NORTHWESTERN UNIVERSITY

# Essays in Labor Economics, Political Economy, and Economic History

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Managerial Economics and Strategy

By
Carlo Medici

EVANSTON, ILLINOIS

June 2024

# Abstract

This dissertation is composed of three essays at the intersection of Labor Economics, Political Economy, and Economic History. The first essay investigates the determinants of the formation and growth of labor unions. The second essay examines the effect of immigration restrictions on the labor market and economic development. The third essay studies the role and consequences of political connections on the careers and performance of public sector employees.

In the first essay, I show that immigration positively affected the emergence of organized labor in the United States. I digitize archival data to construct the first county-level dataset on historical U.S. union membership and use a shift-share instrument to isolate a plausibly exogenous shock to the labor supply induced by immigration, between 1900 and 1920. Counties that received more immigration experienced an increase in the probability of having any labor union, the share of unionized workers, the number of local union branches, and the average branch size. The increase occurred more prominently in counties more exposed to the immigrants' labor competition and harboring less favorable attitudes towards immigration. Taken together, these results indicate that existing workers formed and joined labor unions due to economic and social motivations. The findings shed light on a novel driver of unionization in the early 20$^{\text{th}}$-century United States: in the absence of immigration, the average union density of this period would have been 17% lower. They also identify an unexplored consequence of immigration: the development of institutions that aim to protect workers' status in the labor market.

In the second essay, which is joint work with Joe Long, Nancy Qian, and Marco Tabellini, we examine the impact of the 1882 Chinese Exclusion Act on the economic development of the Western United States. The ban of Chinese immigrants reduced the total number of Chinese workers across all sectors and skills. It had similar negative effects on other workers, including

native-born white workers and European immigrants – the intended beneficiaries of the Act – especially in the manufacturing, railroad, and mining industries. The Act also reduced manufacturing output, productivity, and the number of manufacturing establishments. The adverse economic effects were long-lasting and persisted until at least 1940.

In the third essay, which is joint work with Massimo Pulejo, we analyze the consequences of political connections in the civil service of the United States over more than two centuries. Focusing on the federal judiciary system, where political appointments are the selection method still used today, and leveraging individual-level data on judges and members of Congress from 1789 to the present, we use a difference-in-differences design to compare the careers and performance of judges before and after the senator who recommended their nomination leaves Congress. After losing the connection to their recommender, the probability of a judge being promoted from a district court to a court of appeals decreases by up to 48%. Such impact emerges in years in which judges share partisanship with the incumbent president, and they could thus benefit from the lobbying efforts of their political connection. This event has also sizable consequences on judges' performance: following the recommender's exit from Congress, judges write fewer judicial opinions, of shorter length, and of poorer quality, as proxied by both fewer backward and forward citations. These results are consistent with judges reducing their effort and productivity once their career prospects are drastically hindered.

## Acknowledgements

I am deeply indebted to my committee members, colleagues, friends, and family.

My exceptional group of advisors – Nancy Qian, Joel Mokyr, Matthew Notowidigdo, Nicola Persico, and Edoardo Teso – provided me with invaluable guidance, support, and suggestions. I could not have wished for better mentors.

My research also greatly benefited from the interactions with several faculty members at Northwestern, including Georgy Egorov, Carola Frydman, Walker Hanlon, Leander Heldring, and Silvia Vannutelli. A heartfelt thank you goes to Marco Tabellini for his precious feedback, excitement for research, and moral support over the past few years. Special thanks also to Lorenzo Incoronato and Massimo Pulejo for sharing their PhD journey with mine.

My years in Evanston and Chicago have been made memorable by the presence of many friends, particularly Riccardo Bianchi Vimercati, Devis Decet, Joris Mueller, Ola Paluszynska, Francisco Pareschi, Miguel Talamas, and Tomas Wilner. My first-year study group has provided irreplaceable help and company in one of the most stressful times of the program. My friends from home have made me feel loved throughout these years while being far away; seeing our friendship remain unchanged has been vital to me.

I am grateful to my family, especially my parents, for their unconditional love and support, without which none of this would have been possible.

Finally, my biggest thanks go to Laura Murphy. Her unwavering presence, love, care, and regard have been my main source of serenity over the past five years. Any of my accomplishments are also entirely hers.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

## Closing Ranks: Organized Labor and Immigration

### 1.1 Introduction

Labor unions are among the most important labor market institutions in all advanced economies. Throughout the 20$^{\text{th}}$ century, they have contributed to reducing inequality (Farber et al., 2021), improving working conditions (Rosenfeld, 2019), and influencing policy through extensive political activities (Ahlquist, 2017). Despite ebbs and flows in their membership, labor unions remain central to today's economy. In the U.S., they recently gained prominent victories for several categories of workers, including autoworkers, UPS drivers, and Hollywood writers.[1] In Europe and Canada, where collective bargaining also boasts a long tradition, organized labor continues to expand to previously unorganized sectors and to shape the policy agenda.[2] Given the long-lasting prominence of labor unions, it is perhaps surprising that we have relatively little evidence on the determinants of their emergence and growth. The primary aim of this paper is to address this question with systematic empirical evidence.

The origins of modern organized labor trace back to the Industrial Revolution. One

[1] The tentative agreements of October 2023 between the United Auto Workers and the three largest U.S. automakers (Ford Motor, General Motors, and Stellantis) have been defined as the most generous in decades (Ewing and Boudette, 2023). In August 2023, the Teamsters obtained an agreement with UPS that will allow their full-time drivers to make $170,000 annually in pay and benefits (Hadero and Ott, 2023). Unions' current approval rate is also among the highest recorded since 1965 (McCarthy, 2022), and in 2022, 224,000 workers were involved in work stoppages (Kallas et al., 2022).

[2] Recent examples are the collective bargaining agreements and strikes at Amazon warehouses in several European countries and Canada; the unionization drives at Tesla factories in Germany and Sweden; and, the massive mobilizations of 2023 against pension reforms in France and for better pay in Britain. Across most OECD countries, unions are associated with lower unemployment, higher productivity, and better job quality (OECD, 2019).

prevailing theory of why unions arose during this period stems from the increased capital intensity in industrial production, which shifted bargaining power away from laborers and toward the owners of capital (Foner, 1947; Webb and Webb, 1894). A related hypothesis is that workers organized in response to labor competition (Montgomery, 1979; Taft, 1964), which intensified during this period as boosts to agricultural productivity relieved labor from farming, and both the total population and the urban population share grew.

This paper investigates the second mechanism: the effect of a large and protracted increase in the labor supply on the formation and expansion of labor unions, leveraging the episodes of mass immigration to the United States of the early 20$^{\text{th}}$ century. The effect is *ex ante* ambiguous because it affects both the incentives of workers to organize and the ability of capital owners to undermine organized labor. On the one hand, the increased competition for jobs can motivate workers to organize in response to the economic threats to their employment and wages. On the other hand, a larger labor supply lowers the cost to business owners to replace uncooperative workers and break strikes. Thus, how an increased labor supply impacts unionization is ultimately an empirical question.

The context of the early 20$^{\text{th}}$-century United States provides an ideal setting to answer this question. First, the U.S. economy was already the largest in the world (Bolt and Van Zanden, 2020) and the labor movement experienced its first national expansion at the turn of the century (Foner, 1947). Unions represented workers' interests in the workplace and, at the same time, advocated for significant pro-labor legislation (Goldin, 1994; Mink, 2019). Second, these years witnessed the creation and growth of several labor unions that remain influential today (Stewart, 1926), despite the legal and judicial frameworks of the time allowing employers to easily dismiss and replace unionizing workers (Taft, 1964). Third, this context provides a natural experiment to establish causal identification, given by the large and prolonged influx of European immigrants during this period, often referred to as the Age of Mass Migration (Hatton and Williamson, 1998).

Two main challenges are associated with this study. The first is the need for disaggregated data on the presence and membership of labor unions. The only historical data available record unionization at the state or national level and, therefore, do not allow carrying out analyses across local labor markets. The second challenge is establishing causal effects. For example, the presence of unions may deter immigration. Such reverse causality would result in a negative association between immigrant flows and union presence. Alternatively, both the size of unions and immigration may increase in response to economic growth. Such joint determination would lead to a positive association between unionization and immigrants.

To measure unionization, I hand-collect and digitize archival documents on the quantity, location, and membership of labor union branches across the United States. The main sources of these records are the convention proceedings of the state federations of labor, which report detailed information on the number and location of union branches within each state's territory, along with the names of the delegates sent by each branch to the conventions. I collect these data for the years 1900, 1910, and 1920. To calculate the membership of each local branch, which was never systematically recorded in any historical document, I exploit the different constitutional rules of these state organizations, which specified that local union representation at the conventions be proportional to their membership. I complement these data with proceedings of national unions' annual conventions to improve and validate these measures. The information is then aggregated to the county and year levels, and merged with the historical U.S. Census. These data constitute the first comprehensive dataset measuring historical union presence and density (the share of unionized workers) at the county level in the United States.

To estimate the causal effect of immigration, I use a shift-share instrumental variable (Card, 2001b) to exploit plausibly exogenous variation in the flow of immigrants across counties in each decade. The instrument interacts the 1890 share of immigrants living in a given U.S. county and born in different European countries with the aggregate immigration flows from each country to the United States between 1890 and 1920. This identification strategy is motivated by the empirical regularity that immigrants tend to settle where other migrants from their own country of origin had previously settled, a process known as *chain migration*. The key underlying assumption is that, conditional on controls, the unobserved factors that affected unionization outcomes must not be jointly correlated with the 1890 composition of Europeans' enclaves across U.S. counties and the out-migration patterns from European countries after 1890.[3] I estimate 2SLS regressions that include county and year fixed effects, in addition to baseline county characteristics which are likely correlated with the initial presence of immigrants and the evolution of unionization, such as the urban share of the population and the labor force participation rate, interacted with year dummies.

The main results of this paper show that immigration positively affected the emergence of organized labor. Counties that received more immigrants as a fraction of the population experienced an increase in the probability of having a branch of any labor union, the share of unionized workers, the number of local union branches, and the average branch size. This novel finding documents empirically a new driver of unionization and highlights an unexplored effect of immigration in the labor market. According to the 2SLS estimates, a four percentage point (one standard deviation) increase in immigration raised the share of the unionized workforce by one percentage point. In areas receiving high volumes of immigration – such as New York County (NY) or the smaller Lake (IN) and Kenosha (WI) counties – immigration increased the fraction of union workers by 50–75%. A back-of-the-envelope calculation reveals that in the absence of immigration, the average union density between 1900 and 1920 would have been 17% lower overall. The estimates are robust to a variety sensitivity checks, such as using an alternative instrument that replaces actual immigration flows with plausibly exogenous ones and combining the instrument with a matching strategy.[4]

[3]For a formal discussion of the validity of shift-share designs, see also Adao et al. (2019), Borusyak et al. (2022), Goldsmith-Pinkham et al. (2020), and Jaeger et al. (2018).

[4]Although previous work has argued that this period is particularly suited to the use of shift-share instru-

The findings are also not sensitive to the inclusion of several additional controls, such as the initial size of the immigrant population (total and from each European country), the baseline shares of the labor force in major industries and occupations, and measures of income and economic growth.

In the second part of the paper, I explore the mechanisms driving the expansion of organized labor. First, I investigate the possibility that existing workers joined or created labor unions to respond to the threats posed by immigration to their employment and wages. Theoretically, this reaction should be possible only in occupations with entry barriers (e.g., requiring a certain level of human capital), where incumbent workers are not immediately replaceable and, hence, have an advantage in sustaining a labor union. This is particularly pertinent to the time period studied, when employers frequently resorted to strikebreakers and retaliated against unionizing workers (Foner, 1947; Taft, 1964). Differences in skill requirements across occupations provide a testing ground for this mechanism. Consistent with this hypothesis, immigration strengthened labor unions only in skilled trades. Immigration positively impacted skilled unionization along both the extensive and the intensive margin, as counties became more likely to have unions and saw an increase in union membership. Conversely, immigration had no effect on unions organizing primarily unskilled workers, such as miners, dockworkers, and laborers in the meat-packing or textile industries.

Second, I investigate whether counties in which immigrants more directly competed with existing workers experienced a larger increase in unionization. I construct a measure of exposure to the immigrants' labor market competition, whereby a county is more exposed if occupations prevalent among immigrants entering the United States in each decade are

ments (Abramitzky et al., 2023; Tabellini, 2020), the alternative instrument, which relies on predicted flows using weather shocks across European countries (Sequeira et al., 2020a), allows me to identify causal effects from the exogenous variation in the shocks, while allowing the exposure shares to be endogenous (Borusyak et al., 2022). Moreover, I build on Bazzi et al. (2023) and combine the instrument with a matching exercise, which selects within-state county pairs with the closest levels of union presence in 1890. All the robustness checks are described in Section 1.5.2.

also predominant among the U.S.-born workers of that county. In line with the hypothesis that unionization occurred as a reaction to the economic concerns brought by immigration, unions representing skilled workers expanded more prominently in counties more exposed to labor competition from immigrants. Instead, immigrants' competition slowed the growth of labor unions organizing unskilled workers, whose bargaining power was most weakened by the increased availability of replacement workers.

Third, I explore whether social motivations also contributed to the observed development of labor unions. Given the nativist rhetoric that accompanied the labor movement's support for immigration restrictions throughout the first half of the $20^{\text{th}}$ century (Goldin, 1994; Mink, 2019), one may expect that the cultural dissimilarity of immigrants could provide a further incentive for workers to organize and exclude the newcomers from the labor market. I find evidence consistent with this hypothesis. I show that the increased unionization was more prominent following an inflow of immigrants from Southern and Eastern Europe, whom part of the labor movement considered "slavish, ignorant and unassimilable," and therefore, a threat to American society (Collomp, 1988; Mink, 2019). Further, I show that unionization grew more in places harboring less favorable attitudes towards immigration. In the absence of a direct measure, I use two proxies that likely reflect a county's higher hostility towards immigrants. The first is the historical vote share for the Know Nothing Party, a nativist political party that, in the mid-1850s, ran on an anti-Catholic and anti-Irish platform (Alsan et al., 2020). The second is a measure of residential segregation between U.S.-born individuals and European immigrants. Since residential segregation usually arises either from collective action to exclude minorities or from individuals from the majority group moving away from ethnically mixed neighborhoods (Boustan, 2013), this characteristic likely reflects higher levels of discrimination against immigrants. Using either of these proxies, I find that immigration strengthened organized labor more prominently in counties with higher resentment towards immigrants. These results suggest that non-economic considerations also contributed to the expansion of labor unions.

Next, I rule out several alternative channels that could drive the results. First, I show that the findings are unlikely to be explained by immigrants disproportionately participating in unions. Given that information on the country of origin of individual union members does not exist, I provide suggestive evidence against this alternative explanation by examining the relationship between immigration and the origin and ancestry of local union leaders, inferred from their last names. I document that the share of union leaders with last names that were prevalent among U.S.-born increased overall during this period, and that immigration did not cause an increase in the proportion of immigrants' last names among the local leaders of unions. Moreover, I exploit variation in the strength of labor unions across Europe at the beginning of the 20$^{\text{th}}$ century and document that the inflow of workers from countries with an active labor movement was not responsible for the increased unionization. Second, I show that counties that received more immigration did not experience different economic growth during this time, and therefore, this is unlikely to explain differential trends in unionization. Finally, I find that immigration increased the number of workers in occupations represented by unions. Therefore, the positive effects on union density cannot be explained mechanically by a decrease in the denominator of this measure.

In the last part of the paper, I explore the economic implications of this immigration-induced unionization. Although these findings should not be interpreted as causal, they still provide key insights into short- and medium-run trends associated with a higher presence of organized labor. First, I investigate whether incumbent workers turned to occupations that had union representation in their county, to protect themselves against the economic challenges brought by immigration. I find that immigration increased the share of U.S.-born workers in unionized skilled trades, and, at the same time, reduced their concentration in skilled occupations without local union representation. This finding suggests that U.S.-born workers may have turned to occupations where organized labor could shield them from the potential adverse consequences of immigration. Second, I explore a central economic question related to labor unions: their role in reducing inequality (Card, 2001a; DiNardo et al., 1996). I construct three measures of wage inequality using U.S. Census data from 1940, the first year data on wages were collected. I then investigate their cross-sectional correlation with unionization in 1920, controlling for state fixed effects and the controls in the baseline specification. The results indicate that higher membership in labor unions is associated with lower wage inequality. Third, I examine whether the local patterns of unionization that emerged in the early $20^{\text{th}}$ century, documented for the first time in this paper, persisted until today. I aggregate the data at the metropolitan-area level, to make them consistent with the current measures of unionization from Macpherson and Hirsch (2023), and explore their cross-sectional correlation with the average levels of union density over the first two decades of the $21^{\text{st}}$ century, exactly a century after the period of analysis. Notably, past and present unionization are positively correlated. This suggests that the conditions that favored the initial development of labor unions in the early 1900s may have provided the labor movement with a head start that perdures throughout decades.

In summary, the empirical findings of this paper show that immigration substantially contributed to the emergence and expansion of organized labor in the early $20^{\text{th}}$-century United States. Moreover, the results indicate that existing workers formed and joined labor unions due to both economic and social motivations. The last section of the paper discusses the implications of these results for policy in the contemporary context, as well as related avenues for future research.

**Related literature.** The findings of this paper contribute to several broad literatures. First, they speak to the studies on organized labor, and labor unions more specifically. While a rapidly growing recent empirical literature has studied labor unions and analyzed their impact on a wide range of economic and political outcomes, both in historical and contemporary settings (Ahlquist, 2017; Ash et al., 2019; Barth et al., 2020; Biasi and Sarsons, 2022; Bittarello, 2018; Card, 2001a; Collins and Niemesh, 2019; DiNardo and Lee, 2004; Farber et al., 2021; Feigenbaum et al., 2018; Naidu, 2022; Naidu and Reich, 2018; Rosenfeld and Kleykamp, 2012; Rosenfeld, 2019; Sojourner et al., 2015; Schmick, 2018; Wang and Young, 2022), this study is the first to empirically study the determinants of their early development. The results identify immigration as a key factor that led to the emergence and growth of modern unions during a highly formative period for the American labor movement.

This paper also relates to studies that explore the historical drivers of unionization (Alesina and Glaeser, 2004; Archer, 2010; Asher, 1982; Bernstein, 1954; Briggs, 2001; Burgoon et al., 2010; Brody, 1993; Collomp, 1988; Foner, 1947; Freeman and Medoff, 1984; Griffin et al., 1986; Hannan and Freeman, 1987; Haydu, 1988; Karadja and Prawitz, 2019; Lipset and Marks, 2000; Montgomery, 1979; Moody, 2019; Naidu and Yuchtman, 2016; Olson, 1965; Sezer, 2023; Sombart, 1976; Taft, 1964; Willoughby, 1905; Webb and Webb, 1894; Wolman, 1924), and those that analyze the causes of its decline in recent decades (Acemoglu et al., 2001; Ahlquist and Downey, 2020; Clawson and Clawson, 1999; Farber and Western, 2001; Hirsch, 2008; Scruggs and Lange, 2002; Slaughter, 2007; Southworth and Stepan-Norris, 2009; Wallerstein and Western, 2000). This study advances this literature by identifying an unexplored driver of unionization and shedding light on the channels through which it operates.

The data collection effort of this paper also delivers the first comprehensive county-level dataset on historical union presence and membership in the United States, covering almost the entire country. Although a few existing papers have collected historical information on labor unions, those data are either on extinct organizations whose relevance was limited to the 1880s (Garlock, 2009), only cover a limited set of unions and do not contain information on membership (Schmick, 2018), are not disaggregated below the state level (Farber et al., 2021), or measure unionization only in a handful of states (Downes, 2023). The data introduced in this paper, aggregated at the county level for the analysis, but collected at the city or town level, make a significant advancement in studying geographic patterns of early unionization, and open avenues for future research on the medium- and long-term consequences of organized labor in the United States.

This paper also speaks to the vast literature on immigration. The results are related to the strand of this literature that examines its effects on labor market outcomes (see Abramitzky and Boustan (2017) and Peri (2016) for a review). This paper is the first to document that historical immigration positively affected the emergence and development of one of the most relevant labor market institutions, with heterogeneity in unions' presence and strength that persists until today.

Further, this study relates to the vast literature about the consequences of immigration on domestic workers' employment and wages, which has not reached an agreement on whether immigration has a positive, negative, or null effect (Dustmann et al., 2016). In particular, the findings of this paper are in line with Abramitzky et al. (2023), Card (2001b, 2005, 2009a), Foged and Peri (2016a), Ottaviano and Peri (2012b), and Tabellini (2020), who find negligible or positive impacts on domestic workers. The results of this study suggest that labor unions may play a role in mediating the possible adverse effects of immigration on domestic workers' wages and employment.

Finally, this paper is closely related to the recent political economy studies showing that higher levels of immigration increased the vote share for conservative politicians and support for anti-immigration legislation, both historically (Alsan et al., 2020; Goldin, 1994; Tabellini, 2020) and recently (Barone et al., 2016; Dustmann et al., 2019; Edo et al., 2019; Halla et al., 2017; Mayda et al., 2022; Mendez and Cutillas, 2014; Otto and Steinhardt, 2014). The results of this study identify a novel and unexamined consequence of immigration on the development of institutions that have had – and still have today – vast political influence. Although anecdotal and historical evidence has acknowledged the instrumental role that organized labor played in the introduction of immigration restrictions in the 1920s (Goldin, 1994; Mink, 2019), this paper is the first to empirically estimate a causal and positive effect of immigration on unionization, and document that this was due to both economic and social motivations. Moreover, this paper is related to the work by Alesina and Glaeser (2004), which links the weak labor and socialist movements of the United States to its ethnic diversity. The results of this study shed further light on this phenomenon, showing that reactions to immigration can foster unionization, partly offsetting other opposing forces that may slow down its growth.

**Outline.** The remainder of the paper is organized as follows. Section 1.2 describes the historical background. Section 1.3 presents the data. Section 1.4 introduces the empirical strategy and the instrument for immigration. Section 3.5 presents the main results and a summary of the robustness checks. Section 1.6 sheds light on the mechanisms that are driving the effect. Section 1.7 discusses the economic implications of the findings and long-term trends in unionization. Section 1.8 concludes.

## 1.2 Historical Background

### 1.2.1 Labor Unions at the Turn of the 20$^{\text{th}}$ Century

A new phase for the American labor movement started around the end of the 1880s, as the American Federation of Labor (AFL) became the largest and most influential group of labor unions.[5] By 1890, the main labor organizations that had gained importance during the second half of the 19$^{\text{th}}$ century – the Knights of Labor and the independent railroad workers' movements – had practically disappeared,[6] leaving the field open to new trade

[5]The American Federation of Labor was founded in Columbus, Ohio, on December 8, 1886, and rapidly became the main federation of unions in the country (Foner, 1947).

[6]Scholars have attributed the abrupt decline of these labor unions to a variety of factors, including their lack of a stable and permanent organizational structure, and their overly ambitious political agenda (Taft, 1964; Wolman, 1924).

unions (Wolman, 1924). These years saw the creation of many new organizations, which later became some of the largest national trade unions still active today.[7] Between 1880 and 1920, the total number of union members went from 149,000 to over 4.5 million (Figure 1.1).

The AFL was created as a federation of national unions and organized on the model of craft unionism. This meant that workers were organized based on their particular occupation (or craft).[8] It adopted the policy of *one craft–one union*, according to which each occupation should have only one union representing it. During this period, the unions in the building construction industry became the most stable and largest organizations.[9] This industry was dominated by skilled craftsmen, and characterized by small employing units (Taft, 1964). Only a few unions organized unskilled laborers in industrial settings. The United Mine Workers of America (UMWA) was the largest of these, along with unions representing dock-workers and workers in the meat-packing and textile industries. These sectors, mining in particular, were dominated by large employers who owned and operated several plants or mining sites (Beik, 1996) and strongly opposed unionization efforts (Northrup, 1943).

The AFL-affiliated national unions were organized into branches, called *locals*. The branches were responsible for bargaining agreements directly with individual employers (based on guidelines decided by the national union) to regulate wages, work hours, and conditions of employment. Unions also maintained funds to pay workers' benefits (in the event of strikes, injury, disability, or death), and regulated the terms of apprenticeship within the craft (Stewart, 1926). In most cases, the collective agreements specified that only union members could be employed (*closed-shop* clause). Both mandatory membership and appren-

[7] The International Broterhood of Teamsters, the International Brotherhood of Electrical Workers, the International Association of Machinists, and the United Brotherhood of Carpenters – even now among the 10 largest private sector unions – were established between 1881 and 1903. Moreover, the AFL (now merged with the more recently created CIO) is still the largest federation of labor unions, representing more than 12 million workers (U.S. Department of Labor, 2022).

[8] The main alternative model is *industrial* unionism, in which all workers in the same industry are organized by the same union, regardless of their skill level.

[9] The bricklayers and the carpenters' unions were the dominant organizations among building trades.

ticeships gave unions effective control over which workers could enter the skilled occupations they organized.

Until the mid-1930s, there was no federal law requiring employers to recognize unions or punishing their retaliatory behavior against union members. This situation promoted an environment where company owners, with the support of the courts, made use of strike-breakers, lockouts, retaliatory firing, and other strategies to oppose unions and prevent their organization (Foner, 1947; Taft, 1964).[10]

### 1.2.2 The Age of Mass Migration

Between 1850 and 1920, around 30 million Europeans moved to the United States (Hatton and Williamson, 1998), raising the share of the foreign born population to over 14% (Figure 1.2 and Figure 1.3). The mix of origin countries changed substantially over time. Until 1890, most immigrants were from the United Kingdom, Ireland, Germany, and Scandinavia. Thereafter, as transportation costs decreased (Keeling, 1999), the bulk of immigrants came from the rest of Europe. In 1850, immigrants from Northern or Western Europe were 92% of the foreign-born population, while less than 1% had arrived from Southern, Central, or Eastern Europe. By 1920, these shares were 40% and 43%, respectively (Figure 1.4). Europeans from the new origin regions were different from those who had arrived in the previous decades: they were significantly less skilled, spoke unfamiliar languages, and were not Protestant (Hatton and Williamson, 1998, 2006).

The waves of mass immigration increased enormously the supply of labor, which had already been expanded by the shift of population from rural areas to cities in the 1880s. Often the newly arrived immigrants, eager to earn a livelihood in a new country, made their

[10]Federal legislation of 1898 (Erdman Act) guaranteed the right to unionize only to railroad workers. Several states passed laws in the 1890s prohibiting employers from discharging employees for belonging to a union. However, whenever the labor movement succeeded in obtaining legislation in its favor, courts weakened or entirely wiped out such statutes by declaring the laws unconstitutional (Foner, 1947; Taft, 1964).

first appearance into the American workforce as strikebreakers, hired by business owners in order to undermine the incumbent workers' bargaining power and unionization efforts (Foner, 1947). Over the years, the political climate grew hostile towards European immigrants, based on concerns about labor market competition and xenophobia toward new arrivals (Goldin, 1994). In response, starting in the late 1890s, members of Congress proposed legislation to limit immigration, and in 1917, Congress eventually introduced a literacy requirement for all immigrants.[11] Though immigration temporarily slowed down during World War I, after the end of the war it immediately rose again, resurrecting earlier anti-immigration fears. Consequently, in 1921 Congress passed the Emergency Quota Act and introduced a temporary limit to immigration. In 1924, the National Origins Act made this restriction permanent and more stringent (Abramitzky and Boustan, 2017). The immigration quotas remained in effect for the next 40 years, until they were eliminated in 1965 by the Immigration and Nationality Act.

### 1.2.3 The Labor Movement and Immigration

Organized labor has always been concerned about the potential negative consequences of labor supply expansions, particularly those caused by immigration (Taft, 1964). This is the main reason why it favored immigration restrictions since its inception. In 1881, in the founding meeting of its precursor organization, the AFL adopted a resolution against Chinese laborers and lobbied Congress to ban Chinese immigration through the Chinese Exclusion Act of 1882 (Foner, 1947). In 1885, the labor movement succeeded again when the Alien Contract Labor Law (also known as the Foran Act), which banned the importation of foreigners to perform labor in the United States, was approved.[12] In 1896, in response

[11] One of the first attempts to limit immigration was the legislation introduced by Henry Cabot Lodge, the Republican senator from Massachusetts, which required a literacy test for all potential immigrants. President Cleveland then then vetoed the bill.

[12] Representative Foran, the sponsor of the bill, decried the "large numbers of degraded, ignorant, brutal Italians and Hungarian laborers" for imperiling the racial heights of the republic: "They know nothing of

to the shift of immigration to ethnic and national groups whose schooling levels, skills, and standards of living were substantially below those of previous groups, the AFL endorsed further restrictive measures. It was widely held that Southern and Eastern Europeans lowered wages, dragged down working conditions, were not responsive to the discipline of labor unions, and therefore constituted a threat to the American working man (Mink, 2019; Taft, 1964). The federation vigorously supported further restrictive measures until it obtained the introduction of the 1921 and 1924 nationality quotas (Goldin, 1994).

Throughout this period, the labor movement used increasingly popular racial and eugenics-based arguments to discuss threats to employment and gain momentum in calling for an outright ban on European immigration.[13] Nativism was triggered by the increased presence of foreign laborers, which inundated labor markets, and was intensified by the mounting pressure of mechanization (Mink, 2019; Yellowitz, 1981). These events added credibility to the fears that machines and the new unskilled workers could substitute skilled unionized labor (Olzak, 1989), and led unions to concentrate on securing job control for skilled workers by organizing the workplace and the work process (Mink, 2019). At the same time, the immigration-induced expansion of the labor supply was deemed responsible for weakening unions' bargaining power, by creating a reservoir of potential strikebreakers and freeing employers from the constraints of a tight, unionizing labor market (Montgomery, 1979).

our institutions, our customs, or of the habits and characteristics of our people. [...] They are brought here precisely in the same manner that the Chinese were brought here [...] Being low in the scale of intelligence, they are [...] willing slaves. [...] The fact that American workingmen are vastly superior to these aliens in intelligence, skill, moral and social culture will no doubt be admitted" (Mink, 2019).

[13]Statements made by union men expressing hatred for new immigrants abound. In 1884, a labor leader described Hungarian laborers as a menace because "they work for little or nothing, live on a fare which a Chinaman would not touch, and will submit to any and every indignity which may be imposed on them." Railroad workers in Kankakee, Illinois, objected to: "Italians [...] unloaded in cities from cattle cars; they sleep in huts; they eat stale bread [...] the worst kind of meat and a small amount of rice. [...]. Send them away or we will kill them as one kills mad dogs." American laborers complained that most immigrants were "only scavengers to our country" and that men who could not speak "our language" often beat out natives for jobs." (Asher, 1982).

## 1.3 Data

My study relies on a novel micro-database that combines labor unions' records with labor market and economic outcomes, between 1900 and 1920.

In this section, I describe the data collection effort, the main sources of the data, and present summary statistics and descriptive facts on early unionization in the United States.

### 1.3.1 Dataset on Union Presence and Membership

I assemble the first panel dataset on unionization for the period 1900–1920. This also constitutes the first comprehensive dataset on historical union density measured at the county level in the United States. Most existing studies on modern labor unions in a historical period rely on aggregate national estimates, since microdata on union status were first collected by the Current Population Survey (CPS) only in 1973. There are a few notable exceptions. Schmick (2018) collects data on the presence of local branches of some national unions in the years 1882, 1892, and 1902. However, the dataset contains no information on membership and covers a different set of unions in different years in a time period that precedes the first significant expansion of the labor movement and the largest waves of immigration. Farber et al. (2021) combine survey data, primarily from Gallup, to compute historical levels of union membership for most of the 20$^{\text{th}}$ century. However, their data are not disaggregated below the state level, and only start in 1937, after immigration restrictions had been in place for over a decade and the first national expansion of the American labor movement had occurred. Similarly, Downes (2023) constructs county-level union membership estimates for selected years in the mid-20$^{\text{th}}$ century. However, his data start in 1920 and are limited to five states, hence also unsuitable to study the questions of this paper.

The dataset I assemble to conduct the empirical analysis combines newly digitized historical records on labor unions from several sources.

**Convention proceedings of the state federations of labor.** The main sources of the dataset on unionization are convention proceedings of the state federations of labor, which were state-level subordinate bodies of the AFL. Their functions were mainly legislative and propagandist, and they were composed of representatives from all the local branches of the AFL-affiliated national unions within the state (Stewart, 1926). Local branches (also called local unions, or locals) were a lower level of organization of national unions, and represented workers in either a single employment unit or from several work sites. By 1920, members of AFL unions constituted more than 80% of the total private-sector union membership (Wolman, 1924). Each state federation of labor met annually in conventions to enact legislation and elect general officers. All affiliated local unions were entitled to representation.[14]

I digitize the proceedings of these conventions every 10 years between 1900 and 1920.[15] From these documents, I extract the lists of locals represented at the conventions, along with the union name and branch number, their location, the number of delegates representing them, and the names of such delegates (Figure A.1). Each federation had specific rules to define the number of delegates that could represent a local branch, which often varied over time. Importantly, they established that locals should be represented proportionally to their membership (Figure A.2).[16] I therefore combine the information on the delegates from the convention proceedings with the details on the representation rules contained in

[14]The only exceptions were recently established branches, those that had payments in arrears in the months before the convention (usually three months before), and branches expelled or suspended by their national organization.

[15]In case the proceedings for any of these years were not available, or did not contain the information needed (e.g., location of the union branches), I digitized the analogous document for the convention that took place either the following year or two years later. If those documents were also not suitable or unavailable, I digitized the one of the convention from the previous year or two years before.

[16]The following state federations of labor never adopted a proportional representation in the period 1900–1920: Kansas, Kentucky, Louisiana, Maryland, North Dakota, New Mexico, and Tennessee. For this reason, these states do not enter the sample.

the constitutions of each state federation of labor. Using this information, I construct an estimate of union membership for each local branch. Since the representation rules were often expressed in terms of ranges (e.g., one delegate every 100 members), I use the mid-points of these intervals as the estimates of membership. For example, if the constitution states that a branch is represented by one delegate every 100 members, its membership is estimated to be 50 if one delegate is present at the convention, 150 if two delegates are recorded, and so on. The results are unchanged if membership is estimated using the lower or the upper bound of the intervals instead.

I geocode the location of all the union branches based on their town, village, or city, and retrieve their coordinates. I use the names of each branch's national union to establish which occupations and industries they operated in.[17] Finally, I aggregate the membership of the union branches at the county level to obtain a measure of union membership, both total and by occupation.

**Proceedings of the national conventions of AFL unions.** I complement the data from the state federations of labor with analogous information collected directly by the AFL-affiliated national unions. Similar to the state federations, the AFL-affiliated unions met in national conventions to legislate, elect officers, and set guidelines for the local branches to follow in their bargaining agreements. I digitize the proceedings of these conventions for six of the largest AFL-affiliated unions of this period, every 10 years between 1900 and 1920.[18] The members of these six unions accounted for approximately 40% of the over 100 AFL-affiliated unions' total membership between 1900 and 1920 (Wolman, 1924).[19] I follow

[17]As described in Section 1.2, each AFL national union organized workers in a specific occupation. Their names always indicated the occupations or industries they represented (e.g., United Brotherhood of Carpenters and Joiners, Brotherhood of Painters and Decorators, International Association of Machinists, United Mine Workers of America, etc.).

[18]As above, if suitable documents are not available for 1900, 1910, or 1920, I digitized the analogous documents for the convention that took place in one of these alternative years (in order of preference): one year later, two years later, one year before, or two years before.

[19]These unions are: the Bricklayers, Masons, and Plasterers International Union of America (BMPIU); the International Association of Machinists (IAM); the International Brotherhood of Teamsters (IBT); the

a procedure analogous to the one described for the proceedings of the state federations of labor, and collect data on the lists of local branches, their location, and the names and number of delegates representing them. Next, I construct an estimate of the membership of each of these locals, following the representation rule listed in the convention proceedings or in the constitution of each of these organizations. Finally, I aggregate the data at the county level.

These data sources complement the records from state federations in three main ways. First, they validate the estimates constructed using the main data source. In particular, for the six unions that I observe across both sources, I am able to compare the estimates of union membership and the number of branches. In all cases, the measures display a highly positive correlation (Figure A.3). Nonetheless, some branches may appear in only one of the two types of convention documents I digitize. This may occur because one branch was established too recently before a convention and did not yet qualify to send delegates according to organization-specific rules. Similarly, it could have been formed between the state federation and the national union convention; hence, it could only be observed in one of the two documents. Another possibility is that some delegates may have been erroneously omitted from the roll calls of the meetings.[20] Any of these occurrences would lead to an underestimate of the number of members and/or the number of branches in a given county if only one of the sources was used. Unfortunately, there is no way of knowing with certainty if and how many locals fall into these circumstances, since this information is never systematically reported. However, by combining information from different sources (and collected by different entities), I am able to reduce these instances of mismeasurement. This constitutes the second

International Typographical Union (ITU); the United Brotherhood of Carpenters and Joiners (UBC); and the United Mine Workers of America (UMWA). These are all the unions, among the 10 largest, that systematically and consistently reported information in their convention proceedings about delegates and the local branches they represented, and whose proceedings are available either in physical or digital copy.

[20]Additionally, some locals may have had payments in arrears to either the state federation or the national organization, and therefore did not qualify to send delegates to one of the two conventions.

main contribution of this data source. Third, these additional archival records allow me to expand the time and geographical coverage of the dataset, because some state federations of labor were constituted (and hence convened for the first time) only after 1900.[21] Relying only on the first data source would lead to measure no presence of union branches and zero union membership for counties in states and decades before the first federation of labor's convention. Although the lack of an AFL state subordinate body intrinsically suggests a limited presence of organized labor, it is still possible that some national unions may have already been present in at least some counties of these states. The additional information on the branches (and its delegates) of these six large unions operating throughout the whole U.S. territory in 1900, 1910, and 1920, allows to more accurately measure unionization at the early stages of a state's labor movement.

**Combined data sources.** To construct the final measures of unionization, I combine the information collected from the two sources described above. I first reduce the number of missing observations and misreportings from each of these sources by linearly interpolating the number of union branches and members for counties that are not reported in the convention proceedings of a certain year, but that have representation both in the previous and following decade.[22] Next, for the six unions observed across both types of documents, I com-

[21] The following state federations of labor first convened after 1900, the first year of the empirical analysis: Alabama (1901), Arkansas (1905), Arizona (1912), California (1906), Delaware (1923), Florida (1901), Idaho (1916), Kansas (1907), Louisiana (1913), Maryland (1905), Mississippi (1918), North Carolina (1907), North Dakota (1912), Nebraska (1909), New Hampshire (1902), North Carolina (1907), North Dakota (1912), Nebraska (1909), New Hampshire (1902), New Mexico (1914), Nevada (1921), Oklahoma (1904), Oregon (1902), Rhode Island (1901), South Carolina (1915), South Dakota (1920), Utah (1904), Vermont (1902), Washington (1902), West Virginia (1903), and Wyoming (1909). Consistently with the rest of the data collection, the proceedings of federations constituted in 1901 or 1902 are attributed to the Census year 1900.

[22] Counties may wrongly appear to have no union branches or members in a certain year due to one of the following reasons: error in assigning a locality to the correct county because of homonymous locations, a partial or incorrect reporting of the delegates present at the convention, or county-specific reasons for why no delegate was actually not sent to one of the two conventions. The underlying assumption for this exercise is that a county with union branches and members in, say, 1900 and 1920, will not realistically have zero branches and membership in 1910. I also collect available data for the state federation conventions that took place in 1930 in order to linearly interpolate the data from the first source for the year 1920. Importantly, the results are qualitatively unchanged if this step is not conducted.

pute the number of members and branches in each county and year by averaging the ones from each source. When only one data source reports a positive membership or number of local branches, I use that value in the analysis. Finally, I sum the total number of branches and members across all unions at the county-decade level, and obtain the total number of these quantities in each county over time.

In order to construct measures of union density, I divide the number of union members by the size of the labor force, by occupation and total. For example, the measure of union density for carpenters will be the number of members of the carpenters' union divided by the labor force in carpentry occupations. When computing the overall union density for the county, the total number of union members is divided by the total labor force in occupations within the jurisdiction of the AFL-affiliated unions in existence in the period 1900–1920.[23] Additionally, I construct an indicator for whether a county has any union, the number of union branches within its territory, and their average size, defined as the number of members divided by the number of branches. As a final validation exercise, I compare these measures of union density to those contained in another existing historical dataset. While only aggregated national estimates of union membership exist for the period studied, I ensure that the measures of union density are positively correlated with those calculated at the state level by Farber et al. (2021) using Gallup surveys, starting in 1937 (Figure A.9).

The final dataset contains information on the location and membership of local union branches in over 2,400 counties between 1900 and 1920 (Figure 1.5).[24] Throughout the

[23]As in the rest of the paper, labor force variables are computed restricting the sample to men ages 16–64. The jurisdiction of each union is taken from Stewart (1926) and the complete list of occupations, with corresponding codes and description, is reported in Table A.2. In case the total number of estimated union members exceeds the total labor force, union density is coded to be 1. This is a rare event, which occurs for the main measure of union density only in 12 out of the 5,025 county-year observations of the main sample. In Section 1.5.2, I show that the results are unchanged when using alternative definitions of union density, such as dividing union members by the total labor force in occupations organized both by AFL and non-AFL unions, or by the total non-agricultural labor force, and when excluding outliers.

[24]The counties not part of the sample are those in states whose federations of labor did not have a representation rule for branches proportional to their membership (as previously described), whose convention proceeding are not available, or reported only incomplete records (e.g., no information on the location of

empirical analysis, I restrict the sample to a balanced panel of 1,675 counties, which represent approximately 65% of the total U.S. labor force during this time period.

### 1.3.2 Other Data Sources

**Immigration and population.** The data on county population and on the number of immigrants, by country of origin at the county and national levels, are taken from the decennial U.S. Census of Population. For 1900, 1910, and 1920, I use the full-count Census datasets, made available by IPUMS (Ruggles et al., 2021). For 1890, I use Census datasets aggregated at the county level, made available by the Inter-University Consortium for Political and Social Research (ICPSR) (Haines, 2010b).[25]

**Labor market outcomes.** I compile data on labor force, occupation, and yearly income from the U.S. Census of Population.[26]

**Economic activity.** The county-level data on manufacturing output and establishments (from the Survey of Manufactures) and on the agricultural sector (from the Census of Agriculture) come from Haines (2010b).

**Presidential elections vote shares.** The data for the county-level vote shares in presidential elections are from Inter-University Consortium for Political and Social Research (ICPSR) (1999).

**Railroad network.** Data on the expansion of the railroad network rely on the database compiled by Atack (2016), based on traced lines from historical map images. The database contains the exact placement of railroad lines over time, between 1826 and 1911.

the branch, or no list of delegates altogether). In Section 1.5.2, I show that the results are unchanged when extending the analysis to the whole unbalanced sample of counties.

[25] Since most of the 1890 completed Census forms were lost in a fire, full-count data are unavailable for this Census year.

[26] Due to the unavailability of the labor force participation status in the 1900 full-count Census dataset (Ruggles et al., 2021), I proxy for this variable in that year with an indicator for holding any gainful occupation.

### 1.3.3 Summary Statistics

Figure 1.6 plots union density (the share of the unionized workforce) separately for 1900, 1910, and 1920. Unionization in 1900 was predominantly concentrated in the Northeast and Midwest. By 1920, unions had also spread to many other regions, including the West and selected areas of the South. Across the country, unionization was more prevalent in urban areas, which also received larger immigration flows during this period. Overall, the maps display substantial variation in union density across counties – both within and across states – and over time.

In Table 1.1, I present summary statistics for the main variables on unionization (Panel A), demographic characteristics (Panel B), and labor force (Panel C). The average union membership was slightly short of 255 and the average share of unionized workers 4%. On average, there were almost two union branches per county, with a membership of 30 people per branch. Overall, over a quarter of the observations have positive union membership.

The average number of recent European immigrants – those who entered the United States in the previous decade – as a fraction of the population was 2%. However, this masks substantial heterogeneity across counties, as indicated by the size of the standard deviation.

The average labor force participation was 91%; U.S.-born workers represented 87% of the labor force, while European foreign-born individuals 11%.

## 1.4 Empirical Strategy

### 1.4.1 Baseline Estimating Equation

To study the effects of immigration on unionization, I focus on the three Census years between 1900 and 1920, and I estimate

$$y_{ct} = \beta Imm_{ct} + \theta_c + \tau_t + X_{ct} + u_{ct} \tag{1.1}$$

where $y_{ct}$ is the outcome for county $c$ in Census year $t$, and $Imm_{ct}$ is the number of immigrants as a fraction of the county population. $\theta_c$ and $\tau_t$ are county and year fixed effects, implying that $\beta$ is estimated from changes in the fraction of immigrant labor force within the same county over time. $X_{ct}$ are county-level control variables, which are likely correlated with both the pre-1900 settlement of immigrants and the evolution of unionization over time, measured at baseline and interacted with year fixed effects.[27] Throughout the analysis, standard errors are clustered at the county level, and all variables are harmonized to reflect 1930 county boundaries (Hornbeck, 2010).[28]

In the baseline specification, $Imm_{ct}$ refers to the stock of working-age male European immigrants who entered the United States during the previous decade, as a share of the total working-age male population. Focusing on this definition allows me to more confidently interpret the findings as the consequences of an inflow of new (immigrant) workers into the labor market. All the labor force variables are similarly computed on the sample of men

[27] Whenever available, these variables are measured in 1890. If the 1890 county aggregates of the U.S. Census do not include this information, the variables are taken from the 1880 full-count Census.

[28] Since county boundaries change over time, I maintain consistent geographic units by holding county boundaries constant throughout the sample period. I follow the procedure in Hornbeck (2010) and harmonize all the variables used in the analysis to reflect 1930 county boundaries. This procedure uses area-based weights to harmonize county boundaries across years. Alternative border harmonization procedures that use population-based weights, such as the one in Ferrara et al. (2022), yield almost identical results.

ages 16–64.[29]

### 1.4.2 Instrument for Immigration

Given the hostility of the labor movement towards immigration described in Section 1.2, we may expect immigrants to settle in counties with less unionization, where the chances of being excluded from certain occupations would be lower. This would cause the ordinary least squares (OLS) estimates of equation (1.1) to be biased downwards. By contrast, immigrants may prefer counties with a growing labor movement, to the extent that those labor markets might also present more or better job opportunities. This would bias the OLS estimates upwards. In addition, classical measurement error in the immigration data would attenuate the estimates towards zero.

**Baseline instrument.** To deal with these endogeneity concerns, I construct a shift-share instrument (Card, 2001b). This approach combines two sources of variation. The first is the *share* of European immigrants from country $j$ living in county $c$ as of 1890 (relative to all immigrants from country $j$ in the United States), which I denote as $\alpha^j_{c,1890}$. The second is the change, or *shift*, in the number of European immigrants from country $j$ entering the United States in a given decade, net of those that eventually settled in county $c$, denoted by $O^j_{-ct}$.[30] Formally, the predicted number of immigrants received by county $c$ between Census year $t-10$ and $t$ is given by:

$$\tilde{Z}_{ct} = \sum_j \alpha^j_{c,1890} O^j_{-ct} \tag{1.2}$$

This number is then scaled by county population measured in 1890, $P_{c,1890}$, as the contemporaneous county population would itself be an outcome of immigration.

[29] Over most of the period 1900–1920, union members were almost exclusively men (Wolman, 1924); female labor force participation was only 25% (92% for men). Results are similar when considering all immigrants, regardless of their sex, age, or arrival year.

[30] A similar "leave-out" strategy is also used in Tabellini (2020). See Table A.1 for the list of European origin countries and regions used to construct the shift-share instrument.

Underlying this identification strategy is the empirical regularity that migrants tend to settle where other migrants from their own country of origin had settled previously, a process known as *chain migration*. The pre-1890 migration of Europeans is reflected in the term $\alpha^j_{c,1890}$. I choose 1890 as the base year because it captures many of the important migration networks established in the first part of the Age of Mass Migration, but predates both the peak of immigration flows from Europe and the largest periods of union growth (Figure 1.1 and Figure 1.2).[31] Crucially, 1890 also predates the large compositional shift in immigration that occurred at the turn of the 20$^{\text{th}}$ century (Figure 1.4). As previous work has argued (Abramitzky et al., 2023; Tabellini, 2020), this period is particularly suited to the use of shift-share instruments, not only because of the changes in the quantity of immigration over time, but also variation in the immigrants' country of origins in each decade. Different from Tabellini (2020), who employs an analogous identification strategy to predict immigration between 1910 and 1930, this shift-share instrument exploits the additional variation in the composition of immigration that took place between 1890 and 1900.

Table 1.2 reports first stage coefficients and shows that actual and predicted immigration are highly correlated.

**Identification assumption.** The key identifying assumption behind the instrument described in equation (1.2) is that, conditional on controls, the unobserved factors that affect unionization outcomes must not be jointly correlated with the 1890 composition of Europeans' enclaves across U.S. counties and immigration patterns from European countries after 1890.[32] Previous work has argued that nation-wide shocks that occurred during the period 1900–1920, and which are exogenous to county-specific characteristics, make this setting particularly suited to the use of shift-share instruments (Abramitzky et al., 2023; Tabellini, 2020). In particular, the trend-break in immigration created by WWI lowers the concern

[31] In fact, approximately 70% of the organizations affiliated with the AFL, and in existence before 1920, were founded in 1890 or after (Stewart, 1926).

[32] For theoretical foundations, see Borusyak et al. (2022) and Goldsmith-Pinkham et al. (2020).

that the shift-share instrument may be correlated with shocks jointly affecting local conditions in U.S. counties and immigration patterns from European countries. Moreover, the WWI shock reduces worries about the design being invalidated by the serial correlation in migration flows from the same country to the same U.S. destination (Jaeger et al., 2018).

**Instrument validity.** Nevertheless, although the immigrant networks captured by $\alpha^j_{c,1890}$ predate the time period of the analysis, they may be endogenous with respect to the trajectory of the outcomes of interest. I deal with this concern in several ways. First, I augment the main specification by including interactions between year dummies and county characteristics measured at baseline that might have attracted more immigrants (from each origin country) before 1890, and may have had a time-varying effect on unionization across counties. In the preferred specification, such controls include: (i) the share of the urban population living in county $c$ at baseline, and (ii) the baseline labor force participation rate, defined as the number of individuals in the labor force divided by the total working-age population.[33] The former accounts for the fact that both immigration and labor unions were a predominantly urban phenomenon in this period (Abramitzky and Boustan, 2017; Taft, 1964), and therefore early urbanization levels may have been correlated with both the initial settlement of immigrants and the subsequent evolution of organized labor. Similarly, tighter labor markets likely attracted more immigration early on and affected the growth of labor unions in the beginning of the 20$^{\text{th}}$ century.[34] Second, I directly control for the size of the 1890 European immigrant population, interacted with year dummies. This implies that the effects of immigration are identified exploiting variation only in the ethnic composition of immigrant enclaves across counties, holding constant the size of their foreign-born populations. Since

[33] Consistently with the rest of the paper, both variables are defined restricting to the sample of men ages 16–64.

[34] In Appendix A.1, I show that the estimates are robust to the inclusion of several other baseline county characteristics with a potential effect on both the 1890 levels of the immigrant population and unionization in the subsequent decades (Table A.15), and also that the results do not depend on the inclusion of any of these controls (Table A.16).

the instrument predicts higher immigration to counties with a larger stock of immigrants at baseline, by doing this I also address the concern that a larger 1890 immigrant population may itself have an independent and time-varying effect on unionization. Third, I include interactions between year dummies and the baseline share of immigrants from each European country, $a^j_{c,1890}$, to assuage concerns that the 1890 settlements of specific European groups across U.S. counties might be correlated with both the long-run trends in unionization and the migration patterns of those specific immigrants groups, in each decade between 1890 and 1920.

**Alternative instrument.** In addition, I construct an alternative version of the instrument described in equation (1.2), where I replace the actual immigration flows from each country $j$ with those predicted exploiting variation in weather shocks across European countries over time. This allows me to identify causal effects from the exogenous variation in the shocks, while allowing the exposure shares to be endogenous (Borusyak et al., 2022). I then interact them with the baseline shares of European immigrants from each country $j$ to obtain the alternative instrument. Appendix A.1.1 describes its construction in more detail.

**Matching and shift-share instrument.** Finally, similarly to Bazzi et al. (2023), in Appendix A.1.2 I combine the shift-share instrument of equation (1.2) with a matching exercise. In particular, I select within-state county pairs with similar baseline presence of labor unions, as measured by the number of branches of the Knights of Labor in 1890 as a fraction of the county population.[35] Then, I re-estimate the 2SLS analysis also controlling for fixed effects for the 800+ county pairs interacted with year dummies.

I summarize all other robustness checks in Section 1.5.2, after presenting the main results.

[35] As described in Section 1.2, the Knights of Labor were a federation of unions that was particularly active in the 1880s, and declined after 1890, when the AFL became the dominant federation of unions. For this exercise, I use data from Garlock (2009) to measure union presence as of 1890, when the AFL was only recently established and did not yet have substantial national presence (Foner, 1947).

## 1.5 Main Results

### 1.5.1 The Effect of Immigration on Unionization

In Table 1.3, I investigate the effects of immigration on the formation and growth of labor unions by estimating equation (1.1). I examine four unionization outcomes: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the labor force (column 2);[36] the log number of union branches (column 3);[37] and, the average branch size, defined as the number of members divided by the number of branches (column 4).[38] All regressions include county and year fixed effects, and interactions between year dummies and the baseline urban population share and labor force participation rate (as discussed in Section 1.4.2). I present OLS estimates in Panel A. Although not always precisely estimated, all coefficients are positive. This suggests that counties that received more immigration were also more likely to display higher levels of unionization.

Panels B and C show the reduced form and the 2SLS estimates. The F-stat for weak instruments, reported at the bottom of the table, is always above the conventional levels, and indicates that the instrument is strong. In all cases, the point estimates are positive and statistically significant at either the 5% or the 1% level. The 2SLS estimates imply that a 4 percentage point (1 standard deviation) increase in the share of recent immigrants causes a 6.3 percentage point (24% relative to the mean) higher probability that the county has

[36]My preferred definition of union density is the number of union members divided by the total labor force in occupations covered by the AFL-affiliated national unions during this period. This measure has the main advantage of not being influenced by the relative importance of such occupations in the labor force. In Table A.22, I show that the results hold when using different definitions of the dependent variable.

[37]Since this variable may take value zero if no union branch is observed, throughout the paper I apply the transformation $log(1+x)$ instead of $log(x)$, where $x$ is the number of branches.

[38]To maintain the same sample throughout the table, and for consistency with the other outcomes, I define this variable as zero if the county has no union branch (and, therefore, also no union members). Results are qualitatively similar if restricting only to county-year observations with at least one union branch. See Section 1.6 for a discussion about the effects on the extensive and intensive margins of unionization.

any union (column 1); a higher share of unionized workers by over 1 percentage point, or 29% relative to the sample mean (column 2); 70% more union branches (column 3);[39] and 10 more members per branch, or 35% relative to the sample mean (column 3).

For areas that consistently received high volumes of immigration between 1900 and 1920 – such as large counties like New York (NY), or smaller ones such as Lake (IN) or Kenosha (WI) – immigration increased the fraction of union workers by 50–75% relative to the mean. A back-of-the-envelope calculation, done by comparing the actual level of union density measured in the data to the one predicted by the 2SLS estimates, reveals that in the absence of immigration, the average union density between 1900 and 1920 would have been 17% lower overall.

The difference between OLS and 2SLS estimates indicates that the former are biased downwards, and suggests that European immigrants selected areas where unionization was growing more slowly. This might have happened because, during this period, the vast majority of labor unions actively discriminated against immigrants, precluding them from joining their ranks and the occupations they represented (Asher, 1982). Consistent with the historical evidence, I show in Table A.4 that there is a negative and statistically significant relationship between all four measures of unionization and immigration flows. Another possibility, which may co-occur with the previous one, is that the instrument identifies a local average treatment effect (LATE) for counties that received more European immigrants because of country-of-origin networks, and not because of economic or political characteristics of the destination county. If such immigrants were more likely to generate an increase in unionization – either because of their own preferences, or because of the reactions they would cause among existing workers – this could explain why OLS coefficients are smaller than the 2SLS estimates.

[39] Given that the dependent variable of column 3 is in log, the magnitude of the coefficient can be calculated as follows: $\%\Delta y = 100 \cdot (e^{\beta} - 1)$.

### 1.5.2 Summary of Robustness Checks

I perform several exercises to verify the robustness of the findings. They are summarized visually in Figure 1.7 and Figure 1.8, with more details and formal estimates presented in Appendix A.1.

I show that the results are unchanged when using a version of the instrument that relies on weather shocks in each European country for the period 1890–1920 to predict the flows of European immigration (Table A.13).[40] This alternative identification strategy relies on the observation that the validity of shift-share instruments can be achieved from the exogeneity of the shocks (Borusyak et al., 2022).

Next, building on Bazzi et al. (2023), I combine the shift-share instrument with a matching strategy, which selects within-state county pairs with the closest number of labor unions in 1890 as a fraction of the county population (Table A.14).

Morevoer, I verify that the results are robust to the inclusion of several county characteristics that are likely correlated with the 1890 settlements of European immigrants and the subsequent development of labor unions, measured at baseline and interacted with year dummies (Table A.15). These include an indicator for whether the county was connected to the railroad network, the share of the immigrant (total and European) and Black population, the share of the labor force in the largest industries, the share of the labor force in occupations covered by AFL-affiliated national unions, the average occupational income score, the growth rate of the manufacturing output, the share of land used for farming, and the vote share for the Democratic Party in presidential elections.

Further, I show that the findings are unchanged when using alternative baseline specifications, such as not controlling for any baseline characteristics or including state by year fixed effects (Table A.16); dropping potential outliers (Table A.17); clustering standard errors at

[40] This alternative version of the instrument builds on previous work from Sequeira et al. (2020a) and Tabellini (2020).

the SEA level or using Conley (1999) standard errors to account for spatial correlation (Table A.18); estimating population-weighted regressions (Table A.19); extending the analysis to an unbalanced sample of counties (Table A.20); excluding the South from the estimation sample (Table A.21); and using alternative definitions of union density (Table A.22).

I also re-estimate the baseline specification of Table 1.3 while interacting – one at a time – the initial shares of each immigrant group in the county, i.e., $\alpha^j_{c,1890}$ in equation (1.2), with year dummies (Figure A.8). This exercise is aimed at reducing the concern that combinations of counties and of immigrants from specific European countries might be driving the results by absorbing most of the variation in the data (Goldsmith-Pinkham et al., 2020).[41]

Finally, I check for the absence of pre-trends by regressing the pre-period change in several unionization, population, and economic outcomes on the 1900 to 1920 change in immigration predicted by the instrument (Table A.23). The fact that all these coefficients are never statistically significant indicates that, before 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or other economic variables.

## 1.6 Mechanisms

The results shown so far indicate that counties that received larger inflows of European immigrants between 1890 and 1920 experienced a larger increase in unionization. In this section, I explore the mechanisms that are driving the positive effect of immigration on the emergence and growth of organized labor.

[41] This robustness check also deals with the potential concern that such shares may not be independent of cross-county pull factors related to the initial immigrants' country of origin.

### 1.6.1 Economic Motivations

**Reactions of existing workers.** As described in Section 1.2, unions have always been concerned with labor supply expansions, as they feared that new workers would lower wages, deteriorate working conditions, and induce job scarcity. These worries led existing workers to organize their workplace and limit the access to the labor market (Mink, 2019; Olzak, 1989).[42] Hence, the economic threats posed by immigration may have increased workers' incentives to unionize and keep immigrants out of the labor force. The positive effects presented in Table 1.3 are consistent with this hypothesis.

At the same time, immigration inundated urban labor markets with cheap laborers in search of employment. This, in turn, increased employers' bargaining power by lowering their cost to break strikes and replace workers willing to unionize (Asher, 1982; Mink, 2019; Olzak, 1989). This process was facilitated by the political and legal framework of this period, when workers' right to organize was not granted by law and courts often sided with employers in disputes over the dismissal of unionizing or striking employees (Foner, 1947; Taft, 1964).

As a result, a reaction of existing workers to immigration in the form of unionization can be expected if employees cannot be *immediately* replaced. For example, minimum levels of skill or human capital represent a barrier to entering some occupations and, therefore, can give incumbent workers an advantage in forming and sustaining a labor union. Similarly, employers who do not have a readily available pool of replacement workers will be less likely to successfully oppose unionization efforts and more willing to give in to their employees' demands to form a union.

I leverage differences in the skills required across occupations to test whether immigration had heterogeneous effects between skilled and unskilled workers. The estimates, reported in Table 1.4, indicate that immigration positively impacted all skilled unionization outcomes.

[42] Two of the methods most commonly used by unions to control the access to certain occupations were imposing union membership as a condition of employment and regulating the terms of apprenticeships.

Counties that received larger shares of recently arrived immigrants experienced an increase in the probability of having any union, the share of the unionized workforce, the number of union branches, and their size. Instead, immigration did not affect the expansion of unskilled unions. These results are consistent with the hypothesis that barriers to entering an occupation gave incumbent workers with an advantage in their quest to establish and maintain a labor union since employers could not immediately replace them with newly arrived immigrants.

Most outcomes shown until now, however, do not distinguish between an increase along the extensive or the intensive margin of unionization. In other words, we know that counties that received larger shares of immigrants became more likely to have some union presence. But did immigration also increase the strength of labor unions in already unionized labor markets? I answer this question by restricting the estimation sample to a balanced set of counties that had unions in every decade between 1900 and 1920. This allows me to rule out the possibility that the results simply reflect a comparison between unionized and non-unionized counties. The coefficients, presented in Table A.3, indicate that immigration positively affected skilled unionization also along the intensive margin, increasing the share of unionized workers in always unionized counties.

Taken together, these findings indicate that immigration fostered the emergence and development of labor unions that represented skilled workers. This is consistent with this group of workers having a higher bargaining power with their employers, as their skills provided an entry barrier into their occupations and made them less easily replaceable in the short run. Moreover, these findings indicate that skilled unionization increased as a consequence of immigration both along the extensive and intensive margin.

**Exposure to the immigrants' competition.** One potential alternative explanation for the results just presented is that unions representing skilled workers were able to develop due to an absence of competition between new and existing workers, rather than in reaction to the economic threats brought by the immigrants. In Figure 1.9, I show suggestive evidence in contrast with this hypothesis. I report the prevailing occupations among the immigrants that entered the United States in each decade between 1890 and 1920. Both unskilled (e.g., miners) and skilled (e.g., carpenters, machinists) occupations feature among the most frequent ones.

To formally estimate the effect of immigrant labor market competition on unionization, I interact the main regressor of interest from equation (1.1) with a time-varying measure of a county's exposure to immigrants' competition for jobs.[43] This measure consists of two terms. The first is given by the number of immigrant workers in each occupation $o$ who entered the United States (net of those that settled in county $c$) between $t-10$ and $t$, as a fraction of the total immigrants in the labor force who entered the United States between $t-10$ and $t$. The second is a weight, represented by the share of U.S.-born workers in county $c$ and occupation $o$ in the previous decade:

$$Competition_{c,t} = \sum_{o} \frac{Imm^{o}_{-c,t}}{Imm^{LF}_{-c,t}} \times \frac{USborn^{o}_{c,t-10}}{USborn^{LF}_{c,t-10}} \tag{1.3}$$

The intuition behind this measure is simple: counties where U.S.-born employment (at the beginning of the decade) is concentrated in occupations which are prevalent among recently arrived immigrants are more exposed to labor market competition.

In Table 1.5, I show the results separately for skilled (Panel A) and unskilled (Panel B) unions, where the main regressor of interest is interacted with a standardized version of the measure presented in equation (1.3). In Panel A, the uninteracted estimates are all positive and statistically significant. Remarkably, all the coefficients of the interactions (except for column 1) are also statistically significant. These findings indicate that counties

[43] The logic behind this measure resembles the one employed, among others, by Autor et al. (2020) for import competition from China across U.S. labor markets and by Alsan et al. (2020) for Irish immigrants' labor competition in the 1850s in Massachusetts.

more exposed to the immigrant labor market competition in skilled occupations experienced larger growth in skilled unionization. In contrast, estimates in Panel B show no statistically significant effect on the uninteracted coefficients, while all the estimates of the interaction terms are negative. These results show that, among unskilled workers, increased labor market competition hampered the growth of labor unions instead.

In sum, these findings provide additional evidence for the hypothesis that increased labor competition caused by immigration contributed to the growth of labor unions. Moreover, they indicate that competition fostered unionization only among skilled unions, while it slowed down union growth among unskilled workers. This is consistent with the fact that immigrants were a better and more immediate substitute for unskilled laborers, whose bargaining power got weakened by the increased availability of replacement workers and strikebreakers.

### 1.6.2 Social Motivations

Until now, I have examined the economic channels that have strengthened labor unions as a consequence of immigration. However, one may expect social concerns (e.g., opposition to cultural change) to provide a further incentive for workers to organize and exclude newcomers from the labor market. This possibility is motivated by the nativist rhetoric adopted by the labor movement in this period, and by its vigorous support for immigration restrictions throughout the $20^{\text{th}}$ century (Goldin, 1994; Mink, 2019). At the same time, prominent research has linked the cultural heterogeneity of the U.S. workforce to the country's weak labor movement (Alesina and Glaeser, 2004). In this section, I explore the role of these factors on the development of organized labor.

**Discrimination against culturally distant immigrants.** As described in Section 1.2, not all European immigrants were perceived in the same way. The main worries of the labor movement – and of the nativist movement, more generally – were caused by individuals arriving from Southern and Eastern Europe, who were more culturally distant from U.S.-born residents than the ones who had migrated in large numbers before the 1890s: they spoke non-Germanic languages, were not Protestant, were considered unwilling to assimilate into the American society, and were not responsive to the discipline of labor unions (Goldin, 1994; Higham, 1955; Taft, 1964). If increased unionization was caused in part by xenophobic reactions, the effects should be more prominent in places that received larger shares of more culturally distant immigrants. To test this idea, I estimate

$$y_{ct} = \beta_1 Imm_{ct}^{SE} + \beta_2 Imm_{ct}^{NW} + \theta_c + \tau_t + X_t + u_{ct} \tag{1.4}$$

where $Imm_{ct}^{SE}$ is the fraction of immigrants from Southern or Eastern Europe, and $Imm_{ct}^{NW}$ is the one of immigrants from Northern or Western Europe. Equation (1.4) is estimated using two separate instruments, one for each group, constructed by summing the predicted immigration (as described in Section 1.4.2) from each sending region. I present the results in Table 1.6. As expected, larger increases in unionization are caused by the inflow of immigrants from Southern and Eastern Europe.[44]

**Heterogeneity by attitudes towards immigration.** However, the previous result may conflate economic and cultural concerns, to the extent that immigrants from those areas may have also had lower wage expectations, and made coordination within unions harder due to their higher illiteracy rates and larger linguistic distance than immigrants from Northern and Western Europe. To further explore this channel, I test whether the effects are stronger in counties with worse attitudes towards immigration. In the absence of a direct measure, I use two proxies that likely reflect a county's higher hostility towards immigrants. The first is the historical vote share for the Know Nothing Party, a nativist political party that ran on

[44] The results are unchanged when separately estimating immigration from Protestant and non-Protestant countries (Table A.5).

an anti-Catholic and anti-Irish platform in the mid-1850s (Alsan et al., 2020). The second is a measure of residential segregation between U.S.-born and European immigrants.[45] Since residential segregation usually arises either from collective action to exclude minorities or from individuals from the majority group moving away from ethnically mixed neighborhoods (Boustan, 2013), this characteristic likely reflects higher levels of discrimination against immigrants.

I report the results in Table 1.7, estimating the baseline 2SLS regressions separately for the sample of counties above and below the median vote share for the Know Nothing party and residential segregation, respectively. Using either proxy, I find that immigration strengthened organized labor more prominently in counties with higher resentment towards immigration.

Altogether, these results suggest that non-economic motives also contributed to the expansion of labor unions. Unionization occurred more prominently in counties that received larger shares of culturally distant immigrants, namely those from Southern and Eastern Europe. Moreover, immigration strengthened the American labor movement more in counties that harbored less favorable attitudes towards immigration.

### 1.6.3 Ruling Out Alternative Explanations

**Immigrant-driven unionization.** One alternative explanation for the results is that immigrants may have joined or created labor unions at greater rates than U.S.-born workers. Although data on the individual union members are not available, I exploit the information on the local union representatives described in Section 1.3 to gauge the ethnic composition of their branches. Union delegates can be considered leaders of the organizations they

[45] I construct an index of residential segregation of European immigrants, building on the procedure used in Logan and Parman (2017). The index is constructed using 1880 full-count U.S. Census data, in order to avoid endogeneity concerns. Measuring it after 1890, the baseline year of the instrument, may qualify as a "bad control" (Angrist and Pischke, 2009). For more details on its construction, see Appendix A.3.

represented, as they acted as spokespeople of their local branch at the state and national conventions, and were in charge of making decisions in the name of the members who elected them. For these reasons, their ancestry can be intended as reflecting the ethnic composition of their branch.

As a first step, I use the last names of the delegates to infer their origins, using historical de-anonymized full-count U.S. Census data.[46] Panel A of Figure A.4 shows that, as expected, most of the union leaders were U.S.-born. In Panel B, I break down the shares of delegates by ancestry. Almost all delegates had ancestry from Northern or Western Europe, while very few came from Southern or Eastern Europe.

Although the share of U.S.-born delegates increased – and that of Europeans decreased – over time at the national level, it may still be the case that counties that received more immigrants experienced an increase in the proportion of European leaders. If, for example, newly arrived immigrants joined labor unions *en masse*, we would expect to see an increase in the share of European delegates, as the newcomers would likely obtain the voting power to elect them. To test this, I use the proportion of leaders with last names prevalent among U.S.-born people and Europeans, computed at the county level, as dependent variables in equation (1.1). The coefficients, plotted in Panel A of Figure A.5, indicate that the inflow of immigrants did not increase the proportion of leaders with immigrant last names. The coefficients on the left, estimated on the whole sample of counties, show that immigration increased the share of U.S.-born leaders more than that of immigrants. The ones on the right, computed on the counties where I observe at least a delegate in every year – although imprecisely estimated – suggest that immigration caused a redistribution of delegates in

[46] I describe the procedure I use in Appendix A.2. An alternative approach would be to link individuals to the Census directly, based on the full name. However, most of the unions' convention proceedings only report the delegates' last name and initials, substantially limiting the number of records that could be matched with this method. Moreover, in no occasion do I observe union leaders' year of birth (or age), a key variable usually employed to match people to Census data.

favor of the U.S.-born.[47]

These findings confirm the anecdotal and historical evidence that the observed increase in unionization of this period was not caused by a larger participation of immigrant laborers, but rather by U.S.-born workers (Mink, 2019; Taft, 1964), who maintained the control of labor unions throughout the first 20 years of the 20$^{th}$ century.

**Previous exposure to labor unions.** A second possibility is that immigrants coming from European countries that already had labor unions by the end of the 20$^{th}$ century may have brought into the United States their experience of collective bargaining from their home country, and, in turn, contributed to the growth of unionization in their destination counties. This explanation would be in line with existing work arguing that Europeans who migrated to the United States between 1910 and 1930 promoted spillover of ideologies to U.S.-born individuals (Giuliano and Tabellini, 2022). Although the results just presented already suggest that immigrants' participation in labor unions did not increase upon Europeans' arrival, I test this hypothesis formally, estimating the effect of immigration separately for immigrants coming from countries with or without strong labor unions.[48] The results, shown in Table A.6, rule out this possibility. The coefficients of the share of immigrants from the U.K. and Ireland, the only countries with a strong labor movement at the turn of the 20$^{th}$ century, are never statistically significant; on the contrary, the coefficient for the share of immigrants from the rest of Europe are positive and statistically significant.

**Other economic channels.** Another possibility is that the growth in unionization may have been caused by a differential economic expansion – or contraction – experienced by counties receiving larger shares of immigrants. Alternatively, the observed effect may have been a result of a decrease in the number of individuals working in occupations represented

[47] Analogous conclusions hold when looking at the proportion of union leaders with either Northern/Western European or Southern/Eastern European ancestry (Panel B Figure A.5).

[48] I use data from Crouch (1993) to classify European countries into these two groups. Appendix A.4 provides more information on the data and on labor unions in Europe during this period.

by AFL unions, which would mechanically increase the measure of union density. In Table A.7, I show that this is not the case. Immigration had no effect on economic indicators such as the (male) labor force participation rate or total manufacturing output (measured both as a fraction of the manufacturing labor force in the county, or of the output in the country). Moreover, the effect on the (log of the) total number of workers in occupations covered by skilled unions is actually positive, although imprecisely estimated; if anything, this goes against the estimated effect, as it mechanically reduces union density.

This discussion suggests that the results are unlikely to be driven by the preferences or ideologies brought by immigrants to the United States, or by the effects of immigration on the local economy.

## 1.7 Implications and Discussion

In this section, I provide and discuss some implications related to the immigration-induced unionization in skilled workers' unions. Although not all these findings can be interpreted as causal, they still provide insights on the short- and medium-run trends associated with a higher presence of organized labor.

**Effects on U.S.-born workers' outcomes.** A question unexplored thus far in the paper is whether immigration had any effect on the distribution of occupations among U.S.-born workers. In particular, one may expect U.S.-born workers to turn to unionized occupations, to safeguard themselves from the perceived threats of immigrant competition and cultural differences. I explore this possibility by testing whether immigration had a different impact depending on whether a certain occupation had a positive union membership in the county or not. More specifically, I restrict the attention to occupations within the jurisdiction of the AFL unions, and compute the county shares of U.S.-born workers in occupations with and without local union representation. The results are presented in Table A.8. Consistent with the hypothesis, immigration increased the share of the U.S.-born labor force in skilled occupations that had union representation in the county. On the other hand, the effect on the share of U.S.-born in occupations with no union representation is negative and not statistically significant. Although this explanation is consistent with the historical narrative of the period (Mink, 2019), stating that U.S.-born workers resorted to skilled (craft) unions in response to immigration, these results are also consistent with a different – and potentially complementary – interpretation. In particular, it is possible that union representation may have occurred simultaneously or as a consequence of U.S.-born workers moving to those occupations. Data limitations prevent me from exploring the exact timing. However, the fact that the employment of the U.S.-born did not increase overall across all skilled occupations – but only in those with local union presence – assuages concerns that the observed growth in unionization may be a mere result of an overall employment shift towards skilled occupations.

In addition, consistent with existing evidence in both historical settings (Abramitzky et al., 2023; Tabellini, 2020) and recent times (Card, 2001b, 2005, 2009a; Foged and Peri, 2016a; Ottaviano and Peri, 2012b), immigration did not have negative effects on the labor market outcomes of domestic workers, which I measure with the labor force participation rate and the (log) occupational income score (Table A.9).[49] In light of the increased unionization caused by immigration that this paper documents, these results suggest that labor unions may have mediated the potentially adverse economic consequences on domestic workers of the immigration-induced labor supply expansion.

**Unions and inequality.** Another central economic question that arises from the findings of this paper concerns the consequences of unionization on inequality. Recent evidence (Farber et al., 2021) has documented a causal impact of labor unions in reducing inequality for most of the 20$^{\text{th}}$ century, combining national and state-level survey data on unionization from the

[49]The full-count Census data of this period do not consistently report information on employment status (only in 1910), and information on wages was first collected in 1940 (Ruggles et al., 2021).

mid-1930s onwards. I use data on wages from the U.S. Census of 1940 – the first year in which such information was collected – to compute measures of wage inequality at the county level, and investigate the correlation between them and measures of unionization in 1920 – the last year in the sample. Following the literature (Autor et al., 2008), I measure inequality as the log wage differentials for full-time, full-year workers computed at the following percentiles: 90 to 10; 90 to 50; and 50 to 10.[50] I present the results in Table A.10. The coefficients in Panel A display a negative correlation between the presence of labor unions in the county and wage inequality. Although not causal, these results are consistent with existing studies documenting labor unions' contribution in reducing inequality (Collins and Niemesh, 2019; Farber et al., 2021), and suggest that unions may have done so already in the first four decades of the last century.

**Persistence of unionization.** Further, I examine whether the local patterns of unionization that emerged in the early 20$^{\text{th}}$ century, which I document for the first time in this paper, persisted until today. I aggregate the data on union density between 1900 and 1920 at the metropolitan-area level, to make them consistent with the current measures of unionization from Macpherson and Hirsch (2023), and explore their cross-sectional correlation with the average levels of union density over the first two decades of the 21$^{\text{st}}$ century. The results are presented in Figure A.6. Remarkably, even after controlling for Census division fixed effects, which account for differences in attitudes towards organized labor across areas of the country, past and present unionization are positively correlated. This suggests that the conditions that favored the initial development of labor unions in the early 1900s may have provided the labor movement with a head start that perdures throughout decades.

[50] As in Autor et al. (2008), I exclude self-employed workers, and construct weekly wages focusing on men ages 16–64 years old who worked for at least 40 weeks and at least 35 hours per week.

## 1.8 Conclusion

Despite the enduring relevance of labor unions throughout history and in contemporary society, we lack rigorous empirical evidence regarding the determinants of their origins and early growth. In this paper, I investigate the effects of a large labor supply increase, represented by the mass immigration of the early 20$^{\text{th}}$-century United States, on the development of organized labor. I find that immigration strengthened the labor movement by increasing the probability that a county had any union, the share of unionized workers, the number of unions, and their average membership. I document that both economic and social concerns were responsible for the effect: unions grew due to workers' reactions to the increased labor competition brought by immigrants and to concerns about cultural change.

The findings of this paper quantitatively identify immigration as a novel driver of unionization during the early days of the American labor movement. The estimates imply that in the absence of immigration, the average union density between 1900 and 1920 would have been 17% lower. They also shed light on an unexplored consequence of immigration: the strengthening of institutions that protect incumbent workers' status in the labor market. Notably, this study also broadens our understanding of the multifaceted implications of immigration. It suggests that individuals' reactions to immigration are not confined to political shifts toward conservative parties or the advocacy of anti-immigration policies, as previously emphasized in existing research. Instead, immigration can also spark the development of self-organized institutions with broad political impact, such as labor unions.

While the specific quantitative estimates presented in this paper may pertain to the unique context under examination, its implications carry wider-reaching significance. They underscore the role played by both economic and cultural considerations in shaping labor market dynamics and institutions, suggesting that effective labor market policies should take all these aspects into account. Furthermore, the study provides valuable insights into the factors contributing to the recent resurgence of the labor movement, particularly following a period of challenges for private sector labor unions. The numerous successes achieved by organized labor in various sectors such as automotive, transportation, education, and services over the past few years, as well as the emergence of unionization efforts in previously unorganized multinational corporations like Amazon and Starbucks, encourage new considerations. For example, they suggest that this renewed interest in labor unions may also reflect concerns about job scarcity, arising from a confluence of heightened competition in the labor market (due to significant immigration flows) and rapid technological advancements.

Importantly, the relevance of these findings extends beyond the United States. These results speak to the context of many European countries experiencing a surge in immigration while labor unions continue to wield economic and political influence. Additionally, these findings hold significance for industrializing and recently industrialized countries whose economic transformations parallel those of early 20$^{\text{th}}$-century America. They may also apply more broadly to settings where institutional safeguards for workers' rights to organize and collectively bargain are limited.

Finally, this study paves the way for several promising avenues of future research. First, it prompts further investigation into the drivers of organized labor's growth across different economic contexts and time periods. Second, the comprehensive data collected for this paper will allow researchers to investigate several other questions, such as the long-term consequences of the early 20$^{\text{th}}$-century unionization on the American experience of immigrants, and on the evolution of the U.S. economy, more generally.

# Tables

Table 1.1: Summary Statistics

| | Obs. | Mean | St. Dev. |
|---|---|---|---|
| *Panel A: Unionization* | | | |
| Total Membership | 5,025 | 254.58 | 1,557.79 |
| Union Density | 5,025 | 0.04 | 0.11 |
| Nr. Branches | 5,025 | 1.62 | 5.59 |
| Avg. Branch Size | 5,025 | 29.94 | 71.31 |
| Presence of Any Union | 5,025 | 0.26 | 0.44 |
| *Panel B: Demographics* | | | |
| Share Immigrants (All) | 5,025 | 0.07 | 0.08 |
| Share Immigrants (¡10 years in U.S.) | 5,025 | 0.02 | 0.03 |
| Share Urban Population | 5,025 | 0.18 | 0.24 |
| Total Population | 5,025 | 33,010.55 | 102,216.67 |
| *Panel C: Labor Force (men 16–64)* | | | |
| LF Participation Rate | 5,025 | 0.91 | 0.04 |
| Share of U.S.-Born LF | 5,025 | 0.87 | 0.15 |
| Share of Immigrant LF | 5,025 | 0.11 | 0.12 |

*Notes:* The table presents summary statistics for the over 2,400 counties in the sample described in Section 1.3, in the years 1900, 1910, and 1920. Share Immigrants is the number of European immigrants as a fraction of the population in the county. Union Density is the number of union members divided by the labor force in occupations represented by AFL unions (see Table A.2 for the complete list). Avg. Branch Size is the number of union members divided by the number of branches.

Table 1.2: First Stage of the Instrumental Variable Estimation

| | *Dependent variable:* Share Immigrants | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Predicted Share Immigrants | 0.280*** | 0.258*** | 0.253*** |
| | (0.046) | (0.043) | (0.043) |
| Observations | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.024 | 0.024 | 0.024 |
| Indep. var. mean | 0.027 | 0.027 | 0.027 |
| KP F-stat | 37.28 | 35.33 | 35.14 |
| 1890 Urban Share | No | Yes | Yes |
| 1880 LF Part. Rate | No | No | Yes |

*Notes:* Observations are at the county-decade level. The table reports the first stage of the instrument described in Section 1.4.2. The dependent variable is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The main regressor of interest is the predicted number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the 1890 male population in the county. All regressions include county and year fixed effects. The following controls, interacted with year dummies, are also included: the 1890 share of urban population (column 2); and, the 1880 male labor force participation rate (column 3). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table 1.3: The Effect of Immigration on Organized Labor

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: OLS* | | | | |
| Share Immigrants | 0.167 | 0.027 | 0.885** | 57.633 |
| | (0.239) | (0.056) | (0.345) | (37.767) |
| *Panel B: Reduced Form* | | | | |
| Pred. Share Immigrants | 0.397** | 0.072** | 0.737*** | 65.931** |
| | (0.173) | (0.030) | (0.216) | (28.352) |
| *Panel C: 2SLS* | | | | |
| Share Immigrants | 1.572** | 0.285** | 2.918*** | 260.959** |
| | (0.699) | (0.117) | (0.854) | (110.674) |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.265 | 0.039 | 0.402 | 29.936 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 35.14 | 35.14 | 35.14 | 35.14 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). Panel A shows OLS estimates, where the regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. Panel B shows reduced form estimates, with the instrument described in Section 1.4.2. Panel C shows 2SLS estimates. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table 1.4: Heterogeneous Effects by Workers' Skills

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Skilled (Craft) Unions* | | | | |
| Share Immigrants | 1.456** | 0.239*** | 2.714*** | 250.621*** |
| | (0.646) | (0.083) | (0.792) | (96.708) |
| Observations | 5,025 | 5,025 | 5,024 | 5,025 |
| Dep. var. mean | 0.214 | 0.019 | 1.147 | 21.351 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 35.14 | 35.14 | 35.14 | 35.14 |
| *Panel B: Unskilled (Industrial) Unions* | | | | |
| Share Immigrants | -0.326 | -0.140 | -0.117 | -81.796 |
| | (0.440) | (0.242) | (0.447) | (82.577) |
| Observations | 4,398 | 4,398 | 4,398 | 4,398 |
| Dep. var. mean | 0.134 | 0.084 | 0.545 | 18.582 |
| Indep. var. mean | 0.025 | 0.025 | 0.025 | 0.025 |
| KP F-stat | 99.00 | 99.00 | 99.00 | 99.00 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 1.6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table 1.5: Heterogeneous Effects by Immigrants' Labor Market Competition

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Skilled (Craft) Unions* | | | | |
| Share Immigrants | 1.555** | 0.264*** | 2.987*** | 289.930*** |
| | (0.654) | (0.088) | (0.802) | (104.911) |
| Share Immigrants × Competition | 0.610 | 0.192** | 1.994*** | 275.998** |
| | (0.553) | (0.084) | (0.757) | (122.461) |
| Observations | 5,025 | 5,025 | 5,024 | 5,025 |
| Dep. var. mean | 0.214 | 0.019 | 1.147 | 21.351 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 17.42 | 17.42 | 17.42 | 17.42 |
| SW F-stat (Sh. Imm.) | 38.99 | 38.99 | 39.00 | 38.99 |
| SW F-stat (Sh. Imm. × Competition) | 25.21 | 25.21 | 25.21 | 25.21 |
| *Panel B: Unskilled (Industrial) Unions* | | | | |
| Share Immigrants | 0.197 | -0.015 | 0.544 | 62.253 |
| | (0.481) | (0.274) | (0.509) | (80.107) |
| Share Immigrants × Competition | -0.416** | -0.100 | -0.535*** | -117.235*** |
| | (0.175) | (0.084) | (0.165) | (40.917) |
| Observations | 4,398 | 4,398 | 4,398 | 4,398 |
| Dep. var. mean | 0.134 | 0.084 | 0.545 | 18.582 |
| Indep. var. mean | 0.025 | 0.025 | 0.025 | 0.025 |
| KP F-stat | 41.99 | 41.99 | 41.99 | 41.99 |
| SW F-stat (Sh. Imm.) | 84.64 | 84.64 | 84.64 | 84.64 |
| SW F-stat (Sh. Imm. × Competition) | 88.41 | 88.41 | 88.41 | 88.41 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 1.6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. Competition is a (standardized) measure of the immigrants' labor market competition, based on the prevailing occupations among the immigrants that enter the U.S. in each decade and the ones of the U.S.-born workers in each county in the previous decade, as described in Section 1.6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 1.6: Heterogeneous Effects by Origin of Immigrants

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| Share Immigrants S/E Europe | 1.941 | 0.366* | 3.735** | 383.668* |
| | (1.245) | (0.206) | (1.532) | (214.294) |
| *Standardized coefficient* | *[0.123]* | *[0.090]* | *[0.131]* | *[0.150]* |
| Share Immigrants N/W Europe | 0.769 | 0.110 | 1.136 | -7.702 |
| | (1.444) | (0.313) | (1.567) | (342.994) |
| *Standardized coefficient* | *[0.035]* | *[0.019]* | *[0.028]* | *[-0.002]* |
| Observations | 5,018 | 5,018 | 5,018 | 5,018 |
| Dep. var. mean | 0.265 | 0.039 | 1.627 | 29.978 |
| Indep. var. mean (S/E Europe) | 0.028 | 0.028 | 0.028 | 0.028 |
| Indep. var. mean (N/W Europe) | 0.020 | 0.020 | 0.020 | 0.020 |
| KP F-stat | 16.15 | 16.15 | 16.15 | 16.15 |
| SW F-stat (S/E Europe) | 39.63 | 39.63 | 39.63 | 39.63 |
| SW F-stat (N/W Europe) | 113.34 | 113.34 | 113.34 | 113.34 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from Southern/Eastern Europe or Northern/Western Europe who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 1.4.2 and Section 1.6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

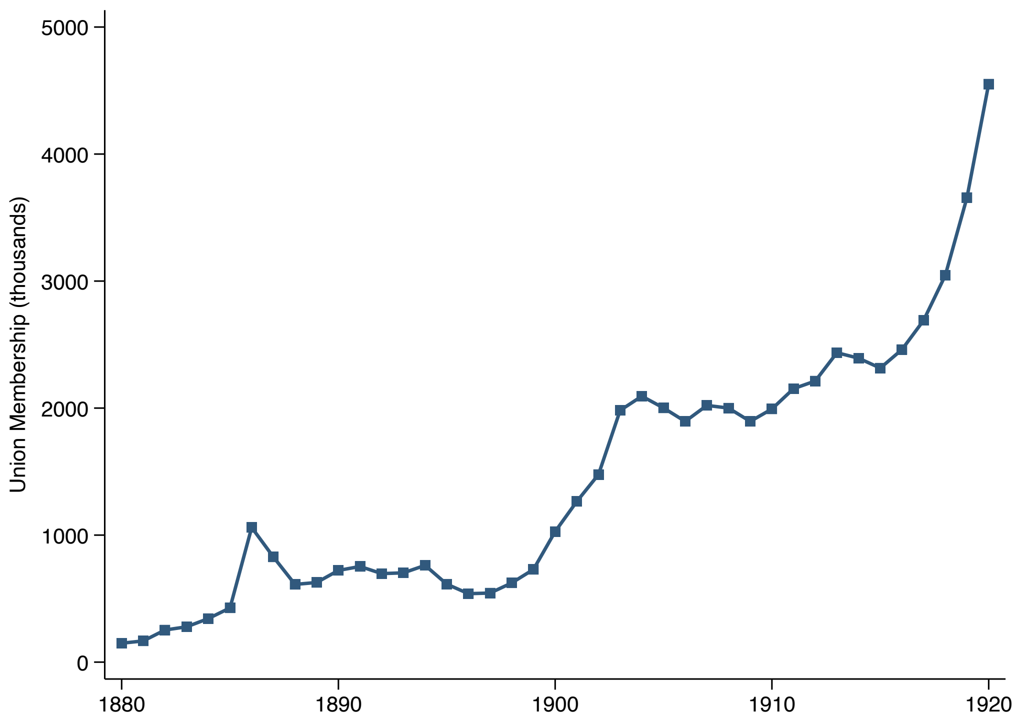
Table 1.7: Heterogeneous Effects by Attitudes Towards Immigration

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Union Presence | Union Density |
| | (1) | (2) | (3) | (4) |
| *Panel A: Vote share Know-Nothing party* | Above median | | Below median | |
| Share Immigrants | 2.019* | 0.342** | 2.024 | -0.254 |
| | (1.156) | (0.158) | (3.041) | (0.362) |
| *Standardized coefficient* | *[0.147]* | *[0.079]* | *[0.142]* | *[-0.088]* |
| Observations | | | | |
| Dep. var. mean | 1,680 | 1,680 | 1,660 | 1,660 |
| Indep. var. mean | 0.257 | 0.050 | 0.346 | 0.040 |
| KP F-stat | 0.014 | 0.014 | 0.020 | 0.020 |
| | 41.83 | 41.83 | 9.19 | 9.19 |
| *Panel B: Index of residential segregation* | Above median | | Below median | |
| Share Immigrants | 3.082** | 0.454** | 0.694 | 0.196 |
| | (1.336) | (0.188) | (0.803) | (0.167) |
| *Standardized coefficient* | *[0.286]* | *[0.162]* | *[0.059]* | *[0.067]* |
| Observations | 2,433 | 2,433 | 2,436 | 2,436 |
| Dep. var. mean | 0.292 | 0.044 | 0.243 | 0.035 |
| Indep. var. mean | 0.028 | 0.028 | 0.020 | 0.020 |
| KP F-stat | 9.81 | 9.81 | 52.87 | 52.87 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. In Panel A, the estimation sample is split around the median of the vote share for the Know Nothing party in the 1856 presidential elections. In Panel B, the estimation sample is split around the median of the index of residential segregation calculated in 1880 and described in Section 1.6 and Appendix A.3. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

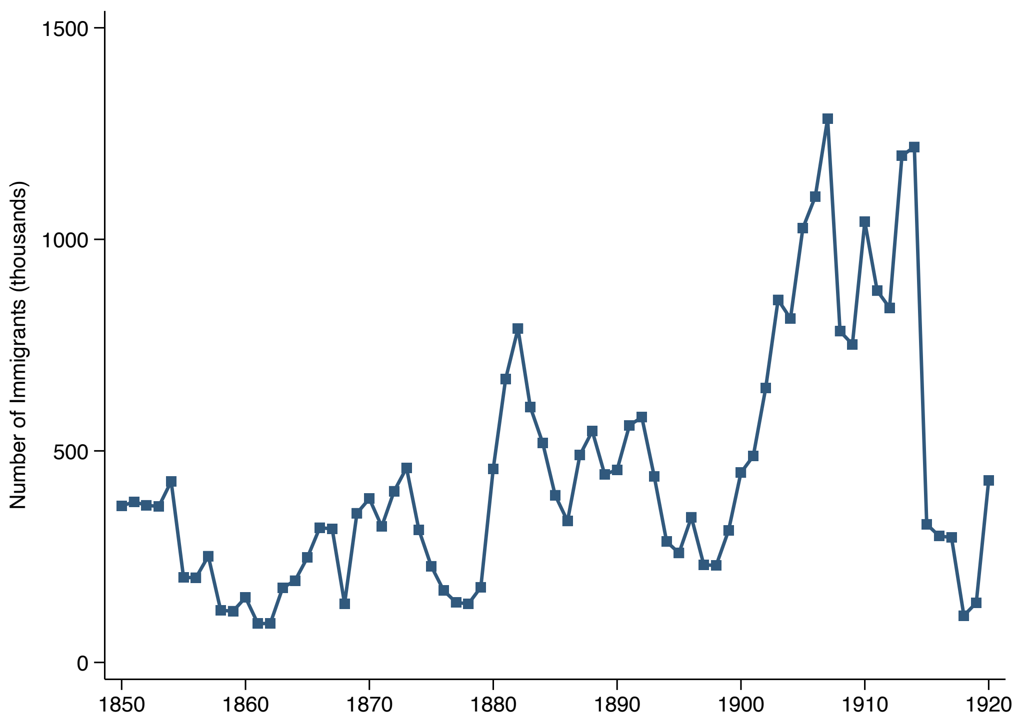
# Figures

Figure 1.1: Estimates of Total Union Membership (1880–1920)



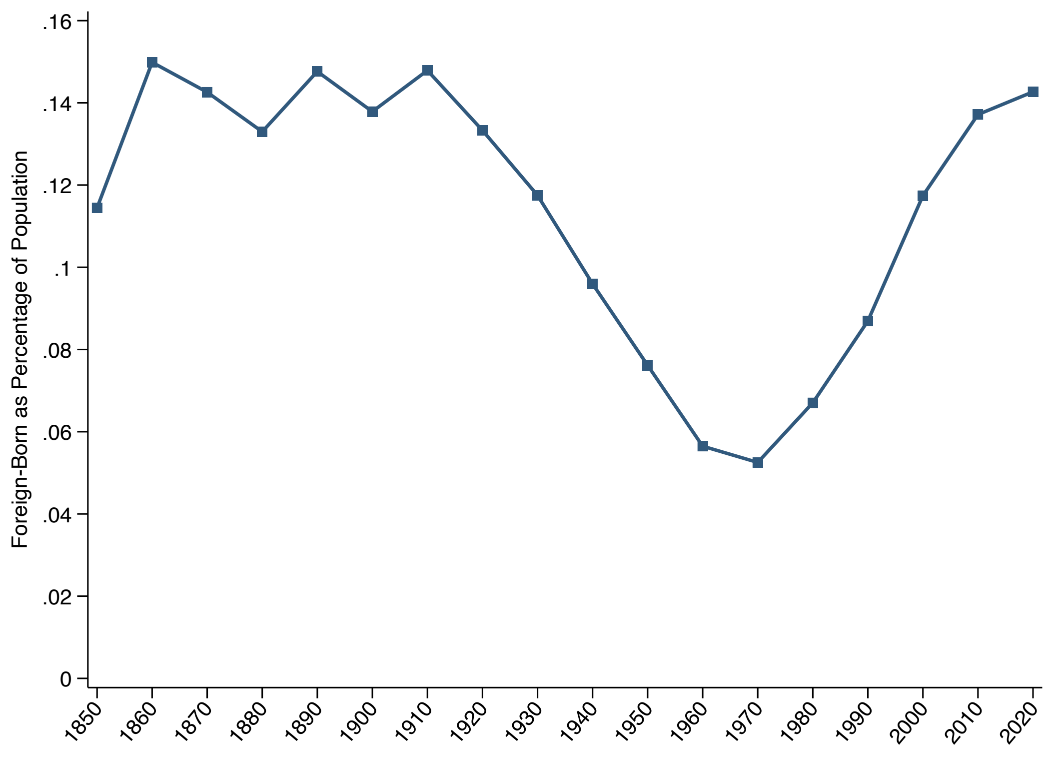

*Notes:* The figure shows the total number of union members in the U.S., between 1880 and 1920. Source: Freeman (1998).

Figure 1.2: Annual Inflow of Immigrants (1850–1920)



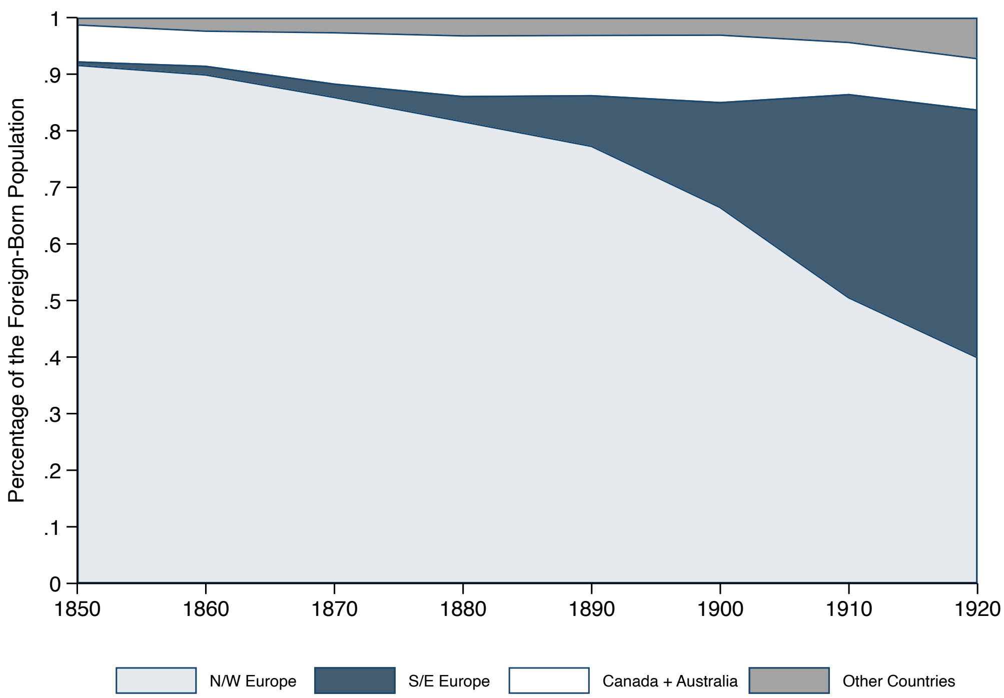

*Notes:* The figure shows the total number of immigrants to the United States, between 1850 and 1920. Source: Immigration Policy Institute.

Figure 1.3: Foreign-Born Stock as a Percentage of the U.S. Population (1850–2020)



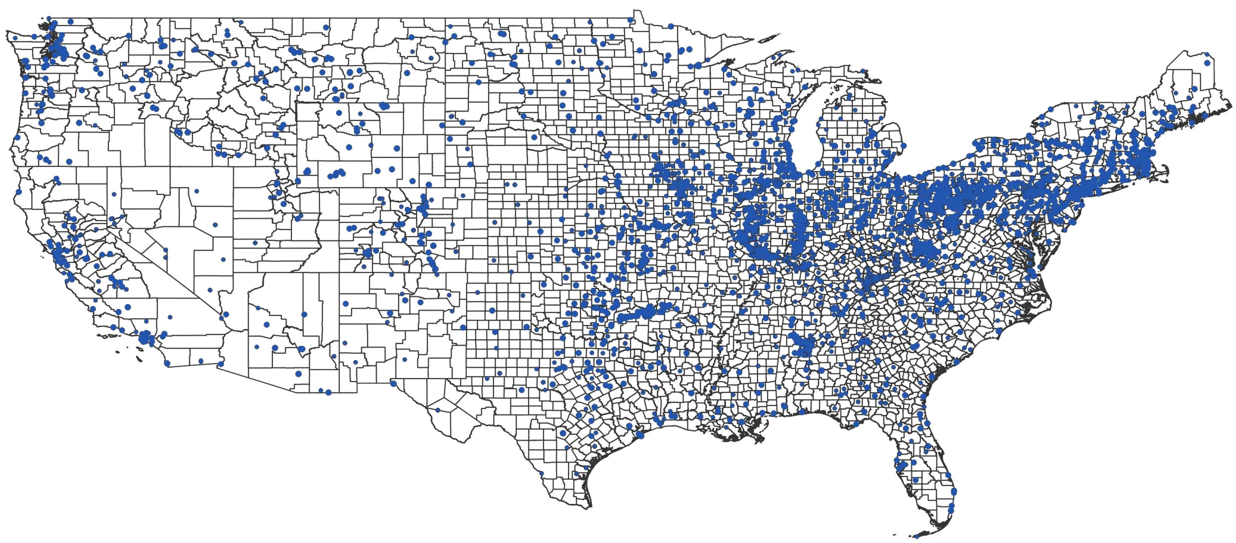

*Notes:* The figure shows the number of foreign-born individuals as a percent of the U.S. population, between 1850 and 2020. Source: Author's calculations from full count and samples of the U.S. Census of Population, made available by IPUMS (Ruggles et al., 2021) and ICSPR (Haines, 2010b).

Figure 1.4: Sending Regions within the Foreign Born Population (1850–1920)



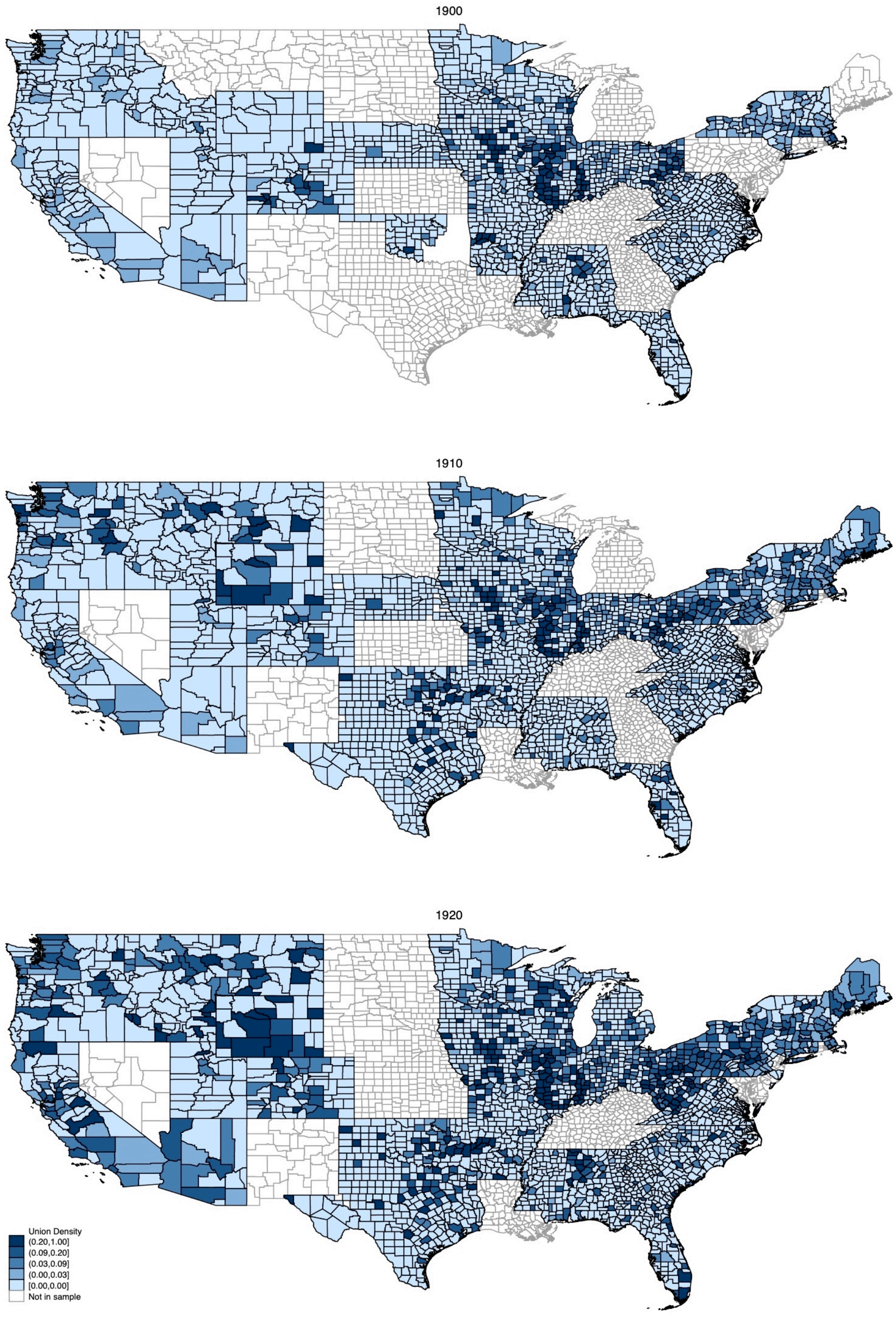

*Notes:* The figure shows the number of foreign-born individuals by region of origin, as a share of the total foreign-born population, between 1850 and 1920. Source: Author's calculations from full count U.S. Census of Population, made available by IPUMS (Ruggles et al., 2021) and ICSPR (Haines, 2010b).

Figure 1.5: Geographic Distribution of Union Branches 1900–1920



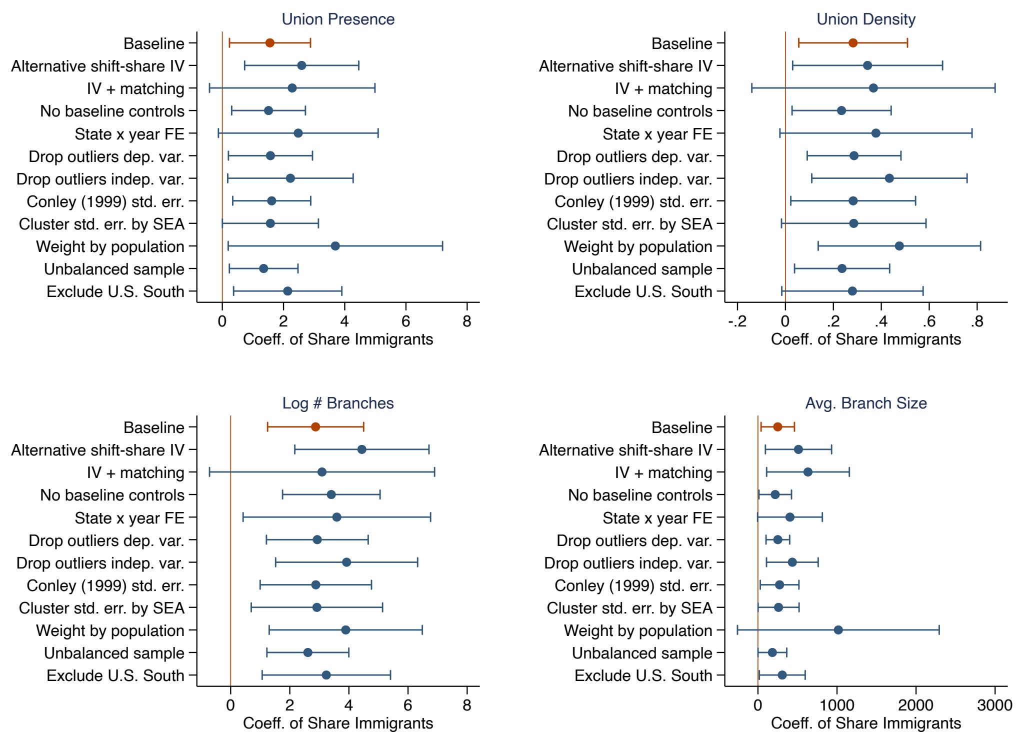
*Notes:* The map plots the all the union branches recorded and geocoded from the newly digitized labor union records described in Section 1.3.

Figure 1.6: Geographic Distribution of Union Density Over Time



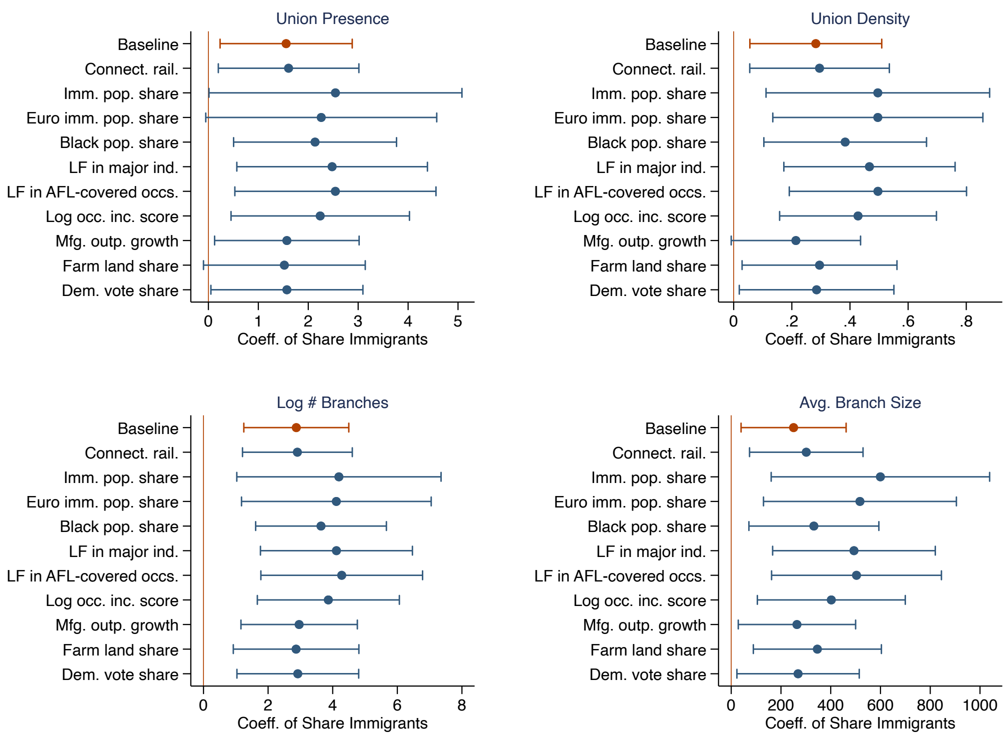

*Notes:* The maps plot the county-level shares of the union membership rate in 1900, 1910, and 1920. The legend shows the deciles with respect to the 1920 distribution. Source: Author's calculations from union convention proceedings, as described in Section 1.3.

Figure 1.7: Summary of Robustness Checks



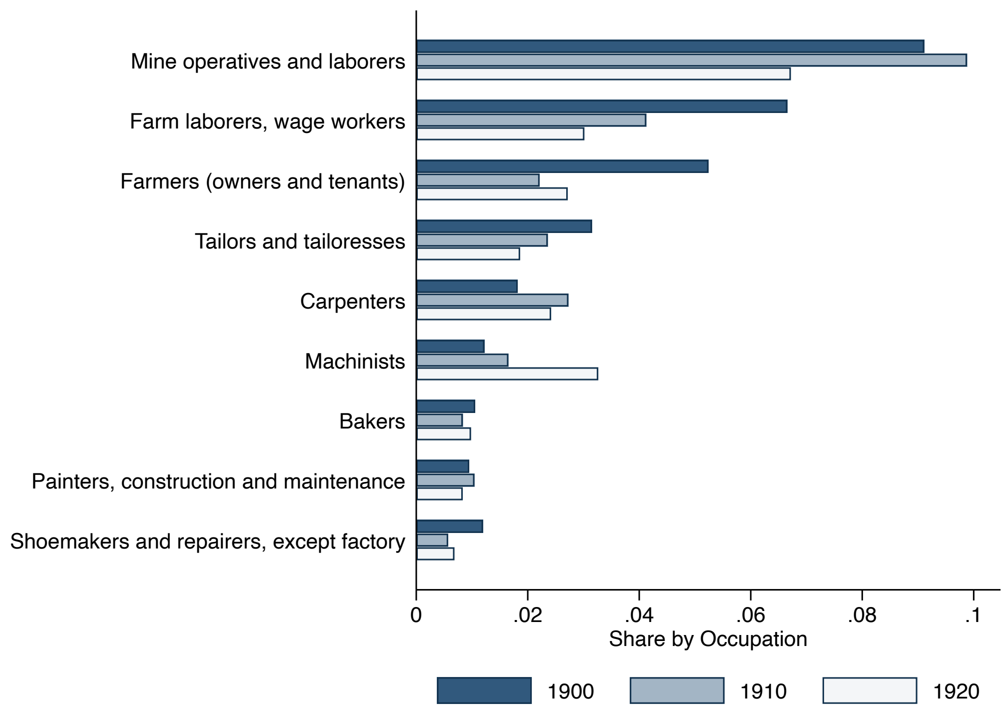

*Notes:* The figure presents a summary of the main robustness checks described in Section 1.5.2. The estimates plotted are the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, the main independent variable of equation (1.1). The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county. For more details and formal estimates, see also Appendix A.1.

Figure 1.8: Robustness Check: Controlling for Additional Baseline Characteristics




*Notes:* The figure plots the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, the main independent variable of equation (1.1), augmenting the specification of Table 1.3 with the variable(s) indicated in each row, measured at baseline and interacted with year dummies. The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county. For more details, see the description of the robustness checks in Section 1.5.2 and the formal estimates presented in Appendix A.1.

Figure 1.9: Prevailing Occupations Among Immigrants 1900–1920




*Notes:* The figure shows the prevailing occupations among recently arrived immigrants, on average between 1900 and 1920. Shares indicate the number of recent (¡ 10 years in the U.S.) immigrants with the reported occupation as a fraction of the total number of recent immigrants. Generic categories not classified by IPUMS (e.g., "laborers (n.e.c.)") are omitted, since they do not identify specific occupations.

# Chapter 2

## The Impact of the Chinese Exclusion Act on U.S. Economic Development

(co-authored with Joe Long, Nancy Qian, and Marco Tabellini)

## 2.1 Introduction

In 1882, the U.S. government enacted the first ban of voluntary immigration of an entire group based on the country of origin or ethnicity. The Chinese Exclusion Act banned individuals born in China from entering the United States and existing Chinese on U.S. soil from re-entering the country or obtaining citizenship. The Act was widely popular in the United States. A central motivation was economic, as many believed that reducing the number of Chinese immigrants would relax resource constraints and improve economic opportunities for white workers. This economic motivation was reinforced by xenophobic fears of the "yellow peril" and strategic political considerations. Indeed, supporting the Act likely increased politicians' popularity among the non-Chinese population, and came at the cost of losing very few, if any, votes, since the Chinese-born individuals were less than 1% of total U.S. population at the time.

The economic impact of the Exclusion Act is *ex ante* ambiguous, even for white workers.[1] On the one hand, it can reduce economic competition from Chinese immigrants, thereby raising wages and employment among white workers. On the other hand, it can lower overall economic prosperity because of the loss of labor and skills, especially if Chinese and white workers are complements in production. Although Chinese workers were a negligible

[1] In 1880, 92% of the population in the western U.S. was white.

share of the U.S. population at the time, they often represented a non-trivial segment of the workforce in several western counties. Hence, whether the net effect of the Act was positive or negative is an empirical question.

Perhaps surprisingly, very little is known about the impact of the Act on non-Chinese workers and aggregate economic production. In this paper, we seek to fill this gap, evaluating the effects of the Exclusion Act on U.S. economic development as well as on population size, employment, and income of Chinese, white, and all groups. We estimate a *difference-in-differences* (DD) regression, combining: variation in the implementation of the Act over time; and, cross-county variation in the intensity of treatment from differences in the size of the Chinese population share prior to the Act. Specifically, we compare the evolution of economic outcomes, before and after the Chinese Exclusion Act, between counties with the 1880 Chinese population share above and below the sample median.

To account for time invariant differences across counties, we include county fixed effects. To control for changes over time that affect all counties within a state similarly, we include state by Census decade fixed effects. Since many earlier Chinese immigrants worked on the Transcontinental railroad and mining, the baseline estimates also control for whether a county was connected to a railroad, and for whether it ever had a mine interacted with decade fixed effects. Causal inference assumes that, conditional on the controls, the economic outcomes of interest would have evolved along parallel trends in counties with a high and a low historical presence of Chinese.

The analysis proceeds as follows. First, we examine the effects of the Act on the Chinese. Consistent with historical narratives that the Act stopped new immigration from China and caused many Chinese workers to leave the U.S., we find that the Act dramatically reduced the size of the Chinese population and labor force. The decline in labor supply occurred in all major sectors – manufacturing, mining, railroads and agriculture – and involved both skilled and unskilled workers. Moreover, the Act led to a steep decline in the income of

Chinese workers.[2]

Second, we examine the non-Chinese and total population. We find that the Act had negative effects for all other workers – white and non-white. The Act had a negative impact on population size, labor supply and incomes in all major economic sectors and across skill levels. We do not find positive effects on any economic outcome for any group of workers. Thus, our analysis suggests that the Act worsened the economic well-being of all workers in the western United States. Our estimates are quantitatively large: relative to counties with below-median 1880 Chinese share, counties with above-median share experienced an approximate decline of 40% in total population and 50% in total labor force.

The dynamic estimates show that there are no pre-trends, and the trend-break occurs soon after the introduciton of the Act. This supports the parallel trends assumption and reduces concerns that our estimates are confounded by spurious correlations. The dynamic estimates also show that the effects persisted over time, until the end of our sample in 1940.

We conduct several additional robustness checks to address the concern of potentially confounding influences from omitted variables. First, we document that results are unchanged if we control for the impact of the Exclusion Act on adjacent counties. This addresses the concern that the main estimates are confounded by geographic spillovers and relocation from counties with high 1880 Chinese share to low 1880 Chinese share.[3] Second, we show that our estimates are robust to the inclusion of interactions between year dummies and a large number of pre-Act county features. The latter include the 1880 share of non-Chinese immigrants, 1880 population, and 1880 labor force in manufacturing and in agriculture. These robustness exercises address the concern that the 1880 Chinese population share was correlated with other characteristics that influenced long-run economic growth. Finally, in order to assuage

[2]The U.S. Census of Population did not collect wages prior to 1940. We follow the literature (Abramitzky et al., 2014), and proxy for income using occupational income scores, which assign to an individual the median income of his job category in 1950.

[3]For example, our estimates may be biased if the Act induced labor or firms to move from counties with high pre-Act Chinese population shares to counties with low pre-Act Chinese population shares.

concerns that our estimates may be driven by specific observations, we omit counties with 1880 shares of Chinese population at the extreme tails of the distribution, and replicate the analysis by excluding San Francisco county from our sample. We present additional robustness tests in the Appendix.

One of the interesting features of the context we study is the structural transformation away from agriculture towards manufacturing. During the early parts of this transformation, rural productivity grew faster than urban productivity (Eckert and Peters, 2022a). To understand how this interacts with the exclusion of the Chinese, we compare counties in the eastern United States, which had virtually no Chinese immigrants in 1880, to otherwise similar counties in the U.S. West that had a Chinese share in 1880 above the sample median. We find that these eastern counties did not experience the same decline as their western analogues after the Act. In contrast, they prospered and grew.

To shed more light on the economic impact of the dramatic demographic change due to the Chinese Exclusion Act, we examine economic production and other indicators of performance. We find that the Act reduced manufacturing output and wages, the number of manufacturing establishments, and the number of mines. This suggests that the depopulation triggered by the Act led to the closure of entire manufacturing and mining establishments, which is consistent with our finding that the Act reduced the number of people working in all sectors and of all skill levels.

We also find that the Act reduced the average value of agricultural inputs: farm land, livestock, and farm machinery, as well as the use of fertilizers. One interpretation for these results is that depopulation reduced demand for agricultural goods (food), thereby lowering the value of agricultural inputs – land, livestock and capital. Another possibility is that the Act reduced the quality of farm land, since Chinese workers were important contributors to complex land improvement projects (e.g., drainage of swamps).[4]

[4]Chang (2003) discusses instances of Chinese workers draining swamps and conducting other engineering

Taken together, the results show that the Chinese Exclusion Act did not lead to any tangible improvement in the economic circumstances of other workers. In fact, the opposite happened, as the loss of skilled and unskilled Chinese workers triggered a cascade of negative economic effects for the economy at large. Our findings imply that Chinese labor was a complement to the labor of natives and other groups. The magnitudes of our findings might be specific to the context of our study. However, the key insight that the loss of economically productive immigrant labor can lead to negative economic consequences if immigrants complement natives is generalizable.

Our paper provides empirical evidence that reducing immigration worsens the economic outcomes of native workers and the overall economy. As such, we add to the large empirical literature studying the impact of immigration on a wide range of outcomes, with some papers finding negligible or positive effects on natives' outcomes (e.g., Card, 2001c; Card, 2009b; Ottaviano and Peri, 2012a; Chassambouli and Peri, 2015; Foged and Peri, 2016b; Sequeira et al., 2020b; Tabellini, 2020) and others finding negative effects (e.g., Borjas, 2003; Borjas, 2005). Our findings also contribute to the literature that studies the effects of in- and out-migration on overall economic activity, productivity, and growth (Burchardi et al., 2020; Chaney and Hornbeck, 2016; Desmet et al., 2018; Peters, 2021).

In addition, we complement a growing literature that evaluates the impact of immigration restrictions on internal migration, natives' outcomes, and on the aggregate economy. Abramitzky et al. (2023) and Clemens et al. (2018) find that the Immigration Acts of the 1920s and the end of the Bracero program in 1964, respectively, did not benefit native workers in any meaningful way. Our results are consistent with, though stronger than, those in these works, as we document that the Chinese Exclusion Act had a steep, negative effect for all workers (including native whites). In this respect, our findings resonate with those obtained by Moser and San (2019), who show that the immigration quotas of the 1920s

activities, often along railroads, that improved farmland.

lowered American science and invention, not only by excluding foreign born scientists but also reducing productivity of the native born ones.

Finally, our paper is related to a recent strand of the literature that analyzes the effects of the Chinese Exclusion Act on the economic and social assimilation of Chinese immigrants and their descendants (Chen and Xie, 2020; Chen, 2015).[5]

The rest of the paper is organized as follows: Section 2.2 discusses the historical background. Section 2.3 presents the empirical strategy. Section 2.4 describes the data. Section 2.5 presents the results. Section 2.6 offers concluding remarks.

## 2.2 Historical Background

### 2.2.1 Chinese Immigrants Before 1882

The Act was the first U.S. policy that banned voluntary immigration of an entire group, and effectively kept Chinese immigration at negligible levels until the Immigration and Nationality Act of 1965. Chinese immigrants first arrived in large numbers to the United States in the 1850s during California's gold rush. From 1870 to 1880, a total of 138,941 Chinese immigrants entered the U.S., which made up around 4.3% of all immigrants during the period (Lee, 2003, p.25). Like other immigrants, the Chinese sought better economic opportunities and a chance to escape political chaos at home. In China, opportunities for upward mobility were limited by the official examination system and widespread corruption (Chang, 2003, pp. 7-9). The Opium Wars (1839-1842, 1856-1860) and the Taiping Rebellion (1850-1864) furthered caused tremendous suffering – famine, poverty – and turmoil (Spence, 1990, pp. 168-175). Although the Qing government opposed its citizens leaving the country, it did

[5]Since our first draft in January 2022, we have learned about a related study by Hoi (2022), who examines the impact of the Act on directly exposed native workers. Our focus is broader, as we consider not only directly exposed, but all workers, and study the effects of Chinese exclusion on several economic indicators – from labor supply to income and wages to productivity – and across sectors.

little to stop emigration in practice. Emigrants left mostly through the southern port of Guangzhou (Canton) and arrived to California.

To come to America, most Chinese immigrants were in one way or another dependent on the Six Companies, an organization of Chinese merchants in America (Spence, 1990, p. 205). In exchange for organization fees, the Six Companies would arrange for a number of services for Chinese immigrants, including temporary lodging, basic healthcare, and assurances that their remains would be sent back to China in the event of an untimely death. In addition, for those who did not have the money to make the voyage to America, which was around six times the average Chinese per capita income at the time, the Six Companies would loan them the money under a form of labor debt contract (Cloud and Galenson, 1987; Galenson, 1984).[6] It was common for families and villages to pool together their money to send one person to the United States, who would then use the saved earning to bring over other (Chang, 2003, p. 18). The organization of the emigration process led Chinese immigrants in the U.S. to having strong social networks, which likely contributed to their success in building businesses.

Since the main port of entry in the United States on the West Coast was San Francisco, most Chinese lived in California and gradually diffused to other nearby states. The Chinese made up around a quarter of all immigrants in California in 1880, followed by the Irish (22%) and the Germans (14%). Most immigrants from China were men. Many were young and single. Those who were married did not bring their spouses with them when they first arrived.

In 1880, about a quarter of the Chinese were employed in some sort of mining. Agriculture and laundering services were the next largest employers of Chinese people, accounting for

[6]The Six Companies had an agreement with steamship companies such that the companies would not sell a ship ticket to a Chinese person unless they could produce a certificate from the Six Companies stating that they had repaid their debt. As most Chinese immigrants during this time intended to return home after accumulating some wealth, this was usually a good enough incentive for people to not run away after coming to America (Cloud and Galenson, 1987)

another ten percent each. Although initially many Chinese came to the US to work on the construction of the First Transcontinental Railroad, its completion in 1869 meant that by 1880 the rail industry only accounted for about 4.5% of Chinese employment. Chinese immigrants comprised of both skilled and unskilled workers. They often – but not exclusively – worked in establishments owned and managed by other Chinese immigrants. Chinese manufacturers of shoes and hats, cigars, for example, dominated the sector in the Western U.S. during this period.

The demand for Chinese labor was very high from American employers. They were seen as a valuable and low cost source of skilled and unskilled labor by mining companies. Experience in railroad construction and mining gave Chinese men useful skills for other large engineering projects. For example, good at dynamiting and transporting large masses of materials, the Chinese built much of the roads along the north Pacific Coast in the 1870s and 1880s. Chinese workers were able to complete physically arduous and complex tasks such as the drainage of agricultural lands and the construction of other land-improvement infrastructure. These were projects that the U.S. government was previously unable to complete because of the lack of willing and able workers. They also worked in lumber mills and made up a significant portion of the labor force in salmon canneries (Pfaelzer, 2008, p. 140). Chinese businesses and workers were seen a key source of tax revenue for local governments, which had few sources of funds during this period. The Chinese were also strategically taxed higher than other workers (Kanazawa, 2005).

### 2.2.2 Anti-Chinese Sentiments and the Chinese Exclusion Act

Hostility towards the Chinese grew as more and more Chinese arrived and a widespread economic depression during the 1870s made jobs scarce (Pfaelzer, 2008). The Chinese were popularly perceived as unskilled or low skilled labor, and many were concerned that Chinese workers took employment opportunities away from and competed down the wages for other workers. Historians estimate that there were four workers per every job in 1871 in California, but Chinese workers were producing 50-75% of the state's boots and shoes; and in 1882, 50-75% of farm labor in some counties was Chinese (Chan, 1986, p. 51-78). Many of the concerns focused on the welfare of white native workers, though hostility was also widespread among European immigrants (Chang, 2003, pp. 116-7).

Economic concerns were accompanied by xenophobia. Many worried about the influence of Chinese immigrants on American culture. The Chinese were typically not Christian, spoke little English, dressed in traditional Chinese robes, and wore their hair in the traditional Manchu queue as mandated by the Qing dynasty. These stark differences led many Americans to believe that a so-called "Yellow Peril" was threatening western civilization.[7] There was a widespread belief among Americans that *all* Chinese women were prostitutes. This view was supported by the American establishment. For example, the American Medical Association conducted a study seeking to link Chinese women to higher rates of venereal disease. Despite finding no substantive evidence to support that hypothesis, the association's president still claimed that "... even boys eight and ten years old have been syphilized by these degraded wretches..." (Chang, 2003, p. 123).

The combination of fears about economic competition and xenophobic sentiments, exemplified by nativist groups such as the Know-Nothings (Higham, 2002), led Congress to pass the Chinese Exclusion Act in 1882. The Act barred all Chinese people from entering the United States, except under very special circumstances (e.g., official diplomats). In addition to the restrictions on new Chinese immigrants, an amendment to the Act in 1884 expanded its scope, banning people of Chinese descent from entering the country. A further amend-

[7] One early proponent of excluding the Chinese, Senator John F. Miller, in a speech to his fellow senators in 1881, called upon them to: "...[preserve] American Anglo-Saxon civilization without contamination or adulteration ... [from] the gangrene of oriental civilization... Why not discriminate? Why aid in the increase and distribution over ... our domain of a degraded and inferior race, and the progenitors of an inferior sort of men?" (Chang, 2003, p. 130)

ment in 1888 prevented immigrants arrived prior to the Act from re-entering the United States.

In practice, these legislative changes meant that no new Chinese could arrive and those who were already in the U.S. could never see their families again, unless they left the United States. Chinese remaining in the U.S. also faced increasing discrimination both through formal and informal channels. For example, the Act prevented Chinese immigrants from becoming naturalized citizens in the same way that the right had been offered to European immigrants, while local governments passed legislation that confiscated the property of the Chinese. There were also many instances of mob violence against the Chinese. These forces led many of the Chinese who stayed to live together in urban areas, where they could organize and better protect themselves. It was during this period that the first "China Town" appeared in San Francisco (in 1900).

The Chinese Exclusion Act was initially viewed as a temporary ten-year measure. It was renewed for ten more years in 1892 with the Geary Act, and then renewed indefinitely in 1902. During the early 20th century, growing anti-immigrant sentiments developed to the point where a more far-reaching immigration restriction was passed by Congress, which, in 1917, introduced a literacy requirement and barred Southeast Asians, South Asians, and Middle Eastern people (those from the so-called "Asiatic Barred Zone") from immigrating to the United States (Goldin, 1994). In 1921 and then, more permanently, in 1924, a new ban introduced a quota on immigration, and fully banned Asian immigrants (Abramitzky and Boustan, 2017). Only in 1943, when China became America's ally in World War II, Congress finally repealed the Exclusion Act. But even then, Chinese immigration was still limited to a mere 105 people a year. It was not until the Immigration and Nationality Act of 1965 that Chinese immigrants were allowed to move to the United States in large numbers again (Lee, 2003, Ch. 3).

## 2.3 Empirical Strategy

The Chinese Exclusion Act drastically reduced the number of Chinese living in the United States. This might have had positive or negative effects for other workers, depending on the characteristics of the excluded individuals and on the degree of complementarity (or substitutability) between immigrant labor and other workers in the economy. On the one hand, the mainstream perception (among native whites) at the time was that Chinese immigrants were mostly low-skilled labor and competed with other workers. If this was true, the Act should have increased economic opportunities for other workers, especially unskilled ones. On the other hand, the Act may have depleted the Western United States of much needed (skilled and unskilled) labor, inducing firms to shut down, and causing long term negative economic consequences across many sectors. In other words, if Chinese and other workers were complements in production, the Act may have hurt other workers.

Our study aims to capture the net effect of the positive and negative forces. We examine the population, labor force, and earnings of Chinese and other workers. We also consider several measures of aggregate economic performance across sectors. To estimate the impact of the Chinese Exclusion Act, we implement a *difference-in-differences* (DD) strategy, and compare outcomes in counties that had 1880 Chinese population shares above and below the sample median (4%) before and after the 1882 Exclusion Act. The empirical strategy assumes that the ban of Chinese immigrants results in a higher loss of Chinese workers – i.e., a higher intensity treatment effect – for counties with more Chinese immigrants prior to the ban. The baseline specification is the following:

$$Y_{ijt} = \alpha + \beta(HighChineseShare_{i,1880} \times 1\{t > 1882\}) + \Gamma X_{ijt} + \varphi_i + \xi_{jt} + \nu_{ijt} \qquad (2.1)$$

where the outcome of interest in county $i$ state $j$ and year $t$, $Y_{ijt}$, is a function of: the interaction of a dummy variable that takes the value of one if the 1880 Chinese population share is above the sample median, $Chinese_{i,1880}$, and an indicator variable equal to one if the time period is after 1882; a vector of controls, $X_{ijt}$; county fixed effects, $\varphi_i$; and state-year fixed effects, $\xi_{jt}$. Standard errors are clustered at the county level.[8]

Since the Census data is observed in each decade except for 1890, the pre-post comparison of outcomes observed in 1880 or earlier versus those observed in 1900 or later includes the effect of all the follow-up legislation that occurred between 1884 and 1900 discussed in the previous section.[9]

County fixed effects control for time invariant differences across counties, such as distance to the San Francisco port. State-year fixed effects control for changes over time that affect all counties within a state similarly. This addresses the fact that economic transformation might have differed across states in ways that may be correlated with both the 1880 Chinese share and the economic outcomes of interest.

Our preferred specification also controls for whether a county was connected to a railroad in a given year and whether there was ever a mine in the county during 1850–1940. Since the latter is a time invariant measure, we interact it with year fixed effects.[10] These controls address the potential concern that the first waves of Chinese immigrants worked in mining and railroad construction, and that the economic development of these sectors may have affected the outcomes of interest even absent the Act. Moreover, the presence of a railroad can affect long run economic development for many other reasons unrelated to the Act (Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021).

The main coefficient of interest is $\beta$. The identification assumption is that, absent the

[8]To address the fact that county boundaries changed over time, we follow standard approaches in the literature (Perlman, 2016), fixing them to 1930.

[9]The 1890 U.S. Census was destroyed by a fire. As noted below, though, we were able to recover a handful of outcomes (e.g., total population) for this year using different sources.

[10]We were unable to find systematic disaggregated data on the presence of mines that varied over time.

Act, the outcomes of interest would have evolved along parallel trends between counties with high and low 1880 Chinese population shares. In other words, we assume that conditional on fixed effects and controls, the interaction of 1880 Chinese population share in the county and the post-1882 dummy variables is uncorrelated with the error term. We will provide evidence to support this assumption after presenting our results. Below, we also show that our findings are robust to interacting additional 1880 variables with year fixed effects, to further relax the parallel trends assumption.

## 2.4 Data

Most of our data come from the U.S. decennial censuses for the period from 1860 to 1940, made available by the Integrated Public Use Microdata Series (Ruggles et al., 2021). In addition, we use county-aggregates from the Census of Manufacturing and of Agriculture (Haines, 2010a; Haines and Rhode, 2018).

The historical data reports each individual's nativity (including that of the parents), country of origin, and race. We define someone to be Chinese if their country of birth is China or if their race is Chinese. Given that Chinese immigrants started arriving in the 1850s, race and country of origin are synonymous for most Chinese adults in the U.S. in 1880. In later censuses, it is possible that U.S. born children from a parent who is Chinese and a parent who is another race choose to report her race as the other race. We will address this by examining the dynamic effects and showing a sharp change in the outcomes immediately after the Act, when this is less likely to be an issue. Moreover, inter-marriage between Chinese and other races was very low during this period.[11] Finally, such classification problem does not affect the interpretation of aggregate outcomes (or those of native white workers).

Our main sample includes the states where the Chinese population is above 1% of the

[11] Over the time period 1870–1940, only 1.7% of married Chinese had a non-Chinese spouse.

total population in 1880: Arizona, California, Idaho, Montana, Nevada, Oregon, Washington, and Wyoming. When analyzing economic outcomes, such as labor supply and earnings, we further restrict attention to working age men (ages 15 to 64).[12]

We define $HighChineseShare$ to be a dummy equal to 1 if the 1880 share of Chinese individuals in county $i$ is above the sample median (4%).

Figure 2.1 plots the population of Chinese immigrants and non-Chinese immigrants in the United States by decade: prior to the Chinese Exclusion Act, both populations grew in a roughly linear fashion. After the Act, the non-Chinese population continued to grow in a roughly linear fashion, while the Chinese population reversed trend. Figure 2.2 maps the spatial distribution of Chinese in 1880 across the counties in our sample, with darker colors corresponding to a higher Chinese share. The map shows that there was significant variation across counties within states in the western part of the country.

Tables 2.1 and 2.2 present descriptive statistics for the main variables for all counties in our sample (Panel A), and for the subsample of counties with high (Panel B) and low (Panel C) 1880 Chinese population share. On average, only 2% of the total population of our sample is Chinese. A comparison of the means in Panels B and C shows that there is little difference in baseline characteristics between counties with a high and low 1880 Chinese population share.

## 2.5 Results

### 2.5.1 Population and Labor Supply

We begin by presenting results for Chinese immigrants in Table 2.3. Columns (1) and (3) of Panel A show that the Exclusion Act drastically reduced the size of the (log) Chinese

[12] Results are unchanged if we use the entire U.S. and/or if we include women.

population and labor force. The coefficients, which are statistically significant at the 1% level, are -1.51 and -1.58, respectively. This implies that, after the Act, a county with 1880 Chinese population share above the median had a Chinese population and labor force approximately 80% lower than a county with 1880 Chinese share below the median.[13] Column (2) suggests that the Act also reduced the share of urban Chinese population, although the point estimate is smaller than for total population, and standard errors are large. Next, in columns (4) to (7), we examine the effects of the Act on the size of the (log) Chinese labor force in each of the major sectors – manufacturing, mining, railroads, and agriculture. In all cases, coefficients are negative and, except for agriculture, they are statistically significant at the 1% or 5% level.

In Panel B, we examine the number of Chinese workers by skill level. The Act had a negative impact on average literacy (column 1), reduced the (log) number of both skilled and unskilled Chinese workers (columns 2 and 3), as well as that of Chinese managers and proprietors (column 4).[14] In column (5), we show that the Act lowered (log) occupational income scores of Chinese workers.[15] This suggests that the Act not only reduced the number of Chinese workers, but also pushed the Chinese who remained in the U.S. in lower paid occupations.

In Tables 2.4 and 2.5, we turn to white and overall population and labor force.[16] Results are strikingly similar to those for Chinese immigrants. Panel A shows that the Exclusion Act had a large, negative effect on population and labor force. This was true for all sectors,

[13] Given that the dependent variable is in log, the magnitude of the coefficients can be calculated as follows: $\%\Delta y = 100 \cdot (e^{\beta} - 1)$.

[14] Skill groups are defined based on individuals' reported occupation following Katz and Margo (2014). In particular, skilled workers include: professionals, managers, craftsmen, clerical and sales occupations. Unskilled occupations include: operatives, laborers, and service workers (both private household and non-household). These groups omit workers employed in agriculture.

[15] As noted above, the U.S. Census did not collect wages prior to 1940. We thus use occupational income scores, which assign to an individual the median income of his job category in 1950 and are often interpreted as a proxy for life-time income.

[16] Variables are defined in the same way as in Table 2.3.

including agriculture (where, instead, the Act had no effect for Chinese workers). As for Chinese immigrants, while the coefficient for share of urban population is negative, it is not statistically significant at conventional levels (column 2).[17] Moreover, as for Chinese immigrants, the Act reduced labor supply across all skill groups, lowered the number of managers and proprietors, and led to a decline in occupational income scores. Table B.2 presents analogous estimates for non-Chinese immigrant population.

Figures 2.3 and 2.4 examine the dynamic effects of the Chinese Exclusion Act, plotting coefficients (with corresponding 95% confidence intervals) from an equation similar to the baseline, where the 1880 Chinese share dummy is interacted with year fixed effects (rather than with the post-1882 dummy), and using 1880 as omitted category. In Figure 2.3, we consider population for Chinese, white natives, white non-natives and all individuals. Reassuringly, and supporting the parallel trends assumption, we find no evidence of pre-trends. Turning to the post-1880 period, we instead observe an immediate decline in the first decade after the Act.[18] Interestingly, the negative effects of the Act keep unfolding until 1920. Since then, population and labor supply remain well below their 1880 level through 1940, for both Chinese immigrants and the county as a whole. This suggests that the exclusion of Chinese immigrants had persistent effects on the economy of Western U.S. counties. Figure 2.4 presents similar trends for occupational income scores for the same four groups of workers.

Taken together, results in this section show that the Chinese Exclusion Act significantly reduced the number of workers from all races, all sectors and all skill levels. Moreover, the reduction in occupational income score implies that, on average, all workers were worse off. The reduction in the number of managers is consistent with an overall reduction of

[17] Note that the point estimates for total population and labor force are smaller (in absolute value) than those in Table 2.3. However, since total population and workforce are larger than those of the Chinese, the implied effect of the Act on population size and on overall labor supply was quantitatively larger for non-Chinese workers.

[18] As noted above, the first year after the Exclusion Act is 1900 for labor force, as the 1890 U.S. Census was destroyed in a fire. However, since data on population (total and by ethnic group) can be obtained also from Haines (2010a), for population regressions, the first post-Act year is 1890.

production (e.g., shutting down factories or factory lines) or a reorganization of production (e.g., reducing the number of managers per worker). We investigate this more in the next section.

The fact that labor force for all races declined in manufacturing, mining, and railroads is consistent with Chinese workers being complements to natives and workers of other races in production. In this sense, it is interesting to note that the Act had little effect on Chinese workers in agriculture, but nonetheless reduced the number workers from other races in the sector. There exist at least two, non-mutually exclusive, explanations for this. The first one is that the decline in total population reduced demand for food production from nearby areas. The second one is that the Chinese were critical in land improvement projects such as draining swamps, such that the Act reduced the amount of arable land. Data limitations prevent us from examining this directly. However, below we examine farm land value and other agricultural variables.

### 2.5.2 Aggregate Economic Outcomes

To shed more light on the drivers of the results estimated in Section 2.5.1, we begin by examining the impact of the Exclusion Act on the manufacturing sector, and present results in Panel A of Table 2.6.[19] Column (1) focuses on (the log of) average wages, which are reported at the county level and cannot be disaggregated by nativity or race. The negative coefficient indicates that the Act reduced average manufacturing wages. Column (2) documents that, in line with the decline in the number of workers we found earlier, (the log of) total manufacturing output declined as well. The estimates are statistically significant at the 1% level, and suggest that, after the Act, manufacturing wages and output were, respectively, 11% and 60% lower in a county with the 1880 Chinese population share above median, compared

[19] The number of observations differs from that in the main sample above because data from the Census of Manufacturing is not available for all counties and years.

to a county with the 1880 Chinese population share below the median.

Column (3) shows that, after the Act, counties with the 1880 Chinese share above median had 61% less establishments than counties with 1880 Chinese share below median. Since the 1880 average number of establishments per county was 35, this implies that, after the Act, counties with 1880 Chinese population share above the median had approximately 21 fewer establishments than counties with 1880 Chinese population share below the median. The estimate in column (4) indicates that there is no change in the number of workers per establishment, consistently with the results of column (3) and of Table 2.5, showing that labor supply fell in all sectors (including manufacturing). These results, together with our earlier findings on the reduction in workers of all sectors and skill levels, suggest that the Act and the subsequent exodus of Chinese workers led to the closure of factories.

Next, in column (5), we turn to the mining sector, which, as of 1880, employed approximately 24% and 12% of the Chinese and the total population in our sample. Since we do not have detailed data on mining output, we use an admittedly crude proxy for whether there is any mine in a county during each decade. This is a dummy variable that equals one if county $i$ in year $t$ has a share of labor force in mining above the sample median. The negative coefficient (statistically significant at the 1% level) suggests that the Act reduced the presence of mines across U.S. counties. Again, this resonates with historical accounts of mine owners expressing concerns that the loss of Chinese labor would force them to shutter their mines.

We have so far shown that the Chinese Exclusion Act had a negative effect on labor supply in all sectors, and reduced productivity and output in manufacturing as well as the probability that counties had active mines. In Panel B of Table 2.6, we turn to the agricultural sector. As shown above, while the Act had no effect on Chinese labor supply in agriculture, it reduced that of other workers in this sector (Tables 2.4 and 2.5). Consistent with the overall drop in labor supply in agriculture, column (1) documents that the Act lowered the (log of the) value of farm land. Columns (2) to (5) show that the Act reduced the (log) value of livestock, the (log of the) value of farm machinery, and (log) average expenditures on fertilizer. The estimates are statistically significant at the 1% level in columns (2) and (4), and at the 5% level in column (5). One interpretation for these results is that the Act lowered the demand for farm products, and led to a corresponding reduction in the value of farm inputs. Another possibility, not in contrast with the previous one, is that the Chinese Exclusion Act reduced the quality of farm land (e.g., from the loss of Chinese workers doing major land improvement), since other inputs are likely to complement land.

Figures 2.5 and 2.6 present the dynamic estimates. As in Section 2.5.1, we find no evidence of pre-trends and we observe a sudden decline in all outcomes from 1890 onwards.

### 2.5.3 Robustness Tests

An important caveat to the interpretation of our estimates is the possibility of geographic spillovers and spatial relocation. For instance, if the Act caused workers and economic activity to move from counties with a high Chinese share in 1880 to counties with a low Chinese share, our results might be confounded by such relocation effect. We address this concern by interacting the average 1880 Chinese share in adjacent counties with our main independent variable. The logic is that, since moving costs increase with distance, on average, workers and firms should be more likely to relocate to nearby counties. Thus, if our results capture relocation to other surrounding counties that have a low Chinese share, controlling for Chinese share in those areas should attenuate our negative findings. Reassuringly, Table 2.7 shows that the main interaction estimates remain unchanged.[20]

The dynamic estimates reported in Figures 2.3–2.6 support the parallel trends assumption, and assuage concerns that our results may be driven by spurious correlations. Never-

[20] Results, not shown for brevity, are similar when replacing the average Chinese share in adjacent counties with that calculated over other counties in the same state.

theless, one may still be concerned that the 1880 location of Chinese immigrants is correlated with other factors that might influence economic development. For this reason, in Table 2.8, we replicate our main analysis controlling for the 1880 share of non-Chinese immigrants interacted with year fixed effects (Panel I) and controlling for the 1880 population, and labor force in manufacturing and agriculture, interacted with year fixed effects (Panel II). Moreover, we drop counties with a Chinese population share above or below the 1st and the 99th percentiles (Panel III), or omit San Francisco county (Panel IV). Reassuringly, results are in line with those reported in Tables 2.3–2.6.[21]

In addition, in Table B.3 we replicate the main results by including also women in the sample, while in Tables B.4 and B.5 we estimate population-weighted regressions in Panel A, and use Conley (1999) adjusted standard errors to account for potential spatial correlation in Panel B. In all cases, results remain similar to the ones presented in Sections 2.5.1 and 2.5.2. In Table B.6, we also verify that results are unchanged when interacting year fixed effects with: *i)* a measure of county market access (Hornbeck and Rotemberg, 2021) measured in 1870; *ii)* distance to New York, the main port of entry for European immigrants; and, *iii)* a dummy equal to one if a county had ever received a homestead until 1880.

Finally, another concern is that our results may be due to pure chance, especially given the limited number of counties part of our main sample. We address this by randomly permuting the independent variable, $HighChineseShare_{i,1880}$, across counties 1,000 times. This allows us to simulate results from randomly assigning whether a county had a share of Chinese population in 1880 above or below the median, and compare the distribution of these estimates to the actual ones presented in Sections 2.5.1 and 2.5.2. Figure B.1 shows that our estimates are unlikely to be generated by chance.

[21]Results, not reported for brevity, are also robust to interacting year dummies with many other variables, such as 1880 population density, manufacturing output, farm value, and geographic coordinates.

### 2.5.4 Placebo Exercise

As noted in the Introduction, allowing counties to be on differential trends based on the size of their 1880 agricultural employment is particularly important because, between 1880 and 1920, the U.S. economy experienced structural transformation, which led to stronger wage and employment growth in initially rural counties (Eckert and Peters, 2022a). Panel II of Table 2.8 reduces concerns that our estimates pick up the spurious correlation between agricultural employment and Chinese settlements in 1880. To more directly tackle this concern, we perform a placebo exercise, focusing on non-Western counties.

Specifically, we first select the best predictors of the Chinese immigrant share in our Western counties using a LASSO procedure.[22] Then, we use these variables to predict the 1880 Chinese immigrant share in non-Western counties, where the actual Chinese population was virtually zero. Finally, we replicate our baseline specification on this sample. Results, reported in Table 2.9, indicate that non-Western counties with a high (predicted) Chinese share in 1880 did not experience a decline in labor force, manufacturing output and number of establishments, or value of farm land. If anything, the opposite is true, suggesting that counties outside of the West with a higher (predicted) Chinese share had higher, rather than lower, growth potentials. These patterns are inconsistent with the possibility that the 1880 Chinese immigrants share were negatively correlated with the baseline agricultural share, *and* that this correlation is responsible for the decline in economic activity estimated in our main analysis above.

[22]LASSO selects the following variables: proxy for mine, non-Chinese immigrant share, employment share in agriculture, mining, railroads, and manufacturing, and the interaction between distance from a major port (San Francisco for the West, New York City for the non-West sample) and a dummy indicating whether the county is connected to the railroad. The variables not selected are: distance from ports, total population, population density, rural population share, average occupational income score, share of literate individuals, manufacturing output, value of farm land, a dummy indicating whether the county is connected to the railroad, and the interaction between the latter and the proxy for having a mine.

## 2.6 Conclusion

The Chinese Exclusion Act of 1882 was introduced both to respond to xenophobic sentiments of the time and to protect the economic livelihoods of white and native workers from Chinese immigrants, who were thought to exert negative pressure on the wages of low skilled workers. However, our analysis shows that the Act failed to achieve its economic goals. Chinese workers were employed in occupations of all skill levels at the time of the Act. Their *en-mass* departure led to an across-the-board economic decline. Manufacturing establishments and mines closed, agricultural land and inputs decreased in values, wages declined, and the population and labor supply of all groups diminished.

Our findings support the notion, discussed by a recent literature, that the Exclusion Act was responsible for the retardation of economic growth in the U.S. West. They are also consistent with the growing body of empirical studies showing the value of immigrants to early 20th century economic growth in the United States (Sequeira et al., 2020b; Ager and Hansen, 2017; Moser and San, 2019) and documenting that immigration restrictions often failed to increase employment and wages among native workers (Abramitzky et al., 2023; Clemens et al., 2018).

# Tables

Table 2.1: Summary Statistics: Chinese Population and Labor Force

| | A. All Counties | | | B. 1880 Ch. Share >= p50 | | | C. 1880 Ch. Share < p50 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. |
| **A. Chinese Population 1880** | | | | | | | | | |
| Chinese Population Share | 231 | 0.06 | 0.08 | 115 | 0.12 | 0.09 | 116 | 0.01 | 0.01 |
| using race | 231 | 0.06 | 0.08 | 115 | 0.12 | 0.09 | 116 | 0.01 | 0.01 |
| using country of origin | 231 | 0.06 | 0.08 | 115 | 0.11 | 0.09 | 116 | 0.01 | 0.01 |
| Age | 209 | 30.94 | 4.25 | 115 | 32.27 | 3.55 | 94 | 29.30 | 4.47 |
| Male Share | 209 | 0.95 | 0.07 | 115 | 0.96 | 0.06 | 94 | 0.95 | 0.07 |
| Chinese/All Immigrant | 231 | 0.21 | 0.18 | 115 | 0.35 | 0.16 | 116 | 0.08 | 0.07 |
| **B. Chinese Labor Force 1880 (Men 15-64)** | | | | | | | | | |
| Chinese/All LF | 231 | 0.12 | 0.12 | 115 | 0.21 | 0.11 | 116 | 0.03 | 0.03 |
| Chinese/All Mfg | 224 | 0.06 | 0.12 | 115 | 0.09 | 0.15 | 109 | 0.02 | 0.05 |
| Chinese/All Mining | 201 | 0.22 | 0.29 | 112 | 0.32 | 0.31 | 89 | 0.10 | 0.19 |
| Chinese/All Railroad | 173 | 0.21 | 0.30 | 98 | 0.26 | 0.33 | 75 | 0.15 | 0.25 |
| Chinese/All Agric | 231 | 0.02 | 0.04 | 115 | 0.04 | 0.06 | 116 | 0.01 | 0.01 |
| Chinese/All Skilled | 231 | 0.03 | 0.05 | 115 | 0.05 | 0.07 | 116 | 0.00 | 0.01 |
| Chinese/All Unskilled | 231 | 0.23 | 0.19 | 115 | 0.38 | 0.16 | 116 | 0.09 | 0.08 |
| Chinese/All Managers | 231 | 0.05 | 0.08 | 115 | 0.09 | 0.10 | 116 | 0.01 | 0.02 |
| Chinese/All Literate | 231 | 0.10 | 0.10 | 115 | 0.17 | 0.10 | 116 | 0.02 | 0.03 |

*Notes:* Observations are at the county and year level. The data are from U.S. Census of 1880.

Table 2.2: Summary Statistics: Population, Labor Force, and Economic Outcomes

| | **A. All Counties** | | | **B. 1880 Chinese Share >= p50** | | | **C. 1880 Chinese Share < p50** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. |
| **A. Population** | | | | | | | | | |
| Total Population | 1,854 | 21,188.42 | 97,559.85 | 945 | 19,927.97 | 54,748.91 | 909 | 22,498.78 | 127,687.26 |
| Urban Share | 1,854 | 0.16 | 0.24 | 945 | 0.18 | 0.26 | 909 | 0.14 | 0.21 |
| Immigrant Share | 1,852 | 0.20 | 0.11 | 943 | 0.23 | 0.12 | 909 | 0.17 | 0.09 |
| Age | 1,641 | 28.05 | 3.33 | 839 | 28.80 | 3.28 | 802 | 27.26 | 3.20 |
| Male Share | 1,641 | 0.60 | 0.08 | 839 | 0.62 | 0.08 | 802 | 0.59 | 0.08 |
| White Share | 1,852 | 0.93 | 0.10 | 943 | 0.92 | 0.09 | 909 | 0.94 | 0.11 |
| Chinese Share | 1,854 | 0.02 | 0.05 | 945 | 0.04 | 0.06 | 909 | 0.01 | 0.01 |
| Other Races Share | 1,641 | 0.05 | 0.09 | 839 | 0.04 | 0.06 | 802 | 0.06 | 0.11 |
| **B. Labor Force (Men 15-64)** | | | | | | | | | |
| Total Labor Force | 1,641 | 7,579.19 | 32,164.88 | 839 | 7,352.50 | 19,638.70 | 802 | 7,816.32 | 41,408.17 |
| Mfg. Share of Labor Force | 1,641 | 0.10 | 0.10 | 839 | 0.11 | 0.10 | 802 | 0.09 | 0.10 |
| Mining Share of Labor Force | 1,641 | 0.10 | 0.15 | 839 | 0.13 | 0.17 | 802 | 0.06 | 0.10 |
| Railroad Share of Labor Force | 1,641 | 0.05 | 0.07 | 839 | 0.05 | 0.07 | 802 | 0.05 | 0.07 |
| Agric. Share of Labor Force | 1,641 | 0.37 | 0.20 | 839 | 0.32 | 0.19 | 802 | 0.42 | 0.19 |
| Share Skilled | 1,641 | 0.25 | 0.10 | 839 | 0.25 | 0.10 | 802 | 0.24 | 0.09 |
| Share Unskilled | 1,641 | 0.36 | 0.17 | 839 | 0.41 | 0.17 | 802 | 0.31 | 0.15 |
| Share Managers | 1,641 | 0.06 | 0.03 | 839 | 0.07 | 0.02 | 802 | 0.06 | 0.03 |
| Share Literate | 1,410 | 0.94 | 0.08 | 724 | 0.94 | 0.06 | 686 | 0.93 | 0.09 |
| **C. Productivity** | | | | | | | | | |
| Income Score | 1,641 | 20.72 | 2.75 | 839 | 21.37 | 2.68 | 802 | 20.05 | 2.67 |
| Mfg. Total Output | 1,476 | 148,893.51 | 948,468.07 | 757 | 153,196.02 | 651,739.47 | 719 | 144,363.62 | 1,183,537.93 |
| Value of Farm Land | 1,240 | 163,082.51 | 337,723.53 | 625 | 164,076.89 | 294,784.96 | 615 | 162,071.97 | 376,619.51 |
| Connected to Railroad | 1,611 | 0.65 | 0.48 | 833 | 0.66 | 0.47 | 778 | 0.63 | 0.48 |

*Notes:* Observations are at the county and year level. The data are from U.S. Censuses between 1860 and 1940.

Table 2.3: Effect on Chinese Individuals

| | Dependent Variable | | | | | | |
|---|---|---|---|---|---|---|---|
| | A. Population and Labor Force Participation | | | | | | |
| | Pop. Total (1) | Urban Share (2) | LF Total (3) | LF Mfg. (4) | LF Mine (5) | LF Rail (6) | LF Agric. (7) |
| *Dependent Variable Mean* | *3.432* | *0.257* | *3.037* | *0.669* | *0.804* | *0.445* | *1.232* |
| *– in 1880* | *4.282* | *0.0500* | *4.182* | *1.201* | *1.890* | *1.044* | *1.608* |
| Post x High Chinese Share | -1.51*** | -0.06 | -1.58*** | -0.31** | -1.45*** | -0.30*** | -0.13 |
| | (0.24) | (0.04) | (0.23) | (0.15) | (0.23) | (0.11) | (0.17) |
| Observations | 1,819 | 1,407 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 |
| | B. Worker Skill Level and Income | | | | | | |
| | Share Literate | Skilled | Unskilled | Managers | Income Score | | |
| *Dependent Variable Mean* | *0.779* | *1.438* | *2.754* | *1.203* | *3.002* | | |
| *– in 1880* | *0.717* | *1.482* | *4.094* | *1.144* | *2.973* | | |
| Post x High Chinese Share | -0.05* | -0.94*** | -1.55*** | -0.70*** | -0.13*** | | |
| | (0.03) | (0.20) | (0.22) | (0.19) | (0.03) | | |
| Observations | 1,213 | 1,611 | 1,611 | 1,611 | 1,368 | | |

*Notes:* Observations are at the county and year level. The dependent variables in Panel A are the log of total population (col. 1), the share of urban population (col. 2), the log of the total labor force (col. 3), or the log of the labor force in the sector stated in the column headings (col. 4 - col. 7). The dependent variables in Panel B are the share of literate (col. 3), the log of total number of workers in the skill category stated in the column headings (col. 4 - col. 6), or the log of the occupational income score (col. 7). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

## Table 2.4: Effect on White Individuals

| | Dependent Variable | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Labor Supply | | | | | | | | | |
| | Pop | Urban Share | Total | Mfg. | Mining | Railroad | Agric. | Share Literate | Skilled | Unskilled | Managers | Income Score |
| | A. White | | | | | | | | | | | |
| *Dependent Variable Mean* | *8.652* | *0.169* | *7.722* | *5.101* | *4.037* | *3.892* | *6.536* | *0.955* | *6.311* | *6.543* | *4.951* | *3.080* |
| *– in 1880* | *7.491* | *0.0498* | *6.649* | *3.928* | *3.323* | *1.821* | *5.510* | *0.933* | *5.019* | *5.531* | *3.674* | *3.028* |
| Post x High Chinese Share | -0.41*** | -0.02 | -0.54*** | -0.51** | -0.68*** | -0.67*** | -0.39*** | -0.02*** | -0.60*** | -0.71*** | -0.62*** | -0.04*** |
| | (0.14) | (0.03) | (0.16) | (0.21) | (0.23) | (0.20) | (0.14) | (0.01) | (0.17) | (0.17) | (0.17) | (0.01) |
| Observations | 1,817 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,611 |
| | B. White Natives | | | | | | | | | | | |
| *Dependent Variable Mean* | *8.436* | *0.169* | *7.386* | *4.761* | *3.634* | *3.525* | *6.225* | *0.976* | *6.049* | *6.120* | *4.654* | *3.092* |
| *– in 1880* | *7.222* | *0.0479* | *6.198* | *3.450* | *2.771* | *1.561* | *5.130* | *0.947* | *4.618* | *4.993* | *3.243* | *3.030* |
| Post x High Chinese Share | -0.37** | -0.02 | -0.47*** | -0.50** | -0.55** | -0.57*** | -0.32** | -0.02*** | -0.57*** | -0.62*** | -0.61*** | -0.04*** |
| | (0.14) | (0.03) | (0.16) | (0.20) | (0.22) | (0.20) | (0.14) | (0.01) | (0.17) | (0.17) | (0.17) | (0.01) |
| Observations | 1,817 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,610 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of total population (col. 1), the share of urban population (col. 2), the log of the total labor force (col. 3), the log of the labor force in the sector stated in the column headings (col. 4 - col. 7), the share of literate (col. 8), the log of total number of workers in the skill category stated in the column headings (col. 9 - col. 11), or the log of the occupational income score (col. 12). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.5: Effect on All Individuals

| | Dependent Variable | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Labor Supply | | | | | | | | | |
| | Pop (1) | Urban Share (2) | Total (3) | Mfg. (4) | Mining (5) | Railroad (6) | Agric. (7) | Share Literate (8) | Skilled (9) | Unskilled (10) | Managers (11) | Income Score (12) |
| | A. All | | | | | | | | | | | |
| *Dependent Variable Mean* | *8.731* | *0.164* | *7.802* | *5.139* | *4.127* | *3.991* | *6.602* | *0.935* | *6.333* | *6.679* | *4.987* | *3.071* |
| *– in 1880* | *7.581* | *0.0492* | *6.797* | *4.002* | *3.646* | *2.129* | *5.543* | *0.911* | *5.053* | *5.853* | *3.732* | *3.029* |
| Post x High Chinese Share | -0.49*** | -0.01 | -0.68*** | -0.55*** | -0.98*** | -0.84*** | -0.42*** | -0.01 | -0.64*** | -0.96*** | -0.67*** | -0.05*** |
| | (0.14) | (0.03) | (0.16) | (0.20) | (0.24) | (0.21) | (0.14) | (0.01) | (0.17) | (0.17) | (0.17) | (0.01) |
| Observations | 1,819 | 1,819 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,611 |
| | B. All Natives | | | | | | | | | | | |
| *Dependent Variable Mean* | *8.581* | *0.167* | *7.423* | *4.782* | *3.644* | *3.539* | *6.278* | *0.957* | *6.057* | *6.167* | *4.661* | *3.084* |
| *– in 1880* | *7.243* | *0.0482* | *6.209* | *3.454* | *2.772* | *1.563* | *5.138* | *0.944* | *4.621* | *5.013* | *3.245* | *3.028* |
| Post x High Chinese Share | -0.39*** | -0.02 | -0.49*** | -0.51** | -0.55** | -0.57*** | -0.34** | -0.03* | -0.58*** | -0.63*** | -0.61*** | -0.04*** |
| | (0.14) | (0.03) | (0.15) | (0.20) | (0.22) | (0.20) | (0.13) | (0.01) | (0.17) | (0.16) | (0.17) | (0.01) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,610 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of total population (col. 1), the share of urban population (col. 2), the log of the total labor force (col. 3), the log of the labor force in the sector stated in the column headings (col. 4 - col. 7), the share of literate (col. 8), the log of total number of workers in the skill category stated in the column headings (col. 9 - col. 11), or the log of the occupational income score (col. 12). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.6: Effect on Manufacturing, Mining, and Agriculture

| | Dependent Variable | | | | |
|---|---|---|---|---|---|
| | A. Manufacturing | | | | |
| | Wage<br>(1) | Total Output<br>(2) | # Establ.<br>(3) | Workers/ Establ.<br>(4) | Proxy for Mine<br>(5) |
| *Dependent Variable Mean* | *2.824* | *8.942* | *72.11* | *2.063* | *0.676* |
| *– in 1880* | *2.477* | *6.929* | *34.50* | *1.370* | *0.722* |
| Post x High Chinese Share | -0.12*** | -0.91*** | -0.93* | -0.02 | -1.43*** |
| | (0.04) | (0.27) | (0.52) | (0.10) | (0.54) |
| Observations | 1,411 | 1,451 | 1,514 | 1,419 | 695 |
| | B. Agriculture | | | | |
| | Farm Land Value | Livestock Value | # Horses | Machinery Value | Fertilizer Expenditure |
| *Dependent Variable Mean* | *10.53* | *9.577* | *5069* | *8.081* | *2.423* |
| *– in 1880* | *8.644* | *8.057* | *2165* | *6.073* | *0.967* |
| Post x High Chinese Share | -0.35** | -0.52*** | -0.06 | -0.51*** | -0.55** |
| | (0.14) | (0.14) | (0.11) | (0.14) | (0.25) |
| Observations | 1,214 | 2,036 | 1,584 | 2,036 | 1,557 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of total population (col. 1), the share of urban population (col. 2), the log of the total labor force (col. 3), the log of the labor force in the sector stated in the column headings (col. 4 - col. 7), the share of literate (col. 8), the log of total number of workers in the skill category stated in the column headings (col. 9 - col. 11), or the log of the occupational income score (col. 12). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.7: Spillovers to Adjacent Counties

| | Dependent Variable | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) Avg. Wage | (2) Total Output | (3) #Mfg Est. | (4) Value Farm Land | (5) Value Livestock | (6) Value Machinery | (7) Avg. Exp. Fertilizers |
| *Dependent Variable Mean* | *2.823* | *8.915* | *59.75* | *10.54* | 9.594 | 8.090 | 2.415 |
| *– in 1880* | *2.475* | *6.896* | *21.54* | *8.636* | 8.057 | 6.072 | 0.960 |
| Post x High Chinese Share | -0.10* | -0.89** | -0.90 | -0.33** | -0.50*** | -0.46*** | -0.32 |
| | (0.05) | (0.35) | (0.56) | (0.15) | (0.15) | (0.16) | (0.27) |
| Post x HCS in Border Counties | -0.06 | -0.02 | -0.12 | -0.06 | -0.02 | -0.10 | -0.48 |
| | (0.06) | (0.36) | (0.35) | (0.17) | (0.17) | (0.19) | (0.32) |
| Observations | 1,397 | 1,436 | 1,499 | 1,203 | 2,018 | 2,018 | 1,543 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of average manufacturing wage (col. 1), the log of the total manufacturing output (col. 2), the log of the number of manufacturing establishments (col. 3, Poisson regression), the log of the value of farm land (col. 4), the log of the value of livestock (col. 5), the log of the value of farming machinery and equipment (col. 6), or the average expenditure for fertilizers (col. 7). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data for the dependent variables in columns (1)-(3) are from the Historical, Demographic, Economic, and Social Data (ICPSR 2896), for the years 1860-1940; the data for the dependent variables in columns (4)-(8) are from the United States Agriculture Data (ICPSR 35206), for the years 1860–1940. Monetary amounts are expressed in thousands of 2020 U.S. dollars (deflated using the Minneapolis Fed 1800–2020 CPI). Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.8: Robustness Checks: Control for 1880 Variables

| | I. Control for Year FE x Other Immigrant Share 1880 | | | II. Control for Year FE x: Population 1880, Mfg LF 1880, Agric LF 1880 | | | III. Omit Top 1% Chinese Share | | | IV. Omit San Francisco County | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Post x High Chinese Share | | Obs. | Post x High Chinese Share | | Obs. | Post x High Chinese Share | | Obs. | Post x High Chinese Share | | Obs. |
| Dependent Variable | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | |
| **A. Chinese LF** | | | | | | | | | | | | |
| (1) Total | -1.66*** | (0.24) | 1,611 | -1.53*** | (0.23) | 1,611 | -1.56*** | (0.23) | 1,597 | -1.60*** | (0.23) | 1,603 |
| (2) Mfg. | -0.38*** | (0.14) | 1,611 | -0.30** | (0.15) | 1,611 | -0.32** | (0.15) | 1,597 | -0.32** | (0.15) | 1,603 |
| (3) Mining | -1.59*** | (0.23) | 1,611 | -1.40*** | (0.23) | 1,611 | -1.44*** | (0.23) | 1,597 | -1.46*** | (0.23) | 1,603 |
| (4) Rail | -0.31*** | (0.11) | 1,611 | -0.31*** | (0.11) | 1,611 | -0.30*** | (0.11) | 1,597 | -0.30*** | (0.11) | 1,603 |
| (5) Income | -0.14*** | (0.03) | 1,367 | -0.14*** | (0.03) | 1,367 | -0.13*** | (0.03) | 1,356 | -0.13*** | (0.03) | 1,359 |
| **B. All LF** | | | | | | | | | | | | |
| (6) Total | -0.51*** | (0.17) | 1,611 | -0.42*** | (0.16) | 1,611 | -0.55*** | (0.16) | 1,597 | -0.55*** | (0.16) | 1,603 |
| (7) Mfg. | -0.46** | (0.23) | 1,611 | -0.38* | (0.21) | 1,611 | -0.50** | (0.21) | 1,597 | -0.51** | (0.21) | 1,603 |
| (8) Mining | -0.69*** | (0.25) | 1,611 | -0.65*** | (0.23) | 1,611 | -0.70*** | (0.23) | 1,597 | -0.68*** | (0.23) | 1,603 |
| (9) Rail | -0.53** | (0.22) | 1,611 | -0.54** | (0.21) | 1,611 | -0.68*** | (0.20) | 1,597 | -0.66*** | (0.20) | 1,603 |
| (10) Income | -0.04*** | (0.02) | 1,611 | -0.04*** | (0.01) | 1,611 | -0.04*** | (0.01) | 1,597 | -0.04*** | (0.01) | 1,603 |
| **C. Manufacturing** | | | | | | | | | | | | |
| (11) Wage | -0.13*** | (0.04) | 1,411 | -0.12*** | (0.04) | 1,411 | -0.12*** | (0.04) | 1,400 | -0.12*** | (0.04) | 1,403 |
| (12) Total Output | -0.92*** | (0.29) | 1,451 | -0.77*** | (0.27) | 1,451 | -0.90*** | (0.27) | 1,440 | -0.91*** | (0.27) | 1,443 |
| (13) # Establ. | -0.58*** | (0.17) | 1,514 | -0.53*** | (0.17) | 1,514 | -0.58*** | (0.16) | 1,502 | -0.58*** | (0.16) | 1,506 |
| **D. Agriculture** | | | | | | | | | | | | |
| (14) Farm Land Value | -0.26* | (0.14) | 1,214 | -0.12 | (0.13) | 1,214 | -0.35** | (0.14) | 1,204 | -0.34** | (0.14) | 1,208 |
| (15) Livestock Value | -0.51*** | (0.14) | 2,036 | -0.45*** | (0.13) | 2,036 | -0.53*** | (0.14) | 2,019 | -0.50*** | (0.14) | 2,026 |
| (16) Machinery Value | -0.46*** | (0.15) | 2,036 | -0.39*** | (0.14) | 2,036 | -0.51*** | (0.15) | 2,019 | -0.50*** | (0.14) | 2,026 |
| (17) Fertilizer Expenditure | -0.50** | (0.25) | 1,557 | -0.48* | (0.25) | 1,557 | -0.55** | (0.25) | 1,544 | -0.55** | (0.25) | 1,550 |

*Notes:* Observations are at the county and year level. The dependent variables in Panel A are the log of the total Chinese labor force (row 1), the log of the Chinese labor force in the sector stated (rows 2-4), or the log of the average occupational income score among Chinese individuals (row 5). The dependent variables in Panel B are the log of the total labor force of all individuals (row 6), the log of the labor force in the sector stated of all individuals (rows 7-9), or the log of the average occupational income score among all individuals (row 10). The dependent variables in Panel C are the log of average manufacturing wage (row 11), the log of the total manufacturing output (row 12), or the number of manufacturing establishments (row 13, Poisson regression). The dependent variables in Panel D are the log of the value of farm land (row 14, the log of the value of livestock (row 15), the log of the value of farming machinery and equipment (row 16), or the average expenditure for fertilizers (row 17). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data for the dependent variables in Panel A and B are from the full count U.S. Censuses, for the years 1860-1940. The data for the dependent variables in Panel C are from the Historical, Demographic, Economic, and Social Data (ICPSR 2896), for the years 1860-1940; the data for the dependent variables in Panel D are from the United States Agriculture Data (ICPSR 35206), for the years 1860–1940. Monetary amounts are expressed in thousands of 2020 U.S. dollars (deflated using the Minneapolis Fed 1800–2020 CPI). Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

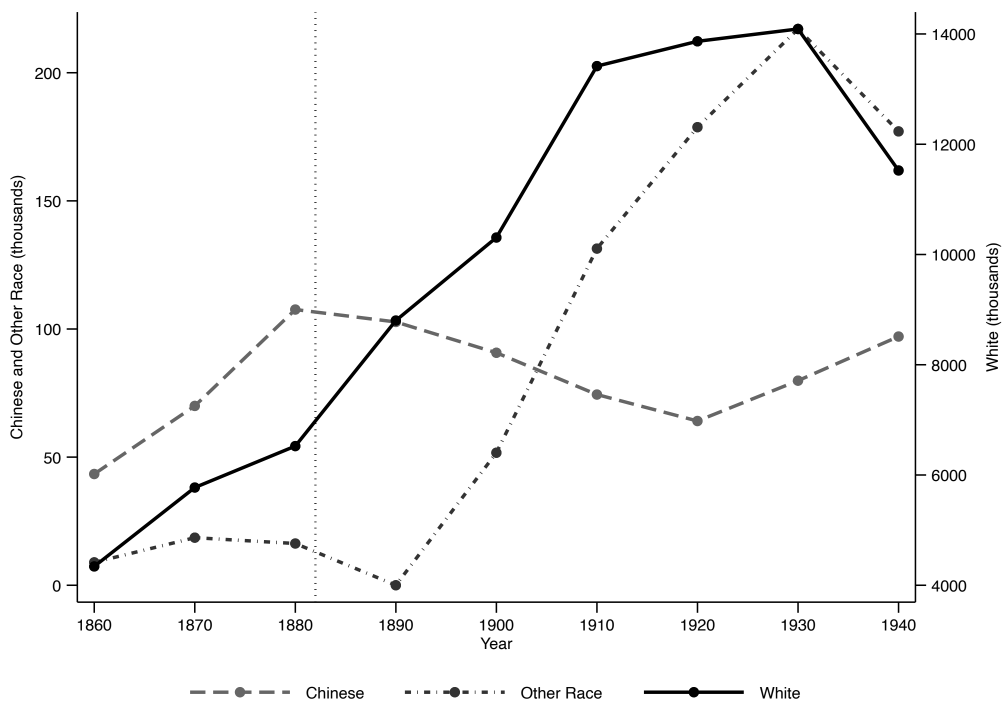
Table 2.9: Placebo Exercise

| | Dependent Variable | | | | | |
|---|---|---|---|---|---|---|
| | Chinese LF (1) | All LF (2) | Non-Chinese Immigrants LF (3) | Mfg. Output (4) | #Mfg Est. (5) | Farm Land Value (6) |
| | **A. Western States (Main Sample)** | | | | | |
| *Dependent Variable Mean* | *3.037* | *7.802* | *6.335* | *8.942* | *72.11* | 10.53 |
| *– in 1880* | *4.182* | *6.797* | *5.554* | *6.929* | *34.50* | 8.644 |
| Post x High Predicted Chinese Share | -1.55*** | -0.64*** | -0.75*** | -1.03*** | -0.75* | -0.29* |
| | (0.22) | (0.14) | (0.16) | (0.26) | (0.43) | (0.15) |
| Observations | 1,611 | 1,611 | 1,611 | 1,451 | 1,514 | 1,214 |
| | **B. All Other States (Placebo Sample)** | | | | | |
| *Dependent Variable Mean* | *0.387* | *8.172* | *4.780* | *9.198* | *108.8* | 11.12 |
| *– in 1880* | *0.215* | *7.761* | *4.882* | *7.832* | *94.86* | 10.21 |
| Post x High Predicted Chinese Share | 0.62*** | 0.25*** | 0.23*** | 0.15 | 0.69*** | 0.08 |
| | (0.07) | (0.05) | (0.08) | (0.10) | (0.07) | (0.08) |
| Observations | 20,054 | 20,054 | 20,054 | 18,454 | 19,078 | 14,992 |

*Notes:* Observations are at the county and year level. The dependent are the log of the total Chinese labor force (col. 1), the log of the total labor force of all individuals (col. 2), the log of the total non-Chinese immigrants labor force (col. 3), the log of the total manufacturing output (col. 4), the number of manufacturing establishments (col. 5, Poisson regression), or the log of the value of farm land (col. 6). The sample in Panel A includes the counties in the main estimation sample; the sample in Panel B includes all the other counties. All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data for the dependent variables are from the full count U.S. Censuses (cols. 1-3), from the Historical, Demographic, Economic, and Social Data (ICPSR 2896) (cols. 4-5), and from the United States Agriculture Data (ICPSR 35206) (col. 6), for the years 1860–1940. Monetary amounts are expressed in thousands of 2020 U.S. dollars (deflated using the Minneapolis Fed 1800–2020 CPI). Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

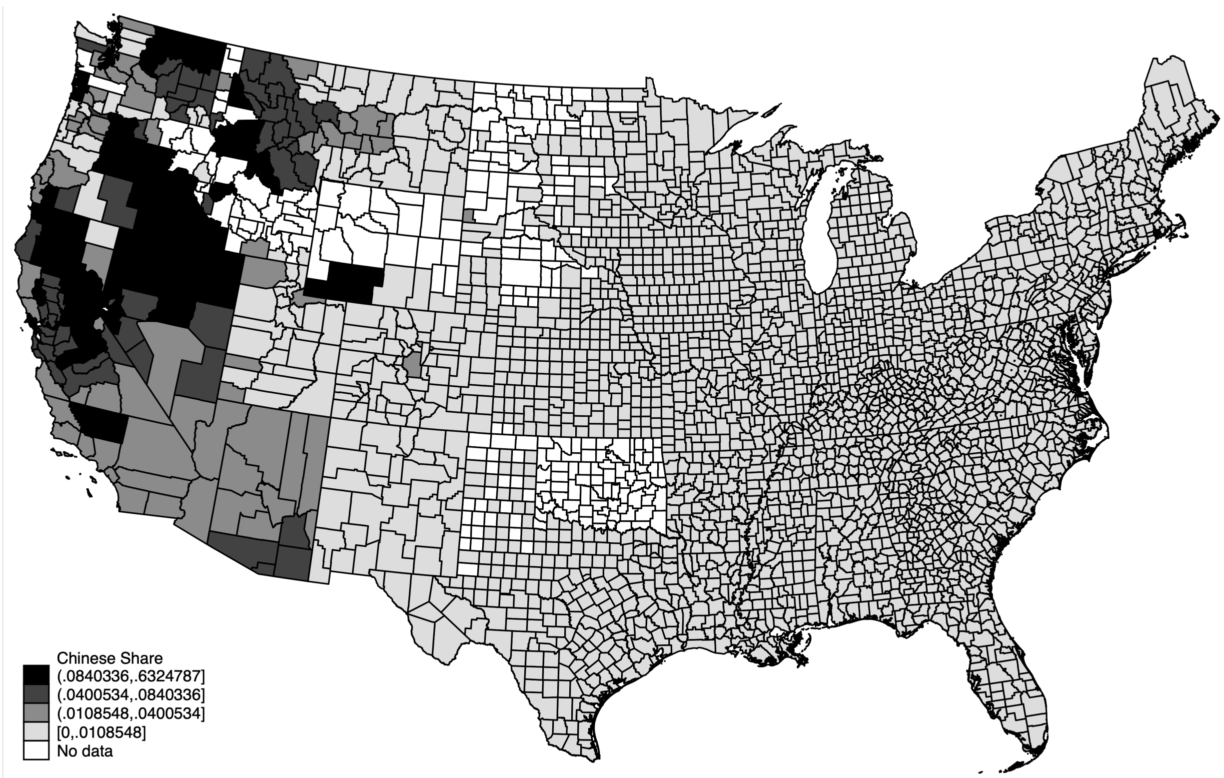
# Figures

Figure 2.1: Evolution of Immigrant Population



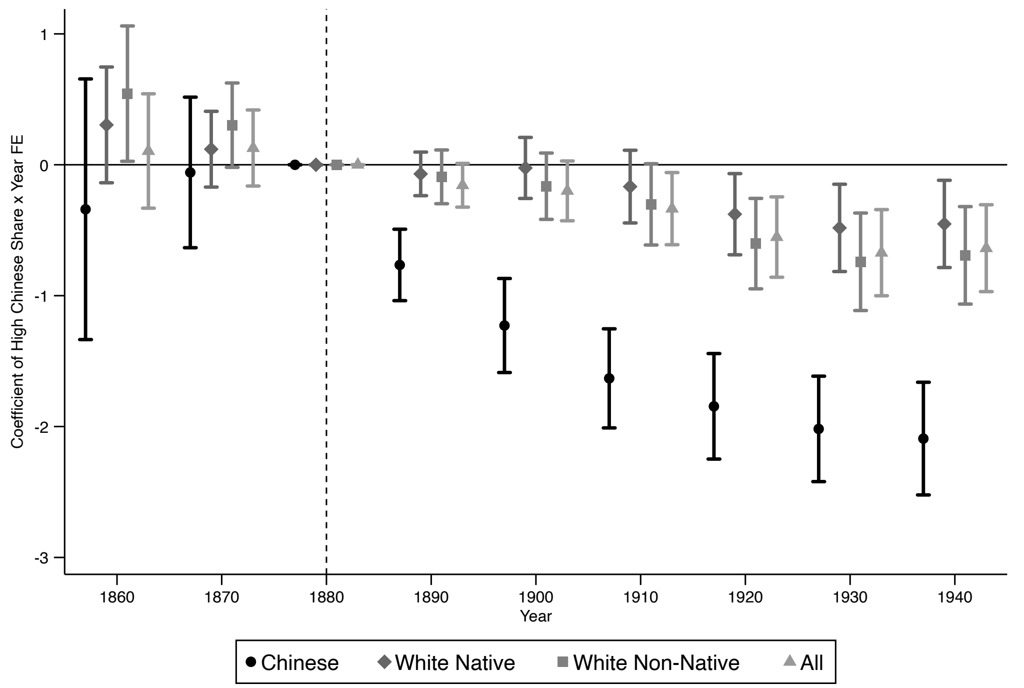

*Notes:* The figure represents the stock of foreign-born individuals in each census year, by race, in the United States. The data are from the full count U.S. Census between 1860 and 1940.

Figure 2.2: Spatial Distribution of Chinese in 1880



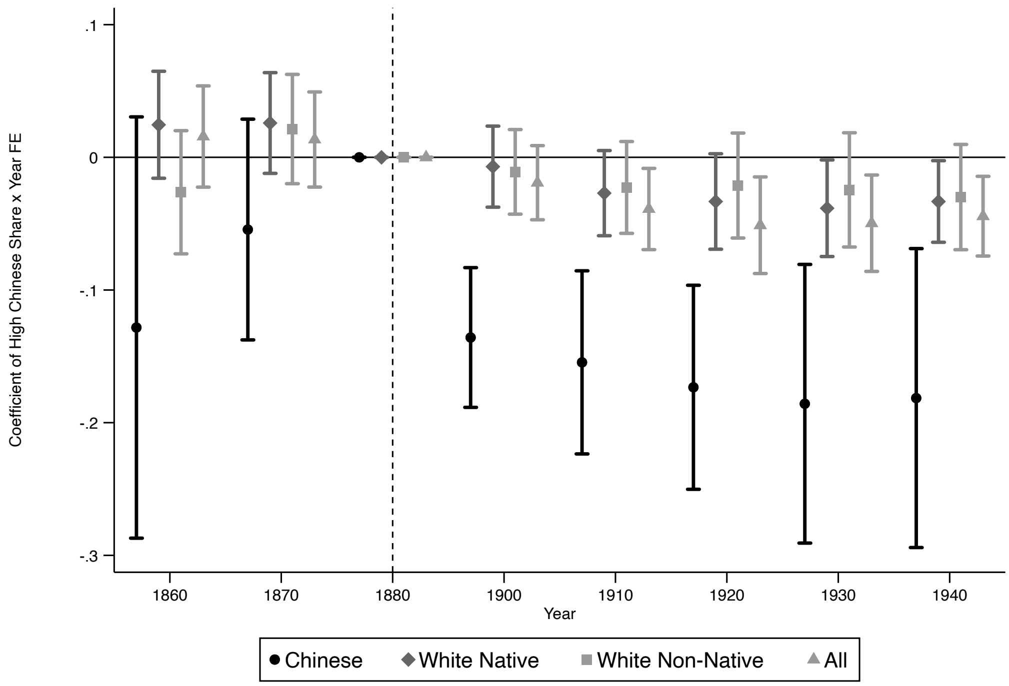

*Notes:* The map represents the 1880 share of Chinese population across U.S. counties. Different colors represent the quartiles of the distribution of Chinese share in the main estimation sample (as described in Section XXX). Lighter colors indicate lower shares, darker colors indicate higher shares.

Figure 2.3: Dynamic Effect on Population



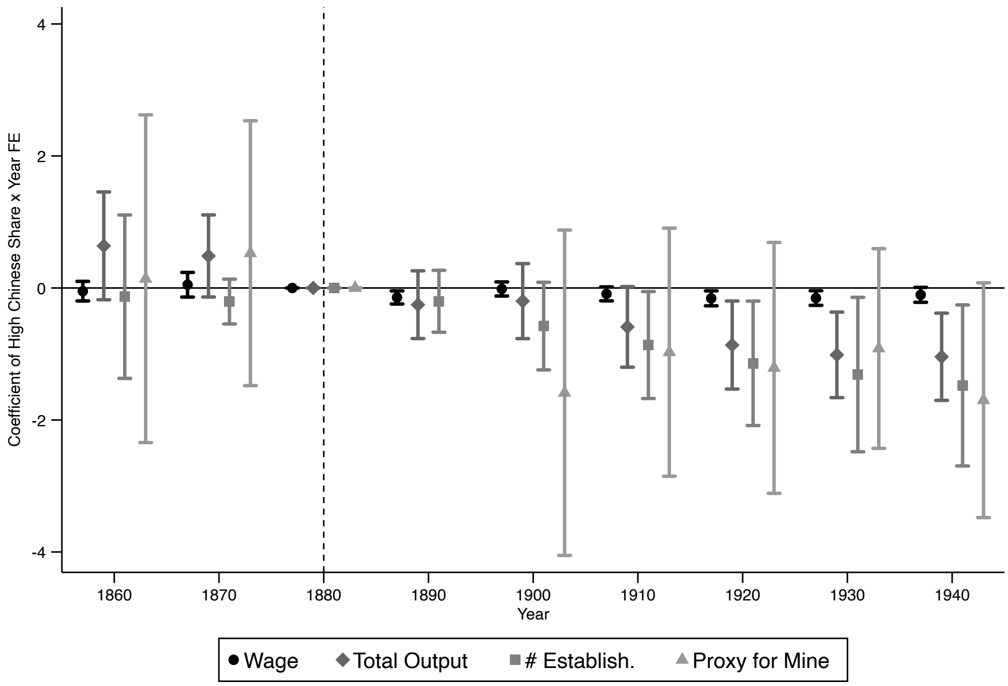

*Notes:* Observations are at the county and decade level. The dependent variable is the log of population The independent variables are the 1880 Chinese share interacted with a vector of time dummy variables. Vertical lines are 95% confidence intervals based on standard errors clustered at the county level. The regression controls for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1870-1940 interacted with decade fixed effects, and county and state-by-decade fixed effects. The data are from the full count U.S. Population Census between 1860 and 1940 (except for the year 1890, where only county-aggregate measures are available).

Figure 2.4: Dynamic Effect on Occupational Income Score



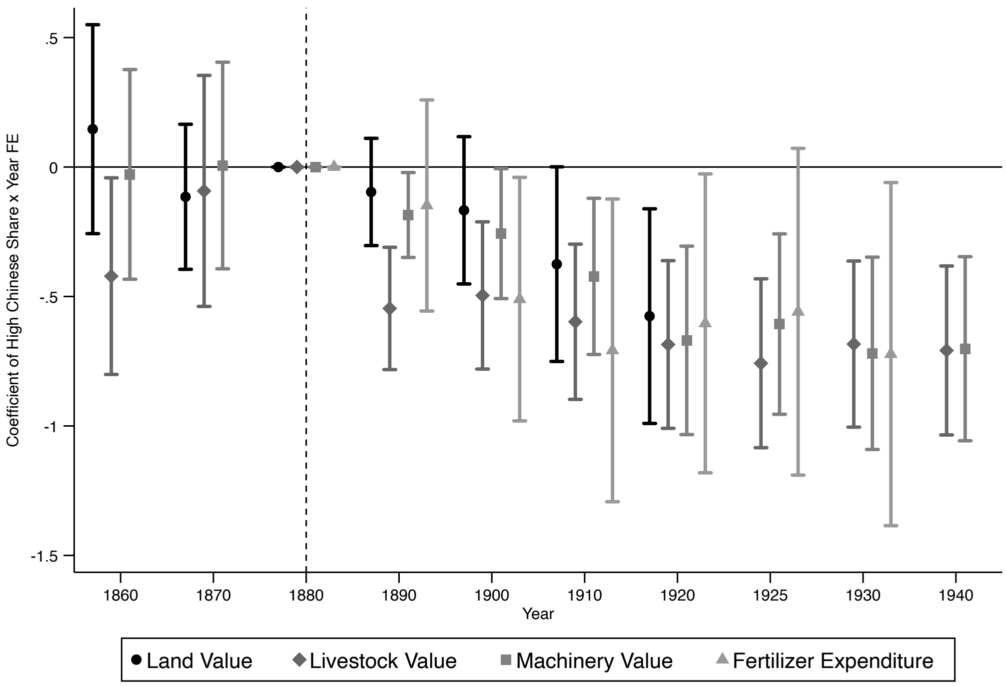

*Notes:* Observations are at the county and decade level. The dependent variable is the log of population The independent variables are the 1880 Chinese share interacted with a vector of time dummy variables. Vertical lines are 95% confidence intervals based on standard errors clustered at the county level. The regression controls for a dummy variable that equals 1 if the county is connected to a railroad in year $t$, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with decade fixed effects, and county and state-by-decade fixed effects. The data are from the full count U.S. Population Census between 1860 and 1940.

Figure 2.5: Dynamic Effect on Manufacturing and Mining




*Notes:* Observations are at the county and decade level. The dependent variables are the log of occupational income score. The independent variables are the 1880 Chinese share interacted with a vector of time dummy variables. Vertical lines are 95% confidence intervals based on standard errors clustered at the county level. The regression controls for a dummy variable that equals 1 if the county is connected to a railroad in year $t$, a dummy variable that equals 1 if the county ever had a mine during 1870-1940 interacted with decade fixed effects, and county and state-by-decade fixed effects. The data are from the full count U.S. Population Census and from the Census of Manufacturing between 1860 and 1940. Missing values for county $i$ at time t are linearly interpolated if data for county $i$ are available for both $t-1$ and $t+1$.

Figure 2.6: Dynamic Effect on Agriculture




*Notes:* Observations are at the county and decade level. The dependent variables are the log of occupational income score. The independent variables are the 1880 Chinese share interacted with a vector of time dummy variables. Vertical lines are 95% confidence intervals based on standard errors clustered at the county level. The regression controls for a dummy variable that equals 1 if the county is connected to a railroad in year $t$, a dummy variable that equals 1 if the county ever had a mine during 1870-1940 interacted with decade fixed effects, and county and state-by-decade fixed effects. The data are from the full count U.S. Population Census and from the Census of Agriculture between 1860 and 1940. Missing values for county $i$ at time $t$ are linearly interpolated if data for county $i$ are available for both $t-1$ and $t+1$.

# Chapter 3

## Political Connections, Careers, and Performance in the Civil Service: Evidence from U.S. Federal Judges

(co-authored with Massimo Pulejo)

## 3.1 Introduction

On May 6th, 2020, the Judiciary Committee of the United States Senate held a confirmation hearing for Judge Justin Walker, Senator Mitch McConnell's handpicked nominee for the appointment to the D.C. Circuit Court. This happened only six months after Judge Walker received judicial commission in the Western District Court of Kentucky, and was rated "Not qualified" by the *American Bar Association* for such role.[1] Judge Richard S. Arnold got his seat in the U.S. Court of Appeals for the Eighth Circuit in February of 1980, only fifteen months after being appointed to the Eastern District Court of Arkansas. On both occasions, he was strongly endorsed by the democratic Senator Dale Bumpers, for whom he had served as both a secretary and a legislative aid for six years.[2]

Within the U.S. Judiciary – and the U.S. Federal Civil Service, more in general – exceptionally rapid promotions are neither a recent nor a party-specific phenomenon (Domnarski, 2009), and raise critical questions. How important are politicians in shaping the careers of civil servants? What consequences do political connections have on the overall functioning and performance of the public sector? If political connections are important not only for the

[1] `https://www.nytimes.com/2020/05/04/us/politics/senate-confirmation-justin-walker.html` `https://www.nytimes.com/2020/06/04/us/judge-justin-walker-nomination-senate.html`.

[2] During the hearing for Judge Arnold's promotion, Senator Bumpers declared that "Richard's ability to understand and express complex issues precisely and succinctly is legendary".

initial appointment but also for the promotion of public sector employees, the disruption of such ties can have an *ex ante* ambiguous effect on their productivity. On the one hand, the loss of political connections may worsen the performance of public sector workers, if they anticipate their career prospects to be impaired by this event. On the other hand, losing a connection may induce public servants to exert more effort on the job, so as to compensate for the lack of political support. Thus, how political connections (or lack thereof) impact the productivity of public sector workers is ultimately an empirical question.

In this paper, we analyze these questions in the context of the civil service of the United States, focusing on the federal judiciary branch, where a highly institutionalized spoils system has been in place for over two centuries. Leveraging individual-level data on the careers and performance of federal judges and congressmen from 1789 to present, we provide an empirical assessment of the role and consequences of political connections between judges and politicians. More specifically, we exploit the exit of senators from Congress as a source of within-judge variation in connectedness. This allows us to establish a causal link between the tenure of senators and the careers and performance of the federal district court judges whose appointment they recommend.

Our difference-in-differences and event-study estimates reveal a strong impact of senators' tenure on judges' career prospects and productivity. Consistent with the mechanisms of federal judicial nominations, such impact emerges in years in which judges share partisanship with the incumbent president, and would thus stand to benefit from the lobbying efforts of their senatorial connection. The effects are concentrated on judges with a unique recommender at the beginning of their career, that is, those who simultaneously lose all ties to incumbent senators as their unique connection leaves Capitol Hill.

Following the exit of their recommender, judges experience a 48% drop in their yearly probability of promotion to the U.S. courts of appeals. In light of a baseline probability of promotion of 1.2% in a given year when the president is of the same political affiliation, the exit of recommenders from office essentially shuts the door to a judge's advancement in the U.S. federal judiciary. To explore the consequences of this event on judges' performance, we then investigate the effects of losing the connection on several productivity measures. We find that losing political connections causes judges to write fewer judicial opinions, of shorter length, and of poorer quality, as proxied by both fewer backward and forward citations. These results are consistent with judges reducing their effort and productivity once their career prospects are drastically hindered.

As we document by means of mediation analyses, these effects apply irrespective of the reason for which a recommender exits office, are homogeneous across judges of different quality, and do not significantly vary by partisan affiliation. In other words, patronage dynamics are likely to be key to the careers and the performance of a large number of district court judges.

To bolster the causality of our results, we implement several identification and robustness checks. First, using an event-study design, we document the absence of significant anticipation effects. Second, we tackle potential issues with standard two-way fixed effects estimators highlighted in recent work (Callaway and Sant'Anna, 2020; Imai and Kim, 2020; Sun and Abraham, 2020; De Chaisemartin and d'Haultfoeuille, 2020). After documenting that negative weights are only mildly affecting our baseline estimates, we show that results are largely similar when using the alternative estimator proposed by De Chaisemartin and d'Haultfoeuille (2020).

Finally, as far as the interpretation of our findings is concerned, we explore and rule out one relevant alternative explanation. Namely, we show that the negative effect of losing connections on judges' careers is virtually identical when a recommender is replaced by a senator of the same party. This confirms that personal connections between senators and judges are what shapes the career perspectives of the latter, as opposed to a generic ideological affinity with incumbent officeholders from their state.

This paper makes several key contributions. First, it adds to the expanding literature on patronage, by studying the effects of this appointment scheme in the context of the federal civil service of a major developed economy. Recent studies (e.g. Xu, 2018; Colonnelli et al., 2020; Gallo and Lewis, 2012) show how patronage appointments can be detrimental for the overall quality and performance of a wide range of organizations. On top of echoing these findings, our results go one step further, by showing that patrons making entry-level appointments can be crucial for determining upper-level nominations, as well. In fact, our analyses reveal that U.S. senators may have a crucial role in the appointment of Court of Appeals judges, even though the Constitution does not give them any formal prerogative in the nomination process. This links the scholarship on patronage with studies about the effect of promotion schemes on incentives and performance (Bertrand et al., 2020; Ke et al., 2018).

In this strand of scholarship, a contribution that is closely related to ours is Spenkuch et al. (2021), which documents the effects of partisanship on composition and turnover across several U.S. public agencies. Our study differs from theirs in two main respects. First, we focus on the judiciary, a sizable part of the U.S. state apparatus which is not part of the analyses in Spenkuch et al. (2021). Second, rather than looking only at the partisan affiliation of civil servants, our study is primarily interested in the personal connections between individual judges and specific senators.

Our findings also provide insights about the relative merits of different appointment procedures of U.S. high-level officials. A host of studies demonstrate how elected public officials may take suboptimal or unfair decisions due to electoral concerns. This has been repeatedly shown to be the case for elected, state-level judges (see e.g. Huber and Gordon, 2004; Besley and Payne, 2005; Berdejó and Yuchtman, 2013). By showing that appointed judges may also face dramatic changes in their incentives, our study casts doubts about the potential for lifetime nominations to solve issues stemming from electoral cycles.

Finally, this article augments our knowledge of the overall functioning of the U.S. Federal Judiciary, and the factors that concur to shaping judicial performance. In this respect, extant contributions have tended to focus on federal judicial bias stemming from judges' party affiliation (Sunstein et al., 2007; Cohen and Yang, 2019) or personal ideology (Schanzenbach and Tiller, 2008). Our study takes a different perspective, and looks at how – through affecting their career incentives – personal connections to specific political officers affect the performance of federal judges. This arguably advances our understanding of judicial behavior, by considering how incentives may change dynamically through the course of a judge's career.

The remainder of the paper is organized as follows. Section 2 gives background information on the U.S. federal court system, with particular regard to the role of home state senators in the process of nomination of district court judges. Section 3 details the sources and features of our data on federal judges and U.S. senators, as well as the procedure carried out to match them. Section 4 presents the empirical strategy adopted to identify the effects of interest. Section 5 illustrates our main results. Section 6 summarizes the upshots of a battery of sensitivity checks. Section 7 concludes and describes ongoing work that we are carrying out on the paper.

## 3.2 Background

Federal courts are in charge of dealing with both civil and criminal cases referred to the potential violation of one or more federal laws. The United States' federal court system consists of three layers: 94 district courts, 13 courts of appeals (also referred to as circuit courts), and the U.S. Supreme Court. Different from state-level judges, who are elected by citizens, federal judges are appointed for life by the President of the United States. However – while formally making the nominations – the president is far from being the only one involved in the process. This is particularly true for the entry-level position in the U.S. federal judiciary, the one of district court judge.

In fact, by a well-established custom, candidates for district court judgeships are put forward by home state senators who are from the same party as the president. Should there be no such senators, the president typically consults with other high-level officials from the state with whom he shares partisanship, such as House representatives (Rutkus, 2016). After vetting the candidate(s) identified by home state senators, the President refers one nominee to the Senate Judiciary committee, which holds a confirmation hearing involving a question and answer session with the candidate.

Following the hearing, the committee reports the candidate to the Senate floor in one of three ways: favorably, unfavorably, or without recommendation. In the overwhelming majority of cases, candidates are reported favorably, and in a relatively quick way.[3] The Senate is then in charge of the final confirmation, which normally takes place by unanimous consent. On top of the U.S. Senate, the only other institution having a say over proposed candidates is the *American Bar Association* (ABA, henceforth), which issues a non-binding evaluation before the nomination is passed on to the Judiciary committee.

Although not enshrined in the Constitution, the practice of accepting names for district judgeships from home state senators has been consistently applied throughout the years, by presidents from all parties. This lead to the association of district court judges with their senatorial recommenders rather than with their nominating president. As effectively summarized by U.S. Attorney General Robert F. Kennedy, "Basically it's senatorial appointment with the advice and consent of the president" (O'Brien, 1986).

Such a practice has not been immune from criticisms, on the grounds that it may favor politically connected candidates over more competent ones. As acknowledged by a U.S.

[3]However, longer confirmation times – and occasional rejections of candidates – have been taking place in more recent decades (see Binder and Maltzman, 2009).

Senator himself, it constitutes an "important source of political patronage" for U.S. senators (Tydings, 1977). Not surprisingly, factors concurring to the identification of candidates by senators include friendship, acquaintance, and family ties, among others (Domnarski, 2009). Furthermore, district judges are often chosen based on their political orientation, and a large majority of them were politically active before being appointed (Carp et al., 2019).

While home state senators are commonly regarded as determining only district court nominations, anecdotal evidence points to their active role in the appointment process of circuit court judges, as well (Domnarski, 2009). Notably, this qualitative evidence is largely corroborated by the official records of Congressional Hearings, which report strong written and oral endorsements of court of appeals nominees on behalf of one or more home state senator. This may imply that they suggest names for direct appointment to the circuit bench from outside the federal court system, or that they favor the promotion of judges that they first recommended for a district court position. The latter type of dynamic – and its implications for the performance of U.S. district court judges – are the object of interest of the present study.

## 3.3 Data

In order to study the impact of senators' tenure on the careers and performance of federal judges, we build a novel dataset combining information on both U.S. federal judges and senators throughout the period 1789-2019.

### 3.3.1 U.S. Federal Judges Data

Data on judges' careers come from the Biographical Directory of Article III Federal Judges compiled by the *Federal Judicial Center* (FJC), the research and education agency of the judicial branch of the United States Government. The directory includes the biographies of judges presidentially appointed to serve during good behavior since 1789 on the U.S. district courts, U.S. courts of appeals, Supreme Court of the United States, and U.S. Court of International Trade, as well as the former U.S. circuit courts, Court of Claims, U.S. Customs Court, and U.S. Court of Customs and Patent Appeals. The FJC data contain information on the full career of federal judges, with the specific dates of each appointment obtained throughout their tenure.

Data on judges' performance come from CourtListener, a free legal research website containing millions of legal opinions from federal (and state) courts, operated by the non-profit Free Law Project. Currently, CourtListener contains information on 9,032,122 legal opinions from federal, state, and specialty courts, from the 1920s until today.

### 3.3.2 U.S. Senators Data

Data on senators are from three sources: the Biographical Directory of the United States Congress,[4] the website voteview.com,[5] and the Roster of Members of the United States Congress compiled by ICPSR.[6] Combining these sources provides us with complete information on the political careers of all U.S. senators, from 1789 to 2019.

### 3.3.3 Matching of the Datasets

In the empirical analysis that follows we focus on the sample of federal judges who, over the 230 years of analysis, were ever appointed as district court judges.[7] We follow their career in the district courts until either their promotion, retirement, resignation, or death – whichever occurs first. In doing so, we also record if and when the senator(s) who recommended their

[4]https://bioguideretro.congress.gov.

[5]https://voteview.com/data.

[6]https://www.icpsr.umich.edu/web/ICPSR/studies/7803.

[7]The following categories are not included in our sample: (i) judges appointed in years in which that State did not have any representative in the Senate yet; (ii) judges in the district courts of DC and Puerto Rico.

nomination left office. To this end, we transform the FJC data into an unbalanced panel at the judge-year level.

In order to identify the senator(s) who recommended the nomination of each federal judge, we match this panel with the data on U.S. senators. In particular, we link each judge to the senator – or pair of senators – who, at the time of their nomination date as district court judge, were occupying the seat(s) corresponding to the state in which they was appointed, and who were of the same party as the nominating president.[8] Finally, given that our treatment of interest is the end of the connection between the judge and their recommending senator, we exclude from our analysis judges who are appointed in states where there is no senator of the same party as the incumbent president at the time of nomination.

The final sample consists of 42,715 judge-year observations, covering 2,155 judges for the time period 1789-2019.[9] Table 3.1 reports summary statistics for a set of judges' characteristics. Approximately 11% of the individuals in the sample get promoted from a district to an appellate court, after an average of 10 years from the first appointment. Figure 3.1 displays the number of promotions in each year, which ranges from a minimum of 0 to a maximum of 7. Approximately half of the judges are appointed by a Democratic president, and half by a Republican one.

[8]The rationale for this matching procedure comes from the process through which senatorial recommendation of federal judges works, as detailed in Section 2 above.

[9]Due to data limitations, the productivity outcomes are measured starting in 1924.

## 3.4 Empirical Strategy

To analyze the effect of political connections on the careers and performance of district court judges, we start by considering the following regression model:

$$y_{it} = \theta_i + \tau_{ts} + \sum_{j=1}^{2} \beta^j \cdot ConnectionLost_{it}^j + x_{it}'\theta + \epsilon_{it} \tag{3.1}$$

where $i$ denotes the judge, $t$ indicates the year, and $s$ the state. The dependent variable, $y_{it}$, is refers to the outcome for district court judge $i$ in year $t$. Since a demotion from an upper to a lower level court is not an option, the appointment to courts of appeals is an absorbing state: if promoted, a district court judge drops out of the sample.

$ConnectionLost_{it}^j$ is a dummy variable that takes value 0 if recommending senator $j \in \{1, 2\}$ is still in office at year $t$, and 1 otherwise.[10] For judges who have one connection at the time of appointment, the equation will only include the term $ConnectionLost_{it}^1$ (such event will henceforth be referred as "unique exit"); for judges who have two connections at the time of appointment, the equation will include both $ConnectionLost_{it}^1$ ("first exit") and $ConnectionLost_{it}^2$ ("second exit"). Being time-varying, each of these variables starts at 0 and then switches to 1 when the recommending senator leaves office. The underlying hypothesis is that senators in office are particularly relevant for recommending district court judges for appointment to the court of appeals (Domnarski, 2009), and that this connection is therefore made obsolete when the senator is no longer in Congress.

The vector $x_{it}$ contains judge-specific, time-varying controls, namely: an indicator taking value 1 if, at year $t$, the president of the United States is of the same party as the one who first appointed judge $i$, and either a full set of dummies for each year of tenure as a district

[10]In the rare cases in which a senator exits Congress temporarily and then enters it again, we consider the judge connected until the year of definitive exit. Results excluding the judges connected to such senators are almost identical, and are available upon request.

court judge or judge-specific linear trends.

The $\beta^j$ coefficient captures the effect of losing the connection with recommending senator $j$ on the career and performance of judge $i$. The terms $\theta_i$ and $\tau_{ts}$ are, respectively, judge and state-by-year fixed effects. Finally, $\epsilon_{it}$ is the error term, which is clustered at the recommending senator(s) level, corresponding to the level of the identifying source of variation.

Our main focus is on $\beta^j$. Interpreting such coefficient as causal requires parallel trends: absent the exit from Congress of the recommending senators, the careers and performance of judges whose political connections are lost and of those still connected would have evolved on parallel paths. In other words, we assume that, conditional on the controls, there is no other variable which is correlated with both the outcome of interest and our main explanatory variables.

Judge fixed effects take account of the fact that judges may be different in several important, time-invariant characteristics, which are likely correlated with both the tenure of their recommending senators and the judges' outcomes of interest (for example, some unobserved component of their ability). State-by-year fixed effects absorb any potential event affecting all the judges of a given state equally on a given year, which may be correlated with both the exit of the recommending senators and the judges' careers and performance. Hence, their inclusion ensures that identification is obtained conditional on shocks common to all judges of a given state in each year. Finally, judge's experience fixed effects allow us to non-parametrically account for the time-varying role of experience, which is plausibly correlated positively with both the judge's probability of being promoted and their productivity, as well as the likelihood that they experience the exit of a recommending senator. The omission of one of these sets of controls would arguably lead to a bias in the estimates of the coefficients of interest.

One relevant variation of equation (3.1) relates to timing. To study how judges' outcomes evolve in the years just before and after the change in $ConnectionLost^j_{it}$, we estimate the following event-study specification:

$$y_{it} = \theta_i + \tau_{ts} + \sum_{j=1}^{2} \sum_{l=-L}^{L} \beta^j_l \cdot Exit^j_{i(t+l)} + x'_{it}\theta + \epsilon_{it} \tag{3.2}$$

where $Exit^j_{it}$ takes values 1 if recommending senator $j$ exits Congress at year $t$, and $l$ flags the years either before or after this event, providing a set of time effects leading up and following the transition period (i.e. the exit).

This allows us both to assess the evolution of the effect over the years following the loss of connection(s) and to check for the absence of pre-trends.

## 3.5 Results

### 3.5.1 Effect on Promotions

Given the potentially different nature of judges who are recommended by one or two senators – possibly reflecting the more or less fragmented political situation of their home states – and the different number of senatorial exits they can experience, we estimate equations (3.1) and (3.2) separately for these two groups of judges.

In Figure 3.2, the coefficients indicate that the exit of the unique recommender implies a reduction in the judge's probability of promotion on year $t$ by 1 percentage point (approximately a 42–48% reduction compared to the average yearly probability of promotion). As expected, the effects are concentrated in years with a president of the same political affiliation as the judge (Figure 3.3), when the latter could mostly benefit from the lobbying efforts of their senatorial connection. In both figures, there is no evidence of pre-trends. Even though some of the coefficients after the exit are not statistically significant the 5% level (which is not surprising given that promotions are a rare event[11] and the demanding specification we are estimating), it is reassuring to see that all the coefficients in the right panel display a negative sign after the treatment. This suggests that, although the promotion probability may not decrease immediately, it also never returns to the pre-treatment level. This is also consistent with the long time that a judicial nomination takes, from the moment in which the candidate is identified by the recommenders to the date of nomination or start of judicial service, which can amount also to two years.[12] In Figure C.5 and Figure C.6, the results are instead much less clear, and we are not able to reject the null hypothesis of no treatment effect or exclude the presence of pre-trends.

There are several possible reasons why the treatment effects differ substantially across the two groups. First, even if two senators of the president's party are present in a certain state when a judge is appointed, not necessarily both of them take part in the selection process. Therefore, for all such cases, the actual treatment effect would be diluted by the null effect of losing the connection with a senator who is not an actual recommender. By focusing instead on those cases in which only one such senator is in office, the probability of incurring in this type of measurement error is drastically reduced. Second, judges who can only count on one recommender for their nomination may be more dependent on that senator for the progression of their careers as well, hence magnifying the (negative) effect of losing such connection on the subsequent probability of promotion, compared to those who can rely on two senators instead. Finally, and related to the previous point, states that have two senators of the same party as the president may also be systematically different from those with only one (e.g., more voters' support for that party, larger share of judges of that

[11]In the sub-sample of the 1,056 judges who have one connection at the time of appointment, we observe a promotion for only 122 of them.

[12]It is not uncommon that, if a vacancy arises, the senator in office at the time is the one in charge of finding the candidate to fill the position, and that, by the time the nomination process is finalized, that senator has already left Congress.

political affiliation who can be promoted, etc.). This can make the role of the recommending senator less crucial for the promotion of district court judges, as well as possibly lead the president to appoint as a court of appeals judge someone who is not already sitting on the federal bench.

Altogether, the results from Figure 3.2 and Figure 3.3 provide supportive evidence for the importance of the recommending senator in promoting district court judges to the upper-level courts. In particular, they indicate that such mechanism is prevalent among judges who are recommended by only one senator. For these reasons, such group of judges – i.e., those for which we are able to precisely estimate a treatment effect – will be the focus of the remainder of the paper.

### 3.5.2 Alternative Explanations and Additional Results

A possible concern, in light of the results shown above, is whether what matters is really the presence of the recommender in Congress, or if instead any senator from that same party is sufficient. To disentangle the role of the recommending senator from the one of the party, we augment the baseline model interacting our explanatory variables with a dummy variable that indicates whether in year $t$ and state $s$ there is any senator of the same party of the judge. The results are illustrated in Figure C.1. Unsurprisingly, the probability of promotion decreases more when no senator in state $s$ is of the same party as the judge (Panel A), and also the marginal effect (Panel B) is larger in magnitude. However, both pairs of coefficients are negative, statistically significant, and not statistically different from each other. This is once again suggestive of the importance for promotions of the recommending senator, whose absence cannot be replace by any senator of the same party.

It is also worth exploring whether the effect is driven by the connection to senators of one party as opposed to another. Figure C.2 shows that this is not the case. The coefficient for Democratic and Republican judges are very similar (Panel A), and the marginal effects of losing the connection are almost identical (Panel B).

In addition, Table C.3 suggests that judges of different quality – as proxied by the rating given by the *American Bar Association* – may benefit differently from being connected to their recommender. The coefficients associated to low-qualified judges[13] consistently display a negative sign – suggesting that the probability of promotion after losing the connection decreases more for such judges. However, the coefficients are all imprecisely estimated[14] and none of them is statistically significant. Therefore, we will only cautiously take this as a slightly suggestive evidence of a negative relationship between quality and importance of connection.

Finally, Table C.4 shows that there is no statistically significant difference in the treatment effects between judges whose recommending senator exits Congress for an unexpected (e.g., electoral defeat) vs. an expected (e.g., retirement) reason.

### 3.5.3 Robustness Checks

In this section, we summarize the upshots of three sets of robustness checks.

We begin by augmenting Equation (3.1) with a full set of judge-specific linear trends.[15] This allows us to control for any characteristics of each judge that evolve linearly over time, and that may correlate with both her chances of being promoted and the tenure of her recommender. As displayed in Table C.1 in the Appendix, the use of this alternative specification does not significantly affect our results.

A second, important thing to check is that our estimates are not significantly affected by

[13] We consider as low-qualified judges those whose ABA rating is either "Not qualified" or "Qualified", as opposed to "Well qualified" or "Very well qualified".

[14] Variation in ABA ratings is generally small, and the information is available only for some of the judges in our sample (756 out of 1,056 judges who have one connection).

[15] This requires removing from (3.1) the full set of indicators for a judge's years of service in the district court, which were included as a way to control non-parametrically for the effects of judicial tenure.

issues associated with two-way fixed effect estimators. In particular, a recent methodological literature has shown how – in difference-in-differences setting with heterogenous treatment timing – two-way fixed effect estimators are a weighted mean of several average treatment effects (ATTs), some of which may receive negative weights (Callaway and Sant'Anna, 2020; Imai and Kim, 2020; Sun and Abraham, 2020; De Chaisemartin and d'Haultfoeuille, 2020). This, in turn, introduces significant biases and interpretation problems. To address this we employ the techniques proposed in De Chaisemartin and d'Haultfoeuille (2020).

First, we use their algorithm to diagnose the extent to which negative weights are actually affecting our baseline estimator in Equation (3.1). The results of this exercise are very reassuring: of 6,704 ATTs, only 365 (5.8%) receive a negative weight. Also, the treatment effect on the weights does not significantly correlate with the moment at which a judge receives her district court appointment, which is arguably the main dimension along which significant heterogeneities in treatment effects might have been plausible.

Next, to further test the robustness of our results, we re-estimate our event study using the $DID_M$ alternative estimator put forward by De Chaisemartin and d'Haultfoeuille (2020). As shown in Figure C.3, the dynamics of our effect of interest closely track those documented in Figure 3.2 and Figure 3.3. In other words, there is no significant evidence that our main result is driven by the choice of a specific estimator.

Finally, we exclude subsets of observations to verify how each of them impacts our estimates. Namely, we first repeat our baseline regression several times, each time excluding judges in the district courts of a given State across all years. We then repeat the same process, but each time removing observations referred to one of the forty-three presidential spells covered by our sample. The upshots of these exercises are illustrated in the Appendix, Figure C.4. As shown by Panel A, our estimate from Equation (3.1) is very stable to the exclusion of federal judges from different states. On the other hand, when it comes to excluding periods referred to different administrations, Panel B of Figure C.4 reveals that excluding observations referred to Ronald Reagan's spell in the White House (1981-1989) significantly reduces the magnitude of our coefficient of interest. This is consistent with President Reagan's exceptional activism in promoting district court judges: of the 310 promotions in our sample, 33 (10.6 %) took place under his presidency, more than any other president in the history of the United States.

### 3.5.4 Effect on Performance

To explore the consequences of (losing) political connections on judges' productivity, we explore the effects on four indicators of judicial performance based on judicial opinions, which represent the main output of judges' work (Ash and MacLeod, 2015, 2024; Posner, 2008). Two of them measure the *quantity* of output: the total number of judicial opinions and the number of words contained in the opinions written in a given year. The other two can be thought of proxy for the *quality* of output: the number of forward and backward citations. Similar to academics, judges observe the number of citations to their opinions, and desire more of them. Importantly, citations are not a measure of whether the decision is correct or not. But, on average, more citations means that a case was more useful to future judges.

Figure 3.4 displays the effect of losing the political connection on the above measures. In all four panels, there is no evidence of pre-trends and the coefficients indicate a statistically significant negative effect on judges' performance. The drop arises immediately after the senator's exit from Congress and persists for at least six years. In particular, the loss of the connection causes a 13% reduction in the number of judicial opinions written and in the opinions' length, 14% fewer citations received, and 15% fewer citations made. Taken together, these results indicate a decline in both the quantity and the quality of judges' output. In light of the negative effects on the probability of promotion to the courts of appeals, these results are consistent with judges reducing their effort and productivity once their career prospects plummet.

## 3.6 Conclusion

In this paper, we have provided evidence that U.S. senators can have a large influence in shaping the careers and affecting the performance of federal judges. In particular, exploiting the exit of senators from Congress as a source of within-judge variation in connectedness, we have shown that losing the tie to their recommending senator reduces the yearly probability of promotion of district court judges by up to 48%. Consistent with the system of political appointments in place in the federal judiciary, and used in many other government offices, such an effect emerges in years in which judges share partisan affiliation with the sitting president, and would thus stand to gain from their connection to a senator. Importantly, this event also worsens judges' productivity and quality of output, as proxied by several indicators of judicial performance.

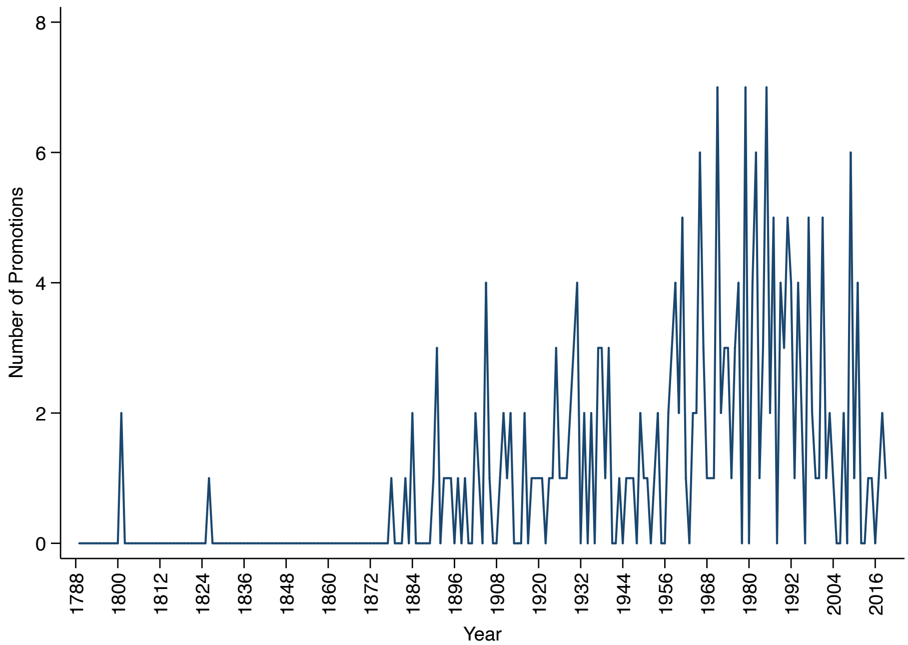
These findings carry important implications for our understanding of the careers and performance of an important category of public sector workers, who have a vital role in the day-to-day functioning of one of the three branches of the U.S. government apparatus. While scholarship has tended to focus on party affiliation and has mostly looked at its impact on sentencing behavior, we have documented how personal connections to specific politicians can affect the chances of judges to access top-level positions within the Federal Court System and be a significant determinant of judges' performance.

Table 3.1: Summary Statistics

| | Mean | Stand. Dev. | Min | Max |
|---|---|---|---|---|
| *Panel A. Cross-Sectional Variables* | | | | |
| Ever Promoted | 0.106 | 0.308 | 0 | 1 |
| Connections at Appointment | 1.510 | 0.500 | 1 | 2 |
| Connections at Promotion | 1.467 | 0.500 | 1 | 2 |
| Total Tenure | 19.82 | 12.22 | 1 | 56 |
| Tenure at Promotion | 9.843 | 5.932 | 1 | 28 |
| *Party of Appointment* | | | | |
| Democratic | 0.484 | 0.500 | 0 | 1 |
| Republican | 0.490 | 0.500 | 0 | 1 |
| Federalist | 0.012 | 0.107 | 0 | 1 |
| Jeffers. Republican | 0.011 | 0.103 | 0 | 1 |
| Whig | 0.003 | 0.057 | 0 | 1 |
| | Mean | Stand. Dev. | Min | Max |
| *Panel B. Time-Varying Variables* | | | | |
| Promoted at Year $t$ (x 100) | 0.536 | 7.302 | 0 | 100 |
| Same-Party President | 0.530 | 0.499 | 0 | 1 |
| Lost Connection (Unique) | 0.631 | 0.483 | 0 | 1 |
| Lost Connection (First) | 0.710 | 0.454 | 0 | 1 |
| Lost Connection (Second) | 0.407 | 0.491 | 0 | 1 |
| Tenure at Year $t$ | 14.18 | 10.20 | 1 | 56 |

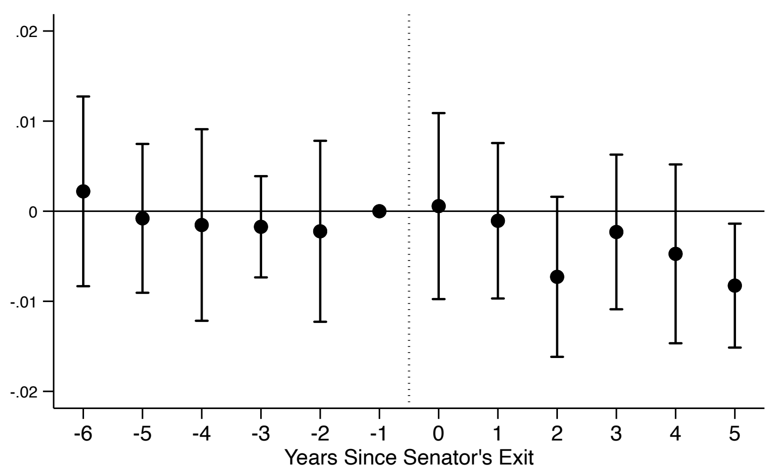
*Notes:* Panel A only includes judges nominated to district court for a state in which there was at least one senator from the same party as the president at the time of nomination. In Panel B, statistics are computed for all the judge-year observations part of our sample, as described in Section 3.3.3.

Figure 3.1: Promotions of District Court Judges in the Period 1789-2019



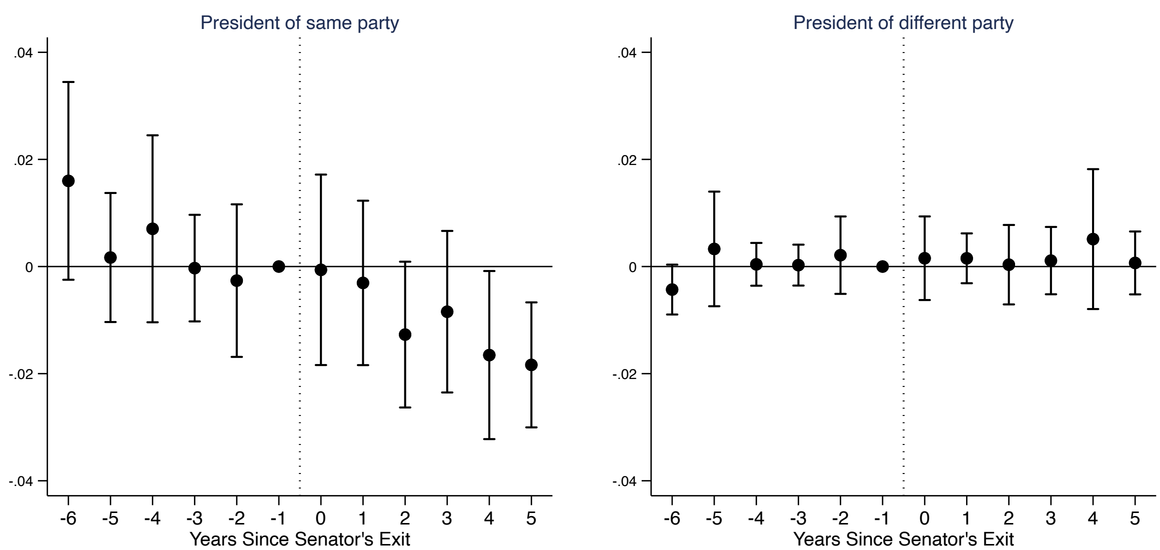

*Notes:* The figure reports the number of federal district court judges, who are part of our sample as described in Section 3.3.3 and got promoted to an appellate court, in every year from 1789 to 2019.

Figure 3.2: Effect on Promotions



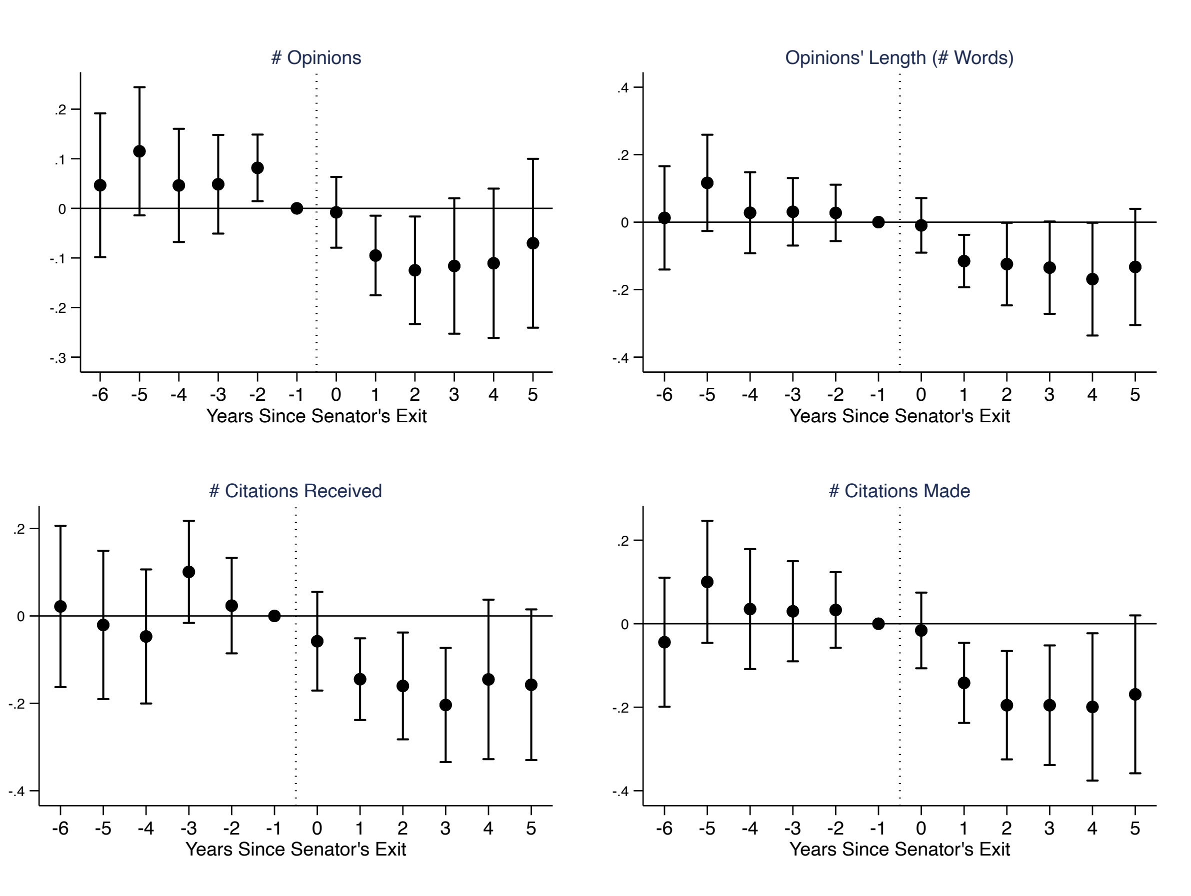

*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

Figure 3.3: Heterogeneous Effects on Promotions by President's Party




*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

Figure 3.4: Effect on Performance




*Notes:* The dependent variables are the number of opinions written by judge $i$ in year $t$ (top left); the number of words in the opinions written by judge $i$ in year $t$ (top right); the number of forward citations for the opinions written by judge $i$ in year $t$ (bottom left); and the number of backward citations for the opinions written by judge $i$ in year $t$ (bottom right). All coefficients are estimated using Poisson regressions. Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

# Bibliography

Abramitzky, R., P. Ager, L. Boustan, E. Cohen, and C. W. Hansen (2023). The effect of immigration restrictions on local labor markets: Lessons from the 1920s border closure. *American Economic Journal: Applied Economics 15*(1), 164–91.

Abramitzky, R. and L. Boustan (2017). Immigration in American Economic History. *Journal of economic literature 55*(4), 1311–45.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122*(3), 467–506.

Acemoglu, D., P. Aghion, and G. L. Violante (2001, December). Deunionization, technical change and inequality. *Carnegie-Rochester Conference Series on Public Policy 55*(1), 229–264.

Adao, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

Ager, P. and C. Hansen (2017). Closing Heaven's Door: Evidence from the 1920s U.S. Immigration Quota Acts. Discussion papers of business and economics, University of Southern Denmark, Department of Business and Economics.

Ahlquist, J. S. (2017, May). Labor Unions, Political Representation, and Economic Inequality. *Annual Review of Political Science 20*(1), 409–432.

Ahlquist, J. S. and M. Downey (2020). The effects of import competition on unionization. *American Economic Journal: Economic Policy*.

Alesina, A. and E. L. Glaeser (2004). *Fighting poverty in the US and Europe: A world of difference*. Oxford University Press, USA.

Alsan, M., K. Eriksson, and G. Niemesh (2020). Understanding the success of the know-nothing party. Technical report, National Bureau of Economic Research.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Archer, R. (2010). *Why is there no labor party in the United States?* Princeton University Press.

Ash, E., B. MacLeod, and S. Naidu (2019). The language of contract: Promises and power in union collective bargaining agreements. *Economics 137*(1), 1–48.

Ash, E. and W. B. MacLeod (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *The Journal of Law and Economics 58*(4), 863–913.

Ash, E. and W. B. MacLeod (2024). Mandatory retirement for judges improved the performance of us state supreme courts. *American Economic Journal: Economic Policy 16*(1), 518–548.

Asher, R. (1982, June). Union Nativism and the Immigrant Response. *Labor History 23*(3), 325–348.

Atack, J. (2016). Historical Geographic Information Systems (GIS) Database of U.S. Railroads for 1826–1911.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review 110*(10), 3139–3183.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics 90*(2), 300–323.

Barone, G., A. D'Ignazio, G. De Blasio, and P. Naticchioni (2016). Mr. rossi, mr. hu and politics. the role of immigration in shaping natives' voting behavior. *Journal of Public Economics 136*, 1–13.

Barth, E., A. Bryson, and H. Dale-Olsen (2020). Union density effects on productivity and wages. *The Economic Journal 130*(631), 1898–1936.

Bazzi, S., A. Ferrara, M. Fiszbein, T. Pearson, and P. A. Testa (2023). The other great migration: Southern whites and the new right. *The Quarterly Journal of Economics 138*(3), 1577–1647.

Beik, M. A. (1996). *The miners of Windber: The struggles of new immigrants for unionization, 1890s-1930s.* Penn State Press.

Berdejó, C. and N. Yuchtman (2013). Crime, Punishment, and Politics: an Analysis of Political Cycles in Criminal Sentencing. *Review of Economics and Statistics 95*(3), 741–756.

Bernstein, I. (1954, June). The growth of American unions. *American Economic Review 2*(2), 131–157.

Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2020). The Glittering Prizes: Career Incentives and Bureaucrat Performance. *The Review of Economic Studies 87*(2), 626–655.

Besley, T. J. and A. A. Payne (2005). Implementation of Anti-Discrimination Policy: does Judicial Selection Matter? *CEPR Discussion Paper*.

Biasi, B. and H. Sarsons (2022). Flexible wages, bargaining, and the gender gap. *The Quarterly Journal of Economics 137*(1), 215–266.

Binder, S. A. and F. Maltzman (2009). *Advice and Dissent: The Struggle to Shape the Federal Judiciary*. Brookings Institution Press.

Bittarello, L. (2018). Organizing Collective Action: Labor Strife in the U.S. in the 1880s. *Unpublished*, 38.

Bolt, J. and J. L. Van Zanden (2020). Maddison style estimates of the evolution of the world economy. a new 2020 update. *Maddison-Project Working Paper WP-15, University of Groningen, Groningen, The Netherlands*.

Borjas, G. (2003). The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics 118*(4), 1335–1374.

Borjas, G. (2005). The labor-market impact of high-skill immigration. *American Economic Review 95*(2), 56–60.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Boustan, L. P. (2010). Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration. *Quarterly Journal of Economics*, 27.

Boustan, L. P. (2013). Racial residential segregation in american cities. Technical report, National Bureau of Economic Research.

Boustan, L. P. (2016). *Competition in the promised land: Black migrants in northern cities and labor markets*, Volume 39. Princeton University Press.

Briggs, V. M. (2001). *Immigration and American unionism*. Cornell University Press.

Brody, D. (1993). *Workers in Industrial America: Essays on the Twentieth Century Struggle* (2nd ed ed.). New York: Oxford University Press.

Burchardi, K. B., T. Chaney, T. A. Hassan, L. Tarquinio, and S. J. Terry (2020). Immigration, innovation, and growth. Technical report, National Bureau of Economic Research.

Burgoon, B., J. Fine, W. Jacoby, and D. Tichenor (2010, December). Immigration and the Transformation of American Unionism. *International Migration Review 44*(4), 933–973.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-Differences with multiple time periods. *Forthcoming, Journal of Econometrics*.

Card, D. (2001a). The effect of unions on wage inequality in the us labor market. *ILR Review 54*(2), 296–315.

Card, D. (2001b). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. (2001c). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. (2005). Is the new immigration really so bad? *The economic journal 115*(507), F300–F323.

Card, D. (2009a). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Card, D. (2009b). Immigration and Inequality. *American Economic Review 99*(2), 1–21.

Carp, R. A., K. L. Manning, L. M. Holmes, and R. Stidham (2019). *Judicial Process in America*. CQ Press.

Chan, S. (1986). *This Bittersweet Soil: The Chinese in California Agriculture, 1860-1910.* Chinese in California Agriculture, 1860-1910. University of California Press.

Chaney, E. and R. Hornbeck (2016). Economic dynamics in the malthusian era: evidence from the 1609 spanish expulsion of the moriscos. *The Economic Journal 126*(594), 1404–1440.

Chang, I. (2003). *The Chinese in America : a narrative history*. Viking New York.

Chassambouli, A. and G. Peri (2015). The Labor Market Effects of Reducing the Number of Illegal Immigrants. *Review of Economic Dynamics 18*(4), 792–821.

Chen, J. (2015). The Impact of Skill-Based Immigration Restrictions: The Chinese Exclusion Act of 1882. *Journal of Human Capital 9*(3), 298 – 328.

Chen, S. and B. Xie (2020). Institutional discrimination and assimilation: Evidence from the chinese exclusion act of 1882.

Clawson, D. and M. A. Clawson (1999). What has happened to the us labor movement? union decline and renewal. *Annual review of sociology 25*(1), 95–119.

Clemens, M. A., E. G. Lewis, and H. M. Postel (2018). Immigration restrictions as active labor market policy: Evidence from the mexican bracero exclusion. *American Economic Review 108*(6), 1468–87.

Cloud, P. and D. Galenson (1987). Chinese immigration and contract labor in the late nineteenth century. *Explorations in Economic History 24*(1), 22–42.

Cohen, A. and C. S. Yang (2019). Judicial Politics and Sentencing Decisions. *American Economic Journal: Economic Policy 11*(1), 160–91.

Collins, W. J. and G. T. Niemesh (2019). Unions and the great compression of wage inequality in the us at mid-century: Evidence from local labour markets. *The Economic History Review 72*(2), 691–715.

Collomp, C. (1988, October). Unions, civics, and National identity: Organized Labor's reaction to Immigration, 1881–1897. *Labor History 29*(4), 450–474.

Colonnelli, E., E. Teso, and M. Prem (2020). Patronage and Selection in Public Sector Organizations. *American Economic Review 110*(10), 3071–3099.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics 92*(1), 1–45.

Crouch, C. (1993). *Industrial relations and European state traditions*. Clarendon Press.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–96.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018). The geography of development. *Journal of Political Economy 126*(3), 903–983.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, 1001–1044.

DiNardo, J. and D. S. Lee (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *The quarterly journal of economics 119*(4), 1383–1441.

Domnarski, W. (2009). *Federal Judges Revealed*. Oxford University Press.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Downes, H. (2023). Did Organized Labor Induce Labor? Unionization and the American Baby Boom. *Working Paper*.

Dustmann, C., U. Schönberg, and J. Stuhler (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives 30*(4), 31–56.

Dustmann, C., K. Vasiljeva, and A. Piil Damm (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies 86*(5), 2035–2091.

Eckert, F. and M. Peters (2022a, September). Spatial structural change. Working Paper 30489, National Bureau of Economic Research.

Eckert, F. and M. Peters (2022b). Spatial structural change. Technical report, National Bureau of Economic Research.

Edo, A., Y. Giesing, J. Öztunc, and P. Poutvaara (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review 115*, 99–143.

Ewing, J. and N. E. Boudette (2023). Autoworkers Score Big Wins in New Contracts with Carmakers. `https://www.nytimes.com/2023/10/30/business/economy/gm-uaw-contract-deal.html`. Accessed: 2023-10-31.

Farber, H. S., D. Herbst, I. Kuziemko, and S. Naidu (2021, April). Unions and Inequality over the Twentieth Century: New Evidence from Survey Data. *The Quarterly Journal of Economics*.

Farber, H. S. and B. Western (2001, September). Accounting for the Decline of Unions in the Private Sector, 1973–1998. *Journal of Labor Research 22*(3), 459–485.

Feigenbaum, J., A. Hertel-Fernandez, and V. Williamson (2018). From the bargaining table to the ballot box: Political effects of right to work laws.

Feng, S., A. B. Krueger, and M. Oppenheimer (2010). Linkages among climate change, crop yields and mexico–us cross-border migration. *Proceedings of the national academy of sciences 107*(32), 14257–14262.

Ferrara, A., P. Testa, and L. Zhou (2022). New Area- and Population-Based Geographic Crosswalks for U.S. Counties and Congressional Districts, 1790-2020. *SSRN Electronic Journal*.

Foged, M. and G. Peri (2016a). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics 8*(2), 1–34.

Foged, M. and G. Peri (2016b). Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data. *American Economic Journal: Applied Economics 8*(2), 1–34.

Foner, P. S. (1947). *History of the Labor Movement in the United States*. New York: International Publishers.

Fouka, V., S. Mazumder, and M. Tabellini (2022). From immigrants to americans: Race and assimilation during the great migration. *The Review of Economic Studies 89*(2), 811–842.

Freeman, R. (1998, January). Spurts in Union Growth: Defining Moments and Social Processes. *NBER Working Paper*.

Freeman, R. and J. Medoff (1984). *What Do Unions Do*. N.Y.: Basic Books.

Galenson, D. (1984). The Rise and Fall of Indentured Servitude in the Americas: An Economic Analysis. *The Journal of Economic History 44*(1), 1–26.

Gallo, N. and D. E. Lewis (2012). The Consequences of Presidential Patronage for Federal Agency Performance. *Journal of Public Administration Research and Theory 22*(2), 219–243.

Garlock, J. (2009). Knights of Labor Assemblies, 1879-1889 (ICPSR 29). *Inter-University Consortium for Political and Social Research [distributor]*.

Giuliano, P. and M. Tabellini (2022). The Seeds of Ideology: Historical Immigration and Political Preferences in the United States. *Working Paper*.

Goldin, C. (1994). The political economy of immigration restriction in the united states, 1890 to 1921. In *The regulated economy: A historical approach to political economy*, pp. 223–258. University of Chicago Press.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Griffin, L. J., M. E. Wallace, and B. A. Rubin (1986). Capitalist resistance to the organization of labor before the new deal: Why? how? success? *American Sociological Review*, 147–167.

Hadero, H. and M. Ott (2023). UPS Workers Approve 5-Year Contract, Capping Contentious Negotiations That Threatened Deliveries. https://apnews.com/article/ups-teamsters-contract-union-voting-b104ca459ddb810f2018c68046d063ed. Accessed: 2023-10-31.

Haines, Michael, F. P. and P. Rhode (2018). United States Agriculture Data, 1840 - 2012. https://doi.org/10.3886/ICPSR35206.v4. Inter-university Consortium for Political and Social Research.

Haines, M. R. (2010a). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [dataset]. https://doi.org/10.3886/ICPSR02896.v3. Inter-university Consortium for Political and Social Research.

Haines, M. R. (2010b). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 (ICPSR 2896).

Halla, M., A. F. Wagner, and J. Zweimüller (2017). Immigration and voting for the far right. *Journal of the European economic association 15*(6), 1341–1385.

Hannan, M. T. and J. Freeman (1987, January). The Ecology of Organizational Founding: American Labor Unions, 1836-1985. *American Journal of Sociology 92*(4), 910–943.

Hatton, T. J. and J. G. Williamson (1995). The Impact of Immigration on American Labor Markets Prior to the Quotas.

Hatton, T. J. and J. G. Williamson (1998). *The age of mass migration: Causes and economic impact*. Oxford University Press on Demand.

Hatton, T. J. and J. G. Williamson (2006). *International migration in the long-run: positive selection, negative selection and policy*. Springer.

Haydu, J. (1988). Employers, unions, and american exceptionalism: Pre-world war i open shops in the machine trades in comparative perspective. *International Review of Social History 33*(1), 25–41.

Higham, J. (1955). Strangers in the land: Patterns of american nativism, 1860–1925. *Rutgers University Press*.

Higham, J. (2002). *Strangers in the land: Patterns of American nativism, 1860-1925*. Rutgers University Press.

Hirsch, B. T. (2008). Sluggish institutions in a dynamic world: Can unions and industrial competition coexist? *Journal of Economic Perspectives 22*(1), 153–176.

Hoi, D. (2022). The Hinge of the Golden Door: Labor Market Impacts of Immigrant Exclusion. Working paper.

Hornbeck, R. (2010). Barbed wire: Property rights and agricultural development. *The Quarterly Journal of Economics 125*(2), 767–810.

Hornbeck, R. and M. Rotemberg (2021). Railroads, market access, and aggregate productivity growth. *University of Chicago Booth School of Business, mimeo*.

Huber, G. A. and S. C. Gordon (2004). Accountability and Coercion: Is Justice Blind When It Runs for Office? *American Journal of Political Science 48*(2), 247–263.

Imai, K. and I. S. Kim (2020). On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. *Forthcoming, Political Analysis*.

Inter-University Consortium for Political and Social Research (ICPSR) (1999). United States Historical Election Returns, 1824-1968 (ICPSR 00001).

Jaeger, D. A., J. Ruist, and J. Stuhler (2018). Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

Kallas, J., K. Ritchie, and E. Friedman (2022). Labor Action Tracker 2022. Technical report, ILR School, Cornell University.

Kanazawa, M. (2005). Immigration, Exclusion, and Taxation: Anti-Chinese Legislation in Gold Rush California. *The Journal of Economic History 65*(3), 779–805.

Karadja, M. and E. Prawitz (2019). Exit, voice, and political change: Evidence from swedish mass migration to the united states. *Journal of Political Economy 127*(4), 1864–1925.

Katz, L. F. and R. A. Margo (2014). Technical change and the relative demand for skilled labor: The united states in historical perspective. In *Human capital in history: The American record*, pp. 15–57. University of Chicago Press.

Ke, R., J. Li, and M. Powell (2018). Managing Careers in Organizations. *Journal of Labor Economics 36*(1), 197–252.

Keeling, D. (1999). The transportation revolution and transatlantic migration, 1850-1914. *Research in Economic History 19*, 39–74.

Lee, E. (2003). *At America's Gates: Chinese Immigration During the Exclusion Era, 1882-1943*. University of North Carolina Press.

Lipset, S. M. and G. Marks (2000). *It Didn't Happen Here: Why Socialism Failed in the United States*. WW Norton & Company.

Logan, T. D. and J. M. Parman (2017). The national rise in residential segregation. *The Journal of Economic History 77*(1), 127–170.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004). European seasonal and annual temperature variability, trends, and extremes since 1500. *Science 303*(5663), 1499–1503.

Macpherson, D. A. and B. T. Hirsch (2023). Five Decades of CPS Wages, Methods, and Union-Nonunion Wage Gaps at Unionstats.com. *Industrial Relations: A Journal of Economy and Society*.

Mayda, A. M., G. Peri, and W. Steingress (2022). The political impact of immigration: Evidence from the united states. *American Economic Journal: Applied Economics 14*(1), 358–389.

McCarthy, J. (2022). U.S. Approval of Labor Unions at Highest Point Since 1965. `https://news.gallup.com/poll/398303/approval-labor-unions-highest-point-1965.aspx`. Accessed: 2023-10-12.

Mendez, I. and I. M. Cutillas (2014). Has immigration affected spanish presidential elections results? *Journal of Population Economics 27*, 135–171.

Mink, G. (2019). *Old labor and new immigrants in American political development: union, party, and state, 1875-1920*. Cornell University Press.

Montgomery, D. (1979). *Workers' control in America: studies in the history of work, technology, and labor struggles*. Cambridge University Press.

Moody, K. (2019). *Tramps & Trade Union Travelers Internal Migration and Organized Labor in Gilded Age America, 1870-1900*. Newburyport: Haymarket Books. OCLC: 1204138258.

Moser, P. and S. San (2019). Immigration, science, and invention: Evidence from the 1920s quota acts. *Unpublished manuscript*.

Naidu, S. (2022). Is there any future for a us labor movement? *Journal of Economic Perspectives 36*(4), 3–28.

Naidu, S. and A. Reich (2018). Collective action and customer service in retail. *ILR Review 71*(4), 986–1001.

Naidu, S. and N. Yuchtman (2016). Labor market institutions in the gilded age of american economic history. Technical report, National Bureau of Economic Research.

Northrup, H. R. (1943). The Negro and the United Mine Workers of America. *Southern Economic Journal*, 313–326.

O'Brien, D. M. (1986). *Storm Center: The Supreme Court in American Politics*, Volume 86. Norton New York.

OECD (2019). *Negotiating Our Way Up.*

Olson, M. (1965). *Logic of Collective Action: Public Goods and the Theory of Groups (Harvard economic studies. v. 124)*. Harvard University Press.

Olzak, S. (1989, May). Labor Unrest, Immigration, and Ethnic Conflict in Urban America, 1880-1914. *American Journal of Sociology 94*(6), 1303–1333.

Ottaviano, G. and G. Peri (2012a). Rethinking the Effect of Immigration on Wages. *Journal of the European Economic Association 10*(1), 152–197.

Ottaviano, G. I. and G. Peri (2012b). Rethinking the effect of immigration on wages. *Journal of the European economic association 10*(1), 152–197.

Otto, A. H. and M. F. Steinhardt (2014). Immigration and election outcomes—evidence from city districts in hamburg. *Regional Science and Urban Economics 45*, 67–79.

Pauling, A., J. Luterbacher, C. Casty, and H. Wanner (2006). Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation. *Climate dynamics 26*, 387–405.

Peri, G. (2016). Immigrants, Productivity, and Labor Markets. *Journal of economic perspectives 30*(4), 3–30.

Perlman, E. R. (2016). *Connecting the periphery: three papers on the developments caused by spreading transportation and information networks in the nineteenth century United States.* Ph. D. thesis, Boston University.

Peters, M. (2021). Market size and spatial growth-evidence from germany's post-war population expulsions. Technical report, National Bureau of Economic Research.

Pfaelzer, J. (2008). *Driven Out: The Forgotten War Against Chinese Americans.* University of California Press.

Posner, R. A. (2008). *How judges think*. Harvard University Press.

Ramankutty, N. and J. A. Foley (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global biogeochemical cycles 13*(4), 997–1027.

Rosenfeld, J. (2019, July). US Labor Studies in the Twenty-First Century: Understanding Laborism Without Labor. *Annual Review of Sociology 45*(1), 449–465.

Rosenfeld, J. and M. Kleykamp (2012, March). Organized Labor and Racial Wage Inequality in the United States. *American Journal of Sociology 117*(5), 1460–1502.

Ruggles, S., C. Fitch, R. Goeken, J. Hacker, M. Nelson, E. Roberts, M. Schouweiler, and M. Sobek (2021). IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. *Minneapolis, MN: IPUMS*.

Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021). IPUMS USA: Version 11.0 [dataset]. `https://doi.org/10.18128/D010.V11.0`. IPUMS.

Rutkus, D. S. (2016). *The Appointment Process for US Circuit and District Court Nominations*. Congressional Research Service.

Schanzenbach, M. M. and E. H. Tiller (2008). Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform. *The University of Chicago Law Review 75*(2), 715–760.

Schmick, E. (2018, September). Collective Action and the Origins of the American Labor Movement. *The Journal of Economic History 78*(3), 744–784.

Scruggs, L. and P. Lange (2002). Where have all the members gone? globalization, institutions, and union density. *the Journal of Politics 64*(1), 126–153.

Sequeira, S., N. Nunn, and N. Qian (2020a, January). Immigrants and the Making of America. *The Review of Economic Studies 87*(1), 382–419.

Sequeira, S., N. Nunn, and N. Qian (2020b). Immigrants and the Making of America. *Review of Economic Studies 87*(1), 382–419.

Sezer, H. (2023). Convicts and Comrades: Coerced Labor's Impact on Early Labor Unions. *Working Paper*.

Slaughter, M. J. (2007). Globalization and declining unionization in the united states. *Industrial Relations: A Journal of Economy and Society 46*(2), 329–346.

Sojourner, A. J., B. R. Frandsen, R. J. Town, D. C. Grabowski, and M. M. Chen (2015). Impacts of unionization on quality and productivity: Regression discontinuity evidence from nursing homes. *ILR Review 68*(4), 771–806.

Solomou, S. and W. Wu (1999). Weather Effects on European Agricultural Output, 1850–1913. *European Review of Economic History 3*(3), 351–373.

Sombart, W. (1976). *Why is There no Socialism in the United States?* International Arts and Sciences Press.

Southworth, C. and J. Stepan-Norris (2009). American trade unions and data limitations: A new agenda for labor studies. *Annual Review of Sociology 35*, 297–320.

Spence, J. D. (1990). *The search for modern China*. W.W. Norton '—&' Co New York.

Spenkuch, J. L., E. Teso, and G. Xu (2021). Ideology and Performance in Public Organizations. *NBER Working Paper*.

Stewart, E. M. (1926). *Handbook of American Trade-Unions: 1926 Edition*. Number 420 in Bulletin of the U.S. Bureau of Labor Statistics. U.S. Government Printing Office.

Sun, L. and S. Abraham (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Forthcoming, Journal of Econometrics*.

Sunstein, C. R., D. Schkade, L. M. Ellman, and A. Sawicki (2007). *Are Judges Political?: An Empirical Analysis of the Federal Judiciary*. Brookings Institution Press.

Tabellini, M. (2020). Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration. *The Review of Economic Studies 87*(1), 454–486.

Taft, P. (1964). *Organized Labor in American History*. Harper and Row Publishers.

Tydings, J. D. (1977). Merit Selection for District Judges. *Judicature 61*, 112.

Wallerstein, M. and B. Western (2000). Unions in decline? what has changed and why. *Annual Review of Political Science 3*(1), 355–377.

Wang, S. and S. Young (2022). Unionization, Employer Opposition, and Establishment Closure. *Essays on Employment and Human Capital, PhD diss. MIT*.

Webb, S. and B. Webb (1894). *The History of Trade Unionism*. Longmans, Green and Company.

Willcox, W. F. (1929). Statistics of Migrations, National Tables, United States. In *International Migrations, Volume I: Statistics*, pp. 372–498. NBER.

Willoughby, W. F. (1905). Employers' associations for dealing with labor in the united states. *The Quarterly Journal of Economics 20*(1), 110–150.

Wolman, L. (1924). The Growth of American Trade Unions 1880-1923.

Xu, G. (2018). The Costs of Patronage: Evidence from The British Empire. *American Economic Review 108*(11), 3170–98.

Yellowitz, I. (1981). Jewish immigrants and the american labor movement, 1900–1920. *American Jewish History 71*(2), 188–217.

# Appendix A

## Appendix Tables and Figures

## Closing Ranks: Organized Labor and Immigration

Table A.1: European Countries and Regions

| | |
|---|---|
| Austria-Hungary | Luxembourg |
| Belgium | Netherlands |
| Czechoslovakia | Norway |
| Denmark | Poland |
| France | Russia |
| Germany | Sweden |
| Greece-Portugal-Spain | Switzerland |
| Ireland | U.K. (England-Scotland-Wales) |
| Italy | |

*Notes:* This table lists the European origin countries and regions used to construct the instrument for immigration described in Section 1.4.2 with 1890 county-level data on the stock of foreign-born individuals from Haines (2010b).

# Table A.2: Occupations Organized by AFL Unions 1900–1920

| Code | Label | Code | Label | Code | Label | Code | Label |
|---|---|---|---|---|---|---|---|
| Skilled | | | | | | | |
| 1 | Actors and actresses | 531 | Heat treaters, annealers, temperers | 591 | Tinsmiths, coppersmiths, and sheet metal workers | 671 | Photographic process workers |
| 4 | Artists and art teachers | 532 | Inspectors, scalers, and graders, log and lumber | 592 | Tool makers, and die makers and setters | 672 | Power station operators |
| 46 | Engineers, mechanical | 533 | Inspectors (n.e.c.) | 593 | Upholsterers | 673 | Sailors and deck hands |
| 47 | Engineers, metallurgical, metallurgists | 534 | Jewelers, watchmakers, goldsmiths, and silversmiths | 594 | Craftsmen and kindred workers (n.e.c.) | 674 | Sawyers |
| 48 | Engineers, mining | 535 | Job setters, metal | 600 | Apprentice auto mechanics | 675 | Spinners, textile |
| 49 | Engineers (n.e.c.) | 540 | Linemen and servicemen, telegraph, telephone, and power | 601 | Apprentice bricklayers and masons | 680 | Stationary firemen |
| 57 | Musicians and music teachers | 541 | Locomotive engineers | 602 | Apprentice carpenters | 681 | Switchmen, railroad |
| 310 | Bookkeepers | 542 | Locomotive firemen | 603 | Apprentice electricians | 682 | Taxicab drivers and chauffers |
| 322 | Dispatchers and starters, vehicle | 543 | Loom fixers | 604 | Apprentice machinists and toolmakers | 683 | Truck and tractor drivers |
| 325 | Express messengers and railway mail clerks | 544 | Machinists | 605 | Apprentice mechanics, except auto | 684 | Weavers, textile |
| 335 | Mail carriers | 545 | Mechanics and repairmen, airplane | 610 | Apprentice plumbers and pipe fitters | 685 | Welders and flame cutters |
| 340 | Messengers and office boys | 550 | Mechanics and repairmen, automobile | 611 | Apprentices, building trades (n.e.c.) | 690 | Operative and kindred workers (n.e.c.) |
| 342 | Shipping and receiving clerks | 551 | Mechanics and repairmen, office machine | 612 | Apprentices, metalworking trades (n.e.c.) | 730 | Attendants, hospital and other institution |
| 350 | Stenographers, typists, and secretaries | 552 | Mechanics and repairmen, radio and television | 613 | Apprentices, printing trades | 731 | Attendants, professional and personal service (n.e.c.) |
| 360 | Telegraph messengers | 553 | Mechanics and repairmen, railroad and car shop | 614 | Apprentices, other specified trades | 732 | Attendants, recreation and amusement |
| 365 | Telegraph operators | 554 | Mechanics and repairmen (n.e.c.) | 615 | Apprentices, trade not specified | 740 | Barbers, beauticians, and manicurists |
| 370 | Telephone operators | 555 | Millers, grain, flour, feed, etc. | 620 | Asbestos and insulation workers | 750 | Bartenders |
| 500 | Bakers | 560 | Millwrights | 621 | Attendants, auto service and parking | 751 | Bootblacks |
| 501 | Blacksmiths | 561 | Molders, metal | 622 | Blasters and powdermen | 752 | Boarding and lodging house keepers |
| 502 | Bookbinders | 562 | Motion picture projectionists | 623 | Boatmen, canalmen, and lock keepers | 753 | Charwomen and cleaners |
| 503 | Boilermakers | 563 | Opticians and lens grinders and polishers | 624 | Brakemen, railroad | 754 | Cooks, except private household |
| 504 | Brickmasons, stonemasons, and tile setters | 564 | Painters, construction and maintenance | 630 | Chainmen, rodmen, and axmen, surveying | 760 | Counter and fountain workers |
| 505 | Cabinetmakers | 565 | Paperhangers | 631 | Conductors, bus and street railway | 761 | Elevator operators |
| 510 | Carpenters | 570 | Pattern and model makers, except paper | 632 | Deliverymen and routemen | 762 | Firemen, fire protection |
| 511 | Cement and concrete finishers | 571 | Photoengravers and lithographers | 633 | Dressmakers and seamstresses, except factory | 764 | Housekeepers and stewards, except private household |
| 512 | Compositors and typesetters | 572 | Piano and organ tuners and repairmen | 634 | Dyers | 770 | Janitors and sextons |
| 513 | Cranemen, derrickmen, and hoistmen | 573 | Plasterers | 635 | Filers, grinders, and polishers, metal | 784 | Waiters and waitresses |
| 514 | Decorators and window dressers | 574 | Plumbers and pipe fitters | 641 | Furnacemen, smeltermen and pourers | 910 | Fishermen and oystermen |
| 515 | Electricians | 575 | Pressmen and plate printers, printing | 642 | Heaters, metal | 950 | Lumbermen, raftsmen, and woodchoppers |
| 520 | Electrotypers and stereotypers | 580 | Rollers and roll hands, metal | 643 | Laundry and dry cleaning operatives | 960 | Teamsters |
| 521 | Engravers, except photoengravers | 581 | Roofers and slaters | 644 | Meat cutters, except slaughter and packing house | | |
| 522 | Excavating, grading, and road machinery operators | 582 | Shoemakers and repairers, except factory | 645 | Milliners | Unskilled | |
| 523 | Foremen (n.e.c.) | 583 | Stationary engineers | 660 | Motormen, mine, factory, logging camp, etc. | 650 | Mine operatives and laborers |
| 524 | Forgemen and hammermen | 584 | Stone cutters and stone carvers | 661 | Motormen, street, subway, and elevated railway | 940 | Longshoremen and stevedores |
| 525 | Furriers | 585 | Structural metal workers | 662 | Oilers and greaser, except auto | 970* | Laborers (n.e.c.)* |
| 530 | Glaziers | 590 | Tailors and tailoresses | 670 | Painters, except construction or maintenance | | |

*Notes:* This table lists the occupations that were represented by AFL-affliated unions during the period 1900–1920, based on the detailed jurisdictions of unions reported by Stewart (1926). Code and Label report, respectively, the occupation code and description for the OCC1950 variable from the full-count Census data made available by IPUMS (Ruggles et al., 2021). Skilled and Unskilled refer to the distinction made for the analysis reported in Table 1.4 and based on the description of the type of workers organized by each union reported by Stewart (1926). For occupational code 970 "Laborers (n.e.c.)", only workers with industry (IND1950) code 437, 446, 449 (textile industry), 406 (meat production industry), and 418 (beverage industry) are considered, consistent with the historical evidence (Stewart, 1926) of which unskilled workers were organized during the period of analysis.

Table A.3: Heterogeneous Effects by Workers' Skills – Intensive Margin

| | *Dependent variable:* | | |
|---|---|---|---|
| | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) |
| *Panel A: Skilled (Craft) Unions* | | | |
| Share Immigrants | 0.963** | 8.368** | 392.368 |
| | (0.373) | (3.246) | (337.532) |
| Observations | 693 | 693 | 693 |
| Dep. var. mean | 0.082 | 7.082 | 102.783 |
| Indep. var. mean | 0.046 | 0.046 | 0.046 |
| KP F-stat | 20.36 | 20.36 | 20.36 |
| *Panel B: Unskilled (Industrial) Unions* | | | |
| Share Immigrants | 0.536 | -0.084 | 78.828 |
| | (1.268) | (1.980) | (408.013) |
| Observations | 276 | 276 | 276 |
| Dep. var. mean | 0.646 | 6.159 | 155.135 |
| Indep. var. mean | 0.039 | 0.039 | 0.039 |
| KP F-stat | 21.42 | 21.42 | 21.42 |

*Notes:* Observations are at the county-decade level. The sample is restricted only to counties that have some union presence in every year they are observed. The dependent variables are: union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 1); the log number of union branches (column 2); or, the average branch size, defined as the number of union members divided by the number of branches (column 3). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 1.6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.4: Unionization and Immigration Flows

| | *Dependent variable:* Share Immigrants | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Union Presence (t-10) | -0.009*** | | | |
| | (0.002) | | | |
| Union Density (t-10) | | -0.025** | | |
| | | (0.011) | | |
| Log # Branches (t-10) | | | -0.009*** | |
| | | | (0.002) | |
| Avg. Branch Size × 100 (t-10) | | | | -0.004*** |
| | | | | (0.002) |
| Observations | 5,020 | 5,020 | 5,020 | 5,020 |
| Dep. var. mean | 0.019 | 0.019 | 0.019 | 0.019 |
| Indep. var. mean | 0.039 | 0.265 | 0.402 | 29.939 |

*Notes:* Observations are at the county-decade level. The dependent variable is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The regressors of interest are the ten-year lag of: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches (multiplied by 100 for expositional purposes) or zero if the county has no labor union (column 4). All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population, the 1880 male labor force participation rate, and the 1890 stock of European immigrants (relative to all European immigrants in the U.S. in that year). Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.5: Heterogeneous Effects by Religion of Immigrants

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| Share Immigrants non-Protestant | 1.825 | 0.343* | 3.656** | 353.184* |
| | (1.215) | (0.201) | (1.493) | (210.117) |
| *Standardized coefficient* | *[0.121]* | *[0.088]* | *[0.134]* | *[0.145]* |
| Share Immigrants Protestant | 1.001 | 0.155 | 1.244 | 51.258 |
| | (1.473) | (0.323) | (1.586) | (360.299) |
| *Standardized coefficient* | *[0.043]* | *[0.026]* | *[0.029]* | *[0.014]* |
| Observations | 5,018 | 5,018 | 5,018 | 5,018 |
| Dep. var. mean | 0.265 | 0.039 | 1.627 | 29.978 |
| Indep. var. mean (non-Protestant) | 0.014 | 0.014 | 0.014 | 0.014 |
| Indep. var. mean (Protestant) | 0.010 | 0.010 | 0.010 | 0.010 |
| KP F-stat | 15.94 | 15.94 | 15.94 | 15.94 |
| SW F-stat (non-Protestant) | 40.78 | 40.78 | 40.78 | 40.78 |
| SW F-stat (Protestant) | 112.20 | 112.20 | 112.20 | 112.20 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from non-Protestant or Protestant European countries who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 1.4.2 and Section 1.6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.6: Heterogeneous Effects by Strength of Labor Movement in Country of Origin

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| Share Immigrants UK-Ireland | -7.307 | -2.822 | -10.995 | -3,304.544* |
| | (8.653) | (1.966) | (11.180) | (1,694.614) |
| *Standardized coefficient* | *[-0.083]* | *[-0.125]* | *[-0.069]* | *[-0.233]* |
| Share Immigrants Other Countries | 1.819** | 0.371** | 3.306*** | 359.524** |
| | (0.873) | (0.144) | (1.080) | (153.123) |
| *Standardized coefficient* | *[0.153]* | *[0.122]* | *[0.154]* | *[0.188]* |
| Observations | 5,018 | 5,018 | 5,018 | 5,018 |
| Dep. var. mean | 0.265 | 0.039 | 1.627 | 29.978 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 14.65 | 14.65 | 14.65 | 14.65 |
| SW F-stat (UK-Ireland) | 32.60 | 32.60 | 32.60 | 32.60 |
| SW F-stat (Other Countries) | 27.36 | 27.36 | 27.36 | 27.36 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from European countries with a strong (UK-Ireland) and weak (other countries) labor movements as of 1870 (see Appendix A.4) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.7: Effect on Local Economic Outcomes

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Labor Force Part. Rate | Mfg. Output (per Worker) | Mfg. Output (Share of Total) | Labor Force in Skilled Unions Occ. |
| | (1) | (2) | (3) | (4) |
| Share Immigrants | -0.036 | -0.158 | -0.001 | 0.955 |
| | (0.079) | (0.552) | (0.008) | (0.925) |
| *Standardized coefficient* | *[-0.035]* | *[-0.014]* | *[-0.015]* | *[0.028]* |
| Observations | 5,025 | 4,932 | 4,932 | 5,025 |
| Dep. var. mean | 0.910 | 1.328 | 0.000 | 6.630 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 35.14 | 34.37 | 34.37 | 35.14 |

*Notes:* Observations are at the county-decade level. Dependent variables are: the male labor force participation rate (column 1); the log of manufacturing output divided by the manufacturing labor force (column 2); the manufacturing output as a share of the total output in the U.S. in that year (column 3); or, the log of the total male labor force in occupations represented by the AFL craft unions (column 4). The dependent variables of columns 2 and 3 for the year 1910, which would otherwise be missing, are linearly interpolated. The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.8: Changes to U.S.-Born Workers' Occupations

| | *Dependent variable:* Share of U.S.-Born LF in AFL-Covered Occupations | | | | | |
|---|---|---|---|---|---|---|
| | With local union branch | | | Without local union branch | | |
| | All | Skilled | Unskilled | All | Skilled | Unskilled |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Share Immigrants | 0.493* | 0.096** | -0.083 | -0.433* | 0.015 | -0.192*** |
| | (0.272) | (0.039) | (0.064) | (0.247) | (0.054) | (0.055) |
| Observations | 5,025 | 5,025 | 4,398 | 5,025 | 5,025 | 4,398 |
| Dep. var. mean | 0.088 | 0.008 | 0.011 | 0.126 | 0.060 | 0.027 |
| Indep. var. mean | 0.024 | 0.024 | 0.025 | 0.024 | 0.024 | 0.025 |
| KP F-stat | 35.14 | 35.14 | 99.00 | 35.14 | 35.14 | 99.00 |

*Notes:* Observations are at the county-decade level. The dependent variables are the shares of U.S.-born workers (men 16–64) in the labor force who are in occupations that have positive union membership in the county (columns 1–3), or no union representation in the county (columns 4–6). All (columns 1 and 4) refers to all occupations covered by an AFL-affiliated national union; Skilled (columns 2 and 5) refers to the occupations covered by the ten largest AFL-affiliated national unions that represented skilled workers; Unskilled (columns 3 and 6) refers to the AFL-affiliated national unions that represented unskilled workers. The sample in each column is restricted to counties that have at least one worker in the indicated set of occupations in every decade between 1900–1920. The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.9: Effect on U.S.-Born Workers' Labor Market Outcomes

| | *Dependent variable:* | |
|---|---|---|
| | Labor Force Participation Rate | (Log) Occupational Income Score |
| | (1) | (2) |
| Share Immigrants | -0.049 | 0.123 |
| | (0.087) | (0.126) |
| Observations | 5,025 | 5,025 |
| Dep. var. mean | 0.905 | 19.137 |
| Indep. var. mean | 0.024 | 0.024 |
| KP F-stat | 35.14 | 35.14 |

*Notes:* Observations are at the county-decade level. The dependent variables are the shares of the labor force participation rate among U.S.-born workers, men 16–64 (column 1), or the log of the average occupational income score among U.S.-born workers (column 2). The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.10: Union Density and Wage Inequality

| | *Dependent variable:* Wage Inequality | | |
|---|---|---|---|
| | 90/10 | 90/50 | 50/10 |
| | (1) | (2) | (3) |
| *Panel A* | | | |
| Union Presence | -0.107*** | -0.079*** | -0.028** |
| | (0.018) | (0.011) | (0.014) |
| Observations | 1,666 | 1,666 | 1,666 |
| Dep. var. mean | 1.890 | 0.852 | 1.038 |
| Indep. var. mean | 0.339 | 0.339 | 0.339 |
| *Panel B* | | | |
| Union Density | -0.118** | -0.116*** | -0.002 |
| | (0.053) | (0.034) | (0.044) |
| Observations | 1,666 | 1,666 | 1,666 |
| Dep. var. mean | 1.888 | 0.851 | 1.037 |
| Indep. var. mean | 0.058 | 0.058 | 0.058 |

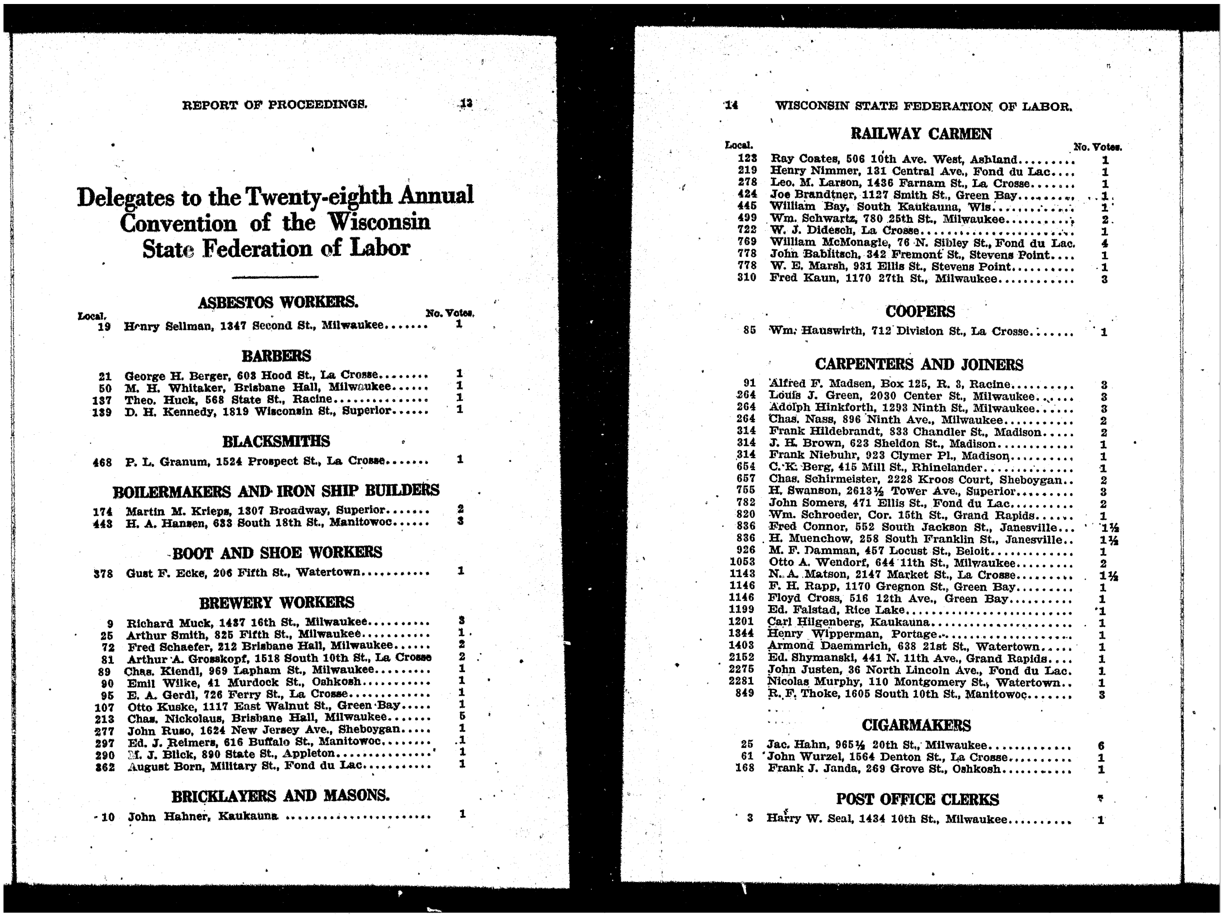
*Notes:* Observations are at the county level. The dependent variables are measures of wage inequality in 1940, proxied by log wage differentials for full-time, full-year workers computed at the following percentiles: 90 to 10 (column 1); 90 to 50 (column 2); or, 50 to 10 (column 3). The main regressors of interest are a dummy for whether the county has positive union membership in 1920 (Panel A), or the share of unionized workers in occupations that are represented by AFL-affiliated national unions (Panel B). All regressions include state fixed effects, and the following controls: the 1890 share of urban population and the 1880 male labor force participation rate. Robust standard errors are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.11: Effect on the Composition of Union Leaders

| | Dependent variable: Share of Leaders | | | | | |
|---|---|---|---|---|---|---|
| | U.S. | N/W Europe | S/E Europe | U.S. | N/W Europe | S/E Europe |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All counties | | | Always unionized counties | | |
| *Panel A: Origin country* | | | | | | |
| Share Immigrants | 1.272** | 0.286 | 0.025 | -0.054 | 0.079 | 0.003 |
| | (0.623) | (0.214) | (0.071) | (0.230) | (0.198) | (0.117) |
| Dep. var. mean | 0.205 | 0.018 | 0.005 | 0.870 | 0.088 | 0.024 |
| *Panel B: Ancestry* | | | | | | |
| Share Immigrants | | 1.396** | 0.198 | | 0.125 | -0.053 |
| | | (0.670) | (0.201) | | (0.280) | (0.272) |
| Dep. var. mean | | 0.204 | 0.023 | | 0.881 | 0.101 |
| Observations | 5,024 | 5,024 | 5,024 | 588 | 588 | 588 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.047 | 0.047 | 0.047 |
| KP F-stat | 35.13 | 35.13 | 35.13 | 21.43 | 21.43 | 21.43 |

*Notes:* Observations are at the county-decade level. The dependent variable is the share of union delegates whose last name is of the origin (Panel A) or ancestry (Panel B) indicated in the column headings. The procedure used to infer the origin or the ancestry is described in Section A.2. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. In columns 1 to 3, the sample includes all counties as in Table 1.3 (in counties with no unionization, both the shares of U.S.-born and of European delegates are set to zero); in columns 4 to 6, the sample is restricted only to counties for which a union delegate is observed in every year. All regressions include county and year fixed effects. The regressions in columns 1 to 3 also include year dummies interacted with the 1890 share of urban population and the 1880 male labor force participation rate. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Figure A.1: Example of Digitized Document on Union Branches and Delegates



REPORT OF PROCEEDINGS. 13

# Delegates to the Twenty-eighth Annual Convention of the Wisconsin State Federation of Labor

## ASBESTOS WORKERS.

| Local. | | No. Votes. |
|---|---|---|
| 19 | Hᵉnry Sellman, 1347 Second St., Milwaukee....... | 1 |

## BARBERS

| | | |
|---|---|---|
| 21 | George H. Berger, 603 Hood St., La Crosse........ | 1 |
| 50 | M. H. Whitaker, Brisbane Hall, Milwaukee...... | 1 |
| 137 | Theo. Huck, 568 State St., Racine............... | 1 |
| 139 | D. H. Kennedy, 1819 Wisconsin St., Superior...... | 1 |

## BLACKSMITHS

| | | |
|---|---|---|
| 468 | P. L. Granum, 1524 Prospect St., La Crosse....... | 1 |

## BOILERMAKERS AND IRON SHIP BUILDERS

| | | |
|---|---|---|
| 174 | Martin M. Krieps, 1307 Broadway, Superior....... | 2 |
| 443 | H. A. Hansen, 633 South 18th St., Manitowoc...... | 3 |

## BOOT AND SHOE WORKERS

| | | |
|---|---|---|
| 378 | Gust F. Ecke, 206 Fifth St., Watertown........... | 1 |

## BREWERY WORKERS

| | | |
|---|---|---|
| 9 | Richard Muck, 1437 16th St., Milwaukee.......... | 3 |
| 25 | Arthur Smith, 825 Fifth St., Milwaukee........... | 1 |
| 72 | Fred Schaefer, 212 Brisbane Hall, Milwaukee...... | 2 |
| 81 | Arthur A. Grosskopf, 1518 South 10th St., La Crosse | 2 |
| 89 | Chas. Kiendl, 969 Lapham St., Milwaukee......... | 1 |
| 90 | Emil Wilke, 41 Murdock St., Oshkosh........... | 1 |
| 95 | E. A. Gerdl, 726 Ferry St., La Crosse............ | 1 |
| 107 | Otto Kuske, 1117 East Walnut St., Green Bay..... | 1 |
| 213 | Chas. Nickolaus, Brisbane Hall, Milwaukee....... | 5 |
| 277 | John Ruso, 1624 New Jersey Ave., Sheboygan..... | 1 |
| 297 | Ed. J. Reimers, 616 Buffalo St., Manitowoc........ | 1 |
| 290 | M. J. Blick, 890 State St., Appleton............... | 1 |
| 362 | August Born, Military St., Fond du Lac........... | 1 |

## BRICKLAYERS AND MASONS.

| | | |
|---|---|---|
| 10 | John Hahner, Kaukauna ....................... | 1 |

14 WISCONSIN STATE FEDERATION OF LABOR.

## RAILWAY CARMEN

| Local. | | No. Votes. |
|---|---|---|
| 123 | Ray Coates, 506 10th Ave. West, Ashland......... | 1 |
| 219 | Henry Nimmer, 131 Central Ave., Fond du Lac.... | 1 |
| 278 | Leo. M. Larson, 1436 Farnam St., La Crosse....... | 1 |
| 424 | Joe Brandtner, 1127 Smith St., Green Bay........ | 1 |
| 445 | William Bay, South Kaukauna, Wis............. | 1 |
| 499 | Wm. Schwartz, 780 25th St., Milwaukee.......... | 2 |
| 722 | W. J. Didesch, La Crosse....................... | 1 |
| 769 | William McMonagle, 76 N. Sibley St., Fond du Lac. | 4 |
| 778 | John Bablitsch, 342 Fremont St., Stevens Point.... | 1 |
| 778 | W. E. Marsh, 931 Ellis St., Stevens Point.......... | 1 |
| 310 | Fred Kaun, 1170 27th St., Milwaukee............ | 3 |

## COOPERS

| | | |
|---|---|---|
| 85 | Wm. Hauswirth, 712 Division St., La Crosse....... | 1 |

## CARPENTERS AND JOINERS

| | | |
|---|---|---|
| 91 | Alfred F. Madsen, Box 125, R. 3, Racine.......... | 3 |
| 264 | Louis J. Green, 2030 Center St., Milwaukee....... | 3 |
| 264 | Adolph Hinkforth, 1293 Ninth St., Milwaukee...... | 3 |
| 264 | Chas. Nass, 896 Ninth Ave., Milwaukee........... | 2 |
| 314 | Frank Hildebrandt, 833 Chandler St., Madison..... | 2 |
| 314 | J. H. Brown, 623 Sheldon St., Madison............ | 1 |
| 314 | Frank Niebuhr, 923 Clymer Pl., Madison.......... | 1 |
| 654 | C. K. Berg, 415 Mill St., Rhinelander.............. | 1 |
| 657 | Chas. Schirmeister, 2228 Kroos Court, Sheboygan.. | 2 |
| 755 | H. Swanson, 2613½ Tower Ave., Superior......... | 3 |
| 782 | John Somers, 471 Ellis St., Fond du Lac.......... | 2 |
| 820 | Wm. Schroeder, Cor. 15th St., Grand Rapids...... | 1 |
| 836 | Fred Connor, 552 South Jackson St., Janesville... | 1½ |
| 836 | H. Muenchow, 258 South Franklin St., Janesville.. | 1½ |
| 926 | M. F. Damman, 457 Locust St., Beloit............ | 1 |
| 1053 | Otto A. Wendorf, 644 11th St., Milwaukee......... | 2 |
| 1143 | N. A. Matson, 2147 Market St., La Crosse......... | 1½ |
| 1146 | F. H. Rapp, 1170 Gregnon St., Green Bay......... | 1 |
| 1146 | Floyd Cross, 516 12th Ave., Green Bay.......... | 1 |
| 1199 | Ed. Falstad, Rice Lake......................... | 1 |
| 1201 | Carl Hilgenberg, Kaukauna..................... | 1 |
| 1344 | Henry Wipperman, Portage..................... | 1 |
| 1403 | Armond Daemmrich, 638 21st St., Watertown..... | 1 |
| 2152 | Ed. Shymanski, 441 N. 11th Ave., Grand Rapids.... | 1 |
| 2275 | John Justen, 36 North Lincoln Ave., Fond du Lac. | 1 |
| 2281 | Nicolas Murphy, 110 Montgomery St., Watertown.. | 1 |
| 849 | R. F. Thoke, 1605 South 10th St., Manitowoc....... | 3 |

## CIGARMAKERS

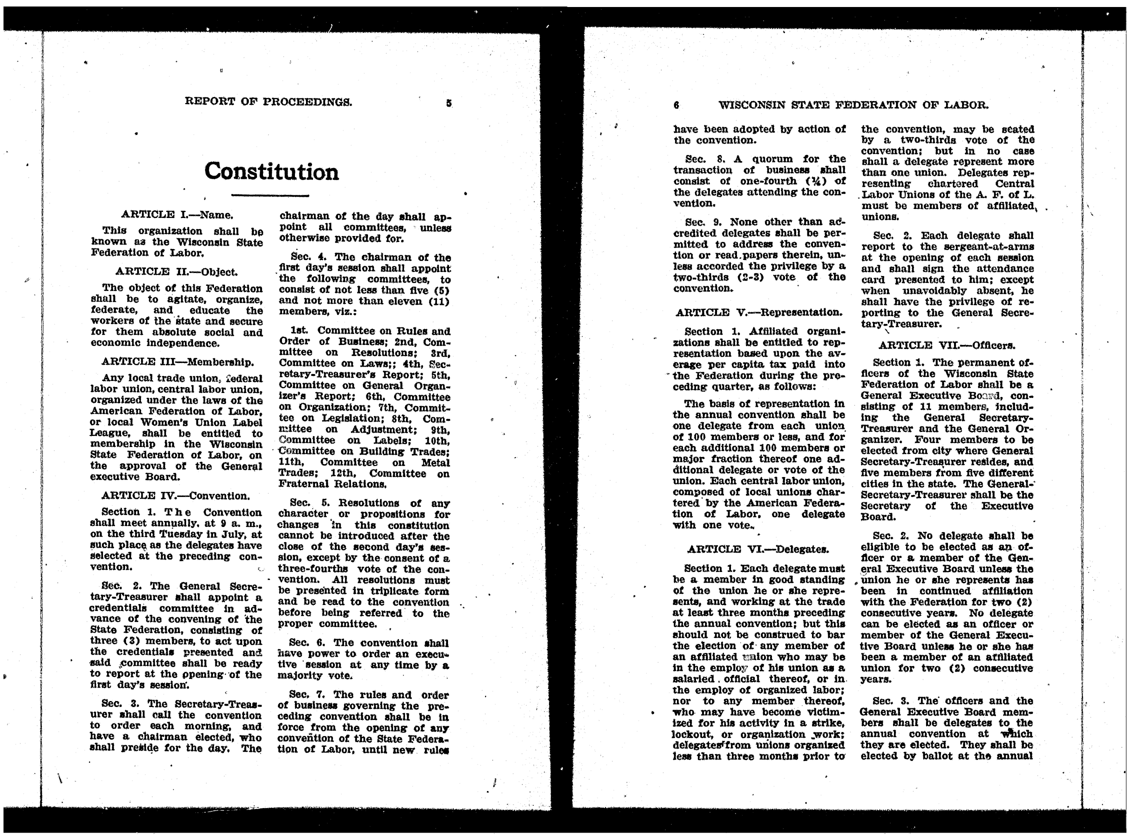
| | | |
|---|---|---|
| 25 | Jac. Hahn, 965½ 20th St., Milwaukee............ | 6 |
| 61 | John Wurzel, 1564 Denton St., La Crosse.......... | 1 |
| 168 | Frank J. Janda, 269 Grove St., Oshkosh........... | 1 |

## POST OFFICE CLERKS

| | | |
|---|---|---|
| 3 | Harry W. Seal, 1434 10th St., Milwaukee.......... | 1 |

*Notes:* The figure shows a digitized document from the proceedings of the state federations of labor's conventions. The documents contain information on the number of branches represented at the conventions, along with information on their delegates.

Figure A.2: Example of Digitized Document on Representation Rules at Conventions



REPORT OF PROCEEDINGS. 5

# Constitution

ARTICLE I.—Name.

This organization shall be known as the Wisconsin State Federation of Labor.

ARTICLE II.—Object.

The object of this Federation shall be to agitate, organize, federate, and educate the workers of the state and secure for them absolute social and economic independence.

ARTICLE III—Membership.

Any local trade union, federal labor union, central labor union, organized under the laws of the American Federation of Labor, or local Women's Union Label League, shall be entitled to membership in the Wisconsin State Federation of Labor, on the approval of the General executive Board.

ARTICLE IV.—Convention.

Section 1. The Convention shall meet annually, at 9 a. m., on the third Tuesday in July, at such place as the delegates have selected at the preceding convention.

Sec. 2. The General Secretary-Treasurer shall appoint a credentials committee in advance of the convening of the State Federation, consisting of three (3) members, to act upon the credentials presented and said committee shall be ready to report at the opening of the first day's session.

Sec. 3. The Secretary-Treasurer shall call the convention to order each morning, and have a chairman elected, who shall preside for the day. The chairman of the day shall appoint all committees, unless otherwise provided for.

Sec. 4. The chairman of the first day's session shall appoint the following committees, to consist of not less than five (5) and not more than eleven (11) members, viz.:

1st. Committee on Rules and Order of Business; 2nd, Committee on Resolutions; 3rd, Committee on Laws;; 4th, Secretary-Treasurer's Report; 5th, Committee on General Organizer's Report; 6th, Committee on Organization; 7th, Committee on Legislation; 8th, Committee on Adjustment; 9th, Committee on Labels; 10th, Committee on Building Trades; 11th, Committee on Metal Trades; 12th, Committee on Fraternal Relations.

Sec. 5. Resolutions of any character or propositions for changes in this constitution cannot be introduced after the close of the second day's session, except by the consent of a three-fourths vote of the convention. All resolutions must be presented in triplicate form and be read to the convention before being referred to the proper committee.

Sec. 6. The convention shall have power to order an executive session at any time by a majority vote.

Sec. 7. The rules and order of business governing the preceding convention shall be in force from the opening of any convention of the State Federation of Labor, until new rules

6 WISCONSIN STATE FEDERATION OF LABOR.

have been adopted by action of the convention.

Sec. 8. A quorum for the transaction of business shall consist of one-fourth (¼) of the delegates attending the convention.

Sec. 9. None other than accredited delegates shall be permitted to address the convention or read papers therein, unless accorded the privilege by a two-thirds (2-3) vote of the convention.

ARTICLE V.—Representation.

Section 1. Affiliated organizations shall be entitled to representation based upon the average per capita tax paid into the Federation during the preceding quarter, as follows:

The basis of representation in the annual convention shall be one delegate from each union of 100 members or less, and for each additional 100 members or major fraction thereof one additional delegate or vote of the union. Each central labor union, composed of local unions chartered by the American Federation of Labor, one delegate with one vote.

ARTICLE VI.—Delegates.

Section 1. Each delegate must be a member in good standing of the union he or she represents, and working at the trade at least three months preceding the annual convention; but this should not be construed to bar the election of any member of an affiliated union who may be in the employ of his union as a salaried official thereof, or in the employ of organized labor; nor to any member thereof, who may have become victimized for his activity in a strike, lockout, or organization work; delegates from unions organized less than three months prior to the convention, may be seated by a two-thirds vote of the convention; but in no case shall a delegate represent more than one union. Delegates representing chartered Central Labor Unions of the A. F. of L. must be members of affiliated unions.

Sec. 2. Each delegate shall report to the sergeant-at-arms at the opening of each session and shall sign the attendance card presented to him; except when unavoidably absent, he shall have the privilege of reporting to the General Secretary-Treasurer.

ARTICLE VII.—Officers.

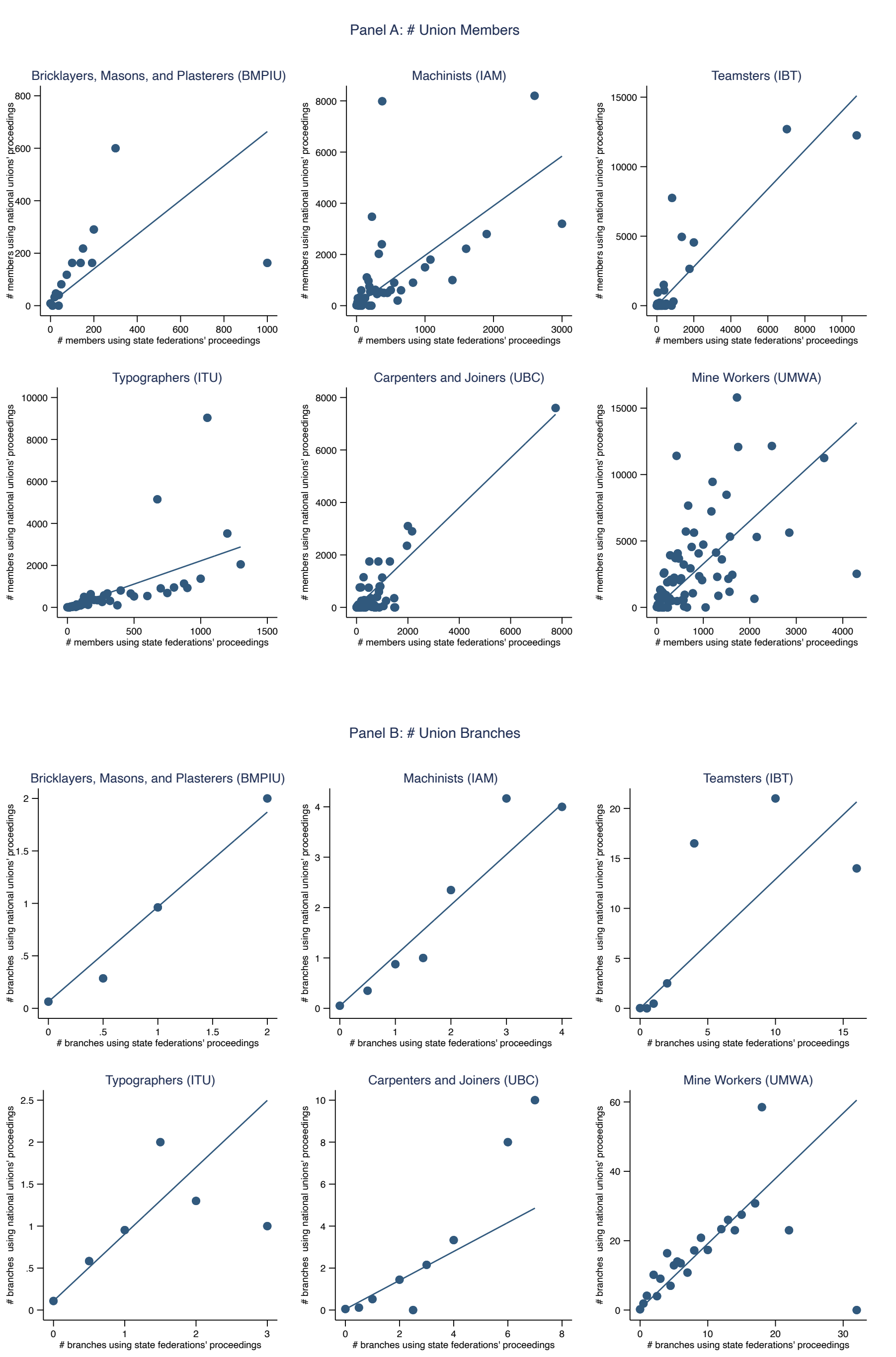
Section 1. The permanent officers of the Wisconsin State Federation of Labor shall be a General Executive Board, consisting of 11 members, including the General Secretary-Treasurer and the General Organizer. Four members to be elected from city where General Secretary-Treasurer resides, and five members from five different cities in the state. The General-Secretary-Treasurer shall be the Secretary of the Executive Board.

Sec. 2. No delegate shall be eligible to be elected as an officer or a member of the General Executive Board unless the union he or she represents has been in continued affiliation with the Federation for two (2) consecutive years. No delegate can be elected as an officer or member of the General Executive Board unless he or she has been a member of an affiliated union for two (2) consecutive years.

Sec. 3. The officers and the General Executive Board members shall be delegates to the annual convention at which they are elected. They shall be elected by ballot at the annual

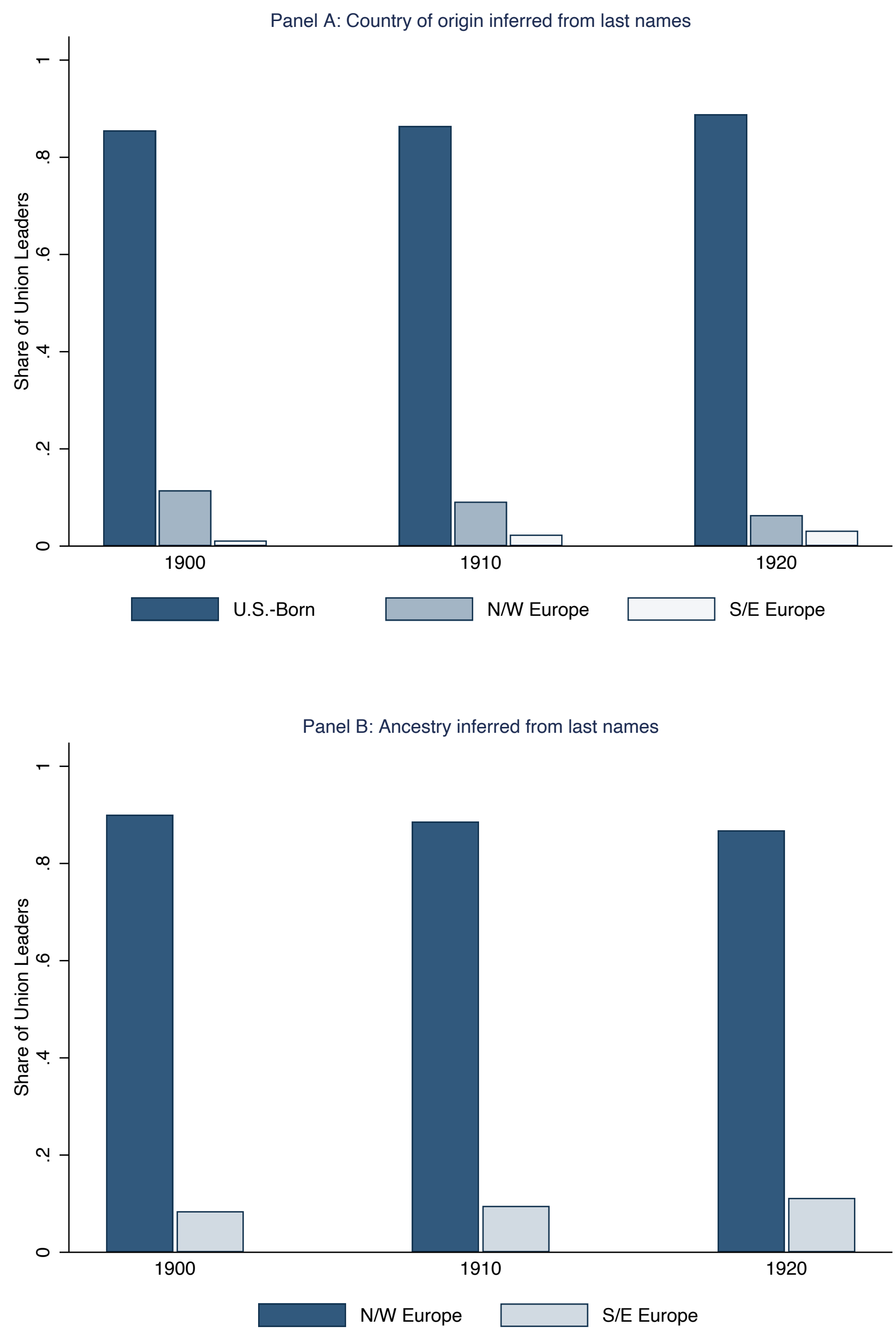
*Notes:* The figure shows a digitized document from the constitutions of the state federations of labor. The documents contain information on the rules that establish the number of delegates that local branches could send to the conventions. The highlighted paragraph on the page on the right provides an example.

Figure A.3: Correlation Between Measures Across Data Sources



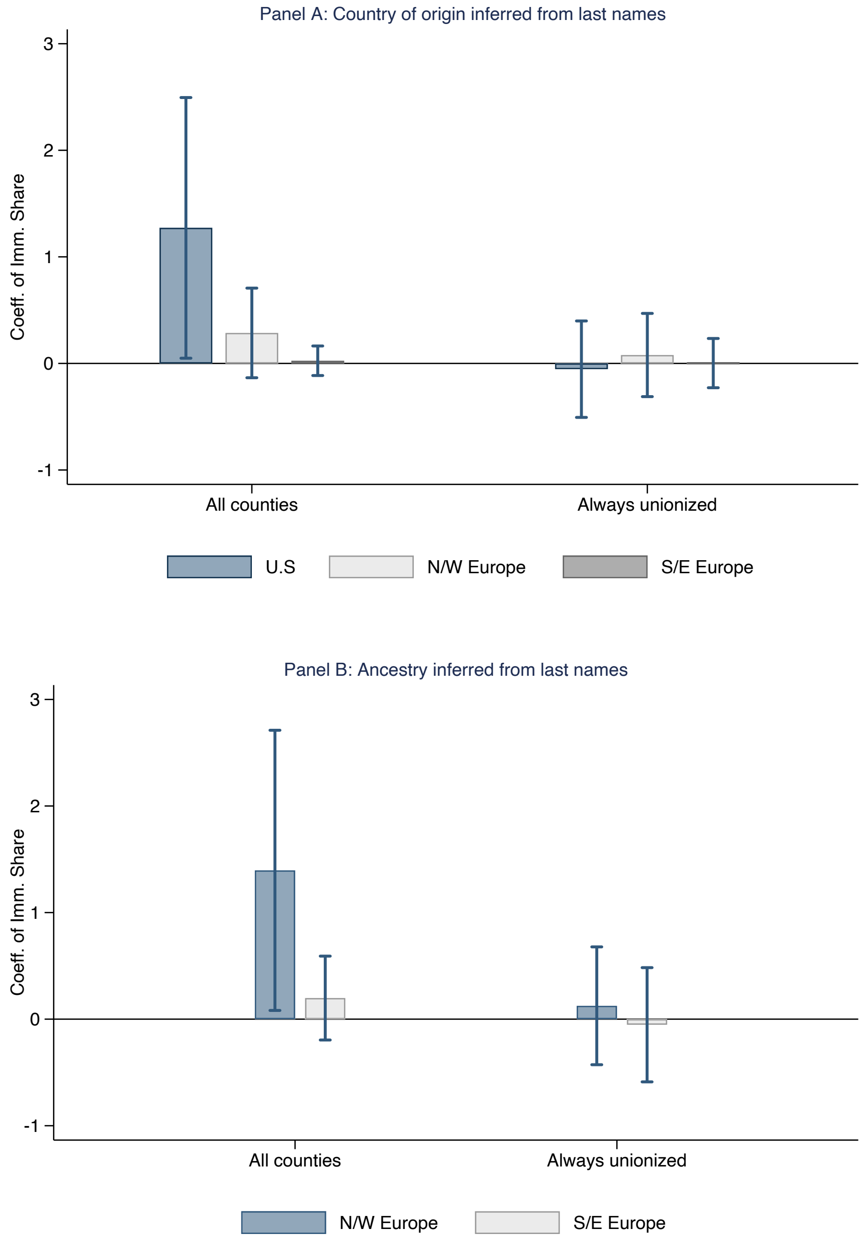

*Notes:* The figure shows binned scatter plots of the county-level union membership estimates (Panel A) and number of union branches (Panel B), constructed using the main data source (convention proceedings of the state federations of labor, on the x-axis) and the complementary data source (convention proceedings of the AFL-affiliated national unions, on the y-axis). Each graph shows the correlation between the two measures for each of the six national unions that are observed in both sources. See Section 1.3 for more details.

Figure A.4: Shares of Union Leaders by Origin and Ancestry



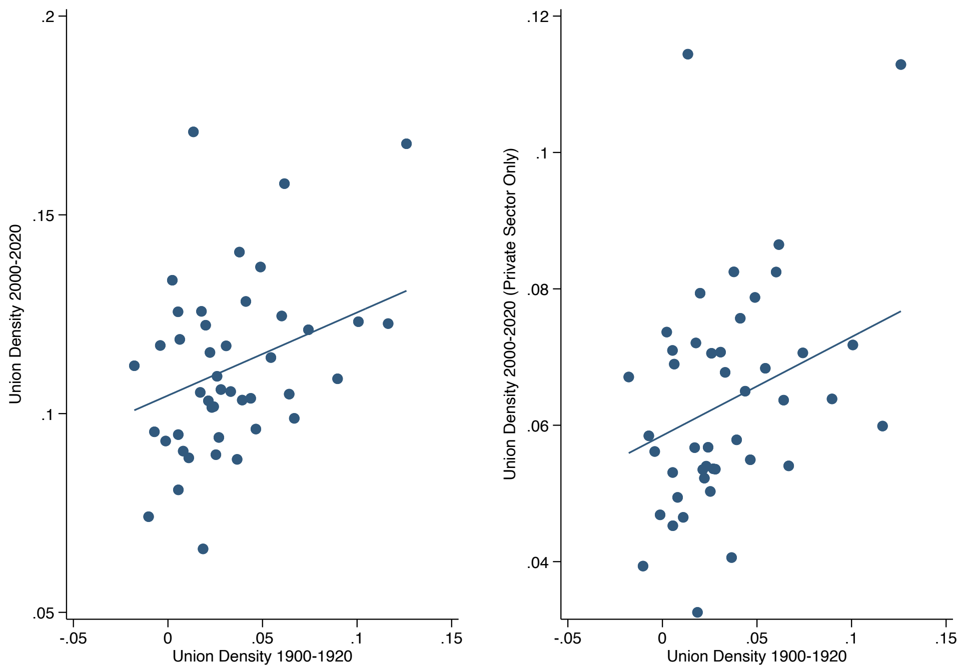

*Notes:* The figure plots the shares of union leaders of U.S.-born, Northern/Western Europe, and Southern/Eastern Europe origin (Panel A) and of Northern/Western and Southern/Eastern Europe ancestry (Panel B), at the beginning of each decade between 1900 and 1920. Union leaders are the delegates sent by the local union branches to the national convention of their union, or to the state conventions of the American Federation of Labor. The country of origin and the ancestry are inferred from delegates' last names, as described in Appendix A.2.

Figure A.5: Effect on the Composition of Union Leaders




*Notes:* Bars plot coefficients (with corresponding 95% confidence intervals) of a 2SLS regression of the share of union leaders with inferred country of origin (U.S., N/W Europe, S/E Europe, in Panel A) or ancestry (N/W and S/E Europe, Panel B) on the share of recently arrived immigrants. Union leaders are the delegates sent by the local union branches to the national convention of their union, or to the state conventions of the American Federation of Labor. On the left, the sample includes all counties as in Table 1.3 (in counties with no unionization, both shares are set to zero); on the right, the sample is restricted only to counties for which a union delegate is observed in every year. The country of origin and the ancestry are inferred from delegates' last names, as described in Appendix A.2 Formal estimates are presented in Table A.11.

Figure A.6: Persistence of Unionization




*Notes:* The figures shows a binscatter of the average levels of union density between 1900–1920 (x-axis) and the average levels of union density between 2000–2020 (y-axis), de-meaned by Census division fixed effects. The left panel shows on the y-axis unionization for both the public and the private sector; the right panel only for the private sector. Current data on union density are from Macpherson and Hirsch (2023), aggregated at the metropolitan-area level.

# A.1 Robustness Checks

## A.1.1 Alternative Shift-Share Instrument

As explained in Section 1.4.2, I replicate the analysis using an alternative instrument that relies on *predicted* flows of European immigration. More specifically, in equation (1.2), I replace the actual number of immigrants from country $j$ entering the U.S. between year $t-10$ and year $t$, with that predicted exploiting variation in weather shocks across European countries over time. This is motivated by previous work which has documented links between agricultural output and weather conditions, both in Europe during the Age of Mass Migration (Hatton and Williamson, 1995; Solomou and Wu, 1999) and in contemporary migration episodes (Feng et al., 2010).

I follow Sequeira et al. (2020a),[1] and estimate a relationship between weather shocks and immigration from each European country (for the period 1900–1920) using the following equation:

$$log(Immigr_{j,t}) = \sum_{s \in S} \sum_{k \in K} \beta_{j,s,k} I^{s,k}_{j,t-1} + u_{j,t} \tag{A.1}$$

where $log(Immigr_{j,t})$ is the log of immigrants from European country $j$ in year $t$; and $I^{s,k}_{j,t-1}$ is a dummy equal to 1 if the average precipitation (or temperature) in season $s \in \{$Spring, Summer, Fall, Winter$\}$ falls in the range $k$. As in Sequeira et al. (2020a), $k$ indexes a set of six weather shock categories: more than 3 standard deviations below the mean; between 2 and 3 standard deviations below the mean; between 1 and 2 standard deviations below the mean; between 1 and 2 standard deviations above the mean; between 2 and 3 standard deviations above the mean; and more than 3 standard deviations above the mean. The

[1] An analogous identification is also used by Tabellini (2020).

omitted category is the one of temperatures (or precipitations) that are within one standard deviation below or above the mean. Since there are six temperature categories and four seasons, there are 24 weather indicators in total.

The data on historical temperatures and precipitations come from Luterbacher et al. (2004) and Pauling et al. (2006), respectively. The data are measured four times annually (once during each season) and approximately at a 55-kilometer spatial resolution. Because the immigration data (from Willcox, 1929) are at the country-level, I average temperatures and precipitations over all grid-cells under cultivation in a country.[2] For this exercise, the sample includes nineteen European countries for which immigration, weather, and crop data are available.[3] In the baseline specification, I consider temperature shocks, but results are unchanged if using precipitations.

I separately estimate equation (A.1) for each European country in the sample. Figure A.7 shows the relationship between actual and predicted log immigration, displaying a strong positive correlation. Then, I predict the log immigrant flows for each country in each year, $\widehat{log(Immigr_{j,t})}$ using the $\widehat{\beta_{j,s,k}}$'s estimated from these regressions. Finally, I aggregate the predicted flows by decade and obtain:

$$\widehat{O}_{jt} = \sum_{t} exp[log(Immigr_{j,t})] \tag{A.2}$$

Table A.12 reports the first stage estimates. Although the F-stat is lower than the one of the main instrument (Table 1.2), it is still always above the conventional levels. Table A.13 shows the main results on the effect of immigration on the four unionization outcomes. Panel A reports the baseline estimates of Table 1.3 using the main instrument, while Panel B displays the estimates from using the alternative instrument based on weather shocks. In

[2]Information on historical land under cultivation is from Ramankutty and Foley (1999).

[3]These are: Austria, Belgium, Denmark, England, France, Germany, Greece, Hungary, Ireland, Italy, the Netherlands, Norway, Portugal, Russia, Scotland, Spain, Sweden, Switzerland, and Wales.

either case, all coefficients are highly statistically significant and positive.

### A.1.2 Matching Exercise

Similar to Bazzi et al. (2023), I conduct a matching exercise. I identify county pairs within the same state that have the closest number of Knights of Labor branches as a fraction of the county population, in 1880 and in 1890. In the absence of comprehensive information on unions affiliated with the American Federation of Labor before 1890, or of complete data on the union membership of the Knights of Labor, this is the best way to proxy for unionization before the time period of the analysis.

I present the results in Table A.14. In Panel A, I re-estimate the baseline specification of Table 1.3 for the counties that can be included in the county-pair strategy.[4] In Panels B and C, I re-estimate equation (1.1), replacing the baseline controls with fixed effects for the 800+ county pairs, interacted with year dummies. In Panel B, counties are matched on the number of Knights of Labor branches in 1890. In Panel C, on the one of 1880. The resulting coefficients identify the effect of immigration inflows on unionization for counties with nearly identical levels of union presence at baseline.[5] Despite the very demanding nature of this specification, reassuringly all the point estimates remain positive, large in magnitude, and similar to the baseline coefficients of Panel A.

[4]Not all counties can be matched in pairs (e.g., when there is an odd number of counties in a state). For this reason, the number of observations for the matching exercise is slightly lower than in the main estimation sample.

[5]In case of equal values of the matching variable, I further match counties on these or additional variables, in the following order: total number of Knights of Labor branches in the county, share of manufacturing labor force, share of agricultural labor force. This is meant to compare counties that have similar labor force composition at baseline. Further ties are then broken arbitrarily by a randomly generated number. Different choices of the ”secondary” matching variables do not affect the results.

### A.1.3 Controlling for Additional Baseline Characteristics

In this section, I address the possibility the instrument described in Section 1.4.2 may predict a higher immigrant share in counties that were already on a trajectory of higher unionization growth, for either economic or political reasons. In Table A.15, I re-estimate the baseline specification by further controlling for several characteristics measured at baseline and interacted with year dummies. This exercise is meant to reduce the concern that factors that are jointly correlated with the 1890 size of immigration and with the development of labor unions between 1900 and 1920 may bias the estimates.

**Connection to the railroad network.** Previous work has shown that, between 1860 and 1920, the timing of the connection to the railroad network had a positive effect on both the inflow of immigrants to a county and on its economic growth in the medium- and long-run (Sequeira et al., 2020a). Therefore, whether a county was crossed by a railroad or not may bias the estimates. To rule out this possibility, I use data from Atack (2016) to construct an indicator for whether each county in the sample was connected to the railroad network as of 1890, and interact this variable with year dummies (Table A.15, column 1).

**Share of immigrant population.** I directly control for the size of the 1890 immigrant population (total and European only), interacted with year dummies. This implies that the effects of immigration are identified exploiting variation only in the ethnic composition of immigrant enclaves across counties, holding constant the size of their foreign born populations. Since mechanically the instrument predicts higher immigration to counties with a larger stock of immigrants at baseline, by doing this I also address the concern that a larger 1890 immigrant population may itself have an independent and time-varying effect on unionization. Despite the highly demanding nature of this specification, all estimates remain statistically significant above the conventional levels (Table A.15, columns 2 and 3).

**Share of immigrant population.** Another potential confounding factor may be represented by the first waves of the Great Migration, which started around 1915 (Boustan, 2016). Although a limited cause of concern given the little overlap with the period studied, I address this possibility by controlling for the shares of Black population in each county in 1890, which will higher immigration rates of Black individuals based on chain migration, as previous work has shown (Boustan, 2010; Fouka et al., 2022). The findings are unchanged (Table A.15, column 4).

**Labor force composition.** I further control for the shares of the labor force in all major industries (agriculture, manufacturing, transportation, trade, manufacturing, and mining) and the share of the labor force in occupations covered by AFL-affiliated national unions in the period 1900–1920, all measured in 1890. These regressions therefore estimate the effect of immigration among counties with similar initial composition of workers across sectors. The results are all positive and statistically significant, and larger in magnitude (Table A.15, columns 5 and 6).

**Average income and economic growth.** Similarly, I control for the initial levels of average income (proxied by the occupational income score) and economic growth (measured by the growth rate of manufacturing output), to reduce any concern that counties with different economic conditions may have attracted more immigration earlier on and also witnessed a different growth of labor unions over time. The estimates are robust to the inclusion of these additional controls (Table A.15, columns 7 and 8).

**Share of farm land.** An additional concern is represented by the structural transformation away from agriculture towards manufacturing that occurred in the U.S. between 1880 and 1920 (Eckert and Peters, 2022b). This may have implied larger growth rates for counties that were rural at the beginning of the time period, with potential implications on the evolution of labor unions too. Although in the baseline specification I already control for the urban share of the population in 1890, I further include interactions between year dummies and the 1890 share of land in farms. The results are almost unchanged (Table A.15, column 9).

**Vote shares for the Democratic Party.** Finally, I control for a measure of the political ideology of each county, namely the average vote shares for the Democratic Party in the presidential elections of 1888 and 1892. Also in this case, all the point estimates are remarkably similar to the baseline estimates.

### A.1.4 Additional Robustness Checks

**Alternative baseline specification.** Table A.16 reports results from using different specifications. In particular, in columns 1 to 3 I estimate less stringent specifications, by gradually including the two controls that are part of the preferred specification (the 1890 share of the urban population and the 1880 labor force participation rate). In columns 4 to 6, I do the same, while also always including state by year dummies, implying that the coefficients are estimated from changes in the fraction of immigrants within the same county over time, compared to other counties in the same state in a given year. The estimates are quantitatively and qualitatively unchanged.

**Drop potential outliers.** I verify that the results are robust to omitting counties with very large and very low levels of the dependent and independent variables, which could be potential outliers. In Table A.17, I re-estimate the baseline results dropping counties with measures of unionization (Panel A) and immigration (Panel B) below the $1^{st}$ and above the $99^{th}$ percentile. Reassuringly, in all cases the coefficients are in line with those reported in Table 1.3.

**Alternative computations of standard errors.** In the baseline specification, standard errors are clustered at the county level. To address potential concerns of spatial correlation, in Table A.18 I verify that the precision of the estimates is unchanged when clustering standard errors at the SEA level (Panel A) and when computing Conley (1999) standard errors (with a 100km bandwidth).

**Population-weighted regressions.** I also re-estimate the results of Table 1.3 weighting the observations by total population, measured in the previous decade (Table A.19). By doing so, the estimates will return the effects for the average county. All coefficients remain positive, and if anything, are larger than the ones estimated with unweighted regressions. Except for column 4, whose coefficient is slightly above the conventional levels of significance, all other estimates are significant at either the 5% or 1% level.

**Alternative samples.** In Table A.20, I relax the restriction of having a balanced sample, and re-estimate the baseline specification on the full sample of counties for which the unions data are available. This yields a larger number of observations: 5,971 against the 5,025 of the baseline specification. All coefficients are remarkably similar to the ones of Table 1.3, both in terms of magnitude and significance. In Table A.21, I re-estimate the baseline regressions omitting the counties in the South. This exercise is motivated by the fact that this region of the U.S. received limited volumes of immigration between 1900 and 1920, and also saw smaller labor unions' activity. Hence, a possible concern is that Southern counties may be driving the positive relationship between immigration and unionization. Reassuringly, even after dropping such counties, all estimates remain positive, statistically significant, and with magnitude similar to the ones reported in Table 1.3 (although less precisely estimated in some cases, due to the smaller sample size).

**Alternative definitions of union density.** The preferred definition of union density used throughout the paper is the number of union members divided by the total labor force in occupations covered by the AFL-affiliated national unions during the period 1900–1920, collected from Stewart (1926). This measure has the main advantage of not being influenced by the relative importance of such occupations in the labor force. In Table A.22, I show that the results are unchanged when using different definitions of the dependent variable. In particular, in column 2, the number of union members is divided by the total labor force in occupations covered by any labor union in existence during this period (regardless of whether it was affiliated with the AFL or not); and, in column 3, by the total labor force in any occupation not in the agricultural industry. As expected, the magnitudes change, but all coefficients remain statistically significant.

**Test of pre-trends.** The validity of the shift-share instrument defined by equation (1.2) rests on the key assumption that counties receiving more immigrants (from each country) before 1890 must not be on different trajectories for the evolution of unionization in subsequent decades (see also Borusyak et al., 2022 and Goldsmith-Pinkham et al., 2020). Although the results of Figure 1.8 already reduce the concerns about this assumption being invalidated, in Table A.23, I test for pre-trends more directly, regressing the pre-period change (from 1880 to 1890) in several outcomes on unionization, population, and economic growth, on the 1900 to 1920 change in immigration predicted by the instrument. Panel A reports the coefficients from reduced form regressions. Panel B display 2SLS estimates, although conclusions from this second set of cofficients should be taken with caution, given the low F-stat. All regressions control for urbanization and labor force participation rate in 1880, in an analogous way to the specification of Table 1.3. The choice of the dependent variables is constrained by data availability. Given the absence of data on union membership before the sample period for which I construct the dataset, and the fact that the American Federation of Labor was constituted only in 1886, I measure unionization with the number of branches of the Knights of Labor (Garlock, 2009); for economic outcomes, I rely on the Census of Manufactures, and use information on the number of workers in manufacturing, the number of manufacturing establishments, and the value of manufacturing output, which are available for both 1880 and 1890 (Haines, 2010b). Reassuringly, no coefficient of Table A.23 is statistically significant. These results indicate that, before 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or in other economic variables.

Table A.12: First Stage of the Alternative Instrumental Variable Estimation

| | *Dependent variable: Share Immigrants* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Predicted Share Immigrants | 0.157*** | 0.142*** | 0.139*** |
| | (0.038) | (0.034) | (0.033) |
| Observations | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.024 | 0.024 | 0.024 |
| Indep. var. mean | 0.084 | 0.084 | 0.084 |
| KP F-stat | 17.44 | 17.20 | 17.66 |
| 1890 Urban Share | No | Yes | Yes |
| 1880 LF Part. Rate | No | No | Yes |

*Notes:* Observations are at the county-decade level. The table reports the first stage of the alternative instrument described in Appendix A.1.1. The dependent variable is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The main regressor of interest is the predicted number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the 1890 male population in the county. All regressions include county and year fixed effects. The following controls, interacted with year dummies, are also included: the 1890 share of urban population (column 2); and, the 1880 male labor force participation rate (column 3). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.13: Alternative Shift-Share Instrument Using Predicted Immigration Flows

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Main instrument* | | | | |
| Share Immigrants | 1.572** | 0.285** | 2.918*** | 260.959** |
| | (0.699) | (0.117) | (0.854) | (110.674) |
| KP F-stat | 35.14 | 35.14 | 35.14 | 35.14 |
| *Panel B: Alternative instrument* | | | | |
| Share Immigrants | 2.594*** | 0.343** | 4.439*** | 513.066** |
| | (0.951) | (0.159) | (1.157) | (213.540) |
| KP F-stat | 17.66 | 17.66 | 17.66 | 17.66 |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.265 | 0.039 | 1.624 | 29.936 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. In Panel A, the instrument used to predict immigration is the one described in Section 1.4.2. In Panel B, the instrument is the one that uses predicted rather than actual immigration flows (predicted using weather shocks in each European country, following Sequeira et al., 2020a), as described in Appendix A.1.1. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.14: Matching Counties with Similar Union Presence at Baseline

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Baseline (matching sample)* | | | | |
| Share Immigrants | 1.522** | 0.234** | 3.427*** | 220.640** |
| | (0.620) | (0.106) | (0.849) | (105.967) |
| KP F-stat | 36.46 | 36.46 | 36.46 | 36.46 |
| *Panel B: Matching on 1890 union presence* | | | | |
| Share Immigrants | 2.285* | 0.367 | 3.092 | 633.162** |
| | (1.377) | (0.259) | (1.939) | (266.600) |
| KP F-stat | 22.95 | 22.95 | 22.95 | 22.95 |
| *Panel C: Matching on 1880 union presence* | | | | |
| Share Immigrants | 2.354 | 0.454 | 3.673* | 599.452* |
| | (1.682) | (0.310) | (2.120) | (329.503) |
| KP F-stat | 13.29 | 13.29 | 13.29 | 13.29 |
| Observations | 4,986 | 4,986 | 4,986 | 4,986 |
| Dep. var. mean | 0.266 | 0.039 | 0.404 | 30.017 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county, year, and, in Panel B and C, county pair by year fixed effects. County pairs are matched within states on the 1890 number of Knights of Labor branches (from Garlock, 2009) divided by county population. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county-pair, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

# Table A.15: Controlling for Additional Baseline Characteristics

| | *Control: Year Dummies Interacted with Baseline Value of* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Dummy Conn. to Railroad (1) | Immigrant Share (Tot.) (2) | Immigrant Share (Euro) (3) | Black Pop. Share (4) | LF Shares in Major Ind. (5) | LF Share in AFL Occ. (6) | Log. Occ. Inc. Score (7) | Mfg. Output Growth (8) | Share of Farm Land (9) | Dem. Vote Share (10) |
| | *Panel A - Dependent variable: Union Presence* | | | | | | | | | |
| Share Immigrants | 1.606** | 2.545** | 2.260* | 2.137** | 2.478** | 2.544** | 2.240** | 1.571** | 1.522* | 1.572** |
| | (0.718) | (1.292) | (1.180) | (0.832) | (0.973) | (1.027) | (0.911) | (0.738) | (0.826) | (0.776) |
| | *Panel B - Dependent variable: Union Density* | | | | | | | | | |
| Share Immigrants | 0.295** | 0.496** | 0.496*** | 0.384*** | 0.467*** | 0.496*** | 0.428*** | 0.214* | 0.295** | 0.285** |
| | (0.122) | (0.196) | (0.184) | (0.143) | (0.150) | (0.155) | (0.138) | (0.113) | (0.136) | (0.136) |
| | *Panel C - Dependent variable: Log # Branches* | | | | | | | | | |
| Share Immigrants | 2.907*** | 4.193*** | 4.113*** | 3.638*** | 4.116*** | 4.279*** | 3.865*** | 2.961*** | 2.867*** | 2.920*** |
| | (0.866) | (1.612) | (1.497) | (1.032) | (1.199) | (1.276) | (1.121) | (0.919) | (0.991) | (0.961) |
| | *Panel D - Dependent variable: Avg. Branch Size* | | | | | | | | | |
| Share Immigrants | 301.831*** | 599.706*** | 517.435*** | 332.322** | 493.471*** | 503.701*** | 402.406*** | 264.257** | 346.412*** | 268.755** |
| | (116.271) | (223.819) | (197.665) | (133.153) | (166.688) | (174.038) | (151.566) | (120.239) | (131.221) | (125.335) |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 | 5,025 | 5,025 | 5,025 | 4,851 | 5,025 | 4,893 |
| KP F-stat | 32.72 | 14.92 | 17.06 | 29.32 | 24.83 | 24.05 | 26.83 | 31.63 | 28.94 | 31.02 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (Panel A); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (Panel B); the log number of union branches (Panel C); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (Panel D). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with the 1890 share of urban population, the 1880 male labor force participation rate, and the following variables: an indicator whether a county was connected to the railroad network in 1890 (column 1); the immigrant share of the population in 1890 (column 2); the European immigrant share of the population in 1890 (column 3); the Black share of the population in 1890 (column 4); the shares of the male labor force in the mining, manufacturing, construction, trade, transportation, and agricultural industries in 1880 (column 5); the share of the male labor force in occupations covered by AFL-affiliated national unions in 1880 (column 6); the log of the average occupational income score in 1880 (column 7); the growth rate of manufacturing output between 1880 and 1890 (column 8); the share of land used in farming in 1890 (column 9); the average vote share for the Democratic Party in the presidential elections of 1888 and 1892 (column 10). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.16: Using Alternative Baseline Specifications

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | *Panel A - Dependent variable: Union Presence* | | | | | |
| Share Immigrants | 1.511** | 1.573** | 1.572** | 2.254* | 2.444* | 2.482* |
| | (0.614) | (0.685) | (0.699) | (1.163) | (1.296) | (1.331) |
| | *Panel B - Dependent variable: Union Density* | | | | | |
| Share Immigrants | 0.234** | 0.275** | 0.285** | 0.270 | 0.365* | 0.378* |
| | (0.105) | (0.116) | (0.117) | (0.181) | (0.200) | (0.204) |
| | *Panel C - Dependent variable: Log # Branches* | | | | | |
| Share Immigrants | 3.405*** | 2.923*** | 2.918*** | 3.704** | 3.574** | 3.592** |
| | (0.841) | (0.836) | (0.854) | (1.558) | (1.574) | (1.617) |
| | *Panel D - Dependent variable: Avg. Branch Size* | | | | | |
| Share Immigrants | 218.850** | 248.637** | 260.959** | 306.912 | 386.158* | 405.173* |
| | (104.895) | (109.293) | (110.674) | (191.680) | (203.537) | (208.878) |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| State-Year FE | No | No | No | Yes | Yes | Yes |
| 1890 Urban Share | No | Yes | Yes | No | Yes | Yes |
| 1880 LF Part. Rate | No | No | Yes | No | No | Yes |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 | 5,025 | 5,025 |
| KP F-stat | 37.28 | 35.33 | 35.14 | 15.75 | 15.15 | 14.98 |

*Notes:* Observations are at the county-decade level. Dependent variables are: the number of union members divided by the male labor force in occupations represented by the American Federation of Labor (Panel A); the log number of union branches (Panel B); the number of union members divided by the number of branches, or zero if no union is present (Panel C); or, an indicator for whether the county has a positive union membership in any occupation (Panel D). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects. Columns 2, 3, 5, and 6 include year dummies interacted with the 1890 share of urban population. Columns 3 and 6 include year dummies interacted with the 1880 male labor force participation rate. Columns 4 to 6 include state by year fixed effects. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.17: Dropping Outliers

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Outliers of dependent variable* | | | | |
| Share Immigrants | 1.573** | 0.286*** | 2.931*** | 251.776*** |
| | (0.700) | (0.100) | (0.878) | (76.279) |
| Observations | 4,966 | 4,966 | 4,968 | 4,969 |
| Dep. var. mean | 0.257 | 0.031 | 1.619 | 25.425 |
| Indep. var. mean | 0.024 | 0.024 | 0.023 | 0.023 |
| KP F-stat | 34.82 | 34.82 | 33.50 | 32.80 |
| *Panel B: Outliers of Share Immigrants* | | | | |
| Share Immigrants | 2.226** | 0.434*** | 3.923*** | 435.089*** |
| | (1.045) | (0.165) | (1.225) | (166.609) |
| Observations | 4,972 | 4,972 | 4,972 | 4,972 |
| Dep. var. mean | 0.262 | 0.039 | 1.619 | 29.585 |
| Indep. var. mean | 0.022 | 0.022 | 0.022 | 0.022 |
| KP F-stat | 23.02 | 23.02 | 23.02 | 23.02 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. Observations below the 1st or above the 99th percentile of the dependent variable (Panel A), or of the independent variable (Panel B), are excluded from the sample. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.18: Computing Standard Errors with Alternative Procedures

| | *Dependent variable:* | | | |
| --- | --- | --- | --- | --- |
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| *Panel A: Clustered by SEA* | | | | |
| Share Immigrants | 1.572** | 0.285* | 2.918** | 260.959** |
| | (0.797) | (0.153) | (1.128) | (131.748) |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.265 | 0.039 | 1.619 | 29.936 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 19.89 | 19.89 | 19.89 | 19.89 |
| *Panel B: Conley (1999), 100km bandwidth* | | | | |
| Share Immigrants | 1.614** | 0.283** | 2.881*** | 274.233** |
| | (0.650) | (0.133) | (0.961) | (124.731) |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.265 | 0.039 | 1.619 | 29.936 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 28.98 | 28.98 | 28.98 | 28.98 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors are shown in parentheses. In Panel A, standard errors are robust and clustered by State Economic Area (SEA). In Panel B, standard errors are computed with the procedure described by Conley (1999) to account for spatial correlation, with a bandwidth of 100km. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.19: Weighting Counties by Population

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| Share Immigrants | 3.695** | 0.476*** | 3.893*** | 1,017.340 |
| | (1.786) | (0.173) | (1.320) | (650.488) |
| Observations | 5,025 | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.265 | 0.039 | 1.619 | 29.936 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 | 0.024 |
| KP F-stat | 28.13 | 28.13 | 28.13 | 28.13 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. Observations are weighted by the total population in the previous decade. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.20: Using an Unbalanced Sample

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence | Union Density | Log # Branches | Avg. Branch Size |
| | (1) | (2) | (3) | (4) |
| Share Immigrants | 1.351** | 0.236** | 2.612*** | 182.677** |
| | (0.572) | (0.101) | (0.706) | (92.610) |
| Observations | 5,971 | 5,971 | 5,971 | 5,971 |
| Dep. var. mean | 0.261 | 0.039 | 1.619 | 30.600 |
| Indep. var. mean | 0.025 | 0.025 | 0.025 | 0.025 |
| KP F-stat | 43.30 | 43.30 | 43.30 | 43.30 |

*Notes:* Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.21: Excluding the South

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Union Presence<br>(1) | Union Density<br>(2) | Log # Branches<br>(3) | Avg. Branch Size<br>(4) |
| Share Immigrants | 2.137** | 0.280* | 3.237*** | 307.841** |
| | (0.901) | (0.150) | (1.106) | (147.858) |
| Observations | 3,180 | 3,180 | 3,180 | 3,180 |
| Dep. var. mean | 0.338 | 0.050 | 1.619 | 40.686 |
| Indep. var. mean | 0.035 | 0.035 | 0.035 | 0.035 |
| KP F-stat | 26.43 | 26.43 | 26.43 | 26.43 |

*Notes:* Observations are at the county-decade level. The estimation sample is restricted to counties in the Northeast, Midwest or West regions. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.22: Alternative Definitions of Union Density

| | *Dependent variable: # Union Members /* | | |
|---|---|---|---|
| | (Baseline)<br>LF in Occ. Covered<br>by AFL Unions<br>(1) | LF in Occ. Covered<br>by Any Union<br>(2) | LF in All<br>Non-Agric. Occ.<br>(3) |
| Share Immigrants | 0.285** | 0.254** | 0.115* |
| | (0.117) | (0.110) | (0.068) |
| Observations | 5,025 | 5,025 | 5,025 |
| Dep. var. mean | 0.039 | 0.036 | 0.021 |
| Indep. var. mean | 0.024 | 0.024 | 0.024 |
| KP F-stat | 35.14 | 35.14 | 35.14 |

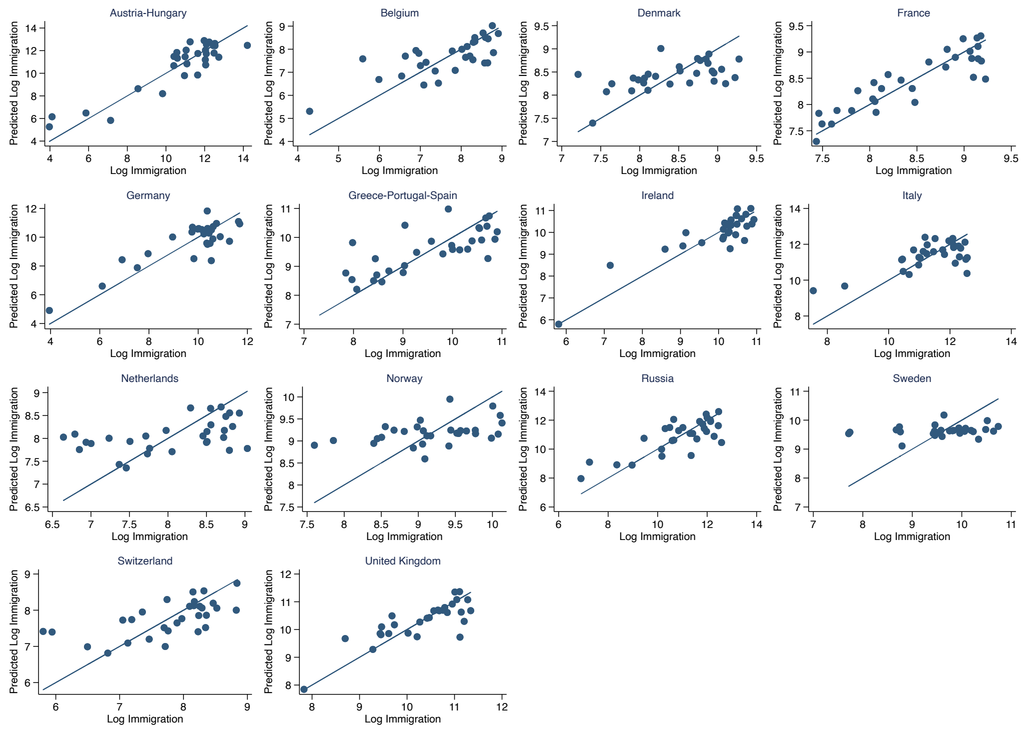
*Notes:* Observations are at the county-decade level. The dependent variables are the number of union members divided by: the total male labor force in occupations represented by the American Federation of Labor (column 1); the total male labor force in occupations represented by any labor union (column 2); the total male labor force in any non-agricultural occupation (column 3). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 1.4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.23: Test of Pre-Trends in Unionization and Economic Outcomes

| | *Dependent variable (1880-1890 change):* | | | | | |
|---|---|---|---|---|---|---|
| | Log # Branches (1) | # Branches / Population (2) | Log Pop. Density (3) | Share Pop. in Mfg. (4) | Log # Mfg. Establ/Worker (5) | Log Mfg. Output/Worker (6) |
| *Panel A: Reduced Form* | | | | | | |
| Pred. Share Immigrants | -0.077 | -0.001 | 0.184 | -0.016 | 0.250 | 0.589 |
| (1900–1920 change) | (0.511) | (0.000) | (0.433) | (0.029) | (0.896) | (0.512) |
| *Panel B: 2SLS* | | | | | | |
| Share Immigrants | -2.292 | -0.019 | 5.493 | -0.437 | 6.647 | 15.654 |
| (1900–1920 change) | (15.195) | (0.022) | (14.538) | (0.882) | (24.243) | (20.463) |
| KP F-stat | 1.28 | 1.28 | 1.28 | 1.61 | 1.60 | 1.60 |
| Observations | 1,675 | 1,675 | 1,675 | 1,651 | 1,648 | 1,648 |

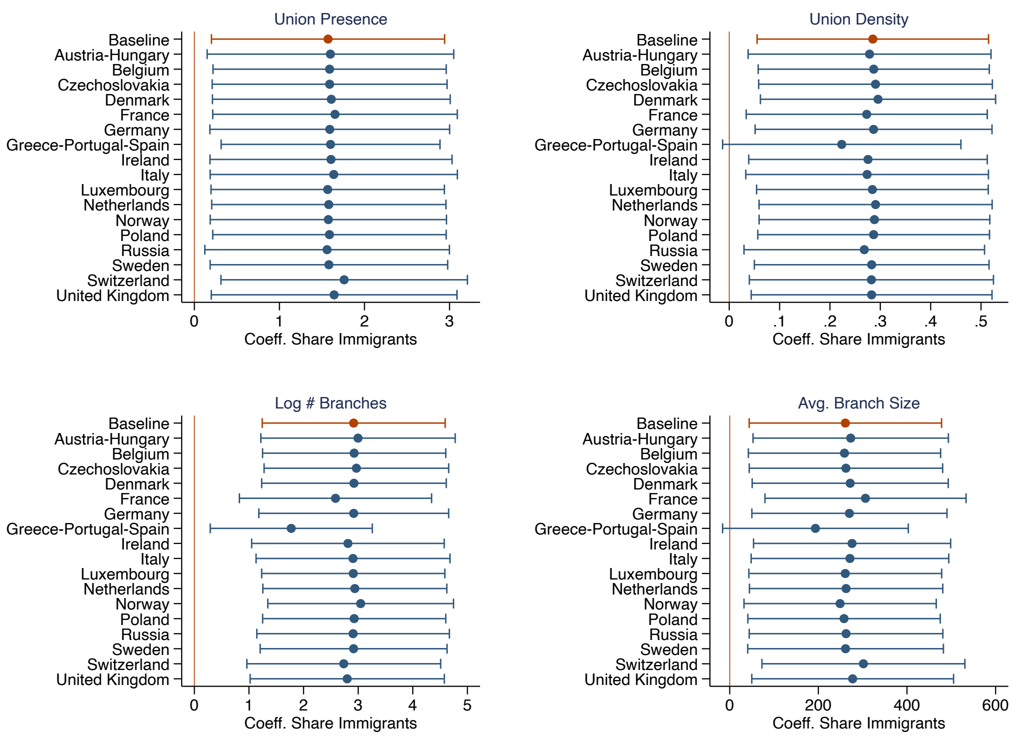
*Notes:* Observations are at the county level. The dependent variables are the 1880–1890 change in: the log of the number of Knights of Labor branches (column 1); the number of Knights of Labor branches divided by population (column 2); the log of population density (column 3); the share of the population employed in manufacturing (column 4); the log of the number of manufacturing establishments divided by the number of workers in manufacturing (column 5); the log of the manufacturing output divided by the number of manufacturing workers (column 6). The regressor of interest is 1900–1920 change in the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county, as predicted by the instrument described in Section 1.4.2. Panel A reports reduced form coefficients; Panel B displays 2SLS estimates. All regressions control for the 1880 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Robust standard errors are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Figure A.7: Actual Versus Predicted Immigration Using Temperature Shocks




*Notes:* The figure displays the correlation between the actual (log) immigrant flows and those predicted using temperature shocks from equation (A.1), separately for the European countries in the sample.

Figure A.8: 2SLS Coefficients, Controlling for Initial Country Shares




*Notes:* The figures plot the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, augmenting the specification reported in Table 1.3 with the 1890 immigrant share from each sending country (relative to all immigrants from that country in the U.S. in that year), separately. The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county.

## A.2 Mapping Delegates' Last Names to Origins and Ancestry

In Section 1.6, I use union delegates' last names to infer their ethnic origins. In this section, I describe how this mapping is constructed.

I start with de-anonymized full count U.S. Census data between 1900 and 1920, which contain information on names and birthplaces of the whole U.S. population. I then restrict the sample to the male population, and classify individuals depending on their country of birth and their ancestry, defined as their country of birth if born abroad, or the country of birth of the father if born in the U.S. from foreign-born father.

Then, I construct two probabilistic mappings: one between the last names and the country of birth, and one between the names and the ancestry. I compute $p_{l,e,t}$, the probability that a person with last name $l$ is of country of birth (ancestry) $e$ in year $t$, as $w_{l,e,t} = \frac{n_{l,e,t}}{N_{l,t}}$, where $n_{l,e,t}$ is the number of individuals with last name $l$ from country of birth (ancestry) $e$ in year $t$, and $N_l, t$ is the total number of individuals with last name $l$ in year $t$. Based on this mapping, for example, the last name Smith in 1900 – the most common name in that year – is 82% U.S.-born, 5% British, and 5% German; Anderson – the eighth most common name – is 46% U.S.-born, 32% Swedish, and 9% Norwegian; and, Murphy is 47% Irish, 45% U.S.-born, and 2% British.

Finally, after standardizing the names (e.g., remove spaces, hyphens, etc.), I match these probabilities to the delegates' last names from the digitized data. After collapsing the data at the county level, I obtain the expected number of delegates of country of birth (ancestry) $e$ in county $c$ and year $t$, which I then use to construct the shares of delegates from each country of birth (ancestry) that I employ in the analysis.

## A.3 Index of Residential Segregation

In Section 1.6, I explore the heterogeneity of the effects of European immigration on unionization, by splitting counties above and below the sample median of the 1880 index of residential segregation of immigrants. In this section, I briefly described how the measure is constructed.[6]

First, I identify next-door neighbors from full-count U.S. Census data. Then, I follow the procedure described in Logan and Parman (2017), and I construct an indicator variable equal to one if a European immigrant has a next-door neighbor who is U.S.-born (from both U.S.-born parents).[7] The sum of this indicator variable across all European households in the county gives the number of European households with a U.S.-born next-door neighbor, $x_c$.

This number is first compared to the expected number that one would see under complete integration, $E(\overline{x_c})$, i.e., a situation in which individuals were randomly assigned within neighborhoods by ethnic group. Then, $x_c$ is compared to the number of immigrants with U.S.-born neighbors that one would observe under complete segregation, $E(\underline{x_c})$, i.e., a situation where the immigrants living next to a U.S.-born would be only the individuals on either end of the immigrant neighborhood.

The index of residential segregation in county c, $\eta_c$, is computed as:

$$\eta_c = \frac{E(\overline{x_c}) - x_c}{E(\overline{x_c}) - E(\underline{x_c})}. \tag{A.3}$$

This segregation measure increases as European residents are more segregated within a county. The measure equals zero in the case of random assignment of neighbors (no

[6]For a more detailed discussion, I refer the reader to Logan and Parman (2017).

[7]The original measure in Logan and Parman (2017) is constructed to compute an index of residential segregation for Black households. In the sample, instead of Black and white, the groups will be: foreign-born Europeans, U.S.-born from U.S.-born parents, and others.

segregation), and equals one in the case of complete segregation.

## A.4 Labor Unions in Europe

Data on the development of labor unions in Europe used in Section 1.6.3 come from Crouch (1993). Estimates on union membership at the country level are available approximately every twenty or thirty years, starting in 1870. In most countries, the right to organize had been gained between 1860 and 1870, and was still often precarious. Similarly to the U.S., organization was limited to the skilled crafts and mining. At the turn of the 20$^{\text{th}}$ century, the only countries with an active and strong labor movement where the U.K. and Ireland. In 1900, there had been some, but limited, union activity also in Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, and Switzerland, although most of it had started only in the year 1900 or after (Crouch, 1993).

In Section 1.6.3, I separately predict (and estimate the impact of) immigration from the U.K. and Ireland (i.e., those with an active labor movement), and all the other European countries in the sample. The idea behind this exercise reflects the fact that individuals emigrating from countries with stronger unions may have been exposed to the experience of collective bargaining by the time they arrived in the U.S., and therefore might have been particularly interested in forming or joining labor unions in their new country. Table A.24 reports union membership at the national level for the years 1870 and 1900.

## A.5 Dataset on Unionization

I provide a validation of the estimates of union density by investigating their correlation with the only other measures available in a historical period. This comes from Farber et al. (2021), who harmonize household-level survey data from Gallup starting in 1937. In Figure

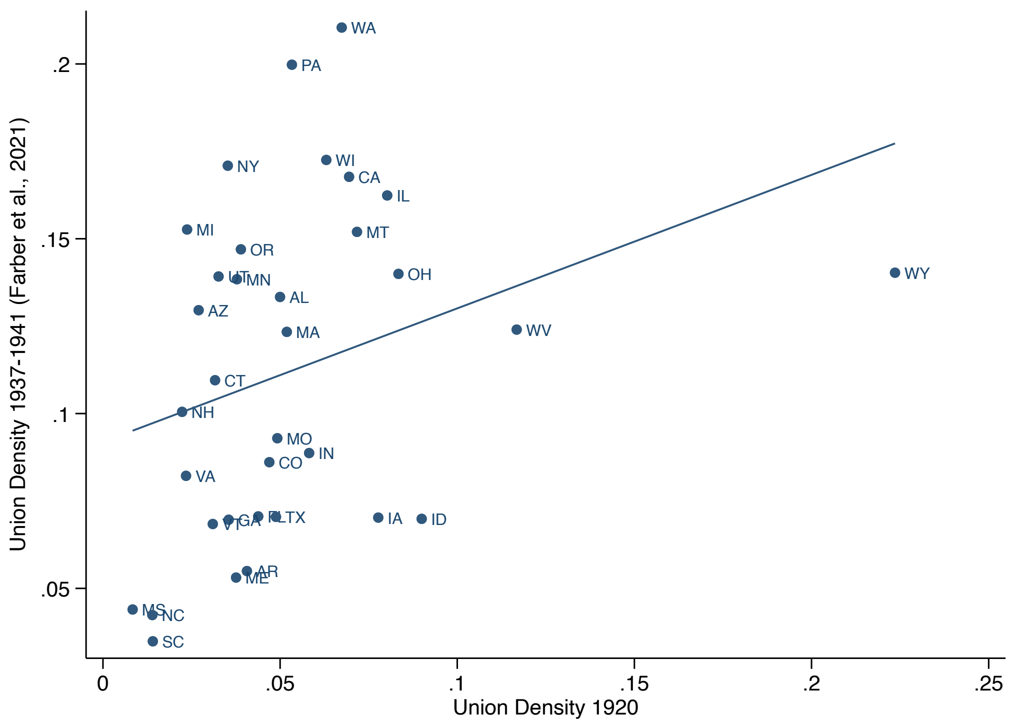
Table A.24: Union Membership in European Countries

| | Members (as % of LF) | |
|---|---|---|
| Country | 1870 | 1900 |
| Austria | 0.28 | 1.00 |
| Belgium | 2.42 | 3.29 |
| Denmark | 0.54 | 8.76 |
| France | 0.20 | 2.99 |
| Germany | 0.39 | 3.40 |
| Italy | n.a. | 3.07 |
| Norway | n.a. | 2.30 |
| Sweden | n.a. | 2.53 |
| U.K. and Ireland | 8.32 | 12.50 |

*Notes:* This table presents estimates of union membership in European countries for the years 1870 and 1900. Data are from Crouch (1993).

A.9, I show a scatter plot between the two measures. Since the data from Farber et al. (2021) are at the state level, I aggregate union membership in the data at the same unit and, to improve comparability with their measure, I divide it by the total non-agricultural labor force in the state. Unfortunately, the two sources do not overlap in time. Hence, I plot on the x-axis thr measure in the last year of observation (Census year 1920) and the measure from Farber et al. (2021) as an average of the first five years of observations (1937–1941). Although the two measures do not agree in levels (and they are not expected to, since by 1937 several industrial unions affiliated with the Congress of Industrial Organizations had been constituted, which represented large masses of workers previously unorganized), the two measures display a positive correlation. The correlation coefficient is over 0.3, and approaches 0.4 once Wyoming (an outlier in the graph) is excluded from the sample.

Figure A.9: Correlation Between Data of This Paper and State-Level Gallup Data




*Notes:* The figure plots a scatter plot for state-level union density measured in 1920 using the newly collected archival data (x-axis) and average union density between 1937–1941 measured using Gallup data as in Farber et al. (2021). See Section 1.3 for more details on the dataset on labor unions I assemble for the period 1900–1920.

# Appendix B

## Appendix Tables and Figures

## The Impact of the Chinese Exclusion Act on U.S. Economic Development

Table B.1: Summary Statistics: Population, Labor Force, and Economic Outcomes (1880)

| | A. All Counties | | | B. 1880 Ch. Share >= p50 | | | C. 1880 Chinese Sh. < p50 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. | Obs. | Mean | Std. Dev. |
| **A. Population 1880** | | | | | | | | | |
| Total Population | 231 | 5,555.64 | 16,520.05 | 115 | 8,067.20 | 22,760.61 | 116 | 3,065.72 | 4,440.89 |
| Urban Share | 231 | 0.05 | 0.15 | 115 | 0.07 | 0.19 | 116 | 0.02 | 0.08 |
| Immigrant Share | 231 | 0.27 | 0.12 | 115 | 0.33 | 0.12 | 116 | 0.21 | 0.09 |
| Age | 231 | 25.96 | 2.23 | 115 | 26.76 | 2.27 | 116 | 25.16 | 1.89 |
| Male Share | 231 | 0.67 | 0.08 | 115 | 0.68 | 0.07 | 116 | 0.66 | 0.09 |
| White Share | 231 | 0.92 | 0.09 | 115 | 0.87 | 0.09 | 116 | 0.97 | 0.04 |
| Chinese Share | 231 | 0.06 | 0.08 | 115 | 0.12 | 0.09 | 116 | 0.01 | 0.01 |
| Other Races Share | 231 | 0.01 | 0.03 | 115 | 0.01 | 0.02 | 116 | 0.02 | 0.04 |
| **B. Labor Force 1880 (Men 15-64 Only)** | | | | | | | | | |
| Total Labor Force | 231 | 2,251.64 | 6,267.85 | 115 | 3,431.57 | 8,625.44 | 116 | 1,081.88 | 1,431.12 |
| Mfg. Share of Labor Force | 231 | 0.07 | 0.07 | 115 | 0.08 | 0.09 | 116 | 0.07 | 0.06 |
| Mining Share of Labor Force | 231 | 0.13 | 0.16 | 115 | 0.19 | 0.18 | 116 | 0.07 | 0.10 |
| Railroad Share of Labor Force | 231 | 0.02 | 0.05 | 115 | 0.03 | 0.06 | 116 | 0.01 | 0.02 |
| Agric. Share of Labor Force | 231 | 0.34 | 0.18 | 115 | 0.29 | 0.16 | 116 | 0.39 | 0.18 |
| Share Skilled | 231 | 0.18 | 0.07 | 115 | 0.18 | 0.07 | 116 | 0.19 | 0.08 |
| Share Unskilled | 231 | 0.42 | 0.17 | 115 | 0.51 | 0.16 | 116 | 0.34 | 0.14 |
| Share Managers | 231 | 0.05 | 0.04 | 115 | 0.05 | 0.02 | 116 | 0.06 | 0.06 |
| Share Literate | 231 | 0.91 | 0.07 | 115 | 0.90 | 0.06 | 116 | 0.92 | 0.07 |
| **C. Productivity 1880** | | | | | | | | | |
| Income Score | 231 | 19.84 | 2.56 | 115.00 | 20.55 | 2.15 | 116.00 | 19.14 | 2.73 |
| Mfg. Total Output | 231 | 15,847.88 | 137,911.46 | 115.00 | 28,330.14 | 194,946.08 | 116.00 | 3,473.23 | 7,440.80 |
| Value of Farm Land | 231 | 39,611.16 | 83,175.07 | 115.00 | 51,148.49 | 98,918.71 | 116.00 | 28,173.29 | 62,219.67 |
| Connected to Railroad | 226 | 0.39 | 0.49 | 114.00 | 0.46 | 0.50 | 112.00 | 0.32 | 0.47 |

*Notes:* Observations are at the county and year level. The data are from U.S. Census of 1880.

Table B.2: Effect on Non-Chinese Immigrant Individuals

| | Dependent Variable | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Labor Supply | | | | | | | | | |
| | Pop. (1) | Urban Share (2) | Total (3) | Mfg. (4) | Mining (5) | Railroad (6) | Agric. (7) | Share Literate (8) | Skilled (9) | Unskilled (10) | Managers (11) | Income Score (12) |
| *Dependent Variable Mean* | *6.917* | *0.164* | *6.335* | *3.722* | *2.942* | *2.952* | *5.085* | *0.906* | *4.703* | *5.317* | *3.495* | *3.044* |
| *– in 1880* | *5.926* | *0.0549* | *5.554* | *3.063* | *2.728* | *1.168* | *4.247* | *0.905* | *3.911* | *4.595* | *2.715* | *3.036* |
| Post x High Chinese Share | -0.69*** | -0.03* | -0.72*** | -0.49*** | -0.63*** | -0.50*** | -0.55*** | -0.02*** | -0.65*** | -0.84*** | -0.63*** | -0.02** |
| | (0.09) | (0.02) | (0.09) | (0.11) | (0.12) | (0.11) | (0.08) | (0.01) | (0.09) | (0.10) | (0.09) | (0.01) |
| Observations | 1,611 | 1,610 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,384 | 1,611 | 1,611 | 1,611 | 1,610 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of total population (col. 1), the share of urban population (col. 2), the log of the total labor force (col. 3), the log of the labor force in the sector stated in the column headings (col. 4 - col. 7), the share of literate (col. 8), the log of total number of workers in the skill category stated in the column headings (col. 9 - col. 11), or the log of the occupational income score (col. 12). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.3: Robustness Check: Include Women in Sample

| | Labor Supply | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total (1) | Mfg. (2) | Mining (3) | Railroad (4) | Agric. (5) | Share Literate (6) | Skilled (7) | Unskilled (8) | Managers (9) | Income Score (10) |
| | | | | | | A. Chinese | | | | |
| *Dependent Variable Mean* | *3.064* | *0.679* | *0.805* | *0.447* | *1.238* | *0.773* | *1.467* | *2.776* | *1.222* | *2.999* |
| *– in 1880* | *4.197* | *1.208* | *1.890* | *1.045* | *1.612* | *0.708* | *1.522* | *4.109* | *1.199* | *2.971* |
| Post x High Chinese Share | -1.60*** | -0.32** | -1.45*** | -0.30*** | -0.14 | -0.06** | -0.94*** | -1.56*** | -0.73*** | -0.12*** |
| | (0.24) | (0.15) | (0.23) | (0.11) | (0.17) | (0.03) | (0.20) | (0.22) | (0.19) | (0.03) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,215 | 1,611 | 1,611 | 1,611 | 1,375 |
| | | | | | | B. White | | | | |
| *Dependent Variable Mean* | *7.836* | *5.165* | *4.048* | *3.911* | *6.569* | *0.953* | *6.498* | *6.710* | *5.097* | *3.072* |
| *– in 1880* | *6.732* | *3.962* | *3.325* | *1.825* | *5.519* | *0.924* | *5.224* | *5.638* | *4.127* | *3.054* |
| Post x High Chinese Share | -0.56*** | -0.51** | -0.68*** | -0.67*** | -0.39*** | -0.03*** | -0.63*** | -0.73*** | -0.66*** | -0.04*** |
| | (0.16) | (0.21) | (0.23) | (0.20) | (0.14) | (0.01) | (0.17) | (0.17) | (0.17) | (0.01) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,611 |
| | | | | | | C. All | | | | |
| *Dependent Variable Mean* | *7.916* | *5.211* | *4.139* | *4.009* | *6.636* | *0.930* | *6.519* | *6.845* | *5.132* | *3.063* |
| *– in 1880* | *6.875* | *4.036* | *3.647* | *2.134* | *5.551* | *0.905* | *5.259* | *5.940* | *4.180* | *3.051* |
| Post x High Chinese Share | -0.69*** | -0.56*** | -0.97*** | -0.84*** | -0.42*** | -0.02 | -0.67*** | -0.97*** | -0.70*** | -0.05*** |
| | (0.16) | (0.20) | (0.24) | (0.21) | (0.14) | (0.01) | (0.17) | (0.17) | (0.17) | (0.01) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 | 1,385 | 1,611 | 1,611 | 1,611 | 1,611 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of the total labor force (col. 1), the log of the labor force in the sector stated in the column headings (col. 2 - col. 5), the share of literate (col. 6), the log of total number of workers in the skill category stated in the column headings (col. 7 - col. 9), or the log of the occupational income score (col. 10). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.4: Robustness Check: Population-Weighted Regressions and Conley Standard Errors (LF Outcomes)

| | Dependent Variable | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Chinese | | | | | All | | | | |
| | LF Total | LF Mfg. | LF Mining | LF Railroad | Income Score | LF Total | LF Mfg. | LF Mining | LF Railroad | Income Score |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| | **A. Population Weighted Regressions** | | | | | | | | | |
| Post x High Chinese Share | -1.28*** | -0.41 | -1.57*** | -0.47* | -0.13*** | -0.83** | -0.90** | -1.46*** | -0.97*** | -0.05** |
| | (0.40) | (0.42) | (0.35) | (0.24) | (0.04) | (0.32) | (0.40) | (0.41) | (0.34) | (0.02) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,367 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 |
| | **B. Conley (1999) Standard Errors, 100km Distance Cutoff** | | | | | | | | | |
| Post x High Chinese Share | -1.58*** | -0.31** | -1.45*** | -0.30** | -0.13*** | -0.68*** | -0.55*** | -0.98*** | -0.84*** | -0.05*** |
| | (0.17) | (0.13) | (0.19) | (0.13) | (0.03) | (0.10) | (0.13) | (0.16) | (0.15) | (0.01) |
| Observations | 1,611 | 1,611 | 1,611 | 1,611 | 1,368 | 1,611 | 1,611 | 1,611 | 1,611 | 1,611 |

*Notes:* Observations are at the county and year level. The dependent variables are the log of the total labor force (cols. 1 and 6), the log of the labor force in the sector stated in the column headings (cols. 2-4 and 7-9), or the log of the occupational income score (cols. 5 and 10). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data are from full count U.S. Censuses between 1860 and 1940. Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.5: Robustness Check: Population-Weighted Regressions and Conley Standard Errors (Economic Outcomes)

| | Dependent Variable | | | | | | |
|---|---|---|---|---|---|---|---|
| | Manufacturing | | | Agriculture | | | |
| | Wage | Total Output | # Establ. | Farm Land Value | Livestock Value | Machinery Value | Fertilizer Expenditure |
| | (1) | (2) | (3) | (4) | (6) | (7) | (8) |
| | **A. Population Weighted Regressions** | | | | | | |
| Post x High Chinese Share | -0.19*** | -1.09** | -1.85*** | -0.68*** | -0.83** | -0.69*** | -1.20** |
| | (0.04) | (0.43) | (0.69) | (0.25) | (0.38) | (0.23) | (0.51) |
| Observations | 1,411 | 1,451 | 1,514 | 1,214 | 2,036 | 2,036 | 1,557 |
| | **B. Conley (1999) Standard Errors, 100km Distance Cutoff** | | | | | | |
| Post x High Chinese Share | -0.12*** | -0.91*** | -0.93* | -0.35** | -0.52*** | -0.51*** | -0.55** |
| | (0.04) | (0.19) | (0.55) | (0.15) | (0.12) | (0.13) | (0.26) |
| Observations | 1,414 | 1,453 | 1,514 | 1,214 | 2,036 | 2,036 | 1,557 |

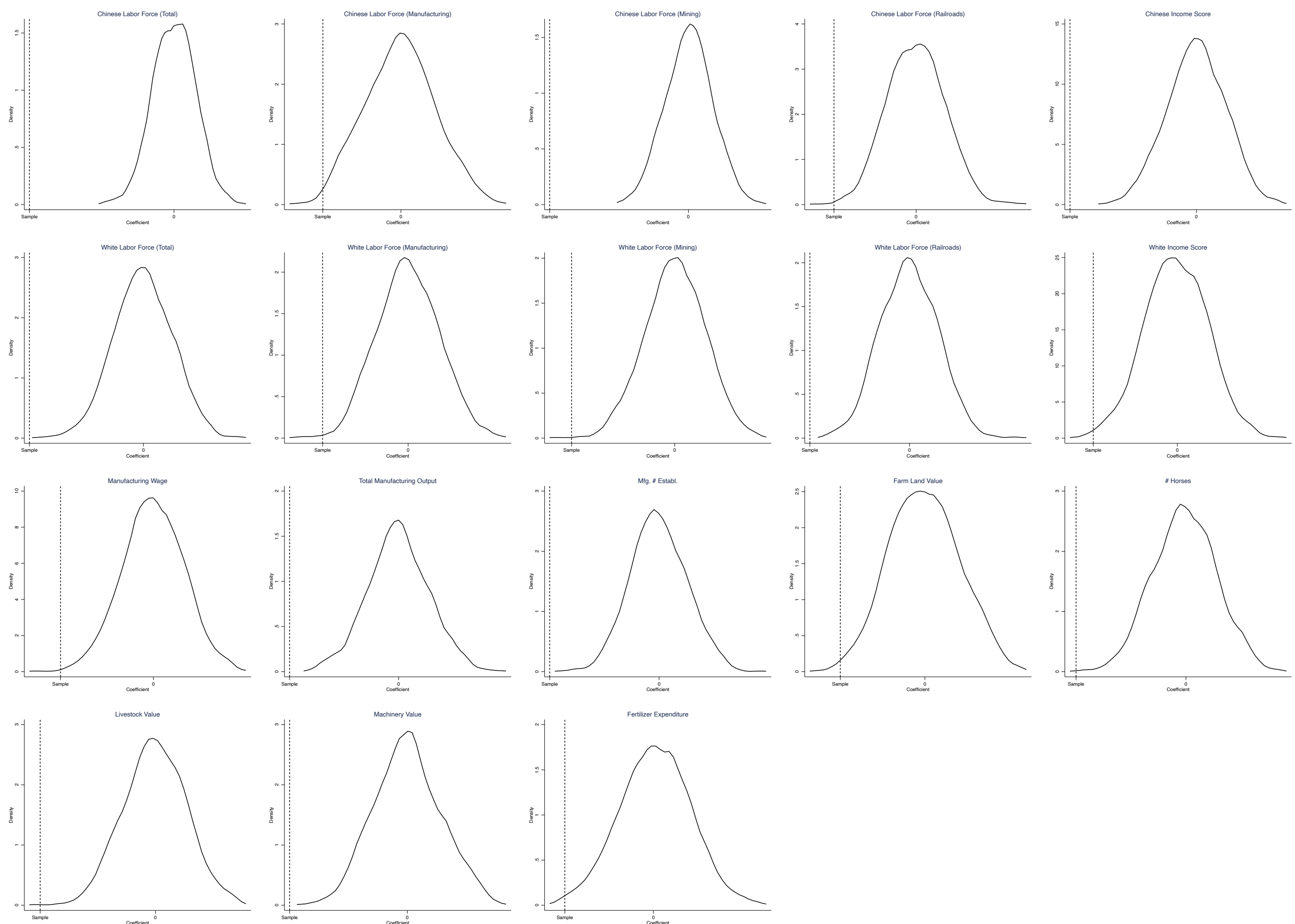
*Notes:* Observations are at the county and year level. The dependent variables are the log of average manufacturing wage (col. 1), the log of the total manufacturing output (col. 2), the number of manufacturing establishments (col. 3, Poisson regression), the log of the value of farm land (col. 4), the log of the value of livestock (col. 5), the log of the value of farming machinery and equipment (col. 6), or the average expenditure for fertilizers (col. 7). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data for the dependent variables in columns (1)-(3) are from the Historical, Demographic, Economic, and Social Data (ICPSR 2896), and from the United States Agriculture Data (ICPSR 35206) in columns (4)-(8), for the years 1860–1940. Standard errors clustered by county in Panel A, or accounting for spatial correlation (Conley 1999) with 100km distance cutoff in Panel B, are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.6: Robustness Check: Control for 1880 Variables

| | | I. Control for Year FE x Market Integration | | | II. Control for Year FE x Distance to NY | | | III. Control for Year FE x Homestead Act | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Post x High Chinese Share | | Obs. | Post x High Chinese Share | | Obs. | Post x High Chinese Share | | Obs. |
| | Dependent Variable | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | |
| | **A. Chinese LF** | | | | | | | | | |
| (1) | Total | -1.57*** | (0.25) | 1,472 | -1.57*** | (0.24) | 1,611 | -1.61*** | (0.23) | 1,611 |
| (2) | Mfg. | -0.40*** | (0.16) | 1,472 | -0.25 | (0.16) | 1,611 | -0.33** | (0.15) | 1,611 |
| (3) | Mining | -1.37*** | (0.24) | 1,472 | -1.46*** | (0.23) | 1,611 | -1.49*** | (0.23) | 1,611 |
| (4) | Rail | -0.33*** | (0.11) | 1,472 | -0.37*** | (0.11) | 1,611 | -0.31*** | (0.11) | 1,611 |
| (5) | Income | -0.11*** | (0.03) | 1,260 | -0.12*** | (0.03) | 1,367 | -0.13*** | (0.03) | 1,367 |
| | **B. All LF** | | | | | | | | | |
| (6) | Total | -0.61*** | (0.17) | 1,472 | -0.50*** | (0.16) | 1,611 | -0.56*** | (0.15) | 1,611 |
| (7) | Mfg. | -0.55** | (0.22) | 1,472 | -0.53** | (0.21) | 1,611 | -0.51** | (0.21) | 1,611 |
| (8) | Mining | -0.76*** | (0.23) | 1,472 | -0.71*** | (0.24) | 1,611 | -0.70*** | (0.23) | 1,611 |
| (9) | Rail | -0.75*** | (0.21) | 1,472 | -0.71*** | (0.21) | 1,611 | -0.67*** | (0.20) | 1,611 |
| (10) | Income | -0.04*** | (0.01) | 1,472 | -0.05*** | (0.01) | 1,611 | -0.04*** | (0.01) | 1,611 |
| | **C. Manufacturing** | | | | | | | | | |
| (11) | Wage | -0.14*** | (0.04) | 1,305 | -0.09** | (0.04) | 1,411 | -0.13*** | (0.04) | 1,411 |
| (12) | Total Output | -0.91*** | (0.28) | 1,341 | -0.87*** | (0.27) | 1,451 | -0.93*** | (0.27) | 1,451 |
| (13) | # Establ. | -1.21** | (0.54) | 1,399 | -1.12** | (0.54) | 1,514 | -0.90* | (0.53) | 1,514 |
| | **D. Agriculture** | | | | | | | | | |
| (14) | Farm Land Value | -0.35** | (0.14) | 1,128 | -0.09 | (0.12) | 1,214 | -0.37*** | (0.14) | 1,214 |
| (15) | Livestock Value | -0.49*** | (0.14) | 1,877 | -0.29** | (0.12) | 2,036 | -0.54*** | (0.14) | 2,036 |
| (16) | Machinery Value | -0.51*** | (0.15) | 1,877 | -0.33** | (0.13) | 2,036 | -0.52*** | (0.14) | 2,036 |
| (17) | Fertilizer Expenditure | -0.60** | (0.28) | 1,430 | -0.66*** | (0.25) | 1,557 | -0.54** | (0.24) | 1,557 |

*Notes:* Observations are at the county and year level. The dependent variables in Panel A are the log of the total Chinese labor force (row 1), the log of the Chinese labor force in the sector stated (rows 2-4), or the log of the average occupational income score among Chinese individuals (col. 5). The dependent variables in Panel B are the log of the total labor force of all individuals (row 6), the log of the labor force in the sector stated of all individuals (rows 7-9), or the log of the average occupational income score among all individuals (col. 10). The dependent variables in Panel C are the log of average manufacturing wage (row 11), the log of the total manufacturing output (row 12), or the log of the number of manufacturing establishments (row 13). The dependent variables in Panel D are the log of the value of farm land (row 14), the log of the number of horses used in farming (row 15), the log of the value of livestock (row 16), the log of the value of farming machinery and equipment (row 17), or the average expenditure for fertilizers (row 19). All regressions control for a dummy variable that equals 1 if the county is connected to a railroad in year t, a dummy variable that equals 1 if the county ever had a mine during 1850-1940 interacted with year fixed effects, and county and state-by-year fixed effects. The data for the dependent variables in Panel A and B are from the full count U.S. Censuses, for the years 1860-1940. The data for the dependent variables in Panel C are from the Historical, Demographic, Economic, and Social Data (ICPSR 2896), for the years 1860-1940; the data for the dependent variables in Panel D are from the United States Agriculture Data (ICPSR 35206), for the years 1860–1940. Monetary amounts are expressed in thousands of 2020 U.S. dollars (deflated using the Minneapolis Fed 1800–2020 CPI). Standard errors clustered by county are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure B.1: Permutation Test




*Notes:* The curves are the distributions of $\beta$ coefficients from 1,000 iterations of equation 2.1 after randomly permuting the variable $HighChineseShare_{i,1880}$ across counties, as explained in Section 2.5.3. The vertical dashed lines correspond to the baseline estimates from Tables 2.3–2.6.

# Appendix C

## Appendix Tables and Figures

## Political Connections, Careers, and Performance in the Civil Service: Evidence from U.S. Federal Judges

Table C.1: Robustness Check: Judge-Specific Linear Trends

| | (1) | (2) |
|---|---|---|
| *ConnectionLost* | -0.23 | 0.20 |
| | (0.26) | (0.33) |
| *ConnectionLost* × *Same-Party President* | | -0.86*** |
| | | (0.29) |
| *Same-Party President* | 0.66*** | 1.26*** |
| | (0.11) | (0.26) |
| *ConnectionLost +* | | -0.66*** |
| *ConnectionLost* × *Same-Party President* | | (0.23) |
| Mean Probability of Promotion | | |
| *(ConnectionLost = 0)* | | |
| *Same-Party President = 0* | 0.16 | 0.16 |
| *Same-Party President = 1* | 1.17 | 1.17 |
| Observations | 20,398 | 20,398 |
| Judge FEs | Y | Y |
| State × Year FEs | Y | Y |
| Judge Linear Trends | Y | Y |

*Notes:* In all models, the dependent variable is an indicator for district judge $i$ being promoted at year $t$. Coefficients, standard errors and baseline means are multiplied by 100 to enhance readability. This sample includes only district court judges who had one connection at the time of appointment. Standard errors clustered by recommending senator(s) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C.2: Heterogeneous Effects on Promotions by Partisanship

| | (1) | (2) | (3) |
|---|---|---|---|
| *ConnectionLost* | -0.12 | 0.36 | 0.52* |
| | (0.27) | (0.28) | (0.28) |
| *ConnectionLost* × *Same-Party President* | | -1.03*** | -1.32*** |
| | | (0.30) | (0.35) |
| *ConnectionLost* × *Republican* | -0.24 | -0.21 | -0.53* |
| | (0.23) | (0.21) | (0.31) |
| *ConnectionLost* × *Republican* × *Same-Party President* | | | 0.55 |
| | | | (0.48) |
| *Republican* × *Same-Party President* | | | 0.26 |
| | | | (0.55) |
| *Same-Party President* | 0.77*** | 1.48*** | 1.36*** |
| | (0.12) | (0.29) | (0.32) |
| *ConnectionLost(Republican)* + *ConnectionLost(Republican)* × *Same-Party Pres.* | | | 0.02 |
| | | | (0.33) |
| Mean Probability of Promotion | | | |
| *(ConnectionLost = 0)* | | | |
| *Same-Party President = 0* | 0.16 | 0.16 | 0.16 |
| *Same-Party President = 1* | 1.16 | 1.16 | 1.16 |
| Observations | 20,395 | 20,395 | 20,395 |
| Judge FEs | Y | Y | Y |
| State × Year FEs | Y | Y | Y |
| Judge's Experience FEs | Y | Y | Y |

*Notes:* In all models, the dependent variable is an indicator for district judge $i$ being promoted at year $t$. Coefficients, standard errors and baseline means are multiplied by 100 to enhance readability. This sample includes only district court judges who had one connection at the time of appointment. Standard errors clustered by recommending senator(s) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C.3: Heterogeneous Effects on Promotions by ABA Rating

| | (1) | (2) | (3) |
|---|---|---|---|
| *ConnectionLost* | 0.09 | 0.67** | 0.47* |
| | (0.27) | (0.27) | (0.28) |
| *ConnectionLost* × *Low ABA* | -0.42 | -0.44 | -0.04 |
| | (0.28) | (0.29) | (0.30) |
| *ConnectionLost* × *Same-Party President* | | -1.24*** | -0.90** |
| | | (0.31) | (0.35) |
| *ConnectionLost* × *Low ABA* × *Same-Party President* | | | -0.71 |
| | | | (0.48) |
| *Low ABA* × *Same-Party President* | | | 0.59 |
| | | | (0.40) |
| *Same-Party President* | 0.77*** | 1.59*** | 1.31*** |
| | (0.13) | (0.30) | (0.29) |
| *ConnectionLost(Low ABA) +* *ConnectionLost(Low ABA)* × *Same-Party Pres.* | | | -0.75* |
| | | | (0.42) |
| Mean Probability of Promotion | | | |
| *(ConnectionLost = 0)* | | | |
| *Same-Party President = 0* | 0.19 | 0.19 | 0.19 |
| *Same-Party President = 1* | 1.09 | 1.09 | 1.09 |
| Observations | 15,457 | 15,457 | 15,457 |
| Judge FEs | Y | Y | Y |
| State × Year FEs | Y | Y | Y |
| Judge's Experience FEs | Y | Y | Y |

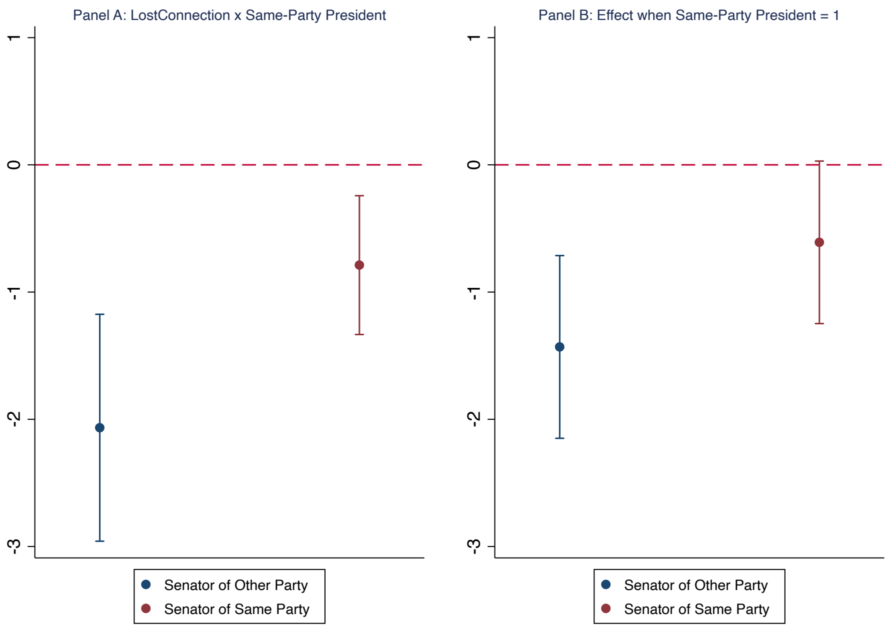
*Notes:* In all models, the dependent variable is an indicator for district judge $i$ being promoted at year $t$. Coefficients, standard errors and baseline means are multiplied by 100 to enhance readability. This sample includes only district court judges who had one connection at the time of appointment. Robust standard errors clustered by recommending senator(s) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C.4: Heterogeneous Effects on Promotions by Type of Senator's Exit

| | (1) | (2) | (3) |
|---|---|---|---|
| *ConnectionLost* | -0.12 | 0.35 | 0.41 |
| | (0.26) | (0.25) | (0.27) |
| *ConnectionLost* × *Same-Party President* | | -1.03*** | -1.13*** |
| | | (0.30) | (0.36) |
| *ConnectionLost* × *Unexpected* | -0.23 | -0.18 | -0.31 |
| | (0.26) | (0.25) | (0.37) |
| *ConnectionLost* × *Unexpected* × *Same-Party President* | | | 0.19 |
| | | | (0.51) |
| *Unexpected* × *Same-Party President* | | | -0.35 |
| | | | (0.49) |
| *Same-Party President* | 0.77*** | 1.59*** | 1.31*** |
| | (0.13) | (0.31) | (0.30) |
| *ConnectionLost(Unexpected)* + *ConnectionLost(Unexpected)* × *Same-Party Pres.* | | | -0.35 |
| | | | (0.49) |
| Mean Probability of Promotion | | | |
| *(ConnectionLost = 0)* | | | |
| *Same-Party President = 0* | 0.16 | 0.16 | 0.16 |
| *Same-Party President = 1* | 1.16 | 1.16 | 1.16 |
| Observations | 20,395 | 20,395 | 20,395 |
| Judge FEs | Y | Y | Y |
| State × Year FEs | Y | Y | Y |
| Judge's Experience FEs | Y | Y | Y |

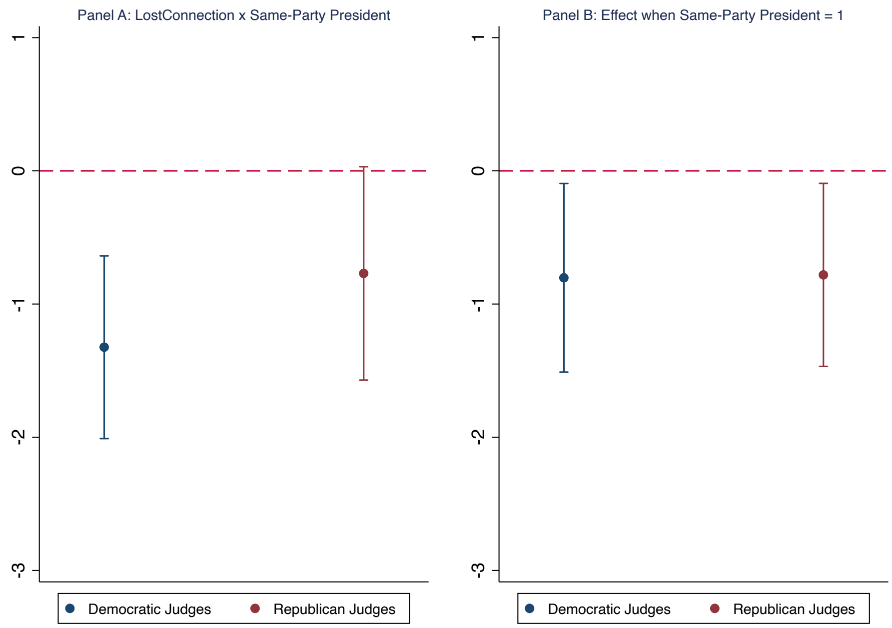
*Notes:* In all models, the dependent variable is an indicator for district judge $i$ being promoted at year $t$. Coefficients, standard errors and baseline means are multiplied by 100 to enhance readability. This sample includes only district court judges who had one connection at the time of appointment. Standard errors clustered by recommending senator(s) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure C.1: Effect of Recommender vs. Party Connection on Promotions



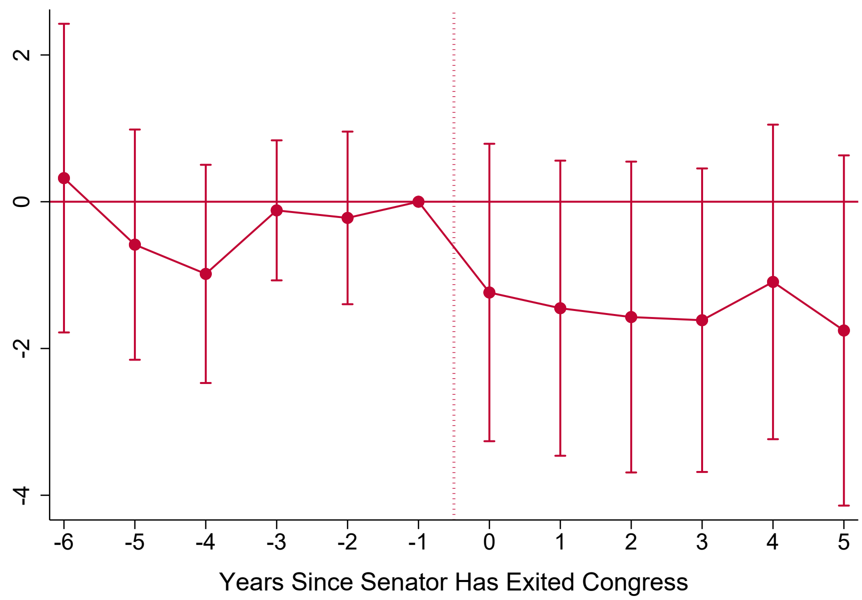

*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending-senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

Figure C.2: Heterogeneous Effects on Promotions by Party Affiliation



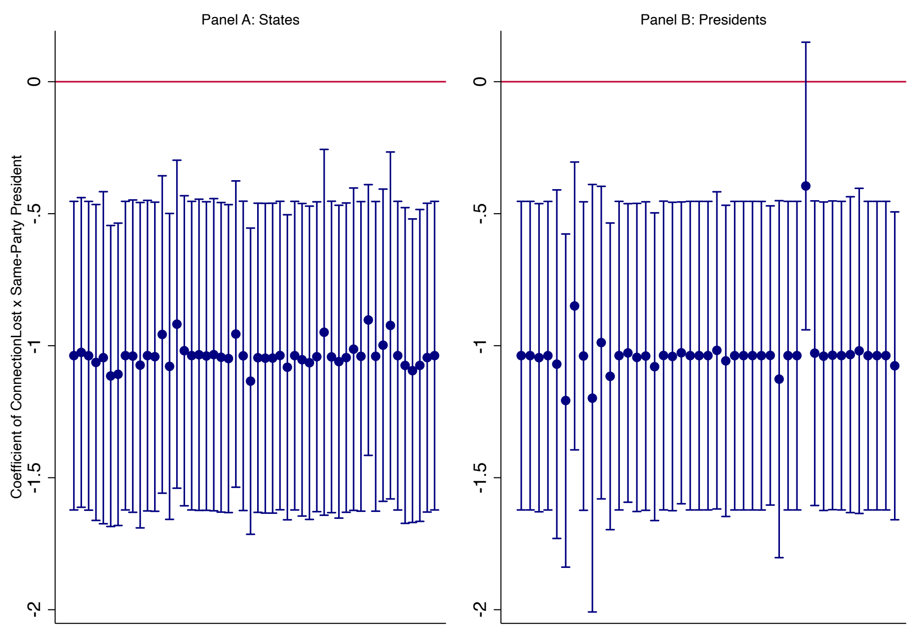

*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending-senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

Figure C.3: Robustness Check: Event Study Using De Chaisemartin and D'Haltfoeuille (2020)



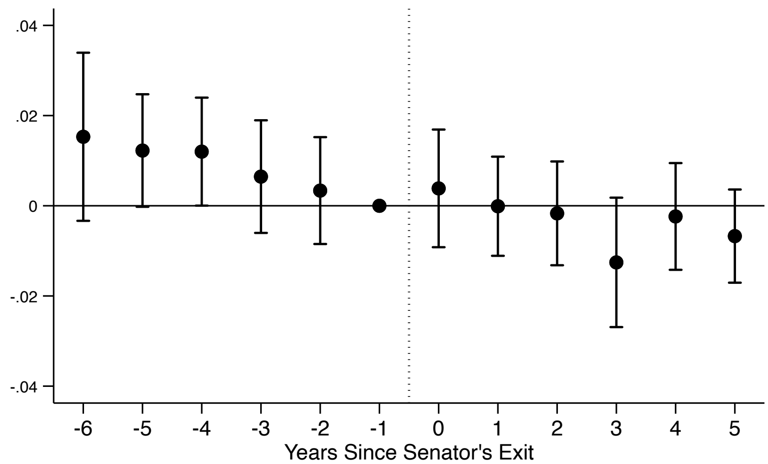

*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Point estimates for the effect of having lost the connection to the recommending senator when the president is of the same party as the judge, retrieved via the $DID_M$ estimator of De Chaisemartin and d'Haultfoeuille (2020). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, year, and judge's experience. This sample includes only district court judges who had one connection at the time of appointment.

Figure C.4: Robustness Check: Excluding States and Presidents



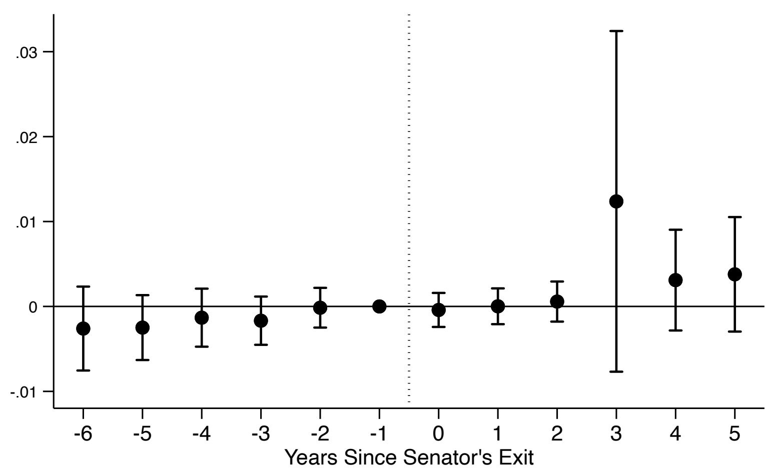

*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$. Point estimates are the marginal effect of losing the connection with the recommending senator when the president is of the same party as the judge. Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level.

Figure C.5: Effect on Promotions – First Exit




*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had two connections at the time of appointment.

Figure C.6: Effect on Promotions – Second Exit




*Notes:* The dependent variable is an indicator for district judge $i$ being promoted at year $t$ (multiplied by 100). Vertical lines are 95% confidence intervals based on robust standard errors clustered at the recommending senator level. Regressions include the following sets of FEs: judge, state by year, and judge's experience. This sample includes only district court judges who had two connections at the time of appointment.