^2 = \widehat{\beta}_1 + \widehat{\beta}_2\frac{Cov(x_1,x_2)}{Var(x_1)}$, we have

$$\phi_1 = \frac{1}{2Var(y)}\left(2\widehat{\beta}_1^2 Var(x_1) + \{Cov(x_1, x_2)\}^2\left[\frac{\widehat{\beta}_2^2}{Var(x_1)} - \frac{\widehat{\beta}_1^2}{Var(x_2)}\right] + 2\widehat{\beta}_1\widehat{\beta}_2 Cov(x_1, x_2)\right).$$

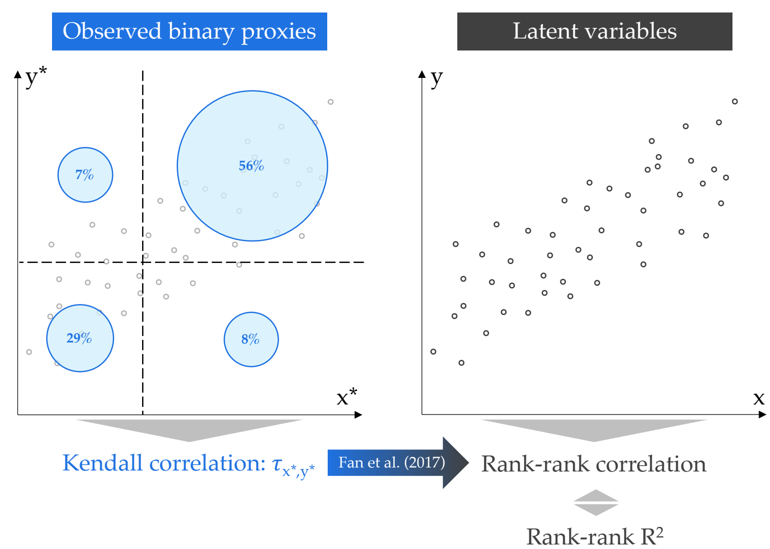
For rank-rank regressions, we have

$$\begin{aligned}\phi_1 &= \widehat{\beta}_1^2 + \frac{1}{2}\left(\widehat{\beta}_2^2 - \widehat{\beta}_1^2\right)\left(\frac{Cov(x_1,x_2)}{Var(y)}\right)^2 + \widehat{\beta}_1\widehat{\beta}_2\frac{Cov(x_1,x_2)}{Var(y)} \\ &= \widehat{\beta}_1^2 + \frac{\widehat{\rho}_{1,2}^2}{2}\left(\widehat{\beta}_2^2 - \widehat{\beta}_1^2\right) + \widehat{\beta}_1\widehat{\beta}_2\widehat{\rho}_{1,2}.\end{aligned}$$

3.3.3 Semi-parametric latent variable method

We use the semi-parametric latent variable method introduced by Fan et al. (2017) to estimate rank-rank mobility (R^2) when only binary proxies of the underlying rank variable are observed. The rank-rank regression of interest is that in equation (III.1).

FIGURE IV.69: Illustrating the Semi-Parametric Latent Variable Method



Notes: This figure illustrates the semi-parametric latent variable method, recovering rank-rank mobility (R^2) in latent variables from observed binary proxies. Assuming that the underlying latent variables are drawn from a joint Gaussian copula distribution, pairwise rank-rank correlations can be identified from Kendall's correlation between the observed binary proxies using the bridging function in (IV.26). Rank-rank regressions can be identified from the pairwise correlation matrix using equations and (IV.27) and (IV.28).

We assume that the dependent and independent variables are drawn from a joint Gaussian copula distribution. That is, we assume that there exists a set of unknown monotonic transformations $f_y, f_1, \cdot\cdot, f_k$ such that $f_y(y_i), f_1(x_{1i}), f_k(x_{ki}) \sim \mathcal{N}(0, \mathbf{\Sigma})$ with $\text{diag}(\mathbf{\Sigma}) = \mathbb{1}$. Because we allow for any monotonic transformation, the assumption that the marginal distributions have zero mean and variance equal to 1 is without loss of generality. Note that the normality assumption does not impose that the latent variables of interest (e.g., human capital) are jointly normally distributed. Rather, it requires that there exists some monotonic transformation of the latent variables that is jointly normally distributed.

Fan et al. (2017) show how to estimate all elements of $\mathbf{\Sigma}$ even if only binary proxies of the rank variables of interest are available. For example, let us consider Σ_{12}, the correlation between $f_y(y_i)$ and $f_1(x_{1i})$. We summarize the more formal arguments by Fan et al. (2017). Three cases are considered. First, that both y_i and x_{1i} are observed. Second, that y_i is observed, but only a binary proxy of x_{1i} is observed. That is, we observe only $\tilde{x}_{1i}$ which is one if x_{1i} is above an arbitrary cut-off and zero otherwise. Third, that only observe binary proxies of each variable are observed.

Case 1: Both rank variables observed. Fan et al. (2017) show that Σ_{12} is an increasing function of the Kendall's rank correlation coefficient τ_{12}. Therefore, observing the ranked variables is sufficient to identify Σ_{12}. Specifically, the "bridging function" between Kendall's rank correlation coefficient and Σ_{12} is

$$\Sigma_{12} = \sin\left(\frac{\pi}{2}\tau_{12}\right). \tag{IV.24}$$

Therefore, our estimate $\hat{\Sigma}_{12}$ is the sample equivalent of equation (IV.24).

Case 2: One rank variable and one binary proxy observed. In this case, we observe $\text{rank}(y_i)$ but we only observe the binary proxy $\tilde{x}_{1i}$. In such cases, Fan et al. (2017) show that

$$\tau_{12} = 4\Phi_2\left(\Delta_2, 0, \frac{\Sigma_{12}}{\sqrt{2}}\right) - 2\Phi\left(\Delta_2\right) \tag{IV.25}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution, $\Phi_2(u, v, t)$ is the CDF of a bivariate normal distribution with correlation coefficient t, evaluated at u and v. Δ_2 is the cut-off value above which the binary proxy is 1 and can be estimated as $\hat{\Delta}_2 = \Phi^{-1}(1 - \bar{x}_1)$ where $\bar{x}_1 \equiv \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_{1i}$. Because equation (IV.25) is strictly increasing in Σ_{12} (see (Fan et al., 2017) for the proof), Σ_{12} is identified as the unique root of equation (IV.25) where τ_{12} and Δ_2 are replaced with their finite sample analogues.

Case 3: Only binary proxies observed. For two binary proxies, the bridging function is

$$\tau_{12} = 2\Phi_2\left(\Delta_1, \Delta_2, \Sigma_{12}\right) - 2\Phi\left(\Delta_1\right)\Phi\left(\Delta_2\right). \tag{IV.26}$$

The right hand side of this equation is increasing in Σ_{12}. Since Δ_1, Δ_2, and τ_{12} can be estimated, Σ_{12} is identified as the unique root of equation (IV.26) where τ_{12}, Δ_1, and Δ_2 are replaced with their finite sample analogues.

The last step of the method is to estimate the parameters and R^2 of equation (III.1) from the pairwise correlations between the underlying random variables that are jointly normal. First, given two jointly normal random variables with correlation ρ, the correlation of their ranks (Spearman's rank correlation ρ_s) is equal to $\rho_s = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho}{2}\right)$. Let $\widehat{\boldsymbol{R}}$ be the rank-rank correlation matrix, i.e. $\widehat{\boldsymbol{R}}_{jl} = \frac{6}{\pi}\sin^{-1}\left(\frac{\hat{\Sigma}_{jl}}{2}\right)$ for each $l, j = 1, \ldots, k+1$. We use that the coefficients and R^2 in rank-rank regressions are identified from the rank-rank correlation matrix (again using that the marginal distributions of all ranked variables are equal). Specifically,

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{R}}_x\right)^{-1}\widehat{\boldsymbol{R}}_{xy} \tag{IV.27}$$

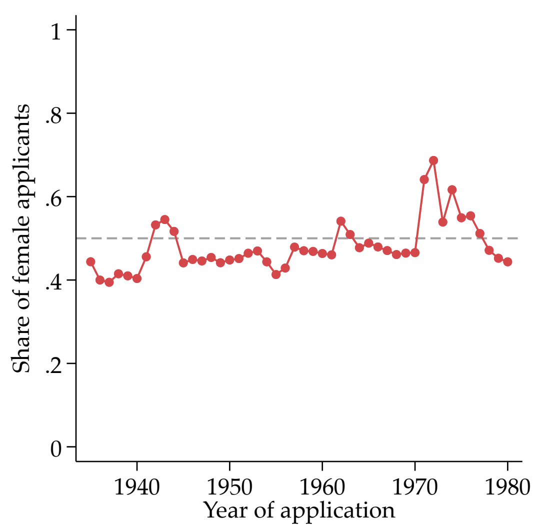
where $\widehat{\boldsymbol{R}}_x$ is a $k \times k$ rank-rank correlation matrix of the independent variables and $\widehat{\boldsymbol{R}}_{xy}$ is a $k \times 1$ vector of rank-correlations between the independent variable and dependent variable. $\hat{\alpha}$ is then computed as $\bar{y} - \widehat{\boldsymbol{\beta}}'\bar{\boldsymbol{x}}$. Similarly, R^2 is estimated as

$$R^2 = \widehat{\boldsymbol{R}}'_{xy}\left(\widehat{\boldsymbol{R}}_x\right)^{-1}\widehat{\boldsymbol{R}}_{xy}. \tag{IV.28}$$

Equations (IV.27) and (IV.28) are numerically equivalent to the rank-rank coefficient vector and R^2 in the case without latent variables (for a proof, see e.g., O'Neill (2021) and impose that the marginal distributions of the variables are identical). From equations (IV.27) and (IV.28), we also see the relation between the slope coefficient and R^2 and in the univariate case discussed in Appendix 3.3.1.1: $\hat{\beta} = \sqrt{R^2}$.

3.4 Data Appendix

FIGURE IV.70: Share of Female Applicants

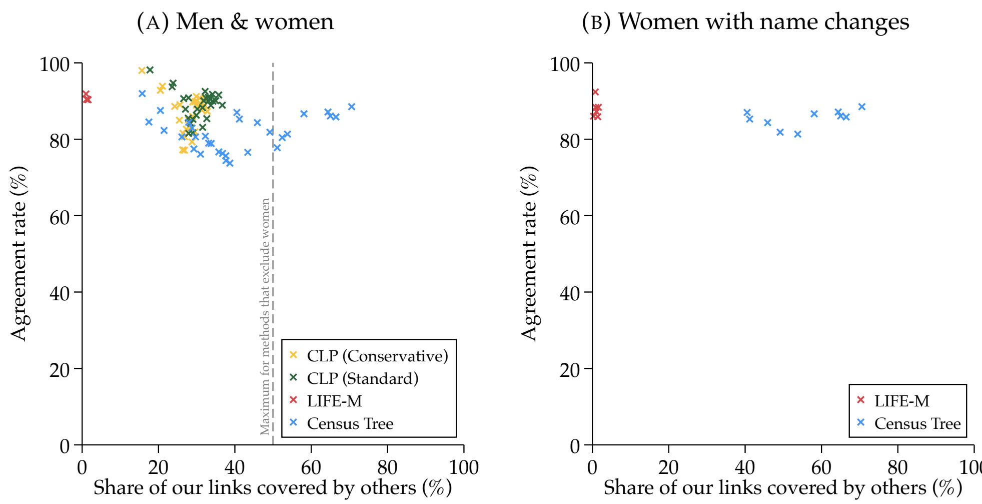


Notes: This figure shows the share of SSN applicants who are female by year of application.

FIGURE IV.71: Sample Balance Prior to Weighting (1850–1920)

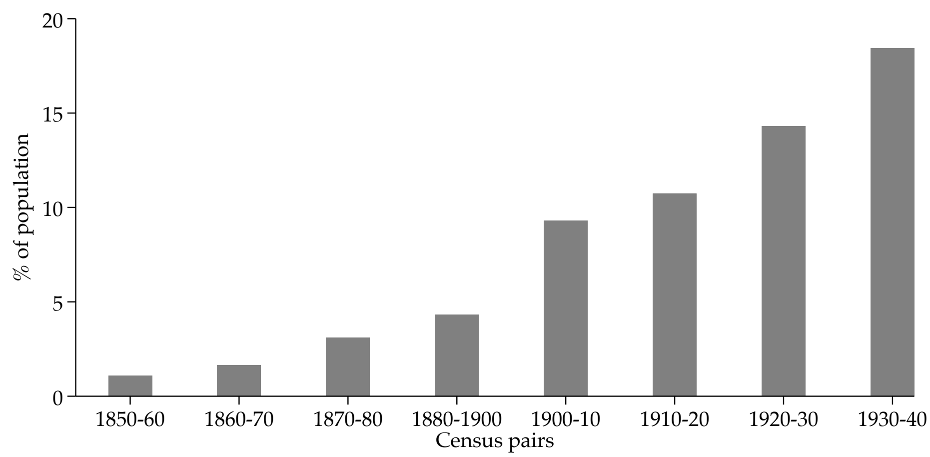
Notes: This figure shows the representativeness of characteristics among individuals who we successfully assign an SSN compared to the full population in each census before 1940. The sample is exceptionally representative compared to existing panels, most notably with respect to sex and race. Because of the large sample sizes, even economically small differences are statistically significant.

FIGURE IV.72: Our New Panel Compared to Existing Data



Notes: This figure compares our linked panel (1850–1940) to those of the Census Linking Project (CLP, Abramitzky et al., 2020), LIFE-M (Bailey et al., 2022), and the Census Tree (Buckles et al., 2023). Each point represents a link from one census decade to another (potentially non-adjacent). The x-axis shows the share of individuals in our panel who were not yet captured by previously existing datasets. The y-axis shows the share of agreement with previously existing datasets on which precise records are linked, conditional on having established any link.

FIGURE IV.73: Fraction of US Population Linked in Our New Panel



Notes: This figure shows the fraction of the full population of men and women that we successfully link from one census decade to the next. Our empirical analysis also leverages links across non-adjacent census pairs, further increasing coverage.

3.5 Linking Procedure

We develop a multi-stage linking process built on the procedural record linkage method developed by Abramitzky et al. (2021b). Our process consists of three stages. 1) linking SSN applications to census records. 2) Identifying the applicant's parents in the census. 3) Tracking these parents' census records over time. With our linking method, we are able to maximize the number of SSN-census links and subsequently build a multigenerational family tree for each linked SSN applicant.

First stage: Applicant SSN ↔ census.

- *Preparing SSN data*: We use a digitized version of the Social Security Number application data from the National Archives and Records Administration (NARA) known as the Numerical Identification Files (NUMIDENT). We harmonize the application, death and claims files to capture all the available information of each SSN record. These data include each applicant's name, age, race, place of birth, and the maiden names of their parents. We recode certain variables to align with census data, for example, we ensure codes for countries of birth, race and sex are consistent across the SSN and Census. Additionally, we apply the ABE name cleaning method to names of applicants and their parents resulting in an "exact" and a NYSIIS cleaned version of all names (Abramitzky et al., 2021a)[8].

- *Preparing Census data*: Within each census decade from 1850 and 1940, we apply the same name cleaning algorithm used to clean the SSN data. Where available, we extract parent and spouse names from each individual's census record to create crosswalks that are later used in the linking process. Each cleaned census decade is subsequently divided into individual birthplace files for easing the computational intensity of the linking procedure.

- *Linking SSN to Census records*: Our goal is to achieve a high linkage rate of SSN applications to the census, while ensuring the accuracy of each link. Our linking algorithm has the following steps:

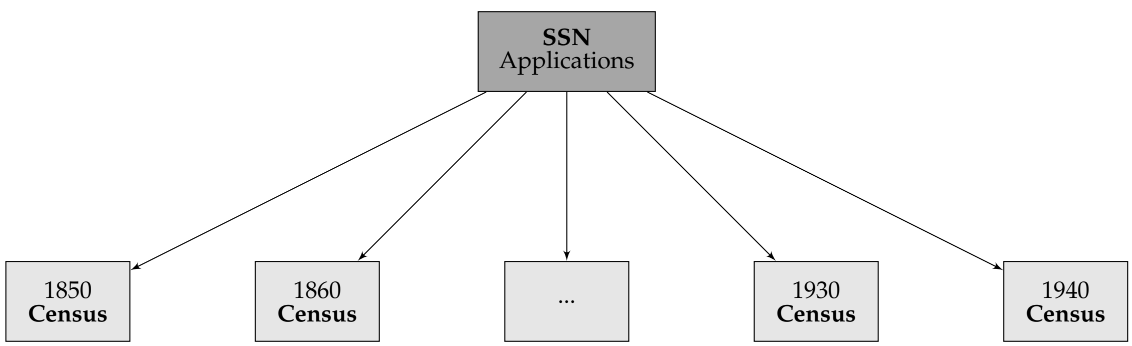
    1. We first create a pool of potential matches by finding all possible links between an SSN application and census record using first and last name (NYSIIS), place of birth, marital status and birth year within a 5-year age band. In the census, we identify marital status from the census variable "marst" or whether her position in the household is described as spouse. In the SSN data, we identify marital status if the applicants last name is different from that of her father.
    2. Once we have established our pool of potential matches, we essentially rerun our linking process. However, we use additional matching variables in order to pin down the most likely correct link among the potential matches. In our first round of this process, we aim to pin down the correct link by matching using the following set of matching characteristics: exact first, middle and last names of both the applicant and their parents, exact birth month (when available), state or country of birth, race, and sex. An SSN application is either uniquely matched to a census record or not.
    3. We attempt a second round of the matching described in point 2. for all SSN applicants who were *not* uniquely matched to a census record. In this round, we keep all matching variables the same, however, we use the phonetically standardized version of the middle name to account for spelling discrepancies. Once again, we separate those SSN applications that were uniquely matched to the census and those that were not.
    4. We repeat this matching process where we remove successfully matched individuals and attempt to rematch unmatched applications from our pool of

[8] The use of the NYSIIS phonetic algorithm helps in matching names with minor spelling differences, as mentioned in Abramitzky et al. (2021a)

potential matches. As we progress through the rounds of linking, the additional matching criteria become less stringent. We allow for misspellings or remove one or more variables in each subsequent iteration until we arrive at the literature standard, which involves only first and last name with spelling variations allowed, state of birth, and year of birth within a 5-year band.

We attempt to match each SSN record to all the census decades available as an individual may appear in the 1900 and 1910 census, for example. For married women applicants, we search for potential census matches using both their maiden and married names. As a result, if we are able to find both records, married women appear in our data twice. We assign these links a slightly altered SSN to differentiate between the married and unmarried SSN-Census link. We do not link married women in the census who are below the age of 16.

FIGURE IV.74: First & Second Linking Stages

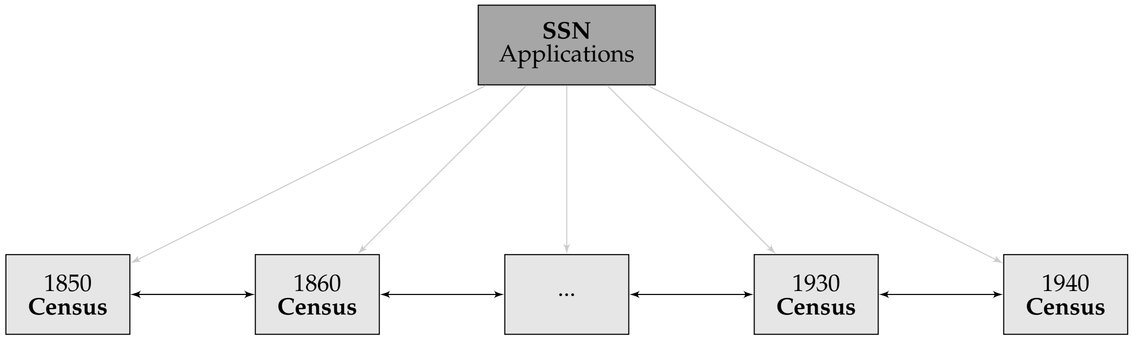


Notes: This figure shows the first and second step of our linking procedure—linking individuals' Social Security Numbers to their census records.

Second stage: SSN applicant parents $\leftrightarrow$ census. Specific birth details for mothers and fathers are not available in the SSN applications meaning we cannot directly link them like we do for the applicants. However, if we can successfully link an SSN applicant to their childhood census record, it is possible to identify and link their parents to other census decades. This process also allows us to identify grandparents. Importantly, we have mother's maiden in the SSN application data, allowing us to link a married mother to her unmarried census record. For parents that we are able to identify in the census from a successful SSN-census link, we apply the same matching procedure described above. However, an important difference is that we do not use parent names (as we no longer have that information), but we are able to use spouse name and information on their parents' birthplace (i.e., the SSN applicant's grandparents birthplace) which is available from the census records. For parents who are not SSN applicants themselves, we create a synthetic identifier similar to an SSN.

Third stage: Census $\leftrightarrow$ census. Having assigned unique SSNs or synthetic identifiers to millions of individuals in the census records, we can link these records over time. We cover all possible pairs of census decades from 1850 to 1940.

FIGURE IV.75: Final Linking Stage



Notes: This figure shows the final step of our linking procedure—linking individuals' census records over time. Once we have linked SSN applications to the census as well as linked their parents where possible (stage one and two), we link individuals across censuses despite potential name changes upon marriage.

3.6 Sample Weight Construction

We use inverse propensity score weights so that our sample is representative of the overall population across key observable characteristics.

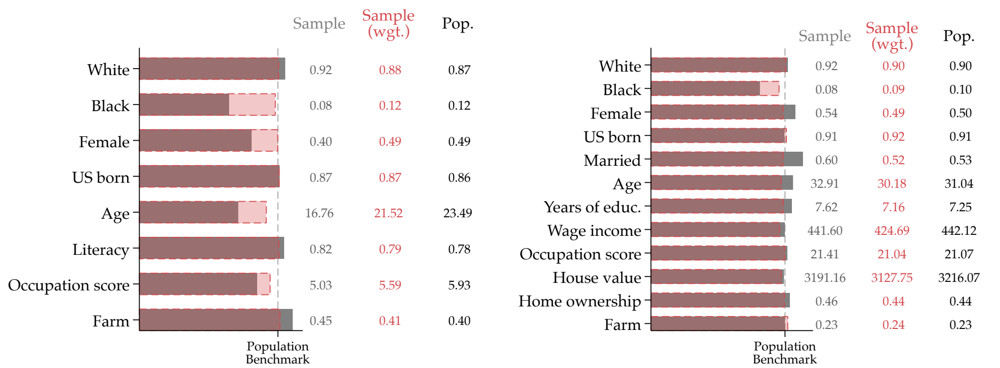
For each census between 1850 to 1940, we create indicator variables for whether (1) we have identified an individual's Social Security Number, (2–4) whether we have been able to measure the economic status of the individual's (2) mother, (3) father, or (4) both parents. Measuring parental economic status may itself involve census linking and does not rely on observing parents in the same census wave.

In a second step, we then divide the population into groups based on their observable characteristics and (non-parametrically) compute the propensity of each group to be included in our sample via indicators (1–4). Those groups are comprised of individuals with equal (i) sex, (ii) race, (iii) age in decades, (iv) region, (v) farm-status, (vi) literacy, (vii) rural-urban status, (viii) state of birth, (ix) homeownership, (x) marital status, (xi) school attendance, (xii) occupational group, and (xiii) industry group.

As the final sample weight, we assign an individual the inverse propensity of being observed in our linked panel given the characteristic-based group to which they belong. We use different sample weights depending on whether we require only the individual to be linked across time (1), observing the person's and their mother's economic status (2), observing the person's and their father's economic status (3), or observing the person's and both of their parents' economic status (4).

FIGURE IV.76: Sample Balance After Inverse Propensity Weighting (1870 & 1940)

(A) 1870 (B) 1940



Notes: This figure shows the representativeness of characteristics among individuals who we successfully assign an SSN compared to the full population in each census before 1940. The sample is exceptionally representative compared to existing panels, most notably with respect to sex and race. Our inverse propensity weights produce an almost perfectly representative sample. Panel A shows the 1870—typically the first year we include in our results—and Panel B shows 1940—the last year of our panel.

Figure IV.76 shows average sample characteristics after applying our new inverse propensity weights. The reweighted sample is almost perfectly representative of the full population in all dimensions, even those not targeted by our reweighting method. For example, wage income and occupational income scores match close to perfectly despite only having included coarse occupation and industry categories in our reweighting procedure. Similarly, housing wealth is not targeted but our reweighted sample closely mirrors the overall population.

Bibliography

ABRAMITZKY, R., L. BOUSTAN, E. JACOME, AND S. PEREZ (2021a): "Intergenerational Mobility of Immigrants in the United States over Two Centuries," *American Economic Review*, 111, 580–608.

ABRAMITZKY, R., L. BOUSTAN, AND M. RASHID (2020): "Census Linking Project: Version 1.0," dataset: `https://censuslinkingproject.org`.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2021b): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

AZEVEDO, J. P., V. SANFELICE, AND M. C. NGUYEN (2012): "Shapley Decomposition by Components of a Welfare Aggregate," .

BAILEY, M. J., P. Z. LIN, S. MOHAMMED, P. MOHNEN, J. MURRAY, M. ZHANG, AND A. PRETTYMAN (2022): "LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database," dataset: `https://doi.org/10.3886/E155186V2`.

BUCKLES, K., A. HAWS, J. PRICE, AND H. WILBERT (2023): "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Working Paper.

FAN, J., H. LIU, Y. NING, AND H. ZOU (2017): "High dimensional semiparametric latent graphical model for mixed data," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79, 405–421.

FOURREY, K. (2023): "A Regression-Based Shapley Decomposition for Inequality Measures," *Annals of Economics and Statistics*, 39–62.

LUNDBERG, S. M. AND S.-I. LEE (2017): "A Unified Approach to Interpreting Model Predictions," Working Paper.

O'NEILL, B. (2021): "Multiple Linear Regression and Correlation: A Geometric Analysis," *arXiv preprint arXiv:2109.08519*.

OWEN, G. (1977): "Values of games with a priori unions," in *Essays in Mathematical Economics and Game Theory*, ed. by R. Heim and O. Moeschlin, New York: Springer.

REDELL, N. (2019): "Shapley Decomposition of R-Squared in Machine Learning Models," Working Paper.

SAAVEDRA, M. AND T. TWINAM (2020): "A machine learning approach to improving occupational income scores," *Explorations in Economic History*, 75, 101304.

SHAPLEY, L. (1953): "A value for n-person games," in *Contributions to the Theory of Games*, ed. by H. Kuhn and A. Tucker, Princeton University Press, vol. 2.