Closing Ranks: Organized Labor and Immigration

Carlo Medici[⋆]

December 2025

Abstract

This paper shows that immigration fostered the emergence of American organized labor. I digitize archival records to assemble the first county-level dataset on historical U.S. unionization and use a shift-share instrument to isolate plausibly exogenous labor supply shocks induced by immigration, between 1900 and 1920. Counties with higher immigration experienced increases in union presence and membership. These effects were more pronounced among skilled workers, particularly in counties more exposed to immigrant labor competition, and in areas with more negative attitudes toward immigrants. The evidence is consistent with existing workers unionizing in response to immigration, driven by economic and social motivations.

Keywords: Unions, Immigration, Labor Market Competition, Discrimination
JEL codes: J15, J5, J7, N31, N32, P1

[⋆]University of California, Los Angeles. carlomedici@ucla.edu.

I am indebted to Joel Mokyr, Matthew Notowidigdo, Nancy Qian, Marco Tabellini, and Edoardo Teso for extensive advice and guidance throughout this project. I thank the editor, three anonymous referees, Devis Decet, Henry Downes, Georgy Egorov, Silvia Farina, Price Fishback, Carola Frydman, Walker Hanlon, Leander Heldring, Joris Mueller, Laura C. Murphy, Santiago Pérez, Nicola Persico, Massimo Pulejo, Hazal Sezer, Miguel Talamas, Silvia Vannutelli, seminar participants at Caltech, CREI, Northwestern, Nottingham, Tilburg, UCLA, UPF, USC, and UW-Milwaukee, and conference participants at the 2024 NBER Summer Institute (DAE), EHA (Pittsburgh 2023 and Sacramento 2024), ASSA (San Antonio 2024), and ASREC (Arlington 2025) Annual Meetings, 2024 NICEP conference, 2024 Junior Economists Meeting, 2024 CEMIR Junior Economist Workshop on Migration Research, 2024 AFD-World Bank Migration and Development Conference, 2024 Dondena Workshop on Public Policy, and 2025 CNEH Conference (Montreal) for helpful comments and conversations; Kellogg Research Support for help in obtaining and digitizing historical data; and Xueyan Li for excellent research assistance.

1 Introduction

Labor unions have long been central institutions in the labor markets of advanced economies. Throughout the 20th century, they played a critical role in reducing inequality (Farber et al., 2021), improving working conditions (Rosenfeld, 2019), shaping policy (Ahlquist, 2017), and influencing political systems (Acemoglu and Robinson, 2013; Kaplan and Naidu, 2024). Despite fluctuations in membership, unions remain pivotal in today's economy (Jäger et al., 2024).[1] Yet, despite their sustained importance, there is surprisingly little evidence on the drivers of unions' emergence and growth. This paper aims to fill this gap with systematic empirical evidence.

The origins of modern organized labor can be traced back to the Industrial Revolution. One prevailing theory for the rise of unions emphasizes the increasing capital intensity of industrial production, which shifted bargaining power from laborers to the owners of capital (Foner, 1947). A complementary hypothesis suggests that workers also organized in response to growing labor competition (Taft, 1964), which intensified during this period alongside population growth, rapid urbanization, and industrial expansion.

This paper investigates the second mechanism: the effect of a large and sustained increase in labor supply, driven by episodes of mass immigration, on the formation and growth of labor unions. The effect is *ex ante* ambiguous, as it influences both workers' incentives to organize and employers' ability to undermine organized labor. On the one hand, increased job competition can motivate workers to unionize in response to economic threats to their employment and wages. On the other hand, a larger labor supply reduces the cost for 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 20th-century United States provides an ideal setting to examine 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). Second, this period witnessed the creation and growth of several unions that remain influential today (Stewart, 1926), despite a legal and judicial framework that enabled employers to easily dismiss and replace unionizing workers (Taft, 1964). Third, it offers a natural experiment for causal identification, driven by the large and sustained influx of European immigrants during what is often referred to as the Age of Mass Migration (Hatton and Williamson, 1998).

[1] In the United States, unions have recently achieved historic victories for various groups of workers, including autoworkers, UPS drivers, and Hollywood writers (Ewing and Boudette, 2023; Hadero and Ott, 2023). In Europe and Canada, where collective bargaining also has a strong tradition, organized labor continues to expand into previously unorganized sectors, shape policy agendas, and improve labor market conditions (OECD, 2019).

This study addresses two key challenges in examining the relationship between immigration and unionization. The first is the need for disaggregated data on the presence and membership of labor unions. The previously available historical data measure unionization at the national or state level, limiting analyses of 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 with economic growth. Such joint determination would lead to a positive association between unions and immigrants.

To measure unionization, I hand-collect and digitize archival documents on the location, quantity, and size 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 every 10 years between 1900 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. 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 membership at the local level in the United States.

To estimate the causal effect of immigration, I employ a shift-share instrumental variable approach (Card, 2001b), leveraging plausibly exogenous variation in immigrant flows across counties over each decade. The instrument combines the 1890 share of immigrants in a given U.S. county and born in different European countries with the aggregate immigration flows from those countries to the United States between 1890 and 1920. This identification strategy builds on the empirical regularity of *chain migration*, whereby immigrants tend to settle in areas where others from their country of origin had previously settled. The key identifying assumption is that, conditional on the controls, the baseline distribution of immigrants from different European countries across counties is uncorrelated with unobserved factors affecting unionization, except through the subsequent migration flows from those countries to the United States.[2] I estimate 2SLS regressions that include county and year fixed effects, as well as interactions between year dummies and baseline economic characteristics.[3]

[2] For a formal discussion of the validity of shift-share designs, see Adao et al. (2019), Borusyak et al. (2022, 2025), Goldsmith-Pinkham et al. (2020), and Jaeger et al. (2018).

[3] The preferred specification includes controls for the baseline share of the population employed in manufacturing and agriculture, and an indicator for the presence of active coal mines (see Section 4.2 for details). The results are robust to adding further baseline characteristics or to using only county and

The main results of this paper show that immigration fostered the emergence of organized labor. Counties that received more immigrants as a fraction of the population experienced an increase in the probability of having labor unions, the number of union branches, the share of unionized workers, and the number of union members per branch. This finding documents empirically a novel 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 probability that a county had labor unions by 11 percentage points (24% relative to the mean), the number of union branches by 0.2 (8% relative to the mean), the share of the unionized workforce by one percentage point (31% relative to the mean), and the number of union members per branch by 19 (40% of the mean). Immigration spurred unionization both at the intensive and extensive margins, as counties with an existing labor movement experienced a growth in union size, while new counties saw the establishment of labor unions. A back-of-the-envelope calculation reveals that in the absence of immigration, the average union density (i.e., the share of unionized workers) between 1900 and 1920 would have been 22% lower. The estimates are robust to a variety of sensitivity checks, including the use of an alternative instrument that replaces realized immigration flows with plausibly exogenous variation driven by conditions abroad, and a matching-style strategy that compares counties with similar levels of union presence at baseline. The findings are also robust to including several baseline controls, such as the initial share of the immigrant population, the shares of the labor force in each industry or occupation, average income, and economic growth.

The second part of the paper examines the mechanisms through which immigration affected unionization. First, I show that the increase in unionization was concentrated among skilled workers, whose specialized expertise, greater control over access to their trades, and limited *immediate* replaceability made them more capable of organizing in response to immigration-induced pressures on wages and employment. By contrast, unionization among unskilled workers, who faced greater substitution from new arrivals, was largely unaffected. Second, I document that counties where immigrants directly competed with U.S.-born workers experienced larger increases in skilled unionization and weaker effects among unskilled workers, consistent with differences in exposure to labor market competition and variation in employers' ability to hire replacement labor. Third, I show that social motivations also contributed to the observed patterns: immigration spurred greater union growth in counties receiving culturally more distant immigrants and in areas with stronger pre-existing hostility toward immigration, consistent with the nativist rhetoric present in parts of the early 20$^{\text{th}}$-century labor movement.

Next, I rule out several alternative channels that could explain the results. First,

year fixed effects, as discussed in Section 5.3.

by inferring the ancestry of local union leaders using their last names, I show that the findings are unlikely to be driven by immigrants disproportionately participating in unions. Second, the inflow of workers from countries with stronger labor movements or support for socialist parties did not contribute to the observed rise in unionization. Third, the results are robust to excluding years surrounding major political and social events, such as World War I and the First Red Scare. Finally, the effects are not amplified by the presence of other labor organizations, nor explained by differences in economic growth across counties.

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, as existing workers formed and joined unions in response to the economic and social challenges posed by immigration. In a longer version of the paper (Medici, 2024),[4] I also examine the broader implications of this immigration-induced unionization. The results indicate that immigration led U.S.-born workers to shift toward unionized occupations, that areas receiving more immigrants in the early 1900s continue to exhibit higher private-sector unionization today, and that early unionization was associated with lower overall wage inequality but higher inequality between immigrant and native-born workers.

Related literature. The results of this paper contribute to several broad literatures. First, they speak to the extensive body of work on organized labor and labor unions. Numerous empirical studies, including many recent ones, have analyzed the impacts of labor unions on a wide range of economic and political outcomes in both historical and contemporary contexts (Ahlquist, 2017; Barth et al., 2020; Biasi and Sarsons, 2022; Card, 2001a; Collins and Niemesh, 2019; DiNardo and Lee, 2004; Dodini et al., 2023; Farber et al., 2021; Feigenbaum et al., 2018; Fishback, 1992, 1995; Lee and Mas, 2012; Naidu and Reich, 2018; Rosenfeld and Kleykamp, 2012).[5] However, much less is known about the drivers of unionization. While existing studies discuss its origins (Alesina and Glaeser, 2004; Bernstein, 1954; Cohen et al., 2016; Foner, 1947; Freeman and Medoff, 1984; Friedman, 1998; Hannan and Freeman, 1987; Karadja and Prawitz, 2019; Lipset and Marks, 2000; Olson, 1965; Schmick, 2018; Sombart, 1976; Taft, 1964; Webb and Webb, 1894; Wolman, 1924), and others focus on its decline in recent decades (Acemoglu et al., 2001; Ahlquist and Downey, 2020; Farber and Western, 2001; Hirsch, 2008; Scruggs and Lange, 2002; Wallerstein and Western, 2000), this paper is the first to provide systematic empirical evidence on the determinants of American unions' early development during the craft-union era of the early twentieth century. The results of this study advance these strands of the literature by identifying immigration as a key

[4]The full paper is available at: https://www.ifo.de/en/cesifo/publications/2024/working-paper/closing-ranks-organized-labor-and-immigration.

[5]See also Jäger et al. (2024) and Kaplan and Naidu (2024) for extensive surveys of this literature.

driver of the emergence and growth of modern unions during a formative period for the American labor movement. Furthermore, they shed light on the channels through which immigration influenced unionization.

A complementary line of research examines the conditions under which workers are able to translate economic pressures into successful collective action. This work emphasizes that institutional constraints and employer opposition pose major barriers to union formation, even when support for unions is high (Naidu, 2022; Naidu and Yuchtman, 2016). Empirical evidence similarly shows that increases in labor market tightness modestly raise support for unions but do not lead to higher union membership (Pezold et al., 2023). The historical evidence presented in this paper contributes to this literature by showing how immigration-driven changes in labor supply interacted with differences in workers' organizing capacity across occupations.

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. Existing data sources on early labor unions are limited: some focus on extinct organizations (Garlock, 2009) or smaller unions (Gregory, 2015), others cover only a subset of unions and lack membership information (Schmick, 2018), are not disaggregated below the state level (Farber et al., 2021), or measure unionization only in a handful of states (Downes, 2024), or beginning only in the 1960s (Lee and Mas, 2012). Collins and Niemesh (2019) similarly use historical data to study mid-century wage inequality and construct a shift-share instrument for unionization, but do not provide direct measures of local union presence or membership. The data introduced in this paper, collected at the union-branch level and aggregated to the county level for analysis, represent a significant advancement for studying geographic patterns of early unionization and enable future work on its medium- and long-term consequences.

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 provides the first systematic evidence that historical immigration fostered the emergence and growth of a key labor market institution.

Further, this study contributes to a large body of work examining how immigration affects native-born 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, 2009), Foged and Peri (2016), Ottaviano and Peri (2012), and Tabellini (2020), who find negligible or positive impacts on native-born workers. The results suggest that labor unions may play a role in mitigating the potential adverse effects of immigration on native-born 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 and recently (see Alesina and Tabellini, 2024 for a review). The results of this study identify a novel consequence of immigration on the development of institutions that have had—and still have today—vast political influence. Although anecdotal and historical evidence have acknowledged the instrumental role that organized labor played in the introduction of immigration restrictions in the 1920s (Goldin, 1994; Mink, 1986), this paper is the first to empirically estimate a causal and positive effect of immigration on unionization, and document that this was due to economic as well as social motivations. Moreover, this paper is related to the work by Alesina and Glaeser (2004), linking the weakness of the U.S. labor and socialist movements to ethnic diversity. The results of this study shed further light on this phenomenon, showing that 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 2 describes the historical background. Section 3 presents the data. Section 4 introduces the empirical strategy and the instrument for immigration. Section 5 presents the main results and a summary of the robustness checks. Section 6 sheds light on the mechanisms that are driving the effect. Section 7 concludes.

2 Historical Background

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

A new phase for the American labor movement began in the late 1880s, marked by the rise of the American Federation of Labor (AFL) as the largest and most influential group of labor unions.[6] By 1890, the main labor organizations that had gained prominence during the second half of the 19$^{\text{th}}$ century—such as the Knights of Labor and the independent railroad workers' movements—had largely faded (Wolman, 1924).[7] This shift cleared the way for the rise of new unions, many of which became some of the largest national trade organizations still active today.[8] Between 1880 and 1920, the total

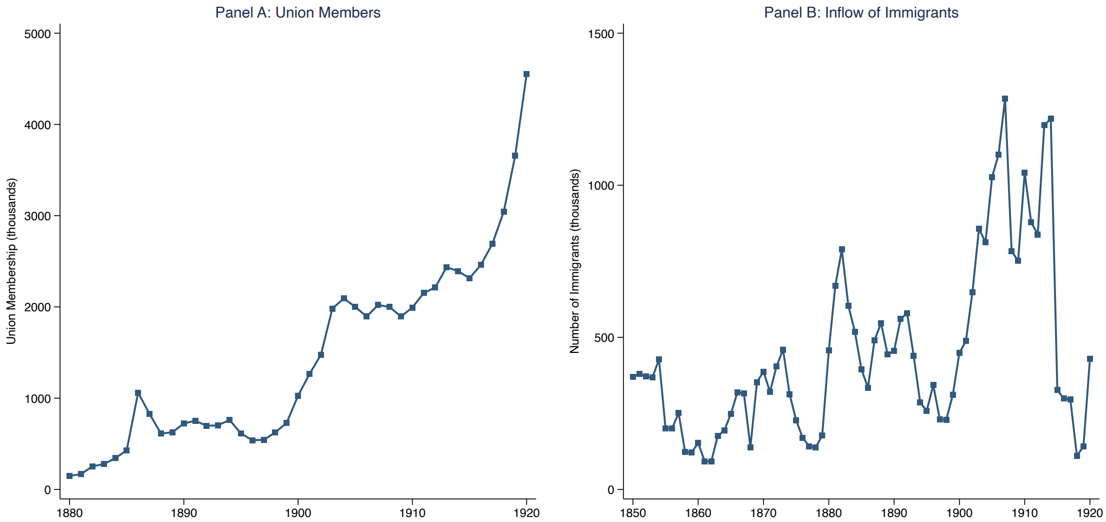
[6]The American Federation of Labor (AFL) was founded in Columbus, Ohio, in 1886 and quickly became the dominant federation of unions (Foner, 1947). An alternative labor organization, the Industrial Workers of the World (IWW), was also established in this period, in 1905. However, its membership remained far more limited than that of the AFL. By 1926, the AFL reported 3,383,997 members, whereas the IWW had only 30,000 (Stewart, 1926). All the results and the data of this paper pertain to AFL unions only. Section 6.3 examines the relationship between the observed effects and the presence of IWW branches from 1906 to 1917, using the only available cross-sectional data on the IWW (Gregory, 2015).

[7]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).

[8]The International Brotherhood of Teamsters, the International Brotherhood of Electrical Workers, the International Association of Machinists, and the United Brotherhood of Carpenters—even now among

number of union members went from 149,000 to over 4.5 million (Figure 1, Panel A).

Figure 1: Trends in Union Membership and Immigration



Notes: The figure shows: in Panel A, the total number of union members in the U.S. between 1880 and 1920 (Freeman, 1998); in Panel B, the inflow of immigrants to the United States between 1850 and 1920 (Immigration Policy Institute).

The AFL was established as a federation of national unions, following the principles of craft unionism. Under this model, workers were organized according to their specific occupation or craft.[9] The AFL adhered to the policy of *one craft–one union*, mandating that each occupation be represented by a single union. During this period, the unions in the building construction industry emerged as the most stable and largest organizations.[10] This industry was dominated by skilled craftsmen, and characterized by small employing units (Taft, 1964). In contrast, only a few unions organized unskilled laborers in industrial settings. The largest of these represented coal miners, as well as longshoremen, teamsters, and laborers in the brewing, 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).

AFL-affiliated unions were organized into branches, called *locals*. These branches were responsible for negotiating agreements with individual employers (based on national union guidelines) to regulate wages, work hours, and employment conditions. Unions also maintained funds to support workers in cases of strikes, injury, disability, or death, and oversaw the terms of apprenticeship within the craft (Stewart, 1926).

Most collective agreements included a *closed-shop* clause, specifying that only union

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 almost 13 million workers (U.S. Department of Labor, 2023).

[9]The primary alternative to craft unionism is *industrial* unionism, which organizes all workers within the same industry into a single union, regardless of skill level.

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

members could be employed. Restricting union membership to U.S.-born workers or American citizens, requiring mandatory membership as a condition of employment, and establishing apprenticeships were common methods used by organized labor to control access to certain occupations (Ignatiev, 1994; Mink, 1986; Stewart, 1926; Witwer, 2002). This effectively gave unions control over which workers could acquire the necessary skills necessary for the (skilled) jobs they organized.

Until the mid-1930s, no federal law required employers to recognize unions or penalized them for retaliating against union members. This environment allowed company owners, with the support of the courts, to make use of strikebreakers, lockouts, retaliatory firing, and other strategies to oppose unions and prevent their organization (Foner, 1947; Taft, 1964).[11] As a result, workers had to battle their employers for union recognition through strikes, walkouts, and violent confrontations, often requiring police intervention (Fishback, 1992; Naidu and Yuchtman, 2016; Olzak, 1989). The establishment or survival of the union was at the heart of many violent strikes across all industries (Taft and Ross, 1969).

2.2 The Age of Mass Migration

Between 1850 and 1920, around 30 million Europeans moved to the United States (Figure 1, Panel B), raising the share of the foreign born population to over 14%. 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), a growing number of immigrants arrived from other parts 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 A.1). 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 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

[11] 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).

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.[12] 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.

2.3 The Labor Movement and Immigration

Organized labor has long been concerned about the potential negative consequences of labor supply expansions, particularly those caused by immigration (Taft, 1964). This fear led to its early support for immigration restrictions. In 1881, at the founding meeting of its precursor organization (named Federation of Organized Trades and Labor Unions), the American Federation of Labor (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 achieved the enactment of the Alien Contract Labor Law (the Foran Act), which banned importing foreigners laborers.[13] In 1896, in response to increased immigration from ethnic and national groups with lower schooling levels, skills, and standards of living, the AFL advocated for further restrictive measures. It was widely believed that immigrants from Southern and Eastern Europe lowered wages, undermined working conditions, and resisted union discipline, posing a threat to the American workforce (Mink, 1986; Taft, 1964).

Within the labor market, unions sought to protect their wages and jobs through restrictive collective bargaining agreements, excluding newcomers or relegating them to the less desirable, lower-paying roles (Asher, 1982).[14] At the policy level, the AFL continued to endorse further restrictive measures, ultimately leading to the introduction of

[12] 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 vetoed the bill.

[13] 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 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, 1986).

[14] As discussed in Section 2.1, unions often restricted membership to U.S.-born workers or American citizens, required union membership for employment, and strictly enforced apprenticeship terms within each trade (Ignatiev, 1994; Mink, 1986; Stewart, 1926; Witwer, 2002).

the 1921 and 1924 nationality quotas (Goldin, 1994).

Throughout this period, the labor movement increasingly relied on racial and eugenics-based arguments to emphasize perceived threats to employment and build support for an outright ban on European immigration.[15] These narratives reinforced fears that newly arrived workers could displace existing labor (Olzak, 1989), prompting unions to focus on securing job control by organizing workplaces and regulating work processes (Mink, 1986). At the same time, the immigration-driven expansion of the labor supply was seen as weakening unions' bargaining power by creating a pool of potential strikebreakers and freeing employers from the constraints of a tight, unionizing labor market (Montgomery, 1979).

3 Data

This study relies on a novel micro-database that combines labor unions' records with labor market outcomes between 1900 and 1920. In this section, I describe the data collection effort, the main sources of the data, and present summary statistics.

3.1 Dataset on Union Presence and Membership

I assemble a novel dataset on unionization for 1900–1920, the first comprehensive local-level dataset on historical U.S. union presence and membership. Existing studies on unions in historical contexts rely largely on aggregate national estimates, as microdata on union status became available only with the Current Population Survey in 1973. The few disaggregated historical data sources are unsuitable for the questions addressed in this paper. Some lack membership data and predate major immigration waves and union expansions (Schmick, 2018), while others focus on extinct organizations (Garlock, 2009) or small unions (Gregory, 2015). Additional sources are not disaggregated below the state level (Farber et al., 2021), are limited to a few states (Downes, 2024), or begin only in the 1960s (Lee and Mas, 2012).

Convention proceedings of the state federations of labor. The main sources for the dataset on unionization are the convention proceedings of the state federations of labor, which served as the state-level organizations of the American Federation of Labor (AFL). Their functions were primarily legislative and propagandist, and they were com-

[15] 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).

posed of representatives from all the local branches of AFL-affiliated national unions within the state (Stewart, 1926). Local branches (also referred to as local unions, or *locals*) formed a lower level of organization within national unions and represented workers either from a single employment unit or across multiple work sites. By 1920, AFL union members accounted for more than 80% of total private-sector union membership (Wolman, 1924). Each state federation of labor convened annually to enact legislation and elect general officers, with all affiliated local unions entitled to representation.[16]

I digitize the proceedings of these conventions every 10 years between 1900 and 1920.[17] From these documents, I extract the lists of union branches (*locals*) represented at the conventions, including their union name, branch number, location, number of delegates, and delegate names (Figure A.2). Each federation established specific rules for determining the number of delegates representing a local branch, which often varied over time. Importantly, these rules specified that representation should be proportional to membership (Figure A.3).[18]

I combine information on delegates from the convention proceedings with representation rules outlined in the constitutions of each state federation of labor to construct estimates of union membership for each local branch. Since representation rules were often expressed in ranges (e.g., one delegate for every 100 members), I estimate membership using the midpoints of these intervals. For example, if the constitution specifies one delegate per 100 members, a branch with one delegate is estimated to have 50 members, two delegates correspond to 150 members, and so on.[19]

I geocode the location of all union branches based on their town, village, or city, and retrieve their geographic coordinates. I use the names of each branch's national union to identify the occupations and industries they operated in.[20] Finally, I aggregate union membership at the county level to construct measures of total membership and membership by occupation.[21]

[16]Exceptions included branches that were only established shortly before the convention and those expelled or suspended by their national organization.

[17]If the proceedings for any of these years were unavailable or lacked the necessary information (e.g., union branch locations), the analogous document from the convention held either in the following year or two years later was digitized. If those documents were also unsuitable or unavailable, the document from the convention that took place in the previous year or two years earlier was used.

[18]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.

[19]The results (not reported for brevity but available upon request) remain unchanged if membership is instead estimated using the lower or upper bound of the intervals.

[20]As described in Section 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.).

[21]The decision to aggregate at the county level alings with the empirical strategy, which employs a shift-share instrument based on spatial variation in immigration measured with the 1890 Census (see Section 4 for details), available only at the county level. In Section 5.3, I also present results based on data

Proceedings of the national conventions of AFL unions. I complement the data from the state federations of labor with analogous information collected directly by AFL-affiliated national unions. Similar to the state federations, these national unions held conventions to legislate, elect officers, and set guidelines for local branches to follow in bargaining agreements. I digitize the proceedings of these conventions for the five largest AFL-affiliated unions (as of 1900) at the same 10-year intervals as the main data.[22] Members of these five unions accounted for more than one-third of the total membership across over 100 AFL-affiliated unions between 1900 and 1920 (Wolman, 1924).[23] Following the same procedure described for the state federations, I collect data on the location, number, and names of delegates representing the local branches. I then estimate each local's membership using the representation rules outlined in the convention proceedings or union constitutions. Finally, I aggregate the data at the county level.

These data sources complement the state federation records in three key ways. First, they validate the estimates constructed using the main data source. In particular, for the five unions observed 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.4).

Second, combining data from national unions helps mitigate measurement error. Some branches may appear in only one type of convention record due to eligibility rules. For example, branches formed too recently before a convention may not have qualified to send delegates under organization-specific rules. Others may have been formed between the conventions of the state federation and the national union, appearing in one document but not the other. Delegates could also have been mistakenly omitted from meeting records.[24] Any of these occurrences could lead to an underestimate of the number of members or branches in a given county if only one data source were used. While it is impossible to determine precisely how many locals were affected—since this information was never systematically reported—combining information from

aggregated at the State Economic Area (SEA) level—groups of contiguous counties within the same state that share similar economic characteristics (Bogue, 1951)—which can be considered analogous to modern labor markets or Commuting Zones (Abramitzky et al., 2023).

[22] As above, if suitable documents are not available for 1900, 1910, or 1920, I digitize 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 earlier, or two years earlier.

[23] These unions are: the Bricklayers, Masons, and Plasterers International Union of America (BMPIU); the International Association of Machinists (IAM); the International Typographical Union (ITU); the United Brotherhood of Carpenters and Joiners (UBC); and the United Mine Workers of America (UMWA). These are also the only unions among the 10 largest that systematically and consistently reported information on delegates and local branches between 1900 and 1920, with proceedings still available either in physical or digital copies.

[24] Branches in arrears on payments to either the state federation or the national organization may also have been excluded from one of the conventions.

independent sources reduces instances of mismeasurement.[25]

Third, these additional records improve the measurement of unionization, particularly in states where federations of labor were established—and held their first conventions—only after 1900.[26] Relying solely on state federation data would imply measuring zero union presence in counties before their respective state federations were founded. While the absence of a state-level body suggests limited organization, some of the largest national unions may already have been active in these areas. Including data on the branches and delegates of these five large unions operating nationwide provides a more accurate measurement of unionization in 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.[27] I first reduce missing observations and misreportings in each source by linearly interpolating the number of union branches and members for counties that are unreported in a given year, but that have representation both in the previous and following decades.[28] Next, for the five unions observed across both sources, I compute the number of members and branches in each county and year by averaging the values reported in each source. When only one source reports positive membership or branch counts, I use that value in the analysis. Finally, I sum the total number of branches and members across all unions at the county-year level to construct measures of unionization over time.

To measure union density, I divide the number of union members by the size of the nonfarm labor force as measured in the U.S. Census (Ruggles et al., 2022).[29] Addition-

[25]Attendance was enforced through fines imposed on branches that sent fewer delegates than they were entitled to. Additionally, branches had no incentive to overreport membership to secure extra delegates, as this would have resulted in higher dues, which were calculated based on membership size.

[26]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.

[27]In Section 5.3, I show that the results are unchanged when using only the main data source (state federations of labor) to measure unionization in a county.

[28]Counties may incorrectly appear to have no union branches or members in a given year for several reasons: error in assigning a locality to the correct county due to homonymous locations, incomplete or inaccurate reporting of delegates present at conventions, or county-specific circumstances that prevented delegates from attending one of the two conventions. The underlying assumption is that a county with union branches and members in, say, 1900 and 1920 is unlikely to have zero branches and members in 1910. I also collect available data from state federation conventions in 1930 to interpolate the data for 1920. Importantly, the results are unchanged if this step is omitted (see Section 5.3).

[29]This includes all men ages 16–64 in nonfarm occupations, excluding managers, proprietors, and private household service workers, as these groups were typically ineligible to unionize during this period (Stewart, 1926; Wolman, 1924). When the estimated number of union members exceeds the labor force,

ally, I construct an indicator for whether a county has any unions, the number of union branches within its territory, and their average size, defined as the number of members divided by the number of branches. All four variables are winsorized at the 1% to remove outliers. As a final validation exercise, I compare these measures of union density to those from an existing historical dataset. While only national-level estimates of union membership are available for the period studied, I verify that my measures are positively correlated with state-level union density estimates by Farber et al. (2021) using Gallup surveys starting in 1937 (Figure H.1).

The final dataset includes information on the location and membership of local union branches in over 2,400 counties between 1900 and 1920.[30]

3.2 Other Main Data Sources

Immigration, population, and labor market outcomes. Data on county population, the number and origin of immigrants, and labor force—total and by occupation—are from the decennial U.S. Census of Population for 1890 (Haines, 2010) and 1900–1920 (Ruggles et al., 2022).[31]

Economic activity. Data on the manufacturing and agricultural sectors come from the Census of Manufacturing (Haines, 2010) and of Agriculture (Haines et al., 2018).

Coal mines. Information on the presence of active coal mines in 1890 is digitized from the *Report on Mineral Industries* for that year (Day, 1892).

The remaining data sources are described in the relevant sections of the paper.

3.3 Summary Statistics

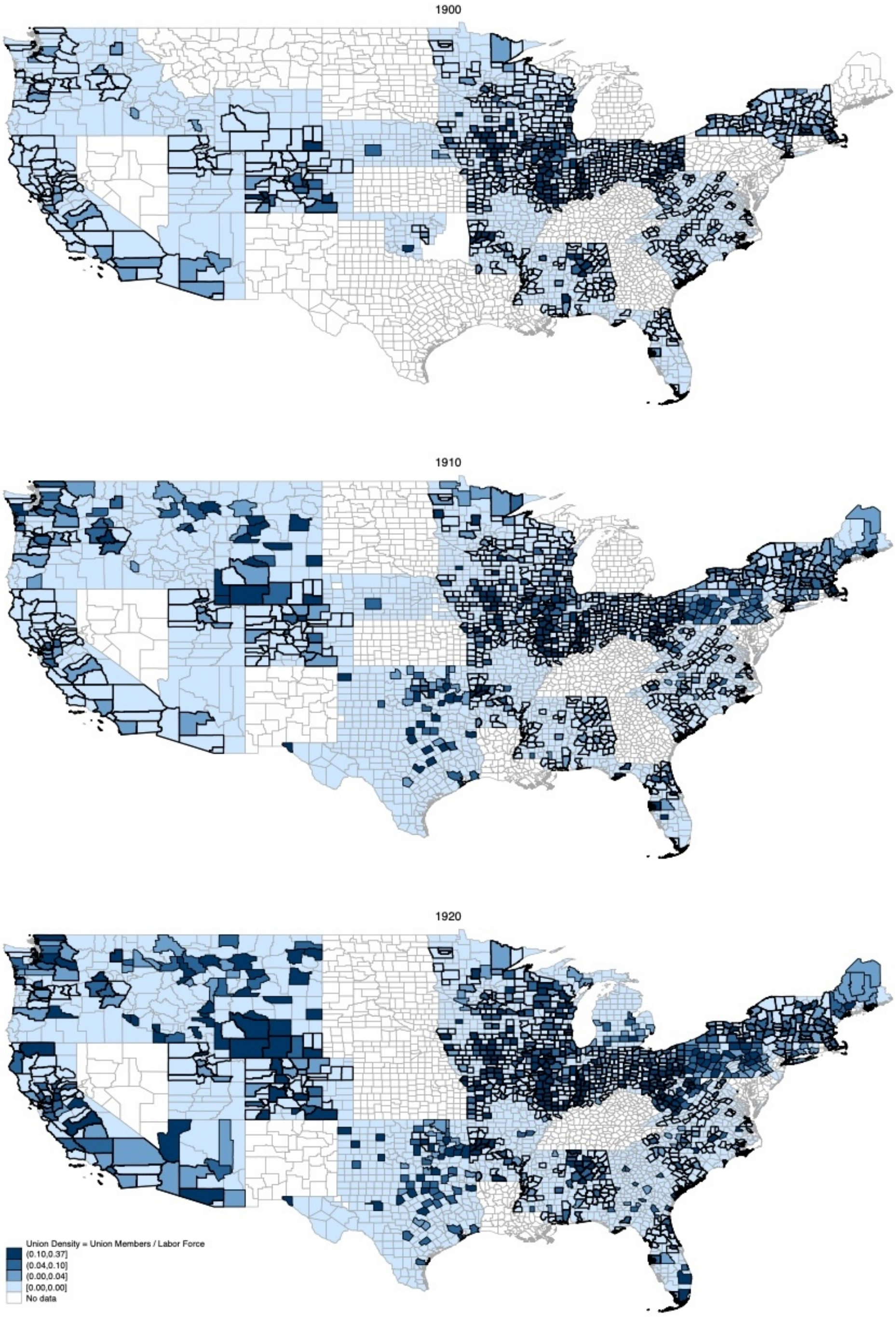
Figure 2 plots union density (the number of union members as a fraction of the labor force) across U.S. counties in 1900, 1910, and 1920. Unionization in 1900 was predominantly concentrated in the Northeast and Midwest. By 1920, unions had also spread to many other regions, including the West and selected 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 the two measures across counties—both within and across states—and over time.

union density is coded as one. This rare event occurs in only three out of the 2,628 county-year observations of the main sample. Section 5.3 also shows that the results are robust to excluding outliers.

[30]The counties not part of the sample are those in states where federations of labor did not adopt proportional representation rules for branches, whose convention proceedings are unavailable, or whose records were incomplete (e.g., lacking branch locations or delegate lists). The main analysis focuses on a balanced panel of 876 urban counties (see Section 3.3). Results are unchanged when applying different sample restrictions or extending the analysis to all available counties (Section 5.3).

[31]Due to the destruction of most of the 1890 Census forms in a fire, full-count data are unavailable for that year, and available information is more limited (e.g., no data on individuals' occupations).

Figure 2: Geographic Distribution of Union Density, 1900–1920



Notes: The maps display county-level union density (i.e., the number of union members as a fraction of the labor force, as defined in Section 3.1) in 1900, 1910, and 1920. The legend represents the deciles of the distribution in 1920. Counties outlined with a thicker border belong to the balanced panel of urban counties used in the main analysis (see Section 3.3). Source: author's calculations from union convention proceedings, as described in Section 3.

In Table 1, I present summary statistics for the main measures of unionization (Panel A), demographic composition (Panel B), labor force (Panel C), and county characteristics at baseline (Panel D). Throughout the paper, I focus on a balanced panel of 876 urban counties between 1900 and 1920, which account for 70% of the male nonfarm labor force in counties where union data are available, and 55% of the same in the entire country.[32]

Table 1: Summary Statistics

	Obs.	Mean	St. Dev.
Panel A: Unionization			
Any Union Present	2,628	0.44	0.50
Number of Union Branches	2,628	2.56	5.30
Union Density (Members / Labor Force)	2,628	0.04	0.08
Union Members per Branch	2,628	48.16	74.38
Panel B: Demographics			
Share of Urban Population	2,628	0.30	0.27
Share of Immigrant Population	2,628	0.09	0.10
Share of European Immigrant Population	2,628	0.08	0.08
Share of European Imm. Pop. (<10 years in U.S.)	2,628	0.02	0.03
Panel C: Labor Force (men ages 16–64)			
Total Labor Force	2,628	14,777.32	44,231.97
Labor Force Participation Rate	2,628	0.91	0.04
Share of U.S.-Born Labor Force	2,628	0.86	0.15
Share of European Immigrant Labor Force	2,628	0.12	0.12
Panel D: Baseline Characteristics (in 1890)			
Share of European Immigrant Population	876	0.10	0.10
Share of Population in Manufacturing	876	0.03	0.04
Share of Population in Agriculture	876	0.45	0.19
Presence of Active Coal Mines	876	0.26	0.44

Notes: The table presents summary statistics for the counties in the main estimation sample described in Section 3. The measures in Panel A, winsorized at the 1% to remove outliers, come from the digitized records of the AFL-affiliated unions described in Section 3.1. The information in Panel B and Panel C is from the full-count Census of Population of 1900, 1910, and 1920 (Ruggles et al., 2022). The data in Panel D are from the 1890 Census of Population and of Manufacturing (Haines, 2010), the 1890 Census of Agriculture (Haines et al., 2018), and the 1890 Report on Mineral Industries (Day, 1892).

On average, 44% of counties have positive union membership, with just under three union branches per county. In the average county, 4% of the labor force are union members, and each union branch has about 50 members.

The average share of the population living in urban areas is 30%. Immigrants make up 9% of the population, with most originating from Europe. Roughly 2% of the total population consist of European immigrants who have entered the United States within the previous decade.

The average county has approximately 15,000 working-age men (16–64 years old)

[32] This sample restriction reflects the fact that both immigration and unionization were predominantly urban phenomena during this time (Foner, 1947; Taft, 1964). A county is classified as urban if, at baseline (i.e., in 1890), it has a share of urban population greater than zero. Additionally, counties with at least one coal mine at baseline are included in the sample, as the coal miners' union was one of the largest labor organizations (Stewart, 1926), and some mining sites were located outside urban areas. In Section 5.3, I show that the results are robust under alternative sample restrictions, such as using a balanced or unbalanced panel of only urban counties or both urban and rural counties.

in the labor force, representing 91% of the male working-age population. Of these, 86% are U.S-born and 12% are European immigrants.

In 1890, a decade before the period analyzed (1900–1920), European immigrants made up an average of 10% of the population. On average, 3% of the population was employed in manufacturing, while 45% of households engaged in farming. Additionally, about one-quarter of the counties in the sample had at least one active coal mine.

4 Empirical Strategy

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

$$y_{ct} = \beta Imm_{ct} + \theta_c + \tau_t + X_{ct} + u_{ct} \tag{1}$$

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. θ_c and τ_t are county and year fixed effects, implying that β is estimated from changes in the fraction of immigrant labor force within the same county over time. X_{ct} are county-level control variables—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. Throughout the analysis, standard errors are clustered at the county level, and all variables are harmonized to reflect 1930 county boundaries (Hornbeck, 2010).[33]

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 (ages 16–64) male population. Focusing on this definition allows for a more confident interpretation of 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 working-age men.[34] The results are unchanged regardless of whether the specification uses the stock of newly arrived immigrants (those entering in the previous decade) or the total stock of immigrants residing in the county, and whether the sample is restricted to working-age men or includes all immigrants regardless of sex or age (see Section 5.3 and Appendix B).

[33]Since county boundaries change over time, I maintain consistent geographic units by holding county boundaries constant throughout the sample period. I use the crosswalks provided by Ferrara et al. (2022) to harmonize all the variables used in the analysis to reflect 1930 county boundaries using area-based weights (Hornbeck, 2010).

[34]Over most of the period 1900–1920, union members were almost exclusively men (Wolman, 1924), and female labor force participation was only 25% (92% for men).

4.2 Instrument for Immigration

Given the hostility of the labor movement towards immigration described in Section 2, union presence in a given county may deter immigration. Such reverse causality would bias the ordinary least squares (OLS) estimates of Equation (1) downwards. Alternatively, immigration and unionization may jointly be determined by a third factor—such as economic growth—introducing the possibility of bias in either direction. 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; Tabellini, 2020). 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), denoted 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}.[35] Formally, the predicted number of immigrants received by county c between Census year $t-10$ and t is given by:

$$\tilde{Z}_{ct} = \sum_{j} \alpha^{j}_{c,1890} O^{j}_{-ct} \tag{2}$$

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.

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 $\alpha^{j}_{c,1890}$. The choice of 1890 as the base year captures many of the key migration networks established during the early part of the Age of Mass Migration, while also preceding both the peak of immigration flows from Europe and the most significant periods of union growth (Figure 1, Panel B).[36] Importantly, 1890 also predates the major shift in the composition of immigration that occurred around the turn of the 20$^{\text{th}}$ century (Figure A.1). As previous work has argued (Abramitzky et al., 2023; Tabellini, 2020), this period is particularly well-suited for the use of shift-share instruments, not only due to changes in the volume of immigration over time but also because of the variation in immigrants' countries of origin in each decade. Different from Tabellini (2020), who employs an analogous identification strategy to predict immigration be-

[35] This leave-out approach addresses the finite sample bias that comes from using own-observation information, which would allow the first stage to load on the endogenous immigrant group-location component of the immigration flows (Goldsmith-Pinkham et al., 2020). See Table A.1 for the list of European origin countries and regions used to construct the shift-share instrument.

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

tween 1910 and 1930, this shift-share instrument leverages additional variation in the composition of immigration that occurred between 1890 and 1900.

Identification assumption. The key identifying assumption behind the instrument described in Equation (2) is that, conditional on the controls, the baseline distribution of immigrants from different European countries across counties is uncorrelated with unobserved factors affecting unionization, except through the subsequent migration flows from those countries to the United States. In other words, absent differences in these immigration flows, counties that were initially more versus less exposed to particular origin countries would have experienced similar trends in unionization over time.[37] Prior studies have argued that nationwide shocks between 1890 and 1920, which were exogenous to county-specific characteristics, make this setting particularly well-suited to the use of shift-share instruments (Abramitzky et al., 2023; Tabellini, 2020). In particular, the trend-break in immigration created by WWI alleviates concerns 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 and the shift in immigrants' composition during this period (Figure A.1) reduce worries about the design being invalidated by serial correlation in migration flows from the same origin to the same U.S. destinations (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. One potential concern is that the economic opportunities attracting more immigrants to a county before 1890 may have also been correlated with long-run trends in unionization. To address this, the preferred specification includes interactions between year dummies and the baseline share of the population employed in each of the two largest industries, agriculture and manufacturing, as well as an indicator for the presence of active coal mines.[38]

These controls account for the fact that both immigration and labor unions were concentrated in areas with better economic and employment opportunities (Abramitzky and Boustan, 2017; Taft, 1964). Specifically, counties with a more substantial manufacturing or mining activity and less agriculture attracted more immigrants early on (Table A.2), and likely experienced greater union growth in the early 20$^{\text{th}}$ century. At the

[37]Suppose that sometime between 1890 and 1920 there was a sudden increase in national migrant inflows from a particular European country (e.g., Austria-Hungary). The assumption would require that, conditional on the controls, counties that in 1890 had a higher share of population from that country would have experienced similar trends in unionization as those with a lower share, absent this differential immigration shock—and similarly for all other European countries. For theoretical foundations on this identification assumption behind shift-share instruments, see Borusyak et al. (2025) and Goldsmith-Pinkham et al. (2020).

[38]Mine operative or laborer was the second most common occupation among recently arrived immigrants between 1890 and 1920, surpassed only by laborers or operatives in manufacturing (Ruggles et al., 2022). Additionally, the mining industry experienced the highest levels of labor unrest during this period (Fishback, 1995; Jeffreys-Jones, 1978).

same time, the controls account for the structural transformation that occurred in the United States between 1880 and 1920, when the share of the male labor force in agriculture declined from 47% to 29%, while manufacturing rose from 12% to 25%. This transformation likely led to higher growth rates in counties that were more rural at the start of the period (Eckert and Peters, 2022), potentially influencing union development. By including these controls, the analysis allows for counties to be on differential trends based on the relative importance of these industries. The results are robust to controlling for various additional county-level variables that could affect both the 1890 immigrant population and subsequent unionization (Figure 4).[39]

Table 2 reports the first stage coefficients, introducing each control incrementally. Across all columns, actual and predicted immigration are positive correlated and all coefficients are statistically significant at the 1% level.

Table 2: First Stage of the Instrumental Variable Estimation

	Dependent variable: Share of Immigrants			
	(1)	(2)	(3)	(4)
Predicted Share of Immigrants	0.336***	0.304***	0.274***	0.273***
	(0.037)	(0.036)	(0.033)	(0.033)
Observations	2628	2628	2628	2628
Dep. var. mean	0.028	0.028	0.028	0.028
Indep. var. mean	0.030	0.030	0.030	0.030
KP F-statistic	80.67	71.09	67.76	67.34
Controls:				
Share of Population in Manufacturing	No	Yes	Yes	Yes
Share of Population in Agriculture	No	No	Yes	Yes
Presence of Coal Mines	No	No	No	Yes

Notes: The observations are at the county-year level. The table reports the first stage of the instrument described in Section 4.2. The dependent variable is the number of European immigrants (men ages 16–64) who entered the United States 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 ages 16–64) who entered the United States 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, measured in 1890 and interacted with year dummies, are also included: the share of the population working in manufacturing (from column 2), the share of the population working in agriculture (from column 3), and an indicator for the presence of active coal mines (column 4). KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Other threats to identification. In robustness analyses, I address other potential threats to identification in several additional ways.[40] First, I control for several county characteristics, measured at baseline and interacted with year dummies, to mitigate concerns about omitted variable bias. This includes directly controlling for the share of the immigrant population in 1890. As a result, the effects of immigration are identified using variation only in the ethnic composition of immigrant enclaves across counties, while

[39] These include, for example, the shares of the labor force in each industry in 1880, when full-count Census data are available, or measures of economic development (e.g., the growth in manufacturing output or the presence of a railroad). Figure 3, row 3, further shows that the findings do not depend on the inclusion of any specific baseline controls. Additional details on the robustness checks are provided in Appendix B.

[40] See Section 5.3 and Appendix B for details on all the robustness exercises.

holding the size of their foreign-born populations constant. This exercise also addresses the concern that a larger immigrant population in 1890 may have an independent, time-varying effect on unionization.

Second, 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 union formation and growth, as well as the subsequent migration patterns of those same groups in each decade between 1890 and 1920 (Goldsmith-Pinkham et al., 2020).[41]

Alternative instrument. Third, I construct an alternative version of the instrument described in Equation (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 exercise addresses the potential concern that realized immigration flows might partly reflect contemporaneous local economic conditions. For instance, if economic booms in certain U.S. regions attracted more migrants from particular European countries, this could generate correlation between the instrument and local (unobserved) shocks to unionization. Using weather-predicted flows isolates variation in immigration that originates abroad and is plausibly exogenous to local U.S. labor market dynamics. I then interact them with the baseline shares of European immigrants from each country j to obtain the alternative instrument. Appendix B.1 describes its construction in detail.

Matching-style exercise and shift-share instrument. Finally, similarly to Bazzi et al. (2023), I combine the shift-share instrument of Equation (2) with a matching-style exercise. In particular, I identify within-state county pairs with similar baseline presence of labor unions in the late 19$^{\text{th}}$ century. Then, I re-estimate the 2SLS analysis including county-pair by year fixed effects. Appendix B.2 provides further details.

I summarize all robustness checks in Section 5.3, after presenting the main results.

5 Main Results

5.1 The Effect of Immigration on Unionization

In Table 3, I present the effects of immigration on the formation and growth of labor unions by estimating Equation (1). I examine four unionization measures: an indicator for whether the county has labor unions (column 1); the number of union branches (column 2);[42] union density, defined as the number of union members as a fraction

[41]In Appendix G, I present the diagnostics on shift-share instruments recommended by Goldsmith-Pinkham et al. (2020), where the estimator is decomposed into several just-identified instrumental variable estimators, each weighted by a so-called "Rotemberg weight".

[42]This variable is transformed using the inverse hyperbolic sine transformation, which is preferred over a log transformation because the variable can take a value of zero when no union branch is observed.

of the labor force (column 3);[43] and the average branch size, defined as the number of union members divided by the number of branches (column 4).[44] All regressions include county and year fixed effects, as well as interactions between year dummies and the baseline share of the population employed in manufacturing and agriculture, and an indicator for the presence of active coal mines (as discussed in Section 4.2). Panel A presents the OLS estimates. All coefficients are positive and statistically significant. This indicates 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-statistic for weak instruments, reported at the bottom of the table, is well-above the conventional level, and indicates that the instrument is strong. All the point estimates are positive and statistically significant at the 5% (column 1) or 1% (columns 2 to 4) level. The 2SLS estimates imply that a four percentage point (one standard deviation) increase in the share of recent immigrants caused an 11 percentage point (24% relative to the mean) higher likelihood that the county had unions (column 1); 0.2 (8% relative to the mean) additional union branches (column 2);[45] a higher share of unionized workers by one percentage point (31% relative to the sample mean, column 3), and 19 more members per branch (40% relative to the sample mean, column 4). In other words, for every 100 immigrants entering the average county, the number of union members increased by nearly 20.[46] A back-of-the-envelope calculation comparing the actual union density in the data to the one predicted by the 2SLS estimates reveals that, in the absence of immigration, the average share of unionized workers between 1900 and 1920 would have been 22% lower.

The difference between OLS and 2SLS estimates indicates that the former are biased

Since the transformation preserves the scale of the variable as the number of branches, the concerns raised by Chen and Roth (2024) regarding the treatment effect's sensitivity to the outcome's units of measurement do not apply. Nonetheless, a decomposition of the estimates for the intensive and extensive margins is presented in Section 5.2. The results are also robust to using a $log(1+x)$ transformation or no transformation at all (estimates unreported for brevity but available upon request).

[43] Throughout the paper, union density is measured as the number of union members as a fraction of the total male labor force in nonfarm occupations, excluding managers, proprietors, and private household service workers, as these groups were typically ineligible to unionize during this period (Stewart, 1926; Wolman, 1924). The results are unchanged when using alternative denominators in the definition of union density, such as including private household workers, farm occupations, or considering the total labor force (estimates not shown for brevity, but available upon request).

[44] To maintain the same sample throughout the analysis, and for consistency with the other measures, this variable is set to zero if the county has no union branch (and, therefore, also no union members). The results are similar when limiting the sample to county-year observations with at least one union branch. See Section 5.2 for a discussion on the effects at the extensive and intensive margins of unionization.

[45] Since the dependent variable in column 2 is transformed using the inverse hyperbolic sine (IHS) transformation, the change in the transformed variable is calculated as $\Delta(IHS(y)) = \beta \cdot \Delta x = 5.245 \cdot 0.04 \approx 0.2$. The corresponding change in the original (non-transformed) units is $\Delta y = sinh(0.2) \approx 0.2$.

[46] In the sample, the average county has 1,037 recently arrived immigrants, a working-age population of 16,074, and a male nonfarm labor force of 9,544. From column 3, this translates to: $0.282 \times \frac{1{,}037}{16{,}074} \times 9{,}544 \approx 174$. Thus, there were roughly 174 additional union members per 1,037 new immigrants.

Table 3: The Effect of Immigration on Organized Labor

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Panel A: OLS				
Share of Immigrants	0.630*	2.089***	0.090*	171.230***
	(0.374)	(0.622)	(0.048)	(58.821)
Panel B: Reduced Form				
Predicted Share of Immigrants	0.716**	1.430***	0.077***	131.170***
	(0.284)	(0.488)	(0.029)	(40.503)
Panel C: 2SLS				
Share of Immigrants	2.626**	5.245***	0.282***	481.007***
	(1.157)	(1.952)	(0.103)	(153.851)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.441	2.563	0.036	48.158
Indep. var. mean	0.028	0.028	0.028	0.028
KP F-statistic	67.34	67.34	67.34	67.34

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. Panel A shows OLS estimates, where the regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. Panel C shows 2SLS estimates. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

downward, and suggests that European immigrants selected counties where unionization was growing more slowly. This is consistent with the historical evidence that many labor unions during this period actively discriminated against immigrants, precluding them from joining their ranks and the occupations they represented (as discussed in Section 2). Table A.3 further shows a negative and statistically significant relationship between unionization and immigration flows. To probe this mechanism more formally, I estimate a control function specification (Garen, 1984; Wooldridge, 2015) that includes the first-stage residual directly in Equation (1). The coefficient on the residual is negative and statistically significant, indicating that the endogenous component of immigration is associated with lower unionization, consistent with downward bias in the OLS estimates (Table A.4). The instrument also corrects attenuation bias due to measurement error in immigration, which would similarly bias OLS toward zero. Finally, because the 2SLS estimate reflects a local average treatment effect (LATE), it may exceed the average effect if immigration had stronger effects in counties more exposed to country-of-origin networks.

5.2 Effect on the Intensive and Extensive Margins

The results in Table 3 do not differentiate between increases along the intensive or extensive margins of unionization. While column 1 shows that immigration increased union presence, the other estimates do not clarify whether the positive effects on unionization are driven by the strengthening of organized labor in already unionized areas, the establishment of unions in new counties, or by both.

To explore this further, I decompose the effect into intensive and extensive margins of unionization. Specifically, I re-estimate columns 2 to 4 of Table 3, interacting the main independent variable with two mutually exclusive indicators within the same regression: one for counties that had at least one labor union branch in 1890 (1,470 counties) and another for counties that did not (1,158 counties).[47] The first interaction captures effect along the intensive margin (growth in places with an existing union presence), while the second measures effects along the extensive margin (the formation of unions where none previously existed).

Table 4 presents the results. All coefficients are positive and statistically significant. For the intensive margin, a four percentage point (one standard deviation) increase in the share of immigrants led to 0.2 more union branches (a 8% increase relative to the sample mean), a 0.9 percentage point rise in union density (a 25% increase relative to the mean), and 19 more members per branch (a 40% increase relative to the mean). For the extensive margin, the same increase in immigration resulted in 0.2 more union branches (a 9% increase relative to the mean), a 1.4 percentage point rise in union density (a 39% increase), and 19 more members per branch (a 40% increase).

This decomposition also provides insights into the relative contribution of the two effects: approximately 39% of the overall impact can be attributed to the intensive margin, while the remaining 61% is driven by the extensive margin.[48] The results indicate that counties with an existing labor movement experienced a growth in union size and new counties saw the establishment of labor unions.

5.3 Summary of Robustness Checks

I conduct several exercises to verify the robustness of the findings. These are visually summarized in Figures 3 and 4, with additional details provided in Appendix B. Each panel presents the coefficients and confidence intervals for the effects of immigration on the four unionization measures analyzed in this paper. For comparison, the coefficients at the top of each panel (in orange) are repeated from Table 3, while those in the

[47] To avoid the "bad-controls" problem (Angrist and Pischke, 2009), the presence of labor union branches is measured in 1890—prior to the study period (1891–1920)—using data on Knights of Labor branches from Garlock (2009).

[48] The relative contributions are determined by the share of counties in each group. This is confirmed by the fact that the weighted average of the two effects corresponds to the overall effect from Table 3.

Table 4: Effects at the Intensive and Extensive Margins

	Number of Union Branches (1)	Union Density (Members / LF) (2)	Union Members per Branch (3)
Intensive Margin:			
Share of Immigrants × (Already Unions = 1)	4.857**	0.227*	476.494***
	(2.059)	(0.117)	(180.057)
Extensive Margin:			
Share of Immigrants × (Already Unions = 0)	5.744**	0.353***	486.807***
	(2.447)	(0.122)	(168.862)
Observations	2,628	2,628	2,628
KP F-statistic	33.31	33.31	33.31
LM F-statistic	35.95	35.95	35.95
SW F-statistic (Already Unions = 1)	68.05	68.05	68.05
SW F-statistic (Already Unions = 0)	57.98	57.98	57.98

Notes: The observations are at the county-year level. The dependent variables are: the (inverse hyperbolic sine of the) number of union branches (column 1); union density, defined as the number of union members divided by the total male nonfarm labor force (column 2); and the number of members per branch, or zero if the county has no union branch (column 3). The mean of the dependent variable in column (1) reflects the average number of union branches, not the transformed value. The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. *Already Unions* is an indicator equal to one if a county has at least one branch of the Knights of Labor in 1890, and zero otherwise. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. LM F-statistic refers to the Lewis-Mertens F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.

numbered rows (in blue) correspond to different robustness exercises.

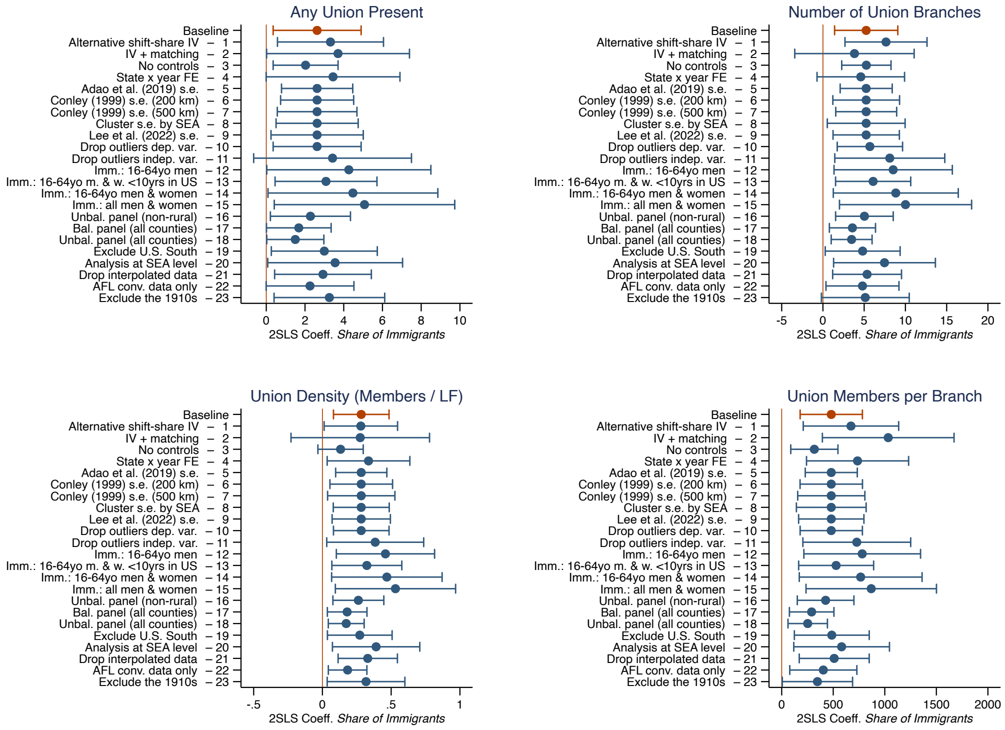
I show that the results are unchanged when using a version of the instrument that relies on weather shocks in each European country to predict the flows of European immigration between 1890 and 1920 (Figure 3, row 1).[49]

Next, building on Bazzi et al. (2023), I combine the shift-share instrument with a matching-style exercise by identifying within-state county pairs that had the closest number of labor unions before 1900 (as a fraction of the county population) and re-estimating the specification including county-pair by year fixed effects (Figure 3, row 2).

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 (Figure 3, rows 3 and 4); clustering standard errors at the State Economic Area (SEA) level, using Conley (1999) standard errors to account for spatial correlation, applying the correction for shift-share estimators proposed by Adao et al. (2019), or computing the tF standard error adjustements following Lee et al. (2022) (Figure 3, rows 5–9); dropping potential outliers (Figure 3, rows 10 and 11); using alternative definitions of the independent variable (Figure 3, rows 12–15); extending the analysis to an unbalanced sample of counties, including rural counties, or excluding

[49] This alternative instrument builds on previous work from Sequeira et al. (2020) and Tabellini (2020).

Figure 3: Summary of Robustness Checks



Notes: The figure presents a summary of the main robustness checks described in Section 5.3. The estimates plotted are the 2SLS coefficients (with corresponding 95% confidence intervals) of *Share of Immigrants* (Imm_{ct}), the main independent variable of Equation (1). The coefficient at the top of each figure (in orange) corresponds to that shown in Table 3. Standard errors are robust and clustered by county. Further details on the robustness checks are presented in Appendix B.

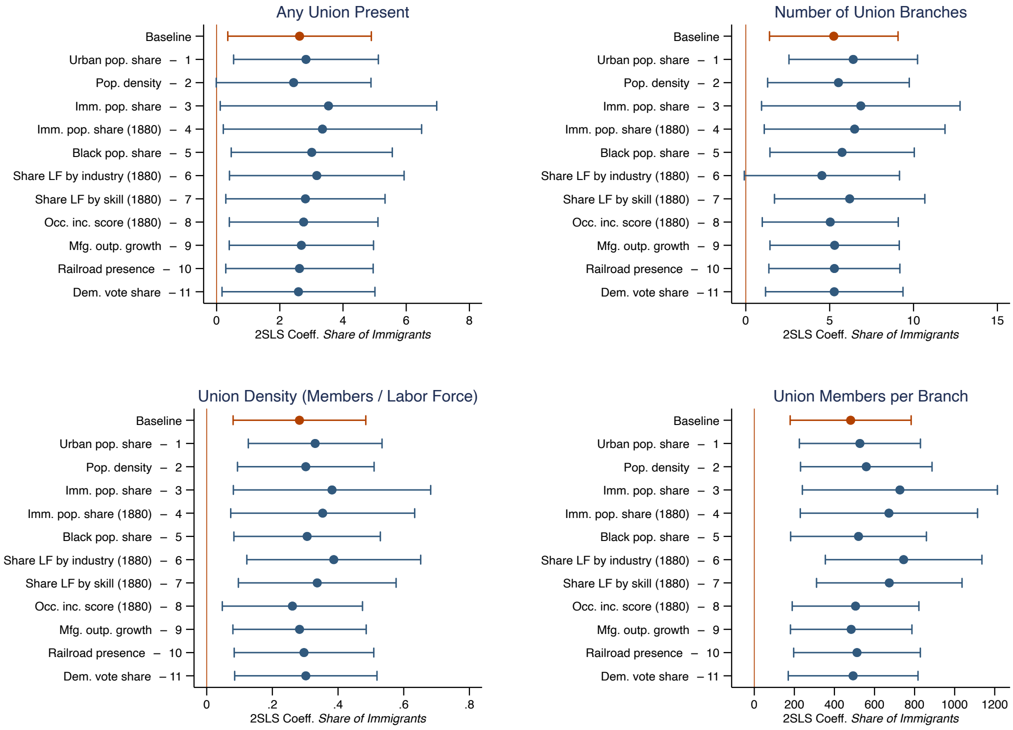
the South from the estimation sample (Figure 3, rows 16–19); performing the analysis at the State Economic Area (SEA) level (Figure 3, row 20); and using different methods for constructing unionization data (Figure 3, rows 21 and 22).

Moreover, 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 (Figure 4).[50] These include the share of urban population and the population density (rows 1 and 2), the share of immigrant population (rows 3 and 4), the share of Black population (row 5), the share of the labor force by industry and skill level (rows 6 and 7), the average occupational income score (row 8), the growth rate of the manufacturing output (row 9), the presence of a railroad (row 10), and the vote share for the Democratic Party in presidential elections (row 11). The exercises presented in rows 3 and 4 are particularly informative, as they identify the effect of immigration from changes in the composition of immigrant inflows, holding constant the overall baseline level of the immigrant population in each county.

Additionally, I interact—one at a time—the initial shares of each immigrant group in the county, i.e., $\alpha^{j}_{c,1890}$ in Equation (2), with year dummies (Figure B.2). This exer-

[50] See Appendix B for the definition of the variables and the omitted variable concern they address.

Figure 4: Robustness Check — Controlling for Additional Baseline Characteristics



Notes: The figure plots the 2SLS coefficients (with corresponding 95% confidence intervals) of *Share of Immigrants* (Imm_{ct}), the main independent variable of Equation (1), augmenting the preferred specification with the variable(s) indicated in each row, measured at baseline and interacted with year dummies. The coefficient at the top of each figure (in orange) corresponds to that shown in Table 3. Standard errors are robust and clustered by county. All robustness checks are described in Section 5.3 and Appendix B.

cise 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).[51]

Finally, I test for the absence of pre-trends by regressing the pre-period changes in unionization, population, and economic outcomes on the subsequent predicted share of immigrants, as determined by the instrument (Table B.2). The lack of statistically significant coefficients indicates that the instrument does not predict more immigration to counties that, before 1890, were already undergoing important changes.

6 Mechanisms

The results presented thus far indicate that counties with higher inflows of immigrants between 1890 and 1920 experienced greater increases in unionization. In this section, I examine the mechanisms underlying this positive effect, presenting evidence consistent

[51] 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. In Appendix G, I present the diagnostics on shift-share instruments recommended by Goldsmith-Pinkham et al. (2020), where the estimator is decomposed into several just-identified instrumental variable estimators, each weighted by a so-called "Rotemberg weight".

with a response by existing workers to the economic and social challenges posed by immigration. Section 6.3 addresses potential alternative explanations for these findings.

6.1 Economic Motivations

As described in Section 2, unions have historically opposed labor supply expansions, fearing downward pressure on wages, deteriorating working conditions, and increased job competition. Scholars have argued that such concerns motivated workers to organize and restrict access to certain jobs (Mink, 1986; Olzak, 1989).[52] Immigration intensified these economic threats, heightening existing workers' incentives to unionize and limit immigrants' entry into the labor force. The results in Table 3 support this hypothesis.

At the same time, workers lacked legal protections to organize. Union recognition often required violent strikes and walkouts (Fishback, 1992; Naidu and Yuchtman, 2016; Olzak, 1989), while courts frequently upheld employers' rights to fire unionizing or striking employees (Foner, 1947; Taft, 1964).[53] Immigration further weakened unions' bargaining power by creating a pool of strikebreakers and increasing employers' ability to replace unionizing workers (Asher, 1982; Montgomery, 1979; Olzak, 1989).

Effects by workers' skills. I test whether the main results reflect workers' responses to the economic threats posed by immigration. Given the competing forces described above, unionization attempts should have been more likely to succeed in occupations with barriers to entry. While efforts to unionize in response to immigration likely occurred across all workers, those in occupations where they could not be *immediately* replaced by immigrants were better positioned to organize and protect their jobs from future competition. Similarly, employers without readily available replacement workers may have faced greater pressure to accommodate unionization efforts.

To test this mechanism, I leverage differences in skill requirements across occupation.[54] The estimates in Table 5 indicate that immigration had positive effects on all

[52]Common methods included restricting membership to U.S.-born workers, requiring union membership as a condition of employment, and regulating apprenticeships to determine who could acquire job-specific skills (Ignatiev, 1994; Mink, 1986; Stewart, 1926; Witwer, 2002).

[53]The introduction or survival of a union was often the central issue of many violent strikes (Taft and Ross, 1969). The mining industry experienced the highest levels of violent unrest and strike-related fatalities (Fishback, 1995; Jeffreys-Jones, 1978).

[54]The classification of skilled and unskilled occupations follows Katz and Margo (2014). Skilled occupations include professional, technical, clerical, sales, and craft workers. Unskilled occupations include operatives, laborers, and service workers. The following AFL-affiliated national unions organized unskilled workers: the Amalgamated Meat Cutters (AMC), the International Brotherhood of Teamsters (IBT), the International Longshoremen Association (ILA), the International Union of United Brewery Workmen of America, the International Union of Mine, Mill and Smelter Workers (IUMMSW), the United Garment Workers of America (UGWA), the United Mine Workers of America (UMWA), and the United Textile Workers of America (UTW). All other AFL unions are classified as skilled (Foner, 1947; Taft, 1964; Stewart, 1926). Since the UMWA organized both skilled and unskilled workers, I allocate its members and branches proportionally based on the 1890 U.S. Census *Report on Mineral Industries*

four unionization measures among skilled workers. Counties with higher immigration experienced increases in union presence, the number of branches, union density, and members per branch. By contrast, the effects on unskilled workers were smaller and statistically insignificant.[55]

Table 5: Heterogeneous Effects by Workers' Skills

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Panel A: Skilled Workers				
Share of Immigrants	2.543**	4.898***	0.753***	463.820***
	(1.162)	(1.760)	(0.237)	(153.119)
Dep. var. mean	0.438	2.119	0.069	47.552
Panel B: Unskilled Workers				
Share of Immigrants	1.302	1.107	-0.001	202.708
	(1.057)	(1.113)	(0.043)	(138.872)
Dep. var. mean	0.208	0.402	0.015	27.273
Observations	2,628	2,628	2,628	2,628
KP F-statistic	67.34	67.34	67.34	67.34

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. In Panel A, the dependent variables refer to skilled workers; in Panel B, to unskilled workers (as detailed in Section 6). The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

These patterns are consistent with the idea that immigration heightened concerns about wage pressures and job security, but workers differed in their ability to respond collectively. Skilled workers, whose specialized expertise and control over access to their trades made them less immediately replaceable, were better able to organize when immigration increased (Asher, 1982; Mink, 1986; Olzak, 1989). Such features facilitated the formation and expansion of their unions to safeguard their positions and restrict entry. By contrast, unskilled workers faced greater replacement risks from inexpensive immigrant labor, which likely undermined their capacity to unionize (Montgomery, 1979). Taken together, the evidence suggests that while immigration generated economic pressures for workers across the skill distribution, only those with sufficient organizing capacity could translate these concerns into successful collective action. This

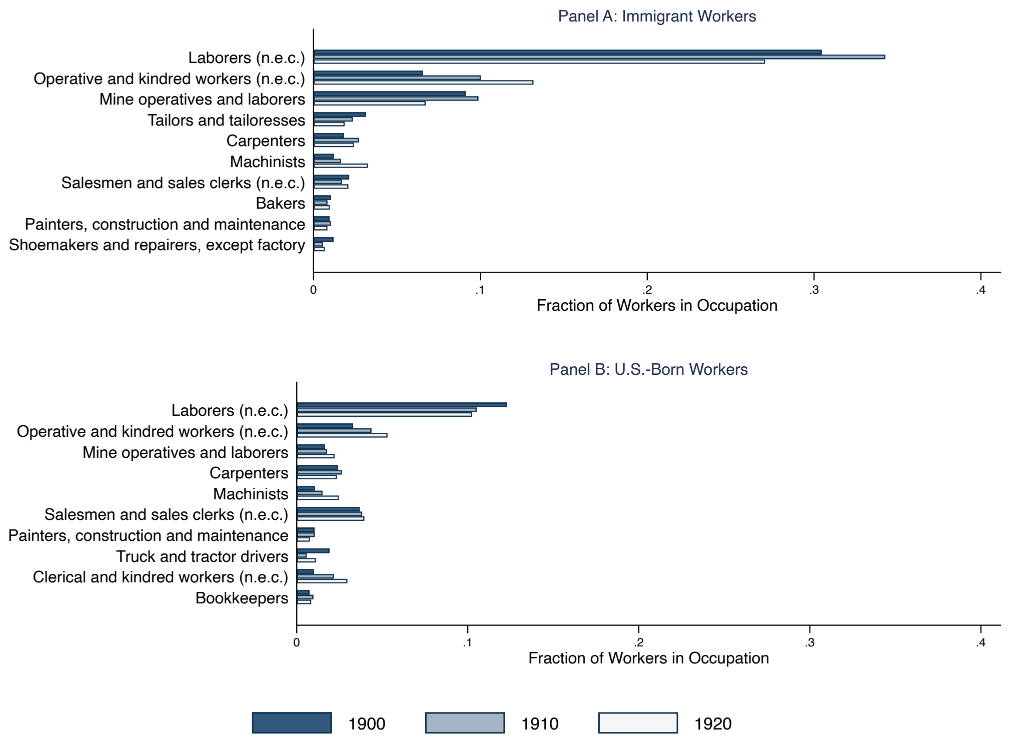
(Day, 1892): 57% skilled (miners and mechanics) and 43% unskilled (laborers and boys younger than 16 years old).

[55] The findings are robust to alternative definitions of skill levels based on apprenticeship requirements (Stewart, 1926) or broader entry barriers involving both human and physical capital. These additional results are unreported for brevity but available upon request.

interpretation is consistent with the notion that institutional frictions and employer opposition shape whether worker demand for collective action results in actual union formation (Naidu, 2022), as well as recent evidence showing that economic pressure alone does not necessarily translate into higher union membership (Pezold et al., 2023).

Exposure to immigrant labor competition. Next, I examine more directly how labor competition pressures influenced union growth. Descriptive evidence reveals an overlap between the occupations of new immigrants and those of existing workers. Both unskilled (e.g., laborers and operatives) and skilled (e.g., carpenters and machinists) occupations were prominently represented among immigrants arriving between 1890 and 1920 (Figure 5).

Figure 5: Most Common Occupations Among Immigrant and U.S.-Born Workers



Notes: The figure displays the ten most common nonfarm and nonmanagerial occupations among European immigrants who entered the United States in the previous ten years (Panel A) and among U.S.-born workers (Panel B) in 1900, 1910, and 1920. The shares indicate the proportion of immigrants or U.S.-born workers in each occupation relative to the total number of workers in the respective group.

To formally test the role of immigrant competition, I interact the main regressor of interest from Equation (1) with a time-varying measure of a county's exposure to immigrant labor competition.[56] The measure consists of two terms. The first term captures the share of immigrant workers in each occupation o who entered the United States (excluding those settling in county c) between $t - 10$ and t, relative to the total immigrant labor force entering the United States (net of c) over the same period. To mitigate endogeneity concerns, this term uses national rather than county-level quantities, as local

[56]This approach mirrors strategies used in studies of import competition (Autor et al., 2020) and historical labor markets (Alsan et al., 2020).

employment in a given occupation may itself be influenced by the level of immigration or unionization. The second term is a weight, represented by the share of U.S.-born workers in county c and occupation o at the start of that decade:[57]

$$Competition_{c,t} = \sum_{o} \frac{Imm^{o}_{-c,t}}{Imm^{LF}_{-c,t}} \times \frac{USborn^{o}_{c,t-10}}{USborn^{LF}_{c,t-10}} \tag{3}$$

The intuition behind this measure is straightforward: counties where U.S.-born workers (at the start of the decade) were concentrated in occupations heavily represented by recently arrived immigrants (in the rest of the country) faced greater exposure to labor competition.

Table 6 presents the results separately for skilled (Panel A) and unskilled (Panel B) workers.[58] Panel A shows that all uninteracted estimates are positive and statistically significant. Moreover, the interaction coefficients are also significant, indicating that counties more exposed to immigrant competition in skilled occupations experienced greater growth in skilled unionization. By contrast, Panel B displays negative coefficients for the interaction terms among unskilled workers, indicating that increased competition hindered union growth in this group. The results are consistent with the interpretation that the competition measure primarily captures exposure to immigrant-induced supply shocks rather than contemporaneous demand shifts.[59]

In sum, immigration-driven labor competition had opposite effects across skill groups—fostering unionization among skilled workers while hindering it among the unskilled. Skilled workers organized to guard against both immediate and future economic threats posed by immigration, whereas unskilled workers faced greater obstacles due to the ready availability of replacement labor, which limited their ability to unionize successfully.

[57] Due to the unavailability of the 1890 full-count Census, the number of U.S.-born workers in each occupation at the beginning of the 1890s refers to the year 1880. The findings are robust to restricting the analysis to the decades starting in 1900 and 1910 (unreported for brevity but available upon request).

[58] Table A.5 reports the results for all workers, interacting the main independent variable with both measures of skilled and unskilled competition. The findings remain unchanged. The results are robust to several alternative constructions of the competition measure, including re-estimating the specification using weather-predicted immigration flows by occupation and origin country, restricting inflows to immigrants who arrived in the United States within the previous five years, and controlling for baseline manufacturing output growth. These exercises confirm that the measure captures exogenous labor-supply pressures rather than contemporaneous demand shifts. For brevity, the corresponding tables are not shown but are available upon request.

[59] The findings remain robust under several alternative constructions of the competition measure designed to further isolate exogenous supply-side variation, such as (i) using weather-predicted immigration flows by occupation and origin country, which rely on shocks abroad rather than U.S. labor demand; (ii) restricting inflows to immigrants who arrived within the previous five years, limiting concerns about post-arrival occupational mobility or skill acquisition; and (iii) controlling for baseline manufacturing output growth, to ensure the measure is not proxying for local industrial demand. For brevity, the corresponding tables are not presented but are available upon request.

Table 6: Heterogeneous Effects by Immigrant Labor Competition

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Panel A: Skilled Workers				
Share of Immigrants [1]	2.228**	3.974***	0.648***	365.007***
	(1.002)	(1.336)	(0.216)	(127.993)
Share of Immigrants x Competition [2]	1.005**	3.123**	0.343**	321.460***
	(0.465)	(1.211)	(0.137)	(120.482)
KP F-statistic	46.40	46.40	46.40	46.40
LM F-statistic	38.44	38.44	38.44	38.44
SW F-statistic [1]	82.34	82.34	82.34	82.34
SW F-statistic [2]	45.65	45.65	45.65	45.65
Panel B: Unskilled Workers				
Share of Immigrants [1]	2.555*	2.591*	0.032	356.032*
	(1.304)	(1.375)	(0.048)	(185.777)
Share of Immigrants x Competition [2]	-0.769***	-0.910***	-0.020	-94.496**
	(0.296)	(0.346)	(0.015)	(40.949)
KP F-statistic	39.20	39.20	39.20	39.20
LM F-statistic	10.97	10.97	10.97	10.97
SW F-statistic [1]	80.32	80.32	80.32	80.32
SW F-statistic [2]	87.97	87.97	87.97	87.97
Observations	2,624	2,624	2,624	2,624

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. In Panel A, the dependent variables refer to skilled workers; in Panel B, to unskilled workers (as detailed in Section 6). The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. *Competition* is a (standardized) measure of immigrant labor competition, based on the prevailing occupations (skilled in Panel A; unskilled in Panel B) among the U.S.-born workers in the county at the beginning of each decade and among the immigrants entering all other U.S. counties during that decade, as detailed in Section 6. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. LM F-statistic refers to the Lewis-Mertens F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.

6.2 Economic and Social Motivations

The results presented thus far have examined the purely economic channels through which immigration strengthened labor unions. However, social concerns, such as opposition to cultural change, may have also motivated workers to organize and exclude newcomers from the labor market. This possibility is underscored by the nativist rhetoric often adopted by the labor movement during this period and its strong advocacy for immigration restrictions throughout the 20th century (Goldin, 1994; Mink, 1986). 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), leaving the net effect of these social forces unclear. This section presents evidence consistent with both economic and social reasons motivating existing workers to unionize in response to immigration.

Heterogeneity by immigrants' origin. As described in Section 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

$$y_{ct} = \beta_1 Imm_{ct}^{SE} + \beta_2 Imm_{ct}^{NW} + \theta_c + \tau_t + X_t + u_{ct} \tag{4}$$

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.[60] Equation (4) is estimated using two separate instruments, one for each group, constructed by summing the predicted immigration (as described in Section 4.2) from each origin region. The results, presented in Table 7, show that statistically significant increases in unionization were driven by the inflow of immigrants from Southern and Eastern Europe.[61]

It is important to note that these results are also consistent with economic explanations. Immigrants from Southern and Eastern Europe may have had lower wage expectations than those from Northern and Western Europe, making them more likely to be perceived as a threat to existing workers' conditions.

Heterogeneity by attitudes towards immigration. To further investigate the role of social factors in driving unionization, I test whether the effects were stronger in counties with more negative attitudes toward immigration. In the absence of a direct measure, I use two proxies that likely reflect a county's hostility towards immigrants. The first is the historical vote share for the Know Nothing Party, a nativist political party that, in the mid-1850s, campaigned nationwide on an anti-Catholic and anti-Irish platform (Alsan et al., 2020).[62] The second is the baseline level of residential segregation between U.S.-born and European immigrants.[63] Since residential segregation usually results from

[60]The classification of European countries follows the one of IPUMS (Ruggles et al., 2022). Northern and Western European countries include: Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Liechtenstein, Luxembourg, Netherlands, Norway, Sweden, Switzerland, and United Kingdom. Southern and Eastern European countries include: Albania, Austria-Hungary, Baltic States, Bulgaria, Czechoslovakia, Gibraltar, Greece, Italy, Malta, Poland, Portugal, Romania, Russia, Spain, Yugoslavia.

[61]The same conclusions apply when alternatively estimating immigration based on whether the origin countries were predominantly Protestant or non-Protestant (results unreported for brevity but available upon request).

[62]In the 1856 presidential election, the party officially ran under the name "American Party".

[63]I construct an index of residential segregation of European immigrants, building on the procedure

Table 7: Heterogeneous Effects by Origin of Immigrants

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of S/E European Immigrants	3.708*	8.222**	0.412**	660.666**
	(2.197)	(4.031)	(0.209)	(305.790)
Standardized coefficient	*[0.234]*	*[0.220]*	*[0.170]*	*[0.278]*
Share of N/W European Immigrants	-0.209	-2.548	-0.058	10.681
	(2.880)	(4.589)	(0.398)	(506.582)
Standardized coefficient	*[-0.008]*	*[-0.042]*	*[-0.015]*	*[0.003]*
Observations	2,628	2,628	2,628	2,628
KP F-statistic	10.42	10.42	10.42	10.42
LM F-statistic	12.35	12.35	12.35	12.35
SW F-statistic (S/E Europe)	22.78	22.78	22.78	22.78
SW F-statistic (N/W Europe)	39.68	39.68	39.68	39.68

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the log-transformed value. The regressors of interest are the number of immigrants (men ages 16–64) from Southern/Eastern Europe or Northern/Western Europe who entered the United States 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 4.2 and Section 6. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. LM F-statistic refers to the Lewis-Mertens F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.

either collective efforts to exclude minorities or the majority group moving away from ethnically mixed neighborhoods (Boustan, 2010, 2013), higher levels of this measure likely reflect greater levels of discrimination against immigrants.

The results are reported in Table 8, where the main independent variable is interacted with an indicator for whether a county has a high historical vote share for the Know Nothing party (Panel A) or a high baseline level of residential segregation (Panel B).[64] Using either proxy, the findings indicate that immigration strengthened organized labor more prominently in counties with higher resentment towards immigration.

Altogether, these findings suggest that social factors, alongside economic motivations, contributed to the expansion of labor unions. Unionization was more prominent in counties that received larger shares of culturally distant immigrants and where attitudes towards immigration were likely more negative.[65] Importantly, these social dy-

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 D.

[64] In Table 8, high is defined as above the first tercile of the sample distribution. The results from interacting the main independent variable with indicators for each tercile are presented in Table A.6.

[65] The findings presented in this section are very similar when focusing solely on skilled workers and they are robust to the inclusion of the immigrant labor competition measure described in Equation (3). These results are not shown for brevity, but they are available upon request.

Table 8: Heterogeneous Effects by Attitudes Towards Immigration

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Panel A: V = Vote share Know-Nothing party				
Share of Immigrants [1]	1.631	-0.031	0.039	378.471
	(1.612)	(2.390)	(0.183)	(306.528)
Share of Immigrants × High V [2]	1.182	5.477**	0.203	162.836
	(1.386)	(2.239)	(0.146)	(246.915)
Sum of coefficients: [1] + [2]	*2.813*	*5.446**	*0.242**	*541.306***
Observations	2,103	2,103	2,103	2,103
KP F-statistic	23.63	23.63	23.63	23.63
LM F-statistic	18.46	18.46	18.46	18.46
SW F-statistic [1]	39.15	39.15	39.15	39.15
SW F-statistic [2]	86.90	86.90	86.90	86.90
Panel B: V = Index of residential segregation				
Share of Immigrants [1]	0.275	1.259	-0.055	-64.595
	(1.327)	(2.139)	(0.162)	(188.162)
Share of Immigrants × High V [2]	2.769**	4.603**	0.403**	643.350***
	(1.173)	(1.919)	(0.159)	(175.594)
Sum of coefficients: [1] + [2]	*3.044***	*5.863****	*0.348****	*578.755****
Observations	2,565	2,565	2,565	2,565
KP F-statistic	31.71	31.71	31.71	31.71
LM F-statistic	18.34	18.34	18.34	18.34
SW F-statistic [1]	48.30	48.30	48.30	48.30
SW F-statistic [2]	74.83	74.83	74.83	74.83

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The coefficients and standard errors for the sum of the two independent variables are reported at the bottom of each panel. The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The regressor of interest is the number of European immigrants (men 16–64) who entered the United States 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 4.2. In Panel A, Share of Immigrants is interacted with indicators for whether the county has a low (first tercile) or high (second or third tercile) historical vote share for the Know-Nothing party (see Section 6 for more details). In Panel B, Share of Immigrants is interacted with indicators for whether the county has low (first tercile) or high (second or third tercile) residential segregation at baseline (see Section 6 and Appendix D for more details). All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. LM F-statistic refers to the Lewis-Mertens F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.

namics are closely intertwined with the economic motivations discussed in Section 6.1. Fears of economic competition from immigrants may have reinforced negative stereotypes or heightened perceptions of declining social standing (Gidron and Hall, 2017). Overall, the evidence supports the role of both economic and social motivations in driving the observed growth of organized labor.

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 higher rates than U.S.-born workers. Although data on the individual union members are not available, I leverage information on local union representatives described in Section 3 to infer the ethnic composition of their branches. Union delegates, as leaders of the organizations they represented, acted as spokespeople for their local branches at state and national conventions and made decisions on behalf of the members who elected them. Consequently, their ancestry serves as a proxy for the ethnic composition of their branch.

As a first step, I infer the origins of the delegates using their last names and historical de-anonymized full-count U.S. Census data.[66] Panel A of Figure A.5 shows that, as expected, most of the union leaders are identified as U.S.-born. In Panel B, I break down the shares of delegates by the inferred 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, counties that received more immigrants might still have experienced a rise in the proportion of European leaders. For instance, if newly arrived immigrants joined labor unions in large numbers, we would expect an increase in the share of European delegates, as the newcomers would likely gain the voting power to elect them. To test this, I use the county-level proportion of leaders with U.S.-born and immigrant last names as dependent variables in Equation (1). The coefficients, presented in Panel A of Table A.7, indicate that the inflow of immigrants did not increase the proportion of leaders with immigrant last names. The coefficients in columns 1 to 3, estimated on the whole sample of counties, show that immigration increased the share of U.S.-born leaders, while having no effect on the one of immigrants. The coefficients in columns 4 to 6, estimated for counties with at least a delegate in every year—although imprecisely estimated—paint a similar picture.[67]

These findings are consistent with the anecdotal and historical evidence that the observed increase in unionization during this period was not driven by a larger participation of immigrant laborers, but rather by U.S.-born (and other existing) workers

[66] By relying solely on last names, the concern of misclassifying an immigrant as a U.S.-born due to the use of American-sounding first names for assimilation (Abramitzky et al., 2020) does not apply. Similarly, potential modifications of union leaders' last names (e.g., Rossi to Ross or Schmidt to Smith) are unlikely to pose a problem. If such modifications were widespread, the newly adapted last names would appear more frequently among immigrants in the Census, thereby increasing the proportion of union delegates classified as Europeans. For details on the procedure used, see Appendix C. An alternative approach would involve linking individuals to the Census directly using their full names. However, many union convention proceedings report only the delegates' last name and initials, substantially limiting the number of records that could be matched. Additionally, the absence of key variables like year of birth (or age) for union leaders further constrains direct matching to Census data.

[67] Analogous conclusions hold when looking at the proportion of union leaders with either Northern/Western European or Southern/Eastern European ancestry (Panel B of Table A.7).

(Mink, 1986; Taft, 1964), who maintained control of labor unions throughout the first two decades of the 20th century. However, these results should be interpreted with some caution. Because the data capture union leaders rather than the broader rank and file, they cannot fully rule out that immigrants contributed to unionization in other ways. In line with this interpretation, many AFL-affiliated unions explicitly barred non-citizens from membership (Foner, 1947; Mink, 1986; Stewart, 1926) and employers often used newly arrived immigrants as strikebreakers to weaken union efforts (Foner, 1947; Olzak, 1989; Taft, 1964). These dynamics help explain how immigration could increase labor conflict and spur union formation without leading to greater immigrant representation among union leaders. The AFL leadership's support for immigration restrictions during this period (Goldin, 1994) further illustrates the craft-protection logic that limited immigrant participation in these organizations.

Immigrants' exposure to unions or socialist ideologies. A second possibility is that immigrants coming from European countries with well-developed labor unions or strong socialist movements at the turn of the 20th century may have brought ideas and experiences from their home countries that contributed to the growth of unionization in their destination counties. This explanation would align with existing research showing that Europeans who arrived in the United States between 1910 and 1930 promoted the spillover of ideologies to U.S.-born individuals (Giuliano and Tabellini, 2022). Although the conservative nature of AFL unions and the results just presented do not support this hypothesis, I formally test it by estimating the effect of immigration separately for immigrants from countries with or without strong labor unions (Table A.8) and with higher or lower support for socialist parties (Table A.9).[68] The results rule out this possibility. The coefficients of the share of immigrants from the United Kingdom and Ireland—the countries with the strongest labor movements at the turn of the 20th century (Crouch, 1993)—and from countries with higher support for socialist parties, are never statistically significant. On the contrary, all the coefficients for immigration from the rest of Europe are positive and nearly always statistically significant.[69]

World War I and the First Red Scare. A third possibility is that the growth in unionization was driven by the abrupt economic and social changes of the late 1910s. First,

[68] As in Section 6.2, the regressions are estimated using two separate instruments, one for each group, constructed by summing the predicted immigration (as described in Section 4.2) from each origin region. I use data from Crouch (1993) to classify European countries into these two groups. Appendix E provides more information on the data and on labor unions in Europe during this period. The data for the vote shares obtained by socialist parties come from Austrian National Library (2024), Mackie and Rose (2016), and Nohlen and Stöver (2010). Appendix F reports more detailed information on the data sources and the socialist vote shares by country for this time period.

[69] The conclusions are unchanged when using alternative classifications of European countries based on the strength of their labor unions, such as considering France as having a strong labor movement in this period (Friedman, 1998), or expanding this definition to any country that had unions as of 1900 (Crouch, 1993): Austria, Belgium, Denmark, France, Germany, Ireland, Italy, Norway, Sweden, and United Kingdom. These results are unreported for brevity but available upon request.

during the two years of U.S. involvement in World War I (1917–1918), the National War Labor Board was established to mediate agreements between labor and management, discouraging strikes and lockouts, while industries involved in war production rapidly expanded. This environment facilitated an overall growth of the labor movement (Wolman, 1924). Second, the First Red Scare (1917–1920) cast immigrants from Southern and Eastern Europe as a threat to American political and social stability. These events may have prompted existing workers to respond more strongly to immigration from these regions or encouraged employers to favor conservative AFL unions over more radical organizations. To ensure that these events are not confounding the estimates, I perform a robustness check excluding the 1910s from the analysis (i.e., removing the effects of immigration during 1911–1920 on unionization in 1920). The resulting estimates, shown in row 23 of Figure 3, remain positive and statistically significant.[70]

Interactions with other labor unions. A fourth possibility is that the growth in unionization was driven by the presence of other prior or contemporaneous labor organizations, such as the Knights of Labor (KoL) or the Industrial Workers of the World (IWW). For instance, the pre-existing presence of the KoL may have facilitated the establishment of new AFL unions by providing a pool of potential members after the KoL's decline in the late 1880s.[71] Similarly, the IWW's radicalism might have accelerated the growth of AFL unions by pushing workers toward—or inducing employers to favor—more conservative organizations. To explore these possibilities, I augment Equation (1) by including an interaction term between the share of immigrants and the number of branches of each of these organizations.[72] The results, presented in Table A.10, indicate that the growth in unionization was not driven by a larger presence of either the KoL (Panel A) or the IWW (Panel B).[73] All uninteracted coefficients remain positive and statistically significant, while the interactions are always small and statistically insignificant. The only exception is column (1) of Panel B, where the interaction is marginally significant but negative, suggesting that, if anything, more IWW branches may have reduced the presence of AFL unions.

Other economic channels. Finally, it is possible that the growth in unionization was driven by differential economic expansion—or contraction—experienced by counties receiving larger shares of immigrants. Table A.11 shows that this is not the case. Immigration had no effect on economic indicators such as the labor force participation rate

[70]Excluding the 1910s immigration inflows and the 1920 unionization outcomes from the analysis does not affect the findings of Table 5 either, confirming much larger effects among skilled workers. These results are unreported for brevity but available upon request.

[71]Alternatively, the results could reflect a re-activation of anti-immigrant sentiment among AFL members, particularly in areas where the KoL were previously established (Cantoni et al., 2019).

[72]The variables for the number of branches are transformed using the inverse hyperbolic sine transformation. The number of KoL branches is measured in 1880 (Garlock, 2009), while IWW branches refer to those active between 1906 and 1917 (Gregory, 2015).

[73]These findings are robust to using an indicator for the presence of any branch instead of the number of branches, and irrespective of whether KoL branches are measured in 1880, 1890, or both.

or the total manufacturing output (either divided by the manufacturing labor force in the county or as a proportion of national output). Furthermore, immigration did not increase the share of the labor force in skilled occupations, ruling out the possibility that the observed effects were driven by the expansion of sectors requiring skilled labor due to the immigration inflows (for example, if immigration had boosted demand for carpenters and construction workers to support new housing development).

This discussion suggests that the results are unlikely to be driven by the preferences or ideologies of immigrants, by major political and economic events of the period, by other labor organizations, or by the effects of immigration on the local economy.

7 Conclusion

Despite the enduring significance of labor unions, rigorous evidence on the determinants of their origins and early growth remains limited. This paper examines how mass immigration in the early 20^{th}-century United States shaped the emergence and development of organized labor. The results reveal that immigration increased union presence, the number of union branches, union density, and average membership, as existing workers formed and joined unions in response to the economic and social challenges posed by immigration.

This study identifies immigration as a key driver of early unionization, estimating that, without immigration, average union density between 1900 and 1920 would have been 22% lower. The findings also highlight an unexplored consequence of immigration: the development of institutions designed to protect incumbent workers' status in the labor market, with effects lasting into the present. Notably, these results broaden our understanding of immigration's implications, showing that responses to immigration extend beyond support for conservative parties or anti-immigration policies, as emphasized by prior research. Immigration can also foster the formation and growth of organizations, such as unions, with broad economic and political impacts.

While the quantitative estimates in this paper are specific to the historical context examined, the findings point to broader mechanisms that remain relevant today. The renewed interest in unions may reflect workers' responses to modern labor market pressures—such as immigration, globalization, and technological change. These dynamics are not unique to the United States. They also resonate with advanced economies facing similar challenges and with industrializing nations undergoing economic transformations comparable to early 20^{th}-century America, particularly in contexts with limited institutional protections for collective bargaining and workers' rights.

At the same time, it is important to recognize that the form of craft-based and often exclusionary unionism examined in this paper is not the only model of labor organization. The mechanisms highlighted here therefore speak most directly to settings where unions emerged to protect the status of incumbent workers. Whether immigration would have fostered different forms of collective organization under alternative institutional environments remains an important and open question.

Relatedly, this study motivates several avenues for future research. First, it underscores the need for further exploration of the drivers of organized labor's growth across different economic contexts and time periods. This includes examining whether similar mechanisms may arise in response to other labor market pressures, such as internal migration or changes in workforce composition, and how institutional or political environments shape the organizational forms that unions take. Second, the comprehensive data collected for this paper provide opportunities to investigate numerous additional questions, such as the long-term consequences of early 20$^{\text{th}}$-century unionization on immigrants' experiences and its broader impact on the U.S. economy and political landscape.

References

Abramitzky, R., Ager, P., Boustan, L., Cohen, E., and Hansen, C. W. (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 Boustan, L. (2017). Immigration in American Economic History. *Journal of economic literature*, 55(4):1311–45.

Abramitzky, R., Boustan, L., and Eriksson, K. (2020). Do Immigrants Assimilate More Slowly Today Than in The Past? *American Economic Review: Insights*, 2(1):125–141.

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

Acemoglu, D. and Robinson, J. A. (2013). Economics Versus Politics: Pitfalls of Policy Advice. *Journal of Economic perspectives*, 27(2):173–192.

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

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

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

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

Alesina, A. and Tabellini, M. (2024). The Political Effects of Immigration: Culture or Economics? *Journal of Economic Literature*, 62(1):5–46.

Alsan, M., Eriksson, K., and Niemesh, G. (2020). Understanding the Success of the Know-Nothing Party. Technical report, National Bureau of Economic Research.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Asher, R. (1982). 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.

Austrian National Library (2024). ANNO: AustriaN Newspapers Online (Historische Österreichische Zeitungen und Zeitschriften Online). https://anno.onb.ac.at.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review*, 110(10):3139–3183.

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

Bazzi, S., Ferrara, A., Fiszbein, M., Pearson, T., and Testa, P. A. (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.

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

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

Bogue, D. J. (1951). *State Economic Areas: A Description of the Procedure Used in Making a Functional Grouping of the Counties of the United States*. US Government Printing Office.

Bolt, J. and Van Zanden, J. L. (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*.

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

Borusyak, K., Hull, P., and Jaravel, X. (2025). A Practical Guide to Shift-Share Instruments. *Journal of Economic Perspectives*, 39(1):181–204.

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

Cantoni, D., Hagemeister, F., and Westcott, M. (2019). Persistence and Activation of Right-Wing Political Ideology.

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

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

Chen, J. and Roth, J. (2024). Logs with Zeros? Some Problems and Solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Chernozhukov, V. and Hansen, C. (2008). The Reduced Form: A Simple Approach to Inference with Weak Instruments. *Economics Letters*, 100(1):68–71.

Cohen, L., Malloy, C. J., and Nguyen, Q. (2016). The Impact of Forced Migration on Modern Cities: Evidence from 1930s Crop Failures. *Available at SSRN 2767564*.

Collins, W. J. and Niemesh, G. T. (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.

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.

Day, D. T. (1892). *Report on Mineral Industries in the United States at the Eleventh Census, 1890*, volume 14. Norman Ross Pub.

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

Dodini, S., Willen, A., and Zhu, J. (2023). The Role of Labor Unions in Immigrant Integration. *Working Paper*.

Downes, H. (2024). Did Organized Labor Induce Labor? Unionization and the American Baby Boom. *Working Paper*.

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

Eckert, F. and Peters, M. (2022). Spatial Structural Change. Technical report, National Bureau of Economic Research.

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

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

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

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

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

Fishback, P. (1995). An Alternative View of Violence in Labor Disputes in the Early 1900s: The Bituminous Coal Industry, 1890–1930. *Labor History*, 36(3):426–456.

Fishback, P. V. (1992). *Soft Coal, Hard Choices: The Economic Welfare of Bituminous Coal Miners, 1890-1930*. Oxford University Press.

Foged, M. and Peri, G. (2016). 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*. International Publishers, New York.

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

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

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

Friedman, G. (1998). *State-Making and Labor Movements: France and the United States, 1876-1914*. Cornell University Press.

Garen, J. (1984). The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica: Journal of the Econometric Society*, pages 1199–1218.

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

Gidron, N. and Hall, P. A. (2017). The Politics of Social Status: Economic and Cultural Roots of the Populist Right. *The British journal of sociology*, 68:S57–S84.

Giuliano, P. and Tabellini, M. (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*, pages 223–258. University of Chicago Press.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Gregory, J. N. (2015). IWW History Project. http://depts.washington.edu/iww/.

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

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

Haines, M., Fishback, P., and Rhode, P. (2018). United States Agriculture Data, 1840-2012 (ICPSR 35206).

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

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

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

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

Higham, J. (1955). 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.

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

Ignatiev, N. (1994). *How the Irish Became White*. Harvard University.

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

Jäger, S., Naidu, S., and Schoefer, B. (2024). Collective Bargaining, Unions, and the Wage Structure: An International Perspective. Technical report, National Bureau of Economic Research.

Jeffreys-Jones, R. (1978). Violence and Reform in American History.

Kaplan, E. and Naidu, S. (2024). Between Government and Market: The Political Economics of Labor Unions. Technical report, National Bureau of Economic Research.

Karadja, M. and Prawitz, E. (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 Margo, R. A. (2014). Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective. In *Human Capital in History: The American Record*, pages 15–57. University of Chicago Press.

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

Lee, D. S. and Mas, A. (2012). Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961–1999. *The Quarterly Journal of Economics*, 127(1):333–378.

Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2022). Valid t-Ratio Inference for IV. *American Economic Review*, 112(10):3260–3290.

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

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

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

Mackie, T. T. and Rose, R. (2016). *The International Almanac of Electoral History*. Springer.

Medici, C. (2024). Closing Ranks: Organized Labor and Immigration. *CESifo Working Paper*, No. 11437.

Mink, G. (1986). *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.

Naidu, S. (2022). Is There Any Future for a US Labor Movement? *Journal of Economic Perspectives*, 36(4):3–28.

Naidu, S. and Reich, A. (2018). Collective Action and Customer Service in Retail. *ILR Review*, 71(4):986–1001.

Naidu, S. and Yuchtman, N. (2016). Labor Market Institutions in the Gilded Age of American Economic History. Technical report, National Bureau of Economic Research.

Nohlen, D. and Stöver, P. (2010). *Elections in Europe*. Nomos Verlagsgesellschaft mbH & Co. KG.

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

OECD (2019). *Negotiating Our Way Up: Collective Bargaining in a Changing World of Work*. OECD Publishing, Paris.

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

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

Pauling, A., Luterbacher, J., Casty, C., and Wanner, H. (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.

Pezold, C., Jäger, S., and Nüss, P. (2023). Labor Market Tightness and Union Activity. Technical report, National Bureau of Economic Research.

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

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

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

Ruggles, S., Fitch, C., Goeken, R., Hacker, J., Nelson, M., Roberts, E., Schouweiler, M., and Sobek, M. (2022). IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. *Minneapolis, MN: IPUMS*.

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

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

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

Solomou, S. and Wu, W. (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.

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.

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.

Taft, P. and Ross, P. (1969). American Labor Violence: Its Causes, Character, and Outcome. *Violence in America: Historical and Comparative Perspectives*, 1:221–301.

U.S. Department of Labor (2023). AFL-CIO's Form LM-2 Labor Organization Annual Report.

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

Webb, S. and Webb, B. (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*, pages 372–498. NBER.

Witwer, D. (2002). Race Relations in the Early Teamsters Union. *Labor History*, 43(4):505–532.

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

Wooldridge, J. M. (2015). Control Function Methods in Applied Econometrics. *Journal of Human Resources*, 50(2):420–445.

ONLINE APPENDIX

A Additional Tables and Figures

Table A.1: European Countries and Regions for Shift-Share Instrument

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 4.2. The stocks of foreign-born individuals by country are from the county-level data of the 1890 Census of Population (Haines, 2010).

Table A.2: Correlation Between Immigration and County Characteristics in 1890

	Dependent variable: Share of European Immigrant Population
Share of Population in Manufacturing	0.692***
	(0.118)
Share of Population in Agriculture	-0.118***
	(0.024)
Presence of Active Coal Mines	0.012**
	(0.006)
Observations	876

Notes: The observations are at the county level for the year 1890. The dependent variable is the number of European immigrants as a fraction of the total population. The independent variables are the share of population employed in manufacturing, the share of population employed in agriculture, and indicator for the presence of active coal mines. Robust standard errors are shown in parentheses. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Table A.3: Unionization and Immigration Flows

	Dependent variable: Share of Immigrants (x 100)			
	(1)	(2)	(3)	(4)
Any Union Present (t-10)	-0.660***			
	(0.252)			
Number of Union Branches (t-10)		-0.372**		
		(0.161)		
Union Density (t-10)			-4.276*	
			(2.418)	
Union Members per Branch (t-10)				-0.004**
				(0.002)
Observations	2,626	2,626	2,626	2,626
Dep. var. mean	2.788	2.788	2.788	2.788
Indep. var. mean	0.441	2.564	0.036	48.142

Notes: The observations are at the county-year level. The dependent variable is the number of European immigrants (men ages 16–64) who entered the United States in the previous decade, as a fraction of the male working-age population in the county (multiplied by 100). The regressors of interest are the ten-year lag of: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the independent variable in column (2) reflects the average number of union branches, not the transformed value. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.4: The Effect of Immigration on Organized Labor – Control Function

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants	2.626**	5.245***	0.282***	481.007***
	(1.053)	(1.820)	(0.104)	(147.736)
First-Stage Residuals of Share of Immigrants	-2.241**	-3.544*	-0.216*	-347.835**
	(1.128)	(2.032)	(0.110)	(163.125)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.441	2.563	0.036	48.158
Indep. var. mean	0.028	0.028	0.028	0.028

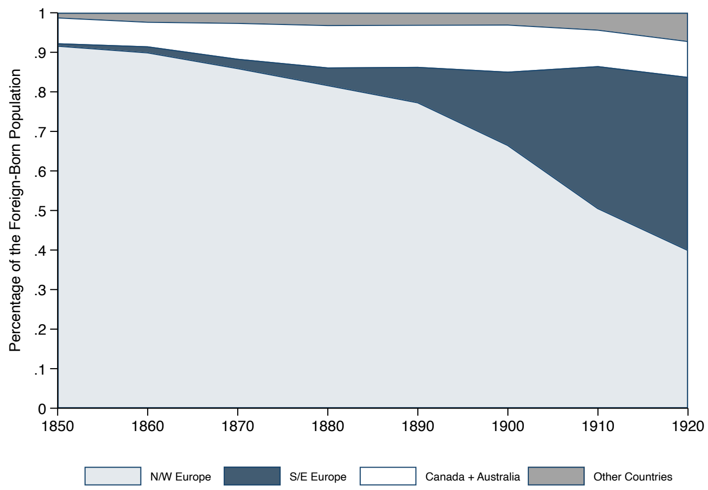
Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. 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 Immigrant Labor Competition

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Share of Immigrants	3.297***	5.666***	0.306***	508.250***
	(1.234)	(1.807)	(0.116)	(165.461)
Share of Immigrants × Skilled Comp.	0.677	3.356**	0.155**	301.723**
	(0.526)	(1.308)	(0.074)	(120.398)
Share of Immigrants × Unskilled Comp.	-0.540**	-0.879*	-0.044	-74.552*
	(0.274)	(0.495)	(0.039)	(43.503)
Observations	2,624	2,624	2,624	2,624
KP F-statistic	32.66	32.66	32.66	32.66
SW F-statistic (Share of Immigrants)	84.62	84.62	84.62	84.62
SW F-statistic (Sh. of Imm. × Sk. Comp.)	42.81	42.81	42.81	42.81
SW F-statistic (Sh. of Imm. × Unsk. Comp.)	50.93	50.93	50.93	50.93

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. *Skilled (Unskilled) Competition* is a (standardized) measure of immigrant labor competition, based on the prevailing skilled (unskilled) occupations among the U.S.-born workers in the county at the beginning of each decade and among the immigrants entering all other U.S. counties during that decade, as detailed in Section 6. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic of the instruments in the three separate first-stage regressions. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Figure A.1: Origin 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., 2022) and ICSPR (Haines, 2010).

Table A.6: Heterogeneous Effects by Terciles of Attitudes Towards Immigration

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
	Panel A: V = Vote share Know-Nothing party			
Share of Immigrants x Low V	1.535	-0.297	0.023	403.200
	(1.610)	(2.460)	(0.185)	(313.316)
Share of Immigrants x Medium V	2.385	4.257	0.174	651.682**
	(1.775)	(3.631)	(0.126)	(285.172)
Share of Immigrants x High V	3.073	6.171*	0.283*	473.972**
	(2.043)	(3.175)	(0.146)	(209.739)
Observations	2,103	2,103	2,103	2,103
KP F-statistic	14.63	14.63	14.63	14.63
SW F-statistic (Low V)	37.92	37.92	37.92	37.92
SW F-statistic (Medium V)	40.74	40.74	40.74	40.74
SW F-statistic (High V)	42.35	42.35	42.35	42.35
	Panel B: V = Index of residential segregation			
Share of Immigrants x Low V	0.289	1.533	-0.043	-57.646
	(1.322)	(2.148)	(0.160)	(188.220)
Share of Immigrants x Medium V	3.208**	9.169***	0.495***	662.753***
	(1.275)	(2.613)	(0.129)	(199.877)
Share of Immigrants x High V	2.897**	2.879	0.216**	502.944***
	(1.416)	(1.873)	(0.104)	(169.165)
Observations	2,565	2,565	2,565	2,565
KP F-statistic	21.09	21.09	21.09	21.09
SW F-statistic (Low V)	48.72	48.72	48.72	48.72
SW F-statistic (Medium V)	62.85	62.85	62.85	62.85
SW F-statistic (High V)	48.84	48.84	48.84	48.84

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The regressor of interest is the number of European immigrants (men 16–64) who entered the United States 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 4.2. In Panel A, *Share of Immigrants* is interacted with indicators for whether the county has a low (first tercile), medium (second tercile), or high (third tercile) historical vote share for the Know-Nothing party (see Section 6 for more details). In Panel B, Share of Immigrants is interacted with indicators for whether the county has low (first tercile), medium (second tercile), or high (third tercile) residential segregation at baseline (see Section 6 and Appendix D for more details). All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic of the instruments in the three 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 A.7: Effect on the Composition of Union Leaders

	Dependent variable: Share of Union 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 of Immigrants	2.382**	0.601	0.046	0.131	-0.142	0.068
	(1.064)	(0.398)	(0.134)	(0.277)	(0.236)	(0.177)
Observations	2,628	2,628	2,628	579	579	579
Dep. var. mean	0.340	0.031	0.009	0.870	0.089	0.024
KP F-statistic	67.34	67.34	67.34	13.98	13.98	13.98
Panel B: Ancestry						
Share of Immigrants		2.678**	0.476		0.006	0.144
		(1.159)	(0.362)		(0.398)	(0.389)
Dep. var. mean		2,625	2,625		579	579
Observations		0.342	0.038		0.881	0.101
KP F-statistic		67.27	67.27		13.98	13.98

Notes: The observations are at the county-year 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 and the ancestry from last names is described in Section C. The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. In columns 1 to 3, the sample includes all counties as in Table 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, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic 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.8: Heterogeneous Effects by Strength of Labor Movement in Country of Origin

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants from UK-Ireland	-4.457	-13.908	-0.872	-2,222.999
	(12.793)	(23.535)	(1.387)	(2,057.525)
Standardized coefficient	[-0.046]	[-0.061]	[-0.059]	[-0.153]
Share of Immigrants from Other Countries	2.861*	5.881**	0.321**	570.784***
	(1.507)	(2.600)	(0.139)	(219.008)
Standardized coefficient	[0.231]	[0.202]	[0.169]	[0.308]
Observations	2,628	2,628	2,628	2,628
KP F-statistic	11.89	11.89	11.89	11.89
SW F-statistic (UK-Ireland)	26.27	26.27	26.27	26.27
SW F-statistic (Other Countries)	44.33	44.33	44.33	44.33

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The regressors of interest are the number of immigrants (men ages 16–64) from European countries with a strong (UK-Ireland) and weak (other countries) labor movements as of 1870 (as detailed in Section 6.3 and Appendix E) who entered the United States 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 4.2 and Section 6.3. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.9: Heterogeneous Effects by Support for Socialist Parties in Country of Origin

	Dependent variable:			
	Any Union Present (1)	Number of Union Branches (2)	Union Density (Members / LF) (3)	Union Members per Branch (4)
Panel A: High Vote Share = Above 20% during 1890-1919				
Share of Immigrants from Countries w/ High Socialist Vote Share	1.013	-1.213	0.045	-194.233
	(2.559)	(4.339)	(0.271)	(392.421)
Standardized coefficient	[0.044]	[-0.023]	[0.013]	[-0.057]
Share of Immigrants from Countries w/ Low Socialist Vote Share	3.582	9.078**	0.424**	881.742***
	(2.675)	(4.117)	(0.213)	(326.762)
Standardized coefficient	[0.190]	[0.204]	[0.147]	[0.312]
KP F-statistic	14.42	14.42	14.42	14.42
SW F-statistic (High Socialist Vote Share)	27.76	27.76	27.76	27.76
SW F-statistic (Low Socialist Vote Share)	38.00	38.00	38.00	38.00
Panel B: High Vote Share = Above 10% during 1890-1919				
Share of Immigrants from Countries w/ High Socialist Vote Share	-1.825	-5.299	-0.156	-298.928
	(2.020)	(3.571)	(0.163)	(288.394)
Standardized coefficient	[-0.131]	[-0.161]	[-0.073]	[-0.143]
Share of Immigrants from Countries w/ Low Socialist Vote Share	25.893*	60.369***	2.577***	4,558.319***
	(14.280)	(21.532)	(0.842)	(1,544.896)
Standardized coefficient	[0.639]	[0.632]	[0.416]	[0.751]
KP F-statistic	3.93	3.93	3.93	3.93
SW F-statistic (High Socialist Vote Share)	12.11	12.11	12.11	12.11
SW F-statistic (Low Socialist Vote Share)	7.98	7.98	7.98	7.98
Observations	2,628	2,628	2,628	2,628

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The regressors of interest are the number of immigrants (men ages 16–64) from European countries with high or low support for socialist parties who entered the United States 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 4.2 and Section 6.3. In Panel A (Panel B) high support is defined as an average vote share for socialist parties between 1890 and 1919 above 20% (10%) (as detailed in Section 6.3 and Appendix F). All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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.10: Heterogeneous Effects by the Number of Other Unions' Branches

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Heterogeneity by Knights of Labor (KoL) Branches				
Share of Immigrants [1]	2.712**	5.141**	0.266**	474.221***
	(1.233)	(2.026)	(0.104)	(159.633)
Share of Immigrants × KoL Branches [2]	-0.329	0.481	0.064	24.687
	(0.381)	(1.086)	(0.058)	(80.178)
Observations	2,616	2,616	2,616	2,616
KP F-statistic	33.21	33.21	33.21	33.21
SW F-statistic [1]	65.84	65.84	65.84	65.84
SW F-statistic [2]	68.78	68.78	68.78	68.78
Panel B: Heterogeneity by Industrial Workers of the World (IWW) Branches				
Share of Immigrants [1]	2.978**	4.314**	0.290**	501.300***
	(1.258)	(1.921)	(0.114)	(169.781)
Share of Immigrants × IWW Branches [2]	-0.403*	1.063	-0.009	-23.157
	(0.208)	(0.917)	(0.031)	(44.224)
Observations	2,628	2,628	2,628	2,628
KP F-statistic	32.27	32.27	32.27	32.27
SW F-statistic [1]	64.05	64.05	64.05	64.05
SW F-statistic [2]	65.37	65.37	65.37	65.37

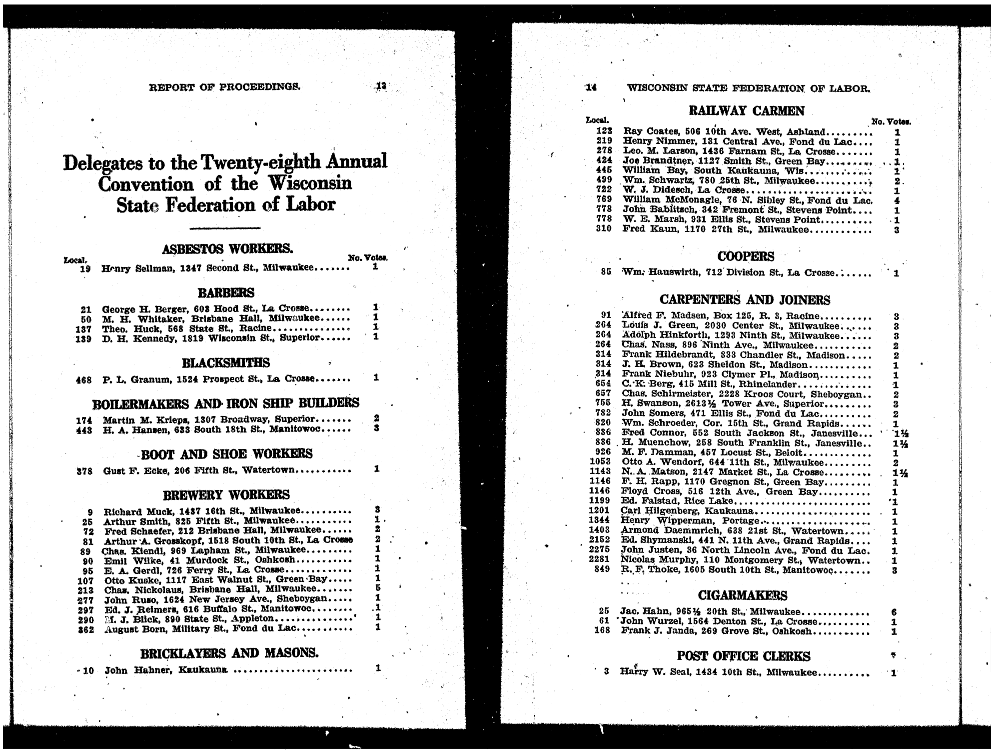
Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the (inverse hyperbolic sine of the) number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor force (column 3); and the number of members per branch, or zero if the county has no union branch (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the transformed value. The main regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. *Kol Branches* (Panel A) and *IWW Branches* (Panel B) is the (inverse hyperbolic sine of the) number of branches of the Knights of Labor in 1880 and of the Industrial Workers of the World between 1906 and 1917, respectively. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. SW F-statistic refers to the Sanderson-Windmeijer F-statistic 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 A.11: Effect on Local Economic Outcomes

	Dependent variable:			
	Labor Force Part. Rate	Mfg. Output (per Worker)	Mfg. Output (Share of U.S. Total)	Share of LF in Skilled Occupations
	(1)	(2)	(3)	(4)
Share of Immigrants	-0.003	-0.866	-0.002	0.104
	(0.098)	(1.094)	(0.015)	(0.110)
Standardized coefficient	[-0.003]	[-0.072]	[-0.031]	[0.048]
Observations	2,628	2,595	2,595	2,628
Outcome mean	0.909	3163.887	0.001	0.201
Imm. Share mean	0.028	0.028	0.028	0.028
KP F-statistic	67.34	66.68	66.68	67.34

Notes: The observations are at the county-year level. The dependent variables are: the male labor force participation rate (column 1); the log of the manufacturing output divided by the manufacturing labor force (column 2); the manufacturing output as a share of the total manufacturing output in the United States in that year (column 3); and the log of the total male labor force in skilled occupations (column 4). The value of manufacturing output for the year 1910, which would otherwise be missing, is linearly interpolated. The dependent variables in columns (2) and (3) are expressed in 2020 USD. The main regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States 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 4.2. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of the population working in manufacturing and in agriculture, and an indicator for the presence of active coal mines. KP F-statistic refers to the Kleibergen-Paap F-statistic 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.

Figure A.2: 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, 683 South 18th St., Manitowoc	3

BOOT AND SHOE WORKERS

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

BREWERY WORKERS

9	Richard Muck, 1487 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
862	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

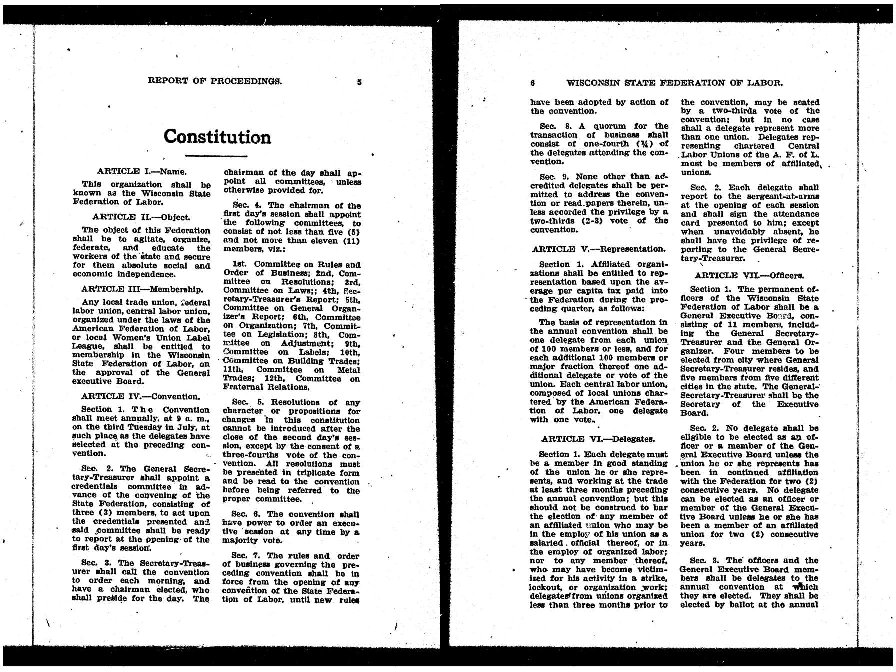
25	Jac. Hahn, 965½ 20th St., Milwaukee	6
61	John Wurzel, 1564 Denton St., La Crosse	1
168	Frank J. Janda, 269 Grove St., Oshkosh	1

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. These documents provide details on the number of branches represented at the conventions, as well as information about their delegates.

Figure A.3: 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.

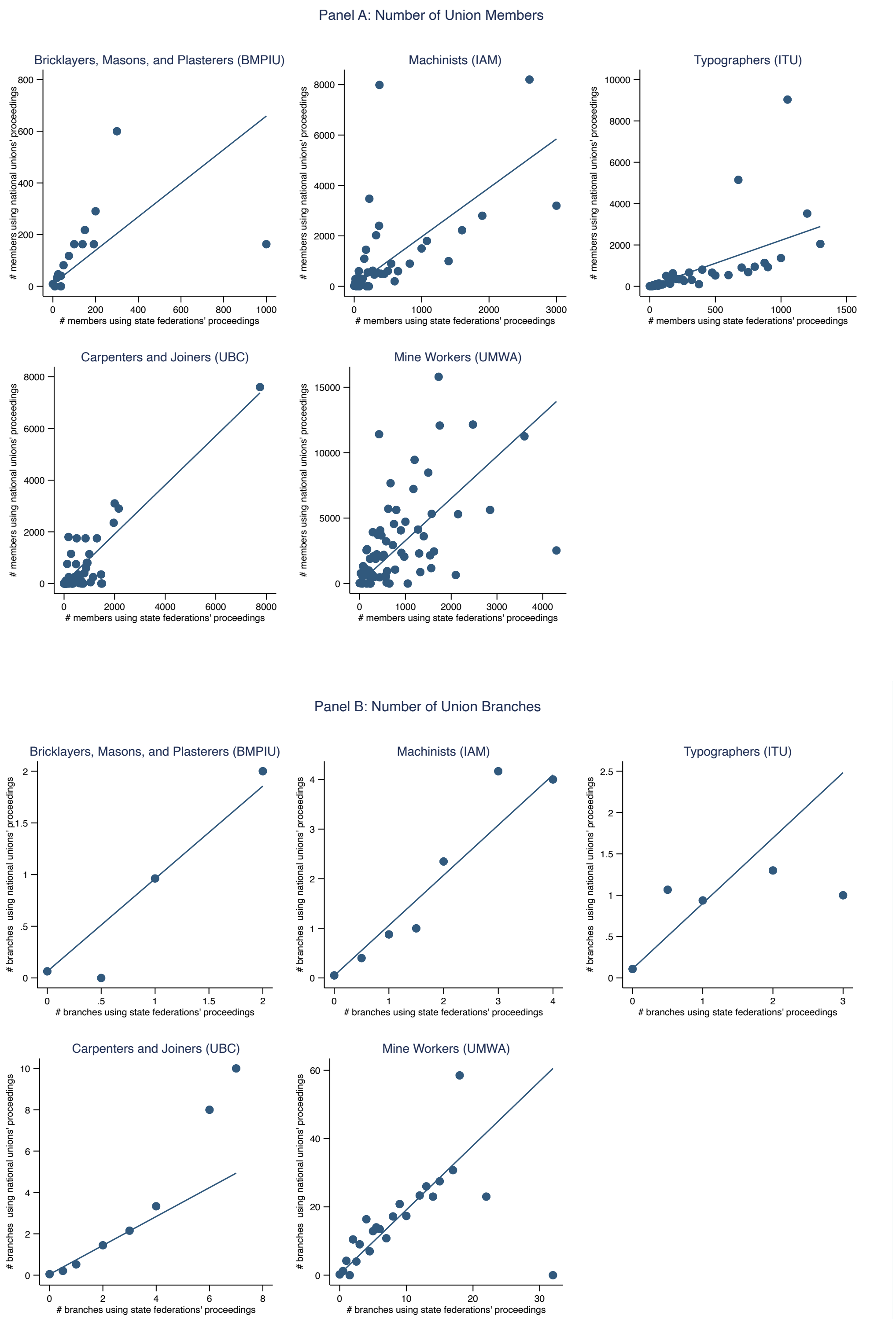
Section 1. The permanent officers of the Wisconsin State Federation of Labor shall be a General Executive Board, consisting of 11 members, including the General Secretary-Treasurer and the General Organizer. Four members to be elected from city where General Secretary-Treasurer resides, and five members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

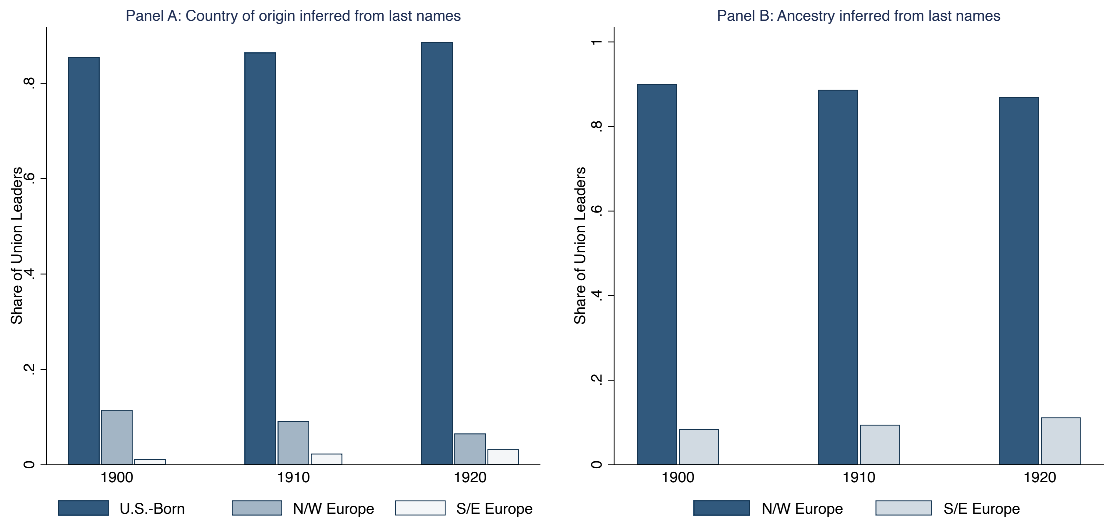
Notes: The figure shows a digitized document from the constitutions of the state federations of labor. These documents outline the rules governing the number of delegates that local branches were entitled to send to conventions. The section titled "Article V — Representation" on the right page provides an example.

Figure A.4: 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 five national unions that are observed in both sources. See Section 3 for more details.

Figure A.5: 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 branches to the state federation of labor convention or to the national convention of their union. The country of origin and the ancestry are inferred from delegates' last names, as described in Appendix C.

B Robustness Checks

B.1 Alternative Shift-Share Instrument

As explained in Section 4.2, I replicate the analysis using an alternative instrument that relies on *predicted* flows of European immigration. More specifically, in Equation (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. (2020),[74] and estimate a relationship between weather shocks and immigration from each European country (for the period 1900–1920) using the following equation:

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

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. (2020), 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 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.[75] For this exercise, the sample includes nineteen European countries for which immigration, weather, and crop data are available.[76] In the baseline specification, I consider temperature shocks, but the results are similar if using precipitations.

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

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

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

I separately estimate Equation (B.1) for each European country in the sample. Figure B.1 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, $log(\widehat{Immigr_{j,t}})$ using the $\widehat{\beta_{j,s,k}}$'s estimated from these regressions. Finally, I aggregate the predicted flows by decade and obtain:

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

Table B.1 reports the first stage estimates. Although the F-stat is lower than the one of the main instrument (Table 2), it is still always above the conventional levels. The coefficients and confidence intervals of the 2SLS estimates using this alternative instrument are presented in Figure 3, row 1. All coefficients are positive and statistically significant.

B.2 Matching-Style Exercise

Similar to Bazzi et al. (2023), I combine the main empirical strategy based on a shift-share instrumental variable with a matching-style exercise.[77] Specifically, I identify county pairs within the same state that had the closest number of union branches as a fraction of the population in 1880.[78] Due to the lack of comprehensive data on unions affiliated with the American Federation of Labor (AFL) before 1890—given that the AFL was established only in 1886—I use data on the Knights of Labor (Garlock, 2009), the largest federation of labor unions in the United States during the 1880s (Friedman, 1998).

I present the results in Figure 3, row 2. In particular, I re-estimate Equation (1), replacing the baseline controls with fixed effects for the county pairs, interacted with year dummies. The resulting coefficients identify the effect of immigration inflows on unionization for counties with nearly identical levels of union presence at baseline.[79] Despite the very demanding nature of this specification, reassuringly all the point estimates remain positive, in some cases are precisely estimated, and are similar to the baseline coefficients from Table 3 (displayed in orange).

[77] This exercise does not involve a nonparametric matching estimator (such as nearest-neighbor or kernel matching). Rather, it simply restricts comparisons to pairs of counties with similar baseline unionization. Identification continues to come entirely from the shift-share instrument.

[78] The results are nearly identical when using the corresponding variable measured in 1890.

[79] In case of equal values of the matching variable, I further match counties on these additional variables, in the following order: total number of union branches in the county, the share of population in manufacturing, and the share of population in agriculture. This is meant to compare counties that have similar economic conditions 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.

B.3 Additional Robustness Checks

Alternative baseline specification. To ensure the robustness of the results to the choice of controls in the preferred specification, I estimate alternative specifications. Specifically, row 3 of Figure 3 presents coefficients from regressions that include only county and year fixed effects, excluding other baseline controls, while row 4 adds state-year fixed effects to the preferred specification. The estimates remain largely unchanged and statistically significant, except for union density in the specification without baseline controls, which is marginally insignificant (p-value = 0.116).

Alternative standard errors. All the results in the paper report standard errors clustered at the county level. In Figure 3, I report standard errors using five alternative procedures: the adjustement proposed by Adao et al. (2019) for shift-share instrumental variables (row 5), the approach of Conley (1999) to account for spatial correlation using a 200km- or 500km-bandwidth (rows 6 and 7), clustering by State Economic Area, i.e., the historical equivalent of modern Commuting Zones (row 8), and the tF standard errors introduced by Lee et al. (2022) (row 9). All coefficients remain statistically significant.

Drop potential outliers. I verify that the results are robust to omitting observations with very large and very low levels of the dependent and independent variables, which could be potential outliers. I re-estimate the baseline results dropping observations with measures of unionization (Figure 3, row 10) and immigration (Figure 3, row 11) below the 1^{st} and above the 99^{th} percentile. Reassuringly, in all cases the coefficients are in line with those at the top of each panel.

Alternative definitions of *Share of Immigrants*. As described in Section 4, the definition of *Share of Immigrants* used in the paper is the number of male working-age (16-64 years old) European immigrants who entered the United States in the previous 10 years as a fraction of the male working-age population. In Figure 3, I show that the results are robust when using alternative definitions of the main independent variable. In row 12, *Share of Immigrants* is defined as the total number of the male working-age European immigrant population, regardless of when they arrived to the United States, as a fraction of the male working age population. In row 13, it is the number of all (men and women) recently entered (within the previous 10 years) working-age European immigrants as a fraction of the total working-age population. In row 14, it is the number of all working-age European immigrants as a fraction of the total working-age population. In row 15, the independent variable is the total number of European immigrants as a fraction of the total population. Across these four rows, the coefficients are positive and statistically significant for all four unionization measures.

Alternative samples. To ensure that the results are not specific to the balanced panel of urban and mining counties (i.e., counties with a positive urban population or at least one coal mine in 1890) used throughout the paper, I re-estimate the preferred specification of Table 3 using alternative samples. The estimates are shown in Figure 3, rows 16–19. In row 16, the sample is an unbalanced panel of only urban or mining counties. In row 17, the sample is a balanced panel of both urban and rural counties. In row 18, the sample is composed of all county-year observations for which the unionization data are available. Finally, in row 19, the counties in the South are excluded from the preferred estimation sample of Table 3. This exercise is motivated by the fact that this region of the United States received low levels of immigration between 1890 and 1920, and also experienced limited organized labor activity. Hence, a possible concern is that Southern counties may be driving the positive relationship between immigration and unionization. Across all four rows, the coefficients are positive and statistically significant.

Analysis at the SEA level. Additionally, I re-estimate the main results from Table 3 using data aggregated at the State Economic Area (SEA) level—groups of contiguous counties within the same state that share similar economic characteristics (Bogue, 1951)—which can be considered analogous to modern labor markets or Commuting Zones (Abramitzky et al., 2023). The coefficients are reported in Figure 3, row 20. Despite the lower number of observations and consequently a lower F-statistic (F-stat = 25.75), the effects of immigration on all four measures of unionization remain positive and statistically significant.

Alternative data construction. In Section 3.1, I described the steps followed to construct the novel dataset on county-level unionization used in this paper. To ensure that the results are not driven by any interpolated value of unionization, I conduct a robustness exercise where I omit such observations (Figure 3, row 21). Additionally, I re-estimate the analysis only relying on the convention proceedings of the state federations of labor, without combining them with the proceedings from the five largest AFL-affiliated national unions (Figure 3, row 22). In both rows, the estimates are very similar to those at the top of each panel and always statistically significant.

Exclude the 1910s. I verify that the results are not driven by the social, economic, and political transformations of the late 1910s, including World War I and the First Red Scare. As discussed in Section 6.3, if these events spurred union growth and the areas most affected also experienced larger immigrant inflows during this period, the main estimates could be confounded. To address this concern, I re-estimate the preferred specification from Table 3, excluding the years 1911–1920 (Figure 3, row 23). The coefficients thus estimate the effect of immigration during 1891–1900 and 1901–1910 on unionization in 1900 and 1910, respectively. The results remain similar in magnitude to the main estimates and statistically significant.

Test of pre-trends. The validity of the shift-share instrument defined by Equation (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, 2025 and Goldsmith-Pinkham et al., 2020). Although the results of Figure 4 already reduce the concerns about this assumption being invalidated, in Table B.2, I test for pre-trends more directly, regressing the pre-period change (between 1880 and 1890) of measures of unionization, population, and economic growth on the subsequent average share of immigrants (in 1900, 1910, and 1920), as predicted by the instrument. All regressions control for the same characteristics as in the rest of the paper, as well as the 1880 share of European immigrant population, to address the concern that correlations between 1890 immigration shares from certain countries of origin and pre-period outcomes may rise because the pre-period outcomes themselves reflect earlier immigration (Borusyak et al., 2025). The choice of the dependent variables is constrained by data availability. Due to the lack of comprehensive data on unions affiliated with the American Federation of Labor (AFL) before 1890—given that the AFL was established only in 1886—I use data on the Knights of Labor (Garlock, 2009), the largest federation of labor unions in the United States during the 1880s (Friedman, 1998). I measure unionization with an indicator for the presence of any union branch (column 1) or the number of union branches (column 2); for population and economic measures, I examine the share of urban population (column 3) and two measures from the Census of Manufacturing: the value of manufacturing output per worker (column 4), and the number of manufacturing establishments per worker (column 5).[80] Reassuringly, no coefficient of Table B.2 is statistically significant. These results indicate that the instrument does not predict more immigration to counties that, before 1890, were already undergoing changes in union presence (columns 1 and 2) or in some other important characteristic (columns 3 to 5).

B.4 Controlling for Additional Baseline Characteristics

This section addresses the possibility that the instrument described in Section 4.2 may predict a higher immigrant share in counties that were already on a trajectory of higher union growth, for either economic or political reasons. Figure 4 displays the 2SLS coefficients and confidence intervals of the effects of immigration on unionization, where the preferred specification of Equation (1) is augmented by interacting several characteristics measured at baseline with year dummies. This exercise is meant to reduce the concern that factors jointly correlated with the 1890 size of immigration and the development of labor unions between 1900 and 1920 may bias the estimates. The coefficients at the top of each panel are repeated from Table 3 for comparison, while the numbered rows correspond to a different set of controls added to the preferred specification.

Share of urban population and population density. Given that immigration and

[80]The dependent variables in columns 4 and 5 are transformed using the inverse hyperbolic sine transformation.

unionization were both primarily concentrated in cities, a potential concern is that higher initial levels of urbanization may have attracted more immigrants before 1890 and contributed to greater union growth over time. To address this possible source of omitted variable bias, I include controls for the share of the population in urban areas or population density in 1890, each interacted with year dummies. The coefficients remain positive and statistically significant (rows 1 and 2).

Share of immigrant population. I also directly control for the share of immigrant population in 1890 (or in 1880), 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 (or 1880) 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 (rows 3 and 4).

Share of Black 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 share of the Black population in each county in 1890, which is associated with 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 (row 5).

Labor force composition. I further control for the shares of the labor force in 1890 in (i) the industries with the largest AFL-affiliated labor unions during the period 1900–1920: mining, construction, and transportation (Stewart, 1926); (ii) and by skill level: low-skilled, mid-skilled, and high-skilled (Katz and Margo, 2014). These regressions therefore estimate the effect of immigration among counties with similar initial size of the labor force in sectors and skills conducive to unionization. The results are all positive and statistically significant, and larger in magnitude (rows 6 and 7).

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 (rows 8 and 9).

Presence of a railroad. 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., 2020). 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. The results are almost unchanged (row 10).

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 (row 11).

Table B.1: First Stage of the Alternative Instrumental Variable Estimation

	Dependent variable: Share of Immigrants			
	(1)	(2)	(3)	(4)
Predicted Share of Immigrants	0.211***	0.183***	0.166***	0.166***
	(0.042)	(0.039)	(0.035)	(0.034)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.028	0.028	0.028	0.028
Indep. var. mean	0.094	0.094	0.094	0.094
KP F-statistic	25.30	22.09	23.12	23.60
Controls:				
Share of Population in Manufacturing	No	Yes	Yes	Yes
Share of Population in Agriculture	No	No	Yes	Yes
Presence of Coal Mines	No	No	No	Yes

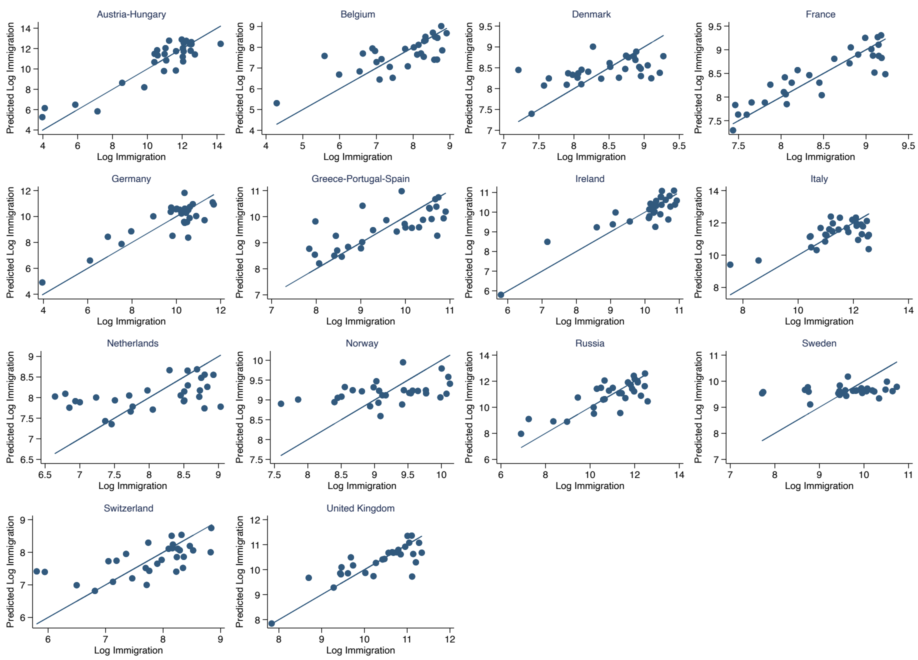
Notes: The observations are at the county-year level. The table reports the first stage of the alternative instrument described in Appendix B.1. The dependent variable is the number of European immigrants (men ages 16–64) who entered the United States in the previous decade, as a fraction of the male working-age population in the county. The main regressor of interest is the predicted number of European immigrants (men ages 16–64) who entered the United States 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, measured in 1890 and interacted with year dummies, are also included: the share of the population working in manufacturing (from column 2); the share of the population working in agriculture (from column 3), and an indicator for the presence of active coal mines (column 4). KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.2: Test of Pre-Trends

	Dependent variable (1880-1890 difference):				
	Any Union Present	Number of Union Branches	Share of Urban Pop.	Mfg. Output per Worker	Mfg. Establish. per Worker
	(1)	(2)	(3)	(4)	(5)
Share of Immigrants	-0.692	-4.101	0.198	-0.543	-0.182
(average 1900–1920)	(1.378)	(2.520)	(0.462)	(1.723)	(5.097)
KP F-statistic	68.17	68.17	68.17	68.17	68.17
Observations	871	871	871	871	871

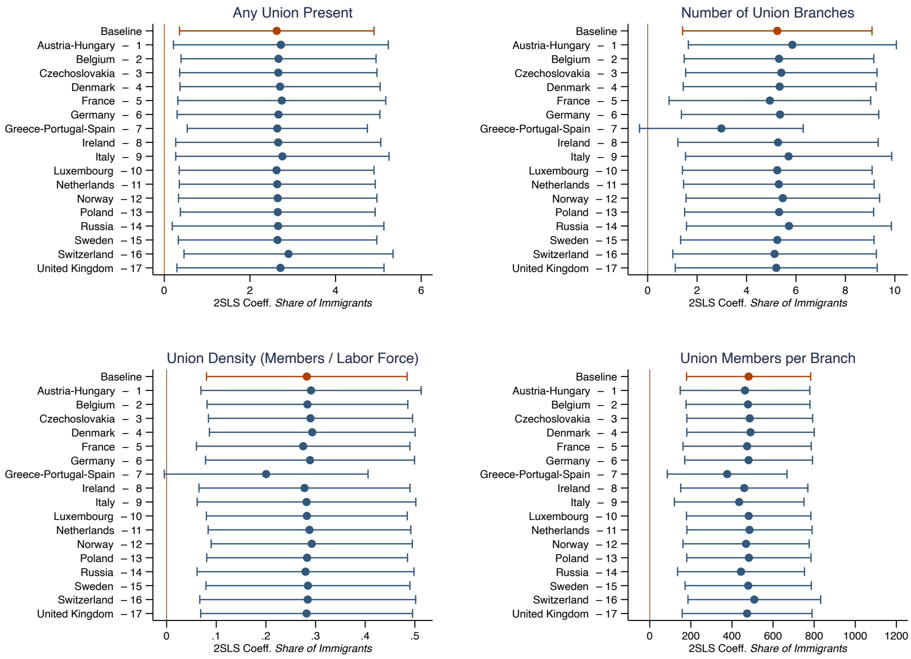
Notes: The observations are at the county level. The dependent variables are the 1890–1880 difference in: an indicator for the presence of any union branch (of the Knights of Labor, column 1); the (inverse hyperbolic sine of the) number of union branches (of the Knights of Labor, column 2); the share of urban population (column 3); the (inverse hyperbolic sine of the) value of manufacturing output divided by the number of manufacturing workers (column 4); and the (inverse hyperbolic sine of the) number of manufacturing establishments divided by the number of manufacturing workers (column 5). The regressor of interest is the number of European immigrants (men ages 16–64) who entered the United States in the previous decade, as a fraction of the male working-age population in the county, averaged over the period 1900–1920, and predicted by the instrument described in Section 4.2, averaged over the period 1900–1920. All regressions include the following controls: the 1890 share of the population working in manufacturing and agriculture, an indicator for the presence of active coal mines in 1890, and the 1880 share of European immigrant population. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Robust standard errors are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Figure B.1: 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 (B.1), separately for each country (or group of countries) used to construct the shift-share instrumental variable.

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



Notes: The figures plot the 2SLS coefficients (with corresponding 95% confidence intervals) of *Share of Immigrants* (Imm_{ct}), augmenting the preferred specification reported in Table 3 with the 1890 immigrant share from each sending country (relative to all immigrants from that country in the United States 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.

C Delegates' Origin and Ancestry Using Last Names

In Section 6, I rely on union delegates' last names to infer their origins and ancestry. This section describes the procedure used.

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 United States 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 last name in that year—is 82% U.S.-born, 5% British, and 5% German (33% Germany ancestry, 31% British ancestry, and 22% Irish ancestry); Anderson—the eighth most last common name—is 46% U.S.-born, 32% Swedish, and 9% Norwegian (60% Swedish ancestry, 18% Norwegian ancestry, and 9% Danish ancestry); and Murphy is 47% Irish, 45% U.S.-born, and 2% British (94% Irish ancestry and 5% British ancestry).

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.

D Index of Residential Segregation

In Section 6, I explore the heterogeneity of the effects of European immigration on unionization with respect to the level of residential segregation in the county in 1880. In this section, I briefly describe how the measure is constructed.[81]

First, I identify next-door neighbors from full-count U.S. Census data as household heads with consecutive household serial numbers within the same enumeration district.[82] 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).[83] 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, η_c, is computed as:

$$\eta_c = \frac{E(\overline{x_c}) - x_c}{E(\overline{x_c}) - E(\underline{x_c})}. \tag{D.1}$$

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

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

[82] Restricting the definition of neighbors to household heads within the same Census page yields similar results (available upon request).

[83] 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.

E Labor Unions in Europe

The data on the presence and membership of labor unions in European countries used in Section 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 United States, organization was mostly prevalent among skilled craftsmen and mining workers. At the turn of the 20th century, the only countries with an active and strong labor movement where the U.K. and Ireland. By 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 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 logic 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 to the United States, and therefore might have been particularly interested in forming or joining labor unions in their new country. Table E.1 reports union membership at the national level for the years 1870 and 1900.

Table E.1: Union Membership Across European Countries, 1870 and 1900

	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).

F Support for Socialist Parties in Europe

In Section 6.3, I separately predict (and estimate the impact of) immigration from Europoean countries with high and low support for socialism, using data from Austrian National Library (2024), Mackie and Rose (2016), and Nohlen and Stöver (2010). The logic behind this exercise reflects the fact that individuals emigrating from countries with stronger socialist parties may have been exposed to the ideas of socialism by the time they arrived to the United States, and therefore might have been particularly interested in continuing that experience in their new country by forming or joining labor unions.

Table F.1 reports the vote shares for socialist parties in national elections held in European countries between 1890 and 1919. In Table A.9, I classify European countries as having high support for socialism if they have an average vote share for socialist parties between 1890 and 1919 above 20% (Panel A) or above 10% (Panel B). Both definitions include the following countries as showing high support for socialism: Austria-Hungary, Denmark, Finland, and Germany; the latter classification also includes: Belgium, France, Italy, Luxembourg, Norway, Sweden, and Switzerland. In addition, the countries part of the Russian Empire are classified as having high support for socialist parties according to either definition.[84]

[84]The results (not shown for brevity, but available upon request), are very similar if such countries are classified as having low support for socialist parties instead.

Table F.1: Socialist Parties Vote Shares in National Elections Across European Countries, 1890–1919

	% Votes for Socialist Parties															
Year	Austria	Belgium	Bulgaria	Denmark	Finland	France	Germany	Iceland	Italy	Luxembourg	Netherlands	Norway	Poland	Sweden	Switzerland	U.K.
1890		0.30					19.70									
1891											1.00					
1892		0.20														
1893						8.40	23.30									
1894		13.20									0.20	0.30				
1895									6.80							1.00
1896		8.50													6.80	
1897	23.13								8.90		3.00	0.60				
1898		21.10				11.30	21.60									
1899			3.50												9.60	
1900	23.39	22.50							13.00			3.00				1.30
1901			5.00	17.10							9.50					
1902		15.00	4.80			10.40								3.50	12.60	
1903				20.40			31.70					9.70				
1904		26.00							21.30							
1905											11.20			9.50	14.70	
1906		6.20		25.40		10.00						16.00				4.80
1907	21.00				37.00		29.00						1.40			
1908		22.60	1.60		38.40									14.60	17.60	
1909				28.70	39.90				19.00		13.90	21.60				
1910		6.70		28.30	40.00	13.10										7.00
1911	25.40		5.00		40.00								1.40	28.50	20.00	
1912		9.30					34.80					26.30				
1913			20.40	29.50	43.11				17.60		18.50					
1914		30.30	11.60			16.80								30.10	10.10	
1915												32.10				
1916					47.30			6.80								
1917					44.80									31.10	30.80	
1918				28.70							22.00	31.60				22.50
1919	40.42	36.60	31.60		38.00	21.20	37.90	6.80	32.30	15.60					23.40	

Notes: This table presents the vote shares obtained by socialist parties in national elections held in European countries between 1890 and 1919. The data are from Austrian National Library (2024), Mackie and Rose (2016), and Nohlen and Stöver (2010). The following parties are included in the count of the socialist vote shares. Austria: Social Democrats (SDP); Belgium: Workers' Party (BWP/POB); Bulgaria: Social Democratic Workers Party; Denmark: Social Democrats; Finland: Social Demoratic Party (SDP); France: Socialist Party (SFIO); Germany: Social Democratic Party (SPD); Iceland: Social Democratic Party; Italy: Socialist Party (PSI); Luxembourg: Social Democratic Party (POSL/LSAP); Netherlands: Social Democratic League (SDB) and Social Democratic Workers' Party (SDAP); Norway: Labor Party (DNA); Poland: Social Democrats; Sweden: Social Democratic Workers' Party (SAP); Switzerland: Social Democrats (PS/SP); United Kingdom: Independent Labour Party (ILP) and Labour Party.

G Diagnostics of the Shift-Share Instruments

In this section, I present the diagnostics on shift-share instruments described by Goldsmith-Pinkham et al. (2020), in addition to the ones highlighted and reported in Section 4.2, Section 5.3, and Appendix B.[85]

Goldsmith-Pinkham et al. (2020) show that shift-share estimators like the one used in this paper are numerically equivalent to a generalized method of moments estimator written as $\widehat{\beta} = \sum_j \widehat{\alpha}_j \widehat{\beta}_j$, where $\widehat{\beta}_j$'s are just-identified immigrant groups (country of origin)-level estimators and $\widehat{\alpha}_j$'s are called "Rotemberg weights". These exercises aim to shed light on the variation captured by the instrumental variable.

G.1 Summary of Rotemberg Weights

The Rotemberg weights of the shift-share estimator, computed with controls and aggregate across time periods, are summarized in Table G.1. Panel A presents descriptive statistics on the positive and negative weights. Only 6% of the weights are negative, so $\widehat{\beta}$ allows for a LATE-like interpretation.[86] Panel B reports the correlations between the Rotemberg weights ($\widehat{\alpha}_j$), the national immigration flows from country j (g_j), the just-identified coefficients estimates ($\widehat{\beta}_j$), the first-stage F-statistics ($\widehat{F}_j$), and the variation in the country-of-origin shares across locations ($var(z_j)$). The top five instruments (out of the 17 total countries of origin, reported in Table A.1) according to the weights are displayed in Panel C. They are: Austria-Hungary, Italy, Russia, Greece-Portugal-Spain, and Sweden. Although the first four are from the so-called "new" European countries of origin (Southern/Eastern Europe), the fifth and the following five are all from Northern/Western Europe (unreported in the table): Norway, United Kingdom, Germany, Denmark, and Ireland. This confirms that the instrument based on the 1890 shares is capable of predicting immigration from a broad set of European countries (as already noticeable from the high F-statistics of both instruments in Table 7). It is also worth noting that none of these weights are particularly large, with the top weight being only 27%.[87] The $\widehat{\beta}_j$ estimates also show considerable heterogeneity across the top five weights. In particular, the fact that the second largest weight has a negative coefficient indicates that the positive effects shown in the main results of the paper are not driven by a narrow set of high-weight countries. Finally, Panel D reports the estimates of β_j for positive and negative weights. Countries with positive weights account for al-

[85]The reported diagnostics are based on regressions where the dependent variable is *Any Union Present*. The results are similar if using any of the other unionization measures analyzed in the paper.

[86]Goldsmith-Pinkham et al. (2020) note that, although the $\widehat{\alpha}_j$'s must sum to one, some of the weights may be negative. In cases where a large share of them are negative, the estimator is unlikely to have a LATE-like interpretation.

[87]For comparison, Goldsmith-Pinkham et al. (2020) show that the top weight in Card (2009), Mexico, is 48%.

most the entire identifying variation (92.8%), and the unweighted mean of the estimates are very similar across positively and negatively weighted $\widehat{\beta}_j$'s.

G.2 Pre-Trends for Origins with the Largest Rotemberg Weights

Table G.2 reports the pre-trends tests for the individual instruments corresponding to the five origin countries with the largest Rotemberg weights. The specification mirros that in Table B.2, but instead of using the aggregate shift-share instrument, each panel uses a shift-share instrument constructed solely from the baseline shares and inflows of a single high-Rotemberg-weight origin. This allows a direct assessment of whether the individual instruments that contribute most to identification exhibit any systematic correlatiion with pre-1890 trends.

None of the instruments systematically predict pre-1890 changes. A small number of coefficients reach statistical significance, but these arise only for origins with comparatively low Rotemberg weights among the top five. Overall, the results provide no evidence of differential pre-trends for the origin-specific instruments most relevant to identification.

G.3 Heterogeneity by Rotemberg Weights

Figure G.1 reports two additional diagnostics that Goldsmith-Pinkham et al. (2020) recommend to explore. Panel A shows the relationship between the estimated $\widehat{\beta}_j$'s (y-axis) and the first-stage F-statistics (x-axis), where the size of the points is proportional to the Rotemberg weights. Circles denote positive weights, while diamonds denote negative weights, and the horizontal dashed line is plotted at the value of the overall $\widehat{\beta}$. Reassuringly, the just-identified instruments with a larger F-statistic tend to be closer to the overall estimate, and have a positive weight. Panel B displays the relationship between the first-stage F-statistics (y-axis) and the Rotemberg weights (x-axis), with the dashed horizontal line marking $F = 10$. The instruments with a positive Rotemberg weight display larger F-statistics, while the complete opposite is true for most of the ones with a very low (or even negative) weight.

Table G.1: Summary of Rotemberg Weights

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.062	-0.015	0.055
Positive	1.062	0.082	0.945

Panel B: Correlations

	$\widehat{\alpha}_j$	g_j	$\widehat{\beta}_j$	$\widehat{F}_j$	var(z_j)
α_k	1				
g_k	0.918	1			
β_k	0.126	-0.070	1		
F_k	0.115	0.124	0.055	1	
Var(z_k)	-0.222	-0.391	0.475	-0.383	1

Panel C: Top five Rotemberg weight countries of origin

	$\widehat{\alpha}_j$	g_j	$\widehat{\beta}_j$	95% CI
Austria-Hungary	0.267	7.55e+05	0.305	(-1.100,1.700)
Italy	0.258	7.36e+05	-1.135	(-4.500,-1.000)
Russia	0.227	6.62e+05	4.447	(-0.500,2.100)
Greece-Portugal-Spain	0.100	1.06e+05	13.431	(-3.200,3.100)
Sweden	0.063	1.20e+05	0.985	(-0.700,2.200)

Panel D: Estimates of β_j for positive and negative weights

	α-weighted sum	Share of overall β	Mean
Negative	0.192	0.072	2.021
Positive	2.461	0.928	1.648

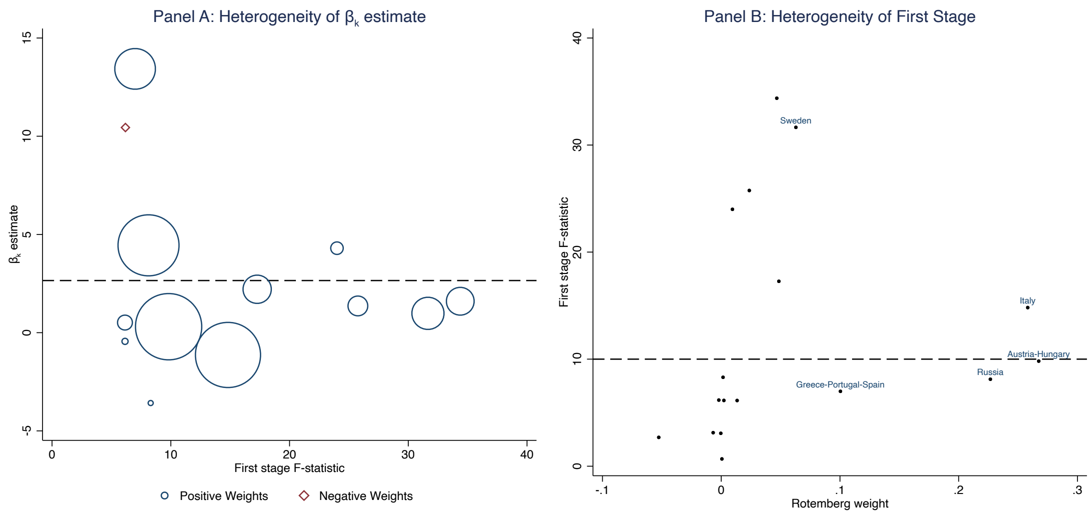
Notes: The table reports statistics about the Rotemberg weights (Goldsmith-Pinkham et al., 2020). In all cases, the reported statistics are about aggregated weights for a given country (or group of countries) of origin across years. The dependent variable is an indicator for whether the county has any labor union (column 1 of Table 3). Panel A reports the share and sum of negative weights. Panel B reports correlations between the weights ($\widehat{\alpha}_j$), the national immigration flows (g_j), the just-identified coefficients estimates ($\widehat{\beta}_j$), the first-stage F-statistic of the country-of-origin share ($\widehat{F}_j$), and the variation in the country-of-origin shares across locations (var(z_j)). Panel C reports the top five countries of origin according to the Rotemberg weights. The g_j is the national number of immigrants from country j, $\widehat{\beta}_j$ is the coefficient from the just-identified regression, and the 95% confidence interval is the weak instrument robust confidence interval using the method from Chernozhukov and Hansen (2008) over a range from -10 to 10. Panel D reports statistics about how the values of β_j vary with the positive and negative Rotemberg weights.

Table G.2: Test of Pre-Trends

	Dependent variable (1880-1890 difference):				
	Any Union Present (1)	Number of Union Branches (2)	Share of Urban Pop. (3)	Mfg. Output per Worker (4)	Mfg. Establish. per Worker (5)
Panel A: Austria-Hungary					
Share of Immigrants	-0.063	-1.607	-0.013	10.649	-20.578
(average 1900–1920)	(4.750)	(8.827)	(2.112)	(8.274)	(14.666)
Panel B: Italy					
Share of Immigrants	-0.214	-10.474	-1.354	0.292	-1.264
(average 1900–1920)	(4.369)	(7.379)	(1.244)	(5.132)	(9.485)
Panel C: Russia					
Share of Immigrants	-18.101	-17.402	11.006	33.433	224.269
(average 1900–1920)	(23.241)	(50.102)	(12.901)	(29.122)	(283.159)
Panel D: Greece, Portugal, Spain					
Share of Immigrants	-7.674	-18.081*	-1.194	-14.497	-28.691**
(average 1900–1920)	(6.423)	(9.636)	(1.884)	(8.891)	(13.960)
Panel D: Sweden					
Share of Immigrants	8.238*	-8.527	3.769**	-6.488	-1.890
(average 1900–1920)	(4.630)	(7.036)	(1.754)	(4.939)	(8.274)
Observations	871	871	871	871	871

Notes: The observations are at the county level. The dependent variables are the 1890–1880 difference in: an indicator for the presence of any union branch (of the Knights of Labor, column 1); the (inverse hyperbolic sine of the) number of union branches (of the Knights of Labor, column 2); the share of urban population (column 3); the (inverse hyperbolic sine of the) value of manufacturing output divided by the number of manufacturing workers (column 4); and the (inverse hyperbolic sine of the) number of the number of manufacturing establishments divided by the number of manufacturing workers (column 5). The regressor of interest is the number of immigrants (men ages 16–64) who entered the United States in the previous decade from the country indicated in the panel title, as a fraction of the male working-age population in the county, averaged over the period 1900–1920, and predicted by the instruments described in Appendix G, averaged over the period 1900–1920. All regressions include the following controls: the 1890 share of the population working in manufacturing and agriculture, an indicator for the presence of active coal mines in 1890, and the 1880 share of European immigrant population. *** p<0.01; ** p<0.05; * p<0.1.

Figure G.1: Heterogeneity by Rotemberg Weights



Notes: Panel A plots the relationship between each instrument's $\widehat{\beta}_j$, first-stage F-statistics and Rotemberg weights. Each point is a separate instrument's estimates (country of origin share). The figure plots the estimated $\widehat{\beta}_j$ for each instrument on the y-axis and the estimated first-stage F-statistic on the x-axis. The size of the points are scaled by the magnitude of the Rotemberg weights, with the circles denoting positive Rotemberg weights and the diamonds denoting negative weights. The horizontal dashed line is plotted at the value of the overall $\widehat{\beta}$. In order to not visually overstate dispersion, the figure reports only instruments with first-stage F-statistic of 5 or above (Goldsmith-Pinkham et al., 2020). Panel B plots the relationship between each instrument's first-stage F-statistic (y-axis) and Rotemberg weights (x-axis). The labeled points correspond to the five countries of origins with the highest Rotemberg weights. The dashed horizontal line is plotted at F = 10.

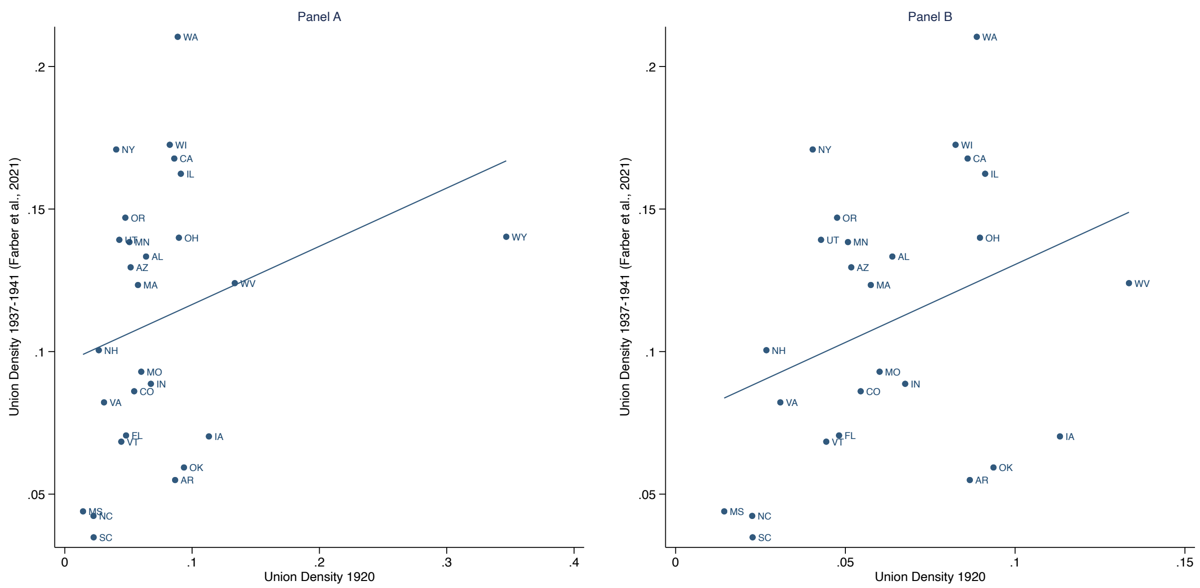
H Validation of Union Density Estimates

In addition to the steps outlined in Section 3.1 to ensure the accuracy and completeness of the collected data, this section provides a validation of the union density estimates. Specifically, I examine their correlation with the only other available measures in a historical period, from Farber et al. (2021), who harmonized household-level survey data from Gallup beginning in 1937.

Figure H.1 presents a scatter plot comparing the two measures, where the data from this paper are aggregated at the state level to match the unit of observation used by Farber et al. (2021). Unfortunately, the two sources do not overlap in time. As a result, the figure plots union density from this paper for 1920 on the x-axis, and the average union density from Farber et al. (2021) for the first five years of available observations (1937–1941) on the y-axis.

While the two measures do not align in levels—unsurprising given the emergence of new industrial unions by 1937, which organized large numbers of previously unorganized workers—they exhibit a positive correlation (Panel A; $\rho = 0.27$, $\beta = 0.2$, $R^2 = 0.07$). This correlation becomes stronger when Wyoming, an outlier, is excluded from the sample (Panel B; $\rho = 0.34$, $\beta = 0.55$, $R^2 = 0.12$).

Figure H.1: Correlation Between Data of This Paper and State-Level Gallup Data



Notes: The scatter plots displays 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 3 for more details on the dataset on labor unions I assemble for the period 1900–1920. Panel A includes all states in the main sample, while Panel B excludes Wyoming.