Closing Ranks: Organized Labor and Immigration

Carlo Medici*

September 25, 2024

Updated regularly. **Click here for the latest version.**

Abstract

This paper shows that immigration fostered the emergence of organized labor in the United States. I digitize archival data to construct the first county-level dataset on historical U.S. union membership and use a shift-share instrument to isolate a plausibly exogenous shock to the labor supply induced by immigration, between 1900 and 1920. Counties with higher immigration 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 increase occurred more prominently among skilled workers, particularly in counties more exposed to labor competition from immigrants, and in areas with less favorable attitudes towards immigration. Taken together, these results are consistent with existing workers forming and joining labor unions for economic as well as social motivations. The findings highlight a novel driver of unionization in the early 20th-century United States: in the absence of immigration, the average share of unionized workers during this period would have been 22% lower. The results also identify an unexplored consequence of immigration: the development of institutions aimed at protecting workers' status in the labor market, with effects that continue into the present.

Keywords: Labor Unions, Immigration

JEL codes: J15, J5, N31, N32, P1

*Brown University. carlo_medici@brown.edu.

First version: August 2023. 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 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 CREI, Northwestern, Nottingham, Tilburg, UPF, and USC, and conference participants at the EHA (Pittsburgh 2023 and Sacramento 2024) and ASSA (San Antonio 2024) Annual Meetings, 2024 NBER Summer Institute (DAE), 2024 NICEP conference, and 2024 Junior Economists Meeting 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 a central institution in the labor markets of advanced economies. Throughout the 20th century, they significantly contributed to reducing inequality (Farber et al., 2021), improving working conditions (Rosenfeld, 2019), and shaping policy through active political engagement (Ahlquist, 2017). Despite fluctuations in membership, unions remain pivotal to today's economy.[1] Yet, given their sustained importance, it is striking how little evidence exists on the factors driving their 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 during this period centers on the increased capital intensity in industrial production, which shifted the bargaining power from laborers to the owners of capital (Foner, 1947). A related hypothesis suggests that workers organized in response to growing labor competition (Taft, 1964), which intensified during this period as boosts to agricultural productivity relieved labor from farming, and both the total population and the urban population share grew.

This paper investigates the second mechanism: the effect of a large and protracted increase in labor supply on the formation and expansion of labor unions, leveraging the episodes of mass immigration to the United States in the early 20th century. The effect is *ex ante* ambiguous, as it influences both workers' incentives to organize and capital owners' 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 answer this question. First, the U.S. economy was already the largest in the world (Bolt and Van Zanden, 2020) and the labor movement experienced its first national expansion at the turn of the century (Foner, 1947). Second, these years witnessed the creation and growth of several labor unions that remain influential today (Stewart, 1926), despite the legal and judicial frameworks of the time allowing employers to easily dismiss and replace unionizing workers (Taft, 1964). Third, this context provides a natural experiment to establish causal identification, given by the large and prolonged influx of European immigrants during this period, often referred to as the Age of Mass Migration (Hatton and Williamson, 1998).

[1] In the U.S., they recently gained historic victories for several categories 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 boasts a long tradition, organized labor continues to expand to previously unorganized sectors, shape the policy agenda, 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 state or national level and, therefore, do not allow for analyses across local labor markets. The second challenge is establishing causal effects. For example, the presence of unions may deter immigration. Such reverse causality would result in a negative association between immigrant flows and union presence. Alternatively, both the size of unions and immigration may increase in response to economic growth. Such joint determination would lead to a positive association between unionization and immigrants.

To measure unionization, I hand-collect and digitize archival documents on the location, quantity, and membership of labor union branches across the United States. The main sources of these records are the convention proceedings of the state federations of labor, which report detailed information on the number and location of union branches within each state's territory, along with the names of the delegates sent by each branch to the conventions. I collect these data 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 density (the share of unionized workers) at the county level in the United States.

To estimate the causal effect of immigration, I use a shift-share instrumental variable (Card, 2001b) to exploit plausibly exogenous variation in the flow of immigrants across counties in each decade. The instrument interacts the 1890 share of immigrants living in a given U.S. county and born in different European countries with the aggregate immigration flows from each country to the United States between 1890 and 1920. This identification strategy is motivated by the empirical regularity that immigrants tend to settle where other migrants from their own country of origin had previously settled, a process known as *chain migration*. The key underlying assumption is that, conditional on controls, the unobserved factors that affected unionization outcomes must not be jointly correlated with the 1890 composition of Europeans' enclaves across U.S. counties and the out-migration patterns from European countries after 1890.[2] I estimate 2SLS regressions that include county and year fixed effects, in addition to baseline county characteristics which are correlated with the initial presence of immigrants, such as the share of population in farming and in manufacturing, the number of coal mines per 1,000 inhabitants, and an indicator for the presence of a railroad in the county, interacted with year dummies.

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

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 10 percentage points (23% of the sample mean), the number of union branches by 17%, the share of the unionized workforce by over one percentage point (39% relative to the mean), and the number of union members per branch by 20 (42% of the mean). Immigration spurred unionization along both the extensive and the intensive margin, as new counties saw the establishment of labor unions and those with an existing labor movement experienced an expansion in its size. 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, such as using an alternative instrument that replaces actual immigration flows with plausibly exogenous ones and combining the instrument with a matching strategy.[3] The findings are also not sensitive to the inclusion of several additional controls, such as the initial size of the immigrant population, the baseline shares of the labor force in major industries and occupations, and measures of income and economic growth.

The second part of the paper examines the mechanisms behind the expansion of organized labor. First, I explore whether existing workers created or joined labor unions in response to the economic challenges posed by immigration. While unionization attempts would be expected across all occupations, as all workers perceived immigrants as a threat to their employment and wages (Asher, 1982; Mink, 1986; Olzak, 1989), they should be more likely to succeed in occupations with barriers to entry. Without laws protecting the right to organize during this period, workers had to battle employers for union recognition through strikes and walkouts, while courts often sided with employers in disputes over the dismissal of unionizing employees (Fishback, 1995; Taft, 1964; Taft and Ross, 1969). In such an environment, incumbent workers who cannot be *immediately* replaced by immigrants should be in a stronger position to unionize and prevent immigrants from acquiring their skills, in order to reduce future job competition. Similarly, employers lacking a readily available pool of replacement workers should be less likely

[3] Although previous work has argued that this period is particularly suited to the use of shift-share instruments (Abramitzky et al., 2023; Tabellini, 2020), the alternative instrument, which relies on predicted flows using weather shocks across European countries (Sequeira et al., 2020), allows me to identify causal effects from the exogenous variation in the shocks, while allowing the exposure shares to be endogenous (Borusyak et al., 2022). Moreover, I build on Bazzi et al. (2023) and combine the instrument with a matching exercise, which selects within-state county pairs with the closest levels of union presence in 1890. All the robustness checks are described in Section 5.3.

to resist their employees' unionization efforts. Differences in skill requirements across occupations provide a testing ground for this mechanism. Supporting this hypothesis, immigration strengthened labor unions in skilled trades, such as craft occupations, while having no statistically significant effect on the unionization of low-skilled workers, such as operatives and laborers.

Second, I investigate whether counties where immigrants directly competed with existing workers experienced a larger increase in unionization. To determine exposure to immigrant labor competition, I measure whether the occupations prevalent among immigrants entering the United States in each decade were also predominant among U.S.-born workers in a given county at the start of that decade. Consistent with the hypothesis that unionization was a response to economic concerns raised by immigration, skilled workers were more likely to unionize in counties with higher exposure to competition from immigrants. In contrast, competition from immigrants slowed union growth among low-skilled workers, whose bargaining power was undermined by the increased availability of inexpensive labor.

Third, I explore whether social motivations also contributed to the observed development of labor unions. Given the nativist rhetoric that accompanied the labor movement's support for immigration restrictions throughout the first half of the 20$^{\text{th}}$ century (Goldin, 1994; Mink, 1986), it is plausible that the cultural dissimilarity of immigrants provided an additional incentive for workers to organize and exclude newcomers from the labor market. I find evidence consistent with this hypothesis. I show that the increased unionization was more prominent following an inflow of immigrants from Southern and Eastern Europe, whom part of the labor movement considered "slavish, ignorant and unassimilable," and therefore, a threat to American society (Collomp, 1988; Mink, 1986). Further, I show that unionization grew more in places with less favorable attitudes towards 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, ran on an anti-Catholic and anti-Irish platform (Alsan et al., 2020). The second is the baseline level of residential segregation between U.S.-born individuals and European immigrants. Since residential segregation usually results from 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 indicate greater levels of discrimination against immigrants. Using either of these proxies, I find that immigration led to greater union growth in counties with higher resentment towards immigrants.

Next, I rule out several alternative channels that could drive the results. First, I show that the findings are unlikely to be explained by immigrants disproportionately participating in unions. Given that information on the country of origin of individual union members does not exist, I provide suggestive evidence against this alternative explanation by examining the relationship between immigration and the origin and ancestry of local union leaders, inferred from their last names. I document that the share of union leaders with last names common among U.S.-born individuals increased overall during this period. At the same time, immigration did not lead to a rise in the proportion of immigrant last names among local union leaders. Second, I exploit variation in the strength of labor unions and socialist parties across Europe at the beginning of the 20th century, and document that the inflow of workers from countries with active labor movements or stronger support for socialism did not contribute to the increase in unionization. Third, I show that counties with higher levels of immigration did not experience different economic growth, as there was no effect on labor force participation, manufacturing activity, or the share of workers in skilled occupations. This indicates that economic expansion or increased demand for skilled labor is unlikely to explain the observed patterns in union growth.

The last part of the paper explores the economic implications of this immigration-induced unionization. Although not all the following findings should be interpreted as causal, they provide key insights into short- and long-run trends associated with the patterns documented in this paper. First, I examine whether U.S.-born workers shifted towards unionized occupations as a means of protection against the perceived economic and social threats brought by immigration. I find that immigration increased the share of U.S.-born workers in unionized occupations and reduced it in non-unionized ones. This finding suggests that U.S.-born workers may have turned to jobs where organized labor could shield them from the potential adverse consequences of immigration. Second, I investigate whether the effects of early 20th-century immigration on unionization persist into the present. Using current union density measures from Macpherson and Hirsch (2023), I aggregate the data at the metropolitan-area level and employ the same shift-share instrument used in the main analysis to assess the impact of historical immigration on present-day private-sector union density. Notably, metropolitan areas that experienced larger waves of immigration continue to show higher unionization today. Consistent with the union growth patterns of the early 20th century, which were largely concentrated in the construction sector, these effects are also concentrated in the construction sector. This suggests that the conditions that facilitated the initial rise of labor unions in the early 1900s provided a lasting advantage to the labor movement which continues into the present. Third, I explore a central economic question concerning labor unions: their relationship with inequality (Card, 2001a; DiNardo et al., 1996). I construct measures of wage inequality and analyze their cross-sectional correlation with unionization.[4] The findings indicate that the presence of labor unions is associated with lower overall wage inequality but higher inequality between U.S.-born and immigrant workers. This

[4]Because the U.S. Census started collecting wage data in 1940, I explore the correlation between inequality measured in 1940 and unionization measured in 1920.

pattern is consistent with labor unions compressing the overall wage distribution while raising disparities between the unionized and nonunionized.

In summary, the empirical findings of this paper show that immigration substantially contributed to the emergence and expansion of organized labor in the early 20th-century United States, with effects persisting over the years. The results are consistent with existing workers forming and joining labor unions for economic as well as social motivations.

Related literature. The results of this paper contribute to several broad literatures. First, they speak to the studies on organized labor, and labor unions more specifically. While a rapidly growing recent empirical literature has studied labor unions and analyzed their impact on a wide range of economic and political outcomes, both in historical and contemporary settings (Ahlquist, 2017; Ash et al., 2019; Barth et al., 2020; Biasi and Sarsons, 2022; Bittarello, 2018; Card, 2001a; Collins and Niemesh, 2019; DiNardo and Lee, 2004; Dodini et al., 2023; Farber et al., 2021; Feigenbaum et al., 2018; Fishback, 1992, 1995; Naidu, 2022; Naidu and Reich, 2018; Rosenfeld and Kleykamp, 2012; Rosenfeld, 2019; Sojourner et al., 2015; Schmick, 2018; Wang and Young, 2022), this paper is the first to study the determinants of their early development with systematic empirical evidence. The results identify immigration as a key factor that led to the emergence and growth of modern unions during a highly formative period for the American labor movement.

This paper also relates to studies that explore the historical drivers of unionization (Alesina and Glaeser, 2004; Archer, 2010; Asher, 1982; Bernstein, 1954; Briggs, 2001; Burgoon et al., 2010; Brody, 1993; Collomp, 1988; Foner, 1947; Freeman and Medoff, 1984; Friedman, 1998; Griffin et al., 1986; Hannan and Freeman, 1987; Haydu, 1988; Karadja and Prawitz, 2019; Lipset and Marks, 2000; Montgomery, 1979; Moody, 2019; Naidu and Yuchtman, 2016; Olson, 1965; Sezer, 2023; Sombart, 1976; Taft, 1964; Willoughby, 1905; Webb and Webb, 1894; Wolman, 1924), and those that analyze the causes of its decline in recent decades (Acemoglu et al., 2001; Ahlquist and Downey, 2020; Clawson and Clawson, 1999; Farber and Western, 2001; Hirsch, 2008; Scruggs and Lange, 2002; Slaughter, 2007; Southworth and Stepan-Norris, 2009; Wallerstein and Western, 2000). This study advances this literature by identifying an unexplored driver of unionization and shedding light on the channels through which it operates.

The data collection effort of this paper also delivers the first comprehensive county-level dataset on historical union presence and membership in the United States, covering almost the entire country. Although a few existing papers have collected historical information on labor unions, those data are either on extinct organizations whose relevance was limited to the 1880s (Garlock, 2009), only cover a limited set of unions and do not contain information on membership (Schmick, 2018), are not disaggregated below the state level (Farber et al., 2021), or measure unionization only in a handful of states (Downes, 2024). The data introduced in this paper, aggregated at the county level for the analysis, but collected at the city or town level, make a significant advancement in studying geographic patterns of early unionization, and open avenues for future research on the medium- and long-term consequences of organized labor in the United States.

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

Further, this study relates to the vast literature about the consequences of immigration on native 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 workers. The results suggest that labor unions may play a role in mitigating the potential adverse effects of immigration on native 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 and unexamined consequence of immigration on the development of institutions that have had – and still have today – vast political influence. Although anecdotal and historical evidence has acknowledged the instrumental role that organized labor played in the introduction of immigration restrictions in the 1920s (Goldin, 1994; Mink, 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), which links the weak labor and socialist movements of the United States to its ethnic diversity. The results of this study shed further light on this phenomenon, showing that reactions to immigration can foster unionization, partly offsetting other opposing forces that may slow down its growth.

Outline. The remainder of the paper is organized as follows. Section 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 discusses the economic implications of the findings and the long-term effects of immigration on unionization. Section 8 concludes.

2 Historical Background

2.1 Labor Unions at the Turn of the 20th Century

A new phase for the American labor movement started around the end of the 1880s, as the American Federation of Labor (AFL) became the largest and most influential group of labor unions.[5] By 1890, the main labor organizations that had gained importance during the second half of the 19th century – the Knights of Labor and the independent railroad workers' movements – had practically disappeared,[6] leaving the field open to new trade unions (Wolman, 1924). These years saw the creation of many new organizations, which later became some of the largest national trade unions still active today.[7] Between 1880 and 1920, the total number of union members went from 149,000 to over 4.5 million (Figure 1).

The AFL was created as a federation of national unions and organized on the model of craft unionism. This meant that workers were organized based on their particular occupation (or craft).[8] It adopted the policy of *one craft–one union*, according to which each occupation should have only one union representing it. During this period, the unions in the building construction industry became the most stable and largest organizations.[9] This industry was dominated by skilled craftsmen, and characterized by small employing units (Taft, 1964). Only a few unions organized unskilled laborers in industrial settings. The United Mine Workers of America (UMWA) was the largest of these, along with unions representing longshoremen, teamsters, and laborers in the brewing, meatpacking, 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 national unions were organized into branches, called *locals*. These branches were responsible for directly 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,

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

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

[7] 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 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).

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

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

disability, or death, and regulated the terms of apprenticeship within the craft (Stewart, 1926).

Most collective agreements included a *closed-shop* clause, specifying that only union 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).[10] 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 (Hatton and Williamson, 1998), raising the share of the foreign born population to over 14% (Figure 2 and Figure 3). The mix of origin countries changed substantially over time. Until 1890, most immigrants were from the United Kingdom, Ireland, Germany, and Scandinavia. Thereafter, as transportation costs decreased (Keeling, 1999), the bulk of immigrants arrived from the rest of Europe. In 1850, immigrants from Northern or Western Europe were 92% of the foreign-born population, while less than 1% had arrived from Southern, Central, or Eastern Europe. By 1920, these shares were 40% and 43%, respectively (Figure 4). Europeans from the new origin regions were different from those who had arrived in the previous decades: they were significantly less skilled, spoke unfamiliar languages, and were not Protestant (Hatton and Williamson, 1998, 2006).

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

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

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 xenophobia toward new arrivals (Goldin, 1994). In response, starting in the late 1890s, members of Congress proposed legislation to limit immigration, and in 1917, Congress eventually introduced a literacy requirement for all immigrants.[11] Though immigration temporarily slowed down during World War I, after the end of the war it immediately rose again, resurrecting earlier anti-immigration fears. Consequently, in 1921 Congress passed the Emergency Quota Act and introduced a temporary limit to immigration. In 1924, the National Origins Act made this restriction permanent and more stringent (Abramitzky and Boustan, 2017). The immigration quotas remained in effect for the next 40 years, until they were eliminated in 1965 by the Immigration and Nationality Act.

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, the AFL adopted a resolution against Chinese laborers and lobbied Congress to ban Chinese immigration through the Chinese Exclusion Act of 1882 (Foner, 1947). In 1885, the labor movement achieved the enactment of the Alien Contract Labor Law (the Foran Act), which banned the importation of foreigners laborers.[12] 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).[13] At the policy level, the AFL contin-

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

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

[13] 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

ued to endorse further restrictive measures, ultimately leading to the introduction of the 1921 and 1924 nationality quotas (Goldin, 1994).

Throughout this period, the labor movement also used increasingly popular racial and eugenics-based arguments to discuss threats to employment and gain momentum in calling for an outright ban on European immigration.[14] Nativist sentiments were triggered by the increased presence of foreign laborers, which inundated labor markets, and was intensified by the mounting pressure of mechanization (Mink, 1986; Yellowitz, 1981). These events added credibility to the fears that machines and the new unskilled workers could substitute skilled unionized labor (Olzak, 1989), and led unions to concentrate on securing job control for skilled workers by organizing the workplace and the work process (Mink, 1986). At the same time, the immigration-induced expansion of the labor supply was deemed responsible for weakening unions' bargaining power, by creating a reservoir of potential strikebreakers and freeing employers from the constraints of a tight, unionizing labor market (Montgomery, 1979).

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 the first panel dataset on unionization for the period 1900–1920. This also constitutes the first comprehensive dataset on historical union density measured at the county level in the United States. Most existing studies on modern labor unions in a historical period rely on aggregate national estimates, since microdata on union status were first collected by the Current Population Survey (CPS) only in 1973. There are a few notable exceptions. Schmick (2018) collects data on the presence of local branches of some national unions in the years 1882, 1892, and 1902. However, the dataset contains no information on membership and covers a different set of unions in different years in a time

each trade (Ignatiev, 1994; Mink, 1986; Stewart, 1926; Witwer, 2002).

[14] 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).

period that precedes the first significant expansion of the labor movement and the largest waves of immigration. Farber et al. (2021) combine survey data, primarily from Gallup, to compute historical levels of union membership for most of the 20th century. However, their data are not disaggregated below the state level, and only start in 1937, after immigration restrictions had been in place for over a decade and the first national expansion of the American labor movement had occurred. Similarly, Downes (2024) constructs county-level union membership estimates for selected years in the mid-20th century. However, his data start in 1920 and are limited to five states, hence also unsuitable to study the questions of this paper.

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

Convention proceedings of the state federations of labor. The main sources of the dataset on unionization are convention proceedings of the state federations of labor, which were state-level subordinate bodies of the AFL. Their functions were mainly legislative and propagandist, and they were composed of representatives from all the local branches of the AFL-affiliated national unions within the state (Stewart, 1926). Local branches (also called local unions, or locals) were a lower level of organization of national unions, and represented workers in either a single employment unit or from several work sites. By 1920, members of AFL unions constituted more than 80% of the total private-sector union membership (Wolman, 1924). Each state federation of labor met annually in conventions to enact legislation and elect general officers. All affiliated local unions were entitled to representation.[15]

I digitize the proceedings of these conventions every 10 years between 1900 and 1920.[16] From these documents, I extract the lists of union branches (*locals*) represented at the conventions, along with the union name and branch number, their location, the number of delegates representing them, and the names of such delegates (Figure A.1). Each federation had specific rules to define the number of delegates that could represent a local branch, which often varied over time. Importantly, they established that locals should be represented proportionally to their membership (Figure A.2).[17] I therefore combine the information on the delegates from the convention proceedings with the details on the representation rules contained in the constitutions of each state federation of labor. Us-

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

[16]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. The complete list of digitized documents by state and year is provided in Table A.3.

[17]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.

ing this information, I construct an estimate of union membership for each local branch. Since the representation rules were often expressed in terms of ranges (e.g., one delegate every 100 members), I use the mid-points of these intervals as the estimates of membership. For example, if the constitution states that a branch is represented by one delegate every 100 members, its membership is estimated to be 50 if one delegate is present at the convention, 150 if two delegates are recorded, and so on.[18]

I geocode the location of all the union branches based on their town, village, or city, and retrieve their coordinates. I use the names of each branch's national union to establish which occupations and industries they operated in.[19] Finally, I aggregate the membership of the union branches at the county level to obtain a measure of union membership, both total and by occupation.

Proceedings of the national conventions of AFL unions. I complement the data from the state federations of labor with analogous information collected directly by the AFL-affiliated national unions. Similar to the state federations, the AFL-affiliated unions met in national conventions to legislate, elect officers, and set guidelines for the local branches to follow in their bargaining agreements. I digitize the proceedings of these conventions for the five largest AFL-affiliated unions (as of 1900) at the same 10-year intervals as the main data.[20] The members of these five unions accounted for more than a third of the over 100 AFL-affiliated unions' total membership between 1900 and 1920 (Wolman, 1924).[21] I follow a procedure analogous to the one described for the proceedings of the state federations of labor, and collect data on the lists of local branches, their location, and the names and number of delegates representing them. Next, I construct an estimate of the membership of each of these locals, following the different representation rules listed in each of the convention proceedings or in the constitutions of these organizations. Finally, I aggregate the data at the county level.

These data sources complement the records from state federations in three main ways. First, they validate the estimates constructed using the main data source. In particular, for the five unions that I observe across both sources, I am able to compare the estimates

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

[19] 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.).

[20] 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 before, or two years before.

[21] 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 in their convention proceedings about their delegates and the local branches they represented between 1900 and 1920, and whose proceedings are still available either in physical or digital copy.

of union membership and the number of branches. In all cases, the measures display a highly positive correlation (Figure A.3). Nonetheless, some branches may appear in only one of the two types of convention documents I digitize. This may occur because one branch was established too recently before a convention and did not yet qualify to send delegates according to organization-specific rules. Similarly, it could have been formed between the state federation and the national union convention; hence, it could only be observed in one of the two documents. Another possibility is that some delegates may have been erroneously omitted from the roll calls of the meetings.[22] Any of these occurrences would lead to an underestimate of the number of members and/or the number of branches in a given county if only one of the sources was used. Unfortunately, there is no way of knowing with certainty if and how many locals fall into these circumstances, since this information is never systematically reported. However, by combining information from different sources (and collected by different entities), I am able to reduce these instances of mismeasurement. This constitutes the second main contribution of this data source. Third, these additional archival records allow me to expand the time and geographical coverage of the dataset, since some state federations of labor were only established (and convened for the first time) after 1900.[23] Relying solely on the main data source would result in measuring zero union presence and membership for counties in states during periods before the first federation of labor conventions. While the absence of an AFL state body does indicate a limited presence of organized labor, some of the largest national unions may still have been active in at least some of these counties. The additional information on the branches and delegates of these five large unions operating nationwide in 1900, 1910, and 1920 allows for 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.[24] I first reduce the number of missing observations and misreportings from each of these sources by linearly interpolating the number of union branches and members for counties that are not reported in the convention proceedings of a certain year, but that have representation both in the

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

[23] 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.

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

previous and following decade.[25] Next, for the five unions observed across both types of documents, I compute the number of members and branches in each county and year by averaging the ones from each source. When only one data source reports a positive membership or number of local branches, I use that value in the analysis. Finally, I sum the total number of branches and members across all unions at the county-year level, and obtain the total number of these quantities in each county over time.

In order to construct a measure of 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).[26] Additionally, I construct an indicator for whether a county has any union, the number of union branches within its territory, and their average size, defined as the number of members divided by the number of branches. All four variables are then winsorized at the 1% to remove outliers. As a final validation exercise, I compare these measures of union density to those contained in another existing historical dataset. While only aggregated national estimates of union membership exist for the period studied, I ensure that the measures of union density are positively correlated with those calculated at the state level by Farber et al. (2021), using Gallup surveys starting in 1937 (Figure H.1).

The final dataset contains information on the location and membership of local union branches in over 2,400 counties between 1900 and 1920.[27]

3.2 Other Data Sources

Immigration and population. The data on county population and on the number of immigrants, by country of origin at the county and national levels, are taken from the decennial U.S. Census of Population. For 1900, 1910, and 1920, I use the full-count Census datasets, made available by IPUMS (Ruggles et al., 2022). For 1890, I use Census datasets

[25]Counties may wrongly appear to have no union branches or members in a certain year due to one of the following reasons: error in assigning a locality to the correct county because of homonymous locations, a partial or incorrect reporting of the delegates present at the convention, or county-specific reasons for why no delegate was actually not sent to one of the two conventions. The underlying assumption for this exercise is that a county with union branches and members in, say, 1900 and 1920, will not realistically have zero branches and membership in 1910. I also collect available data for the state federation conventions that took place in 1930 in order to linearly interpolate the data from the first source for the year 1920. Importantly, the results are unchanged if this step is not conducted (see Section 5.3).

[26]This includes all men ages 16–64 in nonfarm occupations, except for managers, proprietors, and private household service workers, as these groups were typically ineligible to unionize during this period (Stewart, 1926; Wolman, 1924). In case the total number of estimated union members exceeds the labor force, union density is coded to be one. This is a rare event, which occurs for the main measure of union density only in three out of the 2,628 county-year observations of the main sample. In Section 5.3, I also show that the results are not sensitive to the exclusion of outliers.

[27]The counties not part of the sample are those in states whose federations of labor did not have a representation rule for branches proportional to their membership (as previously described), whose convention proceeding are not available, or reported only incomplete records (e.g., no information on the location of the branch, or no list of delegates altogether). Throughout the main analysis, I will restrict the sample to a balanced panel of 876 urban counties (see Section 3.3); the results are unchanged when using different sample restrictions or extending the analysis to all counties (Section 5.3).

aggregated at the county level, made available by the Inter-University Consortium for Political and Social Research (ICPSR) (Haines, 2010).[28]

Labor market outcomes. I compile data on labor force, occupation, and yearly income from the U.S. Census of Population.[29]

Economic activity. The county-level data on the manufacturing and agricultural sectors come from the Census of Manufacturing (Haines, 2010) and the Census of Agriculture (Haines et al., 2018), respectively.

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

Coal mines. I digitize the information on the location and number of coal mines from the Report on Mineral Industries of the 1890 U.S. Census (Day, 1892).

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

3.3 Summary Statistics

Figure 5 plots the number of union branches, and Figure 6 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.

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.[30]

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

[29]Due to the unavailability of the labor force participation status in the 1900 full-count Census dataset (Ruggles et al., 2022), I proxy for this variable in that year with an indicator for holding any gainful occupation.

[30]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

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) 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), the average share of European immigrants in the population was 10%. On average, 45% of families worked in farming, and 3% of the population in manufacturing. There were about 1.5 coal mines per 10,000 inhabitants, and 95% of counties were crossed by a railroad.

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, which are likely correlated with both the pre-1900 settlement of immigrants and the evolution of unionization over time, measured at baseline and interacted with year fixed effects. Throughout the analysis, standard errors are clustered at the county level, and all variables are harmonized to reflect 1930 county boundaries (Hornbeck, 2010).[31]

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

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.

[31] 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).

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.[32]

4.2 Instrument for Immigration

Given the hostility of the labor movement towards immigration described in Section 2, we may expect immigrants to settle in counties with less unionization, where the chances of being excluded from certain occupations would be lower. This would cause the ordinary least squares (OLS) estimates of equation (1) to be biased downwards. By contrast, immigrants may prefer counties with a growing labor movement, to the extent that those labor markets might also present more or better job opportunities. This would bias the OLS estimates upwards. In addition, classical measurement error in the immigration data would attenuate the estimates towards zero.

Baseline instrument. To deal with these endogeneity concerns, I construct a shift-share instrument (Card, 2001b; 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), which I denote as $\alpha^j_{c,1890}$. The second is the change, or *shift*, in the number of European immigrants from country j entering the United States in a given decade, net of those that eventually settled in county c, denoted by O^j_{-ct}.[33] 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 peri-

[32]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). The results are very similar when considering all immigrants, regardless of their sex, age, or arrival year (Section 5.3).

[33]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.

ods of union growth (Figure 1 and Figure 2).[34] Importantly, 1890 also predates the major shift in the composition of immigration that occurred around the turn of the 20$^{\text{th}}$ century (Figure 4). 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 between 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 controls, the unobserved factors that affect unionization outcomes must not be jointly correlated with the 1890 composition of Europeans' enclaves across U.S. counties and immigration patterns from European countries after 1890.[35] Previous work has argued that nation-wide shocks that occurred during the period 1900–1920, and which are exogenous to county-specific characteristics, make this setting particularly suited to the use of shift-share instruments (Abramitzky et al., 2023; Tabellini, 2020). In particular, the trend-break in immigration created by WWI lowers the concern that the shift-share instrument may be correlated with shocks jointly affecting local conditions in U.S. counties and immigration patterns from European countries. Moreover, the WWI shock reduces worries about the design being invalidated by the serial correlation in migration flows from the same country to the same U.S. destination (Jaeger et al., 2018).

Instrument validity. Nevertheless, although the immigrant networks captured by $\alpha^j_{c,1890}$ predate the time period of the analysis, they may be endogenous with respect to the trajectory of the outcomes of interest. I deal with this concern in several ways. First, I augment the main specification by including interactions between year dummies and county characteristics that were correlated with immigrants' settlements (from each origin country) in 1890, and which may have had a time-varying effect on unionization across counties. In the preferred specification, such controls include: (i) the share of families in farming, (ii) the share of the population employed in manufacturing, (iii) the number of coal mines per 1,000 people, and (iv) an indicator for the presence of a railroad in the county. 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). Counties with lower farming activity, substantial manufacturing and coal mining, and connections to the railroad network attracted more immigrants early on (Table A.2), and likely experienced greater union growth in the early 20$^{\text{th}}$ century.[36]

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

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

[36] In Appendix B, I show that the results are robust to the inclusion of various county-level controls that

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

Second, I directly control for the size of the 1890 immigrant population, interacted with year dummies. This implies that the effects of immigration are identified exploiting variation only in the ethnic composition of immigrant enclaves across counties, holding constant the size of their foreign-born populations. Since the instrument predicts higher immigration to counties with a larger stock of immigrants at baseline, by doing this I also address the concern that a larger 1890 immigrant population may itself have an independent and time-varying effect on unionization. Third, I include interactions between year dummies and the baseline share of immigrants from each European country, $a^{j}_{c,1890}$, to assuage concerns that the 1890 settlements of specific European groups across U.S. counties might be correlated with both the long-run trends in unionization and the migration patterns of those specific immigrants groups, in each decade between 1890 and 1920 (Goldsmith-Pinkham et al., 2020).[37]

Alternative instrument. In addition, 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 allows me to identify causal effects from the exogenous variation in the shocks, while allowing the exposure shares to be endogenous (Borusyak et al., 2022). I then interact them with the baseline shares of European immigrants from each country j to obtain the alternative instrument. Appendix B.1 describes its construction in more detail.

Matching and shift-share instrument. Finally, similarly to Bazzi et al. (2023), in Appendix B.2 I combine the shift-share instrument of equation (2) with a matching exercise. In particular, I select within-state county pairs with similar baseline presence of labor unions, as measured by the number of branches of the Knights of Labor in the late 19th century as a fraction of the county population.[38] Then, I re-estimate the 2SLS analysis also controlling for fixed effects for the county pairs interacted with year dummies.

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

could influence both the 1890 immigrant population and subsequent unionization (Table B.4). I also show that the findings are not dependent on the inclusion of any of these baseline controls (Table B.5)

[37] 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".

[38] As described in Section 2, the Knights of Labor were a federation of unions that was particularly active in the 1880s, and declined after 1890, when the AFL became the dominant federation of unions. For this exercise, I use data from Garlock (2009) to measure union presence as of 1880 and 1890, when the AFL had either not yet or only recently been established (Foner, 1947).

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 (log) number of union branches (column 2);[39] union density, defined as the number of union members as a fraction of the labor force (column 3);[40] and the average branch size, defined as the number of union members divided by the number of branches (column 4).[41] All regressions include county and year fixed effects, and interactions between year dummies and the baseline share of the population working in farming and in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county (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-stat 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 a 10 percentage point (23% relative to the mean) higher probability that the county had unions (column 1); a 17% increase in the number of union branches (column 2);[42] a higher share of unionized workers by 1.2 percentage points, or 32% relative to the sample mean (column 3); and 20 more members per branch, or 42% 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.[43] A back-of-the-

[39] Since this variable may take value zero if no union branch is observed, I apply the transformation $log(1 + x)$ instead of $log(x)$, where x is the number of branches. The results are very similar when using an inverse hyperbolic sine transformation instead.

[40] 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 (Table B.9).

[41] 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.

[42] Given that the dependent variable of column 2 is log-transformed, the magnitude of the coefficient can be calculated as follows: $\%\Delta y = 100 \cdot (e^{\beta} - 1) = 100 \cdot (e^{3.985 \times 0.04} - 1) \approx 17\%$.

[43] In the sample, the average number of recently arrived immigrants was 1,037 (4% of the population) and the average male nonfarm labor force was 16,074. From column 3, this translates to: $0.04 \times 0.293 \times 16,074 \approx 188$. Thus, there were roughly 188 additional union members per 1,037 new immigrants.

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 downwards, and suggests that European immigrants selected areas where unionization was growing more slowly. This might have happened because, during this period, the vast majority of labor unions actively discriminated against immigrants, precluding them from joining their ranks and the occupations they represented (as discussed in Section 2). Consistent with the historical evidence, Table A.6 shows that there is a negative and statistically significant relationship between all four measures of unionization and immigration flows. Moreover, the instrument addresses any attenuation bias caused by measurement error in the independent variable, which would lead to downward-biased OLS estimates. Finally, the identification strategy estimates a local average treatment effect (LATE) for counties that received more European immigrants because of country-of-origin networks. If these immigrants were more likely to drive an increase in unionization, this would contribute to the 2SLS estimates being larger than the OLS coefficients.

5.2 Effect on the Intensive and Extensive Margins

The results presented in Table 3 do not differentiate between an increase along the extensive or the intensive margin of unionization. In other words, they show that counties receiving larger shares of immigrants were more likely to have unions (column 1) and experienced an increase in the number of union branches and members (columns 2 to 4). However, the findings do not clarify whether this is driven solely by the establishment of unions in new areas or if immigration also strengthened organized labor in already unionized labor markets.

To explore this further, I first restrict the estimation sample to counties that had unions in every decade between 1900 and 1920. This approach allows for a more specific investigation of the results along the intensive margin. All coefficients, shown in Panel A of Table 4, are positive and statistically significant. Specifically, a five percentage point (one standard deviation) increase in the share of immigrants leads to a 33% increase in the number of branches, a 2.7 percentage point rise in union density (a 31% increase relative to the mean), and 39 more members per branch (a 35% increase relative to the mean). Similarly, to focus on the extensive margin, I re-estimate equation (1) for counties that did not consistently have unions between 1900 and 1920, with the results presented in Panel B of Table 4. Again, all coefficients are positive and statistically significant.

These two additional sets of estimates indicate that immigration positively impacted unionization along both the intensive and extensive margins. New counties saw the establishment of labor unions, while those with an existing labor movement experienced an expansion in its size.

5.3 Summary of Robustness Checks

I perform several exercises to verify the robustness of the findings. They are summarized visually in Figure 7 and Figure 8, with more details and formal estimates presented in Appendix B.

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 (Table B.2).[44] This alternative identification strategy relies on the observation that the validity of shift-share instruments can be achieved from the exogeneity of the shocks (Borusyak et al., 2022).

Next, building on Bazzi et al. (2023), I combine the shift-share instrument with a matching strategy, which selects within-state county pairs with the closest number of labor unions in 1880 or 1890 as a fraction of the county population (Table B.3).

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 (Table B.4). These include the share of the total immigrant and Black population, the share of the labor force in highly unionized industries and by skill level, the average occupational income score, the growth rate of the manufacturing output, the share of land used for farming, and the vote share for the Democratic Party in presidential elections.

Further, I show that the findings are unchanged when using alternative baseline specifications, such as not controlling for any baseline characteristics or including state by year fixed effects (Table B.5); clustering standard errors at the State Economic Area (SEA) level, using Conley (1999) standard errors to account for spatial correlation, or applying the correction for shift-share estimators proposed by Adao et al. (2019) (Table B.6); dropping potential outliers (Table B.7); using alternative definitions of the independent and dependent variables (Table B.8 and Table B.9); extending the analysis to an unbalanced sample of counties or excluding the South from the estimation sample (Table B.10); performing the analysis at the State Economic Area (SEA) level (Table B.11); and using different methods for constructing unionization data (Table B.12 and Table B.13).

I also re-estimate the preferred specification of Table 3 while interacting – one at a time – the initial shares of each immigrant group in the county, i.e., $\alpha^j_{c,1890}$ in equation (2), with year dummies (Figure B.2). This exercise is aimed at reducing the concern that combinations of counties and of immigrants from specific European countries might be

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

driving the results by absorbing most of the variation in the data (Goldsmith-Pinkham et al., 2020).[45]

Finally, I check for the absence of pre-trends by regressing the pre-period change of a set of unionization, population, and economic outcomes on the average 1900–1920 immigration predicted by the instrument (Table B.14). The fact that all the coefficients are not statistically significant indicates that, as of 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or other economic variables.

6 Mechanisms

The results shown so far indicate that counties with higher inflows of European immigration between 1890 and 1920 experienced a larger increase in unionization. In this section, I shed light on the mechanisms behind the positive effect of immigration on the emergence and growth of organized labor.

6.1 Economic Motivations

Reactions of existing workers. As described in Section 2, unions have long been concerned with labor supply expansions, fearing that an influx of new workers would lower wages, worsen working conditions, and create job scarcity. Scholars have argued that these concerns led existing workers to organize their workplaces and restrict access to the labor market (Mink, 1986; Olzak, 1989).[46] The economic threats posed by immigration heightened workers' incentives to unionize and limit immigrants' entry into the labor force. The positive effects presented in Table 3 are consistent with this hypothesis.

At the same time, since the law did not protect the right to organize, workers had to battle their employers for union recognition through violent strikes and walkouts (Fishback, 1992; Naidu and Yuchtman, 2016; Olzak, 1989), while courts often sided with employers in cases involving the dismissal of unionizing or striking employees (Foner, 1947; Taft, 1964).[47] As a result, when immigration inundated urban labor markets with inexpensive laborers seeking employment, employers' ability to break strikes and replace

[45] 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".

[46] The most common methods used by unions to control access to certain occupations included restricting membership to U.S.-born workers or American citizens, requiring union membership as a condition of employment, and regulating apprenticeship terms, thereby *de facto* determining who could acquire the skills necessary for specific jobs (Ignatiev, 1994; Mink, 1986; Stewart, 1926; Witwer, 2002).

[47] The introduction of a union or its survival was the issue in many of the violent strikes across all industries (Taft and Ross, 1969).

workers willing to unionize increased (Asher, 1982; Montgomery, 1979; Olzak, 1989).

Given these opposing forces, while unionization attempts would be expected across all occupations, as all workers perceived immigration as an economic threat (Asher, 1982; Mink, 1986; Olzak, 1989),[48] they should be more likely to succeed in occupations with barriers to entry. In these jobs, existing workers could not be *immediately* replaced by immigrants, and should have been in a stronger position to unionize and prevent immigrants from acquiring their skills, in order to reduce future job competition. Similarly, employers without a readily available pool of replacement workers should have been less likely to resist unionization efforts and more inclined to accommodate their employees' demands to form a union.

I leverage differences in the skill requirements across occupations to test whether immigration had heterogeneous effects on skilled and unskilled workers.[49] The estimates, reported in Table 5, indicate that immigration positively impacted all four unionization measures among skilled workers. Counties with higher shares of recently arrived immigrants experienced increases in the likelihood of having labor unions, the number of union branches, the share of the unionized workforce, and the number of union members per branch. By contrast, immigration had smaller – and statistically insignificant – effects on the unionization of unskilled workers.[50]

Similarly to Panel A of Table 4, Table A.4 restricts the estimation sample to a balanced set of counties that had skilled (Panel A) and unskilled (Panel B) unions in every decade between 1900 and 1920. The coefficients indicate that immigration positively affected skilled unionization also along the intensive margin, increasing the share of unionized workers in always unionized counties. Table A.5 re-estimates the analysis presented in Panel B of Table 4, to focus on the effect along the extensive margin separately by skill. All the coefficients are positive and statistically significant for skilled workers, while they

[48] The mining industry, where unions also represented unskilled workers, experienced the highest levels of violent labor unrest and mortality in strikes during this period (Fishback, 1995; Jeffreys-Jones, 1978).

[49] The classification of occupations based on their skill levels 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 in and around coal mines, I proportionally allocate its members and branches based on the distribution of skilled and unskilled workers in the coal mining industry, as reported in the 1890 U.S. Census Report on Mineral Industries (Day, 1892): 57% skilled (miners and mechanics) and 43% unskilled (laborers and boys younger than 16 years old).

[50] Table A.7 shows that the results are nearly unchanged when using an alternative definition of skilled and unskilled workers, based on whether the national union representing their occupation had established apprenticeship terms (Stewart, 1926). Additionally, adopting a broader definition of barriers to entry – encompassing both human and physical capital requirements – does not affect the findings (these results are not reported for brevity but are available upon request.)

are much smaller and, with the exception of column 4, not statistically significant for unskilled workers.

Taken together, these findings show that immigration fostered the emergence and development of labor unions in skilled occupations. Although not immediately replaceable by newcomers, skilled workers unionized to prevent immigrants from acquiring the specialized skills needed for their jobs and to discourage employers from adopting machines that would allow unskilled workers to perform their tasks at a lower cost (Asher, 1982; Mink, 1986; Olzak, 1989). These skills acted as barriers to entry, giving skilled workers an advantage in establishing new unions and expanding existing ones. In contrast, unskilled workers' attempts at unionization in response to immigration likely failed, as new, inexpensive laborers could more easily replace them (Montgomery, 1979).

Exposure to immigrant labor competition. One potential alternative explanation for the results just presented is that unions representing skilled workers were able to develop due to an absence of competition between new and existing workers, rather than in reaction to the economic threats brought by the immigrants. In Figure 9, I show suggestive evidence in contrast with this hypothesis. I report the prevailing occupations among the immigrants that entered the United States in each decade between 1890 and 1920. Both unskilled (e.g., mine laborers) and skilled (e.g., carpenters, machinists) occupations feature among the most frequent ones.

To formally estimate the effects of immigrant competition for jobs on unionization, I interact the main regressor of interest from equation (1) with a time-varying measure of a county's exposure to immigrant labor competition.[51] The measure consists of two terms. The first is given by the number of immigrant workers in each occupation o who entered the United States (net of those that settled in county c) between $t-10$ and t, as a fraction of the total immigrants in the labor force who entered the United States between $t-10$ and t. The second is a weight, represented by the share of U.S.-born workers in county c and occupation o at the beginning of that decade:[52]

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

The intuition behind this measure is simple: counties where the U.S.-born labor force (at the beginning of the decade) is concentrated in occupations heavily represented by recently arrived immigrants are more exposed to labor competition.

In Table 6, I present the results separately for skilled (Panel A) and unskilled (Panel B)

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

[52] 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 results are robust to restricting the sample to the decades starting in 1900 and 1910 only.

workers.[53] In Panel A, the uninteracted estimates are all positive and statistically significant. Remarkably, all the coefficients of the interactions are also statistically significant, indicating that counties more exposed to immigrant labor competition in skilled occupations experienced larger growth in skilled unionization. Conversely, Panel B shows negative coefficients for the interaction term among unskilled workers, indicating that increased competition hindered the growth of labor unions in this group.

In sum, these findings provide evidence that increased labor competition due to immigration had opposite effects in skilled and unskilled occupations: it fostered unionization among skilled workers while hindering union growth among the unskilled. This evidence is consistent with skilled workers forming and joining unions in response to the economic threats posed by immigration. In contrast, unskilled workers were unable to do so due to the increased availability of replacement labor, which weakened their bargaining power and limited their capacity to organize successfully.

6.2 Social Motivations

The results shown so far have examined the purely economic channels that strengthened labor unions as a consequence of immigration. However, social concerns, such as opposition to cultural change, may have provided an additional incentive for workers to organize and exclude newcomers from the labor market. This possibility is suggested by the nativist rhetoric adopted by the labor movement during this period and its strong support 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), making the net effect of these social forces ambiguous. In this section, I explore the role of these factors on the development of organized labor.

Discrimination against culturally distant immigrants. As described in Section 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

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

[53] Table A.8 presents the results for all workers, interacting the main independent variable with both measures of skilled and unskilled competition. The findings are unchanged.

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.[54] 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. I present the results in Table 7. As expected, larger increases in unionization were caused by the inflow of immigrants from Southern and Eastern Europe.[55]

Heterogeneity by attitudes towards immigration. The previous results may, however, conflate economic and cultural concerns. Immigrants from Southern and Eastern European may have had lower wage expectations and may have made coordination within unions more difficult due to higher illiteracy rates and greater linguistic differences compared to those from Northern and Western Europe. To further explore this channel, I test whether the effects were stronger in counties with worse attitudes towards immigration. In the absence of a direct measure, I use two proxies that likely reflect a county's hostility towards immigrants. The first is the historical vote share for the Know Nothing Party, a nativist political party that, in the mid-1850s, ran on an anti-Catholic and anti-Irish platform (Alsan et al., 2020). The second is the baseline level of residential segregation between U.S.-born and European immigrants.[56] Since residential segregation usually results from 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 baseline level of residential segregation (Panel B).[57] Using either proxy, the findings indicate that immigration strengthened organized labor more prominently in counties with higher resentment towards immigration.

Altogether, these results suggest that non-economic motives also contributed to the expansion of labor unions. Unionization occurred more prominently in counties that received larger shares of culturally distant immigrants, namely those from Southern and Eastern Europe. Moreover, immigration led to greater union growth in counties with less

[54] The classification of European countries follows the one made by 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.

[55] The results are unchanged when separately estimating immigration from Protestant and non-Protestant countries (Table A.9).

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

[57] 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.10.

favorable attitudes towards immigration.[58]

6.3 Ruling Out Alternative Explanations

Immigrant-driven unionization. One alternative explanation for the results is that immigrants may have joined or created labor unions at greater rates than U.S.-born workers. Although data on the individual union members are not available, I exploit the information on the local union representatives described in Section 3 to gauge the ethnic composition of their branches. Union delegates can be considered leaders of the organizations they represented, as they acted as spokespeople of their local branch at the state and national conventions, and were in charge of making decisions in the name of the members who elected them. For these reasons, their ancestry can be intended as reflecting the ethnic composition of their branch.

As a first step, I use the last names of the delegates to infer their origins, using historical de-anonymized full-count U.S. Census data.[59] Panel A of Figure A.4 shows that, as expected, most of the union leaders were U.S.-born. In Panel B, I break down the shares of delegates by ancestry. Almost all delegates had ancestry from Northern or Western Europe, while very few came from Southern or Eastern Europe.

Although the share of U.S.-born delegates increased – and that of Europeans decreased – over time at the national level, 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 proportion of leaders with U.S.-born and immigrant last names, calculated at the county level, as dependent variables in equation (1). The coefficients, presented in Panel A of Table A.13, 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 on the sample counties where at least a delegate is observed in every year – although imprecisely estimated – paint a similar picture.[60]

[58] Since the results in Section 6.1 indicate that the increase in unionization is driven by skilled workers, Table A.11 and Table A.12 re-estimate the analysis of this section only for this group. The findings are very similar to those shown in Table 7 and Table 8.

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

[60] 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.13).

These findings confirm the anecdotal and historical evidence that the observed increase in unionization of this period was not caused by a larger participation of immigrant laborers, but rather by U.S.-born workers (Mink, 1986; Taft, 1964), who maintained the control of labor unions throughout the first 20 years of the 20th century.

Previous exposure to labor unions or socialist ideologies. A second possibility is that immigrants coming from European countries that already had well-developed labor unions by the end of the 20th century, or where socialist ideas were more popular, may have brought into the United States their ideas and experience from their home country, and, in turn, contributed to the growth of unionization in their destination counties. This explanation would be in line with existing work arguing that Europeans who migrated to the United States between 1910 and 1930 promoted spillover of ideologies to U.S.-born individuals (Giuliano and Tabellini, 2022). Although the results just presented already suggest that immigrants' participation in labor unions did not increase upon Europeans' arrival, I test this hypothesis formally, estimating the effect of immigration separately for immigrants coming from countries with or without strong labor unions (Table A.14) and with higher or lower support for socialist parties (Table A.15).[61] The results rule out this possibility. The coefficients of the share of immigrants from the United Kingdom and Ireland, the only countries with a strong labor movement 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 the share of immigrants from the rest of Europe are positive and nearly always statistically significant.[62]

Other economic channels. Another possibility is that the growth in unionization was driven by differential economic expansion – or contraction – experienced by counties receiving larger shares of immigrants. In Table A.16, I show that this is not the case. Immigration had no effect on economic indicators such as the labor force participation rate 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

[61] 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.

[62] 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.

and construction workers to support new housing development).

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

7 Implications and Discussion

In this section, I discuss some implications related to the immigration-induced unionization in skilled workers' unions. Although not all the following findings can be interpreted as causal, they provide insights on the short- and medium-run trends associated with the patterns documented in this paper.

Effects on U.S.-born workers' outcomes. A question unexplored thus far is whether immigration affected the occupational distribution among U.S.-born workers. One might expect U.S.-born workers to shift towards unionized occupations as a means of protection against the perceived economic and social threats brought by immigration. I examine this possibility by testing whether immigration had a different impact depending on the presence of unions within specific occupations in each county. More specifically, I focus on occupations under the jurisdiction of AFL unions and calculate the county-level shares of U.S.-born workers in both unionized and non-unionized occupations. The results, presented in Table A.17, support the hypothesis: immigration increased the share of U.S.-born workers in unionized occupations and reduced it in non-unionized ones. While this finding aligns with historical accounts that U.S.-born workers resorted to skilled (craft) unions in response to immigration (Mink, 1986), it is also consistent with an alternative – and potentially complementary – interpretation: union representation may have developed alongside or as a result of U.S.-born workers moving into these jobs. Although data limitations prevent an analysis of the exact timing, the fact that the overall employment of U.S.-born workers did not increase across all skilled occupations – but only in those with local union representation – suggests that the observed growth in unionization is not merely a result of a general employment shift toward skilled occupations.

Additionally, consistent with evidence from both historical (Abramitzky et al., 2023; Tabellini, 2020) and recent settings (Card, 2001b, 2005, 2009; Foged and Peri, 2016; Ottaviano and Peri, 2012), immigration did not negatively affect the labor market outcomes of native workers, measured by the labor force participation rate and the (log) occupational income score (Table A.18).[63] In light of the main results of the paper, these findings suggest that labor unions may have mitigated the potential economic impact of immigration on native workers.

Persistence of unionization. Furthermore, I examine whether the effects of early 20$^{\text{th}}$-

[63]The full-count Census data of this period do not consistently report information on employment status (only in 1910), and information on wages was first collected in 1940 (Ruggles et al., 2022).

century immigration on unionization continue into the present. Using current union density measures from Macpherson and Hirsch (2023), I aggregate the data at the metropolitan-area level and employ the same shift-share instrument described in Section 4.2, averaged between 1900 and 1920, to assess the impact of historical immigration on the average share of unionized private-sector workers over the first two decades of the 21st century. Notably, metropolitan areas that experienced larger immigration waves still exhibit higher union density today (Table A.19). According to the 2SLS estimates, a four percentage point (one standard deviation) increase in historical immigration results in an almost three percentage point increase in unionized workers today, roughly 34% of the sample mean. Consistent with the early 20th-century union development, primarily in the construction sector rather than manufacturing, past immigration has a lasting positive effect on today's union density in construction (a 48% increase relative to the mean) but no significant effect in manufacturing. This suggests that the conditions fostering union growth in the early 1900s provided a lasting advantage to the labor movement that has persisted over time.

Unions and inequality. Another central economic question arising from the findings of this paper concerns the impact of unionization on inequality. Recent evidence (Farber et al., 2021) has documented a causal impact of labor unions in reducing inequality for most of the 20th century, combining national and state-level survey data on unionization from the mid-1930s onwards. To investigate this relationship, I use wage data from the 1940 U.S. Census – the first year such information was collected – to compute county-level measures of wage inequality and examine their correlation with the presence of labor unions in 1920, the last year in the sample. Following the literature (Autor et al., 2008), I measure inequality using log wage differentials at the 90th to 10th, 90th to 50th, and 50th to 10th percentiles for full-time, full-year workers.[64] The results are shown in Panel A of Table A.20, where the coefficients indicate that union presence was negatively correlated with wage inequality. In Panel B, I examine the correlation between union presence and log wage differentials between U.S.-born and immigrant workers at the 90th, 50th, and 10th percentiles. In all cases, the coefficients are positive, though not always precisely estimated. These findings, while not causal, align with existing studies showing that labor unions contribute to reducing overall wage inequality (Collins and Niemesh, 2019; Farber et al., 2021), while also raising disparities between unionized (U.S.-born) and nonunionized (immigrants).

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

8 Conclusion

Despite the enduring significance of labor unions throughout history and in contemporary society, we lack rigorous empirical evidence on the determinants of their origins and early growth. In this paper, I investigate the effects of a large labor supply increase, represented by the mass immigration of the early 20$^{\text{th}}$-century United States, on the emergence and development of organized labor. I find that immigration fostered union growth by increasing the probability that a county had unions, the number of union branches, the share of unionized workers, and their average membership. The findings are consistent with existing workers forming and joining unions for economic as well as social motivations: unions grew due to reactions to the increased labor competition brought by immigration and to concerns about cultural change.

The findings of this paper quantitatively identify immigration as a novel driver of unionization during the early days of the American labor movement. The estimates imply that in the absence of immigration, the average union density between 1900 and 1920 would have been 22% lower. They also shed light on an unexplored consequence of immigration: the strengthening of institutions that protect incumbent workers' status in the labor market, with lasting effects into the present. Notably, this study also deepens our understanding of the broader implications of immigration. It suggests that individuals' reactions to immigration are not confined to political shifts toward conservative parties or the advocacy of anti-immigration policies, as emphasized by previous research. Instead, immigration can also foster the development of self-organized institutions with broad political impact, such as labor unions.

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

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

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

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.

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.

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.

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

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

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.

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

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.

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

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.

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.

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

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

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

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.

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

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

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.

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

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

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`. Accessed: 2023-10-31.

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.

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

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.

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

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. Accessed: 2023-10-31.

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.

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

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

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.

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

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

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

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.

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.

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

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.

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

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*.

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.

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.

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

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

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

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.

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

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. https://olmsapps.dol.gov/query/orgReport.do?rptId=873493&rptForm=LM2Form.

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

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

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.

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

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.

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

Tables

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	2.59	5.33
Union Density (Members / Labor Force)	2,628	0.04	0.08
Union Members per Branch	2,628	48.12	73.61
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 Farming	876	0.45	0.19
Share of Population in Manufacturing	876	0.03	0.04
Number of Coal Mines (per 1,000 people)	876	0.14	0.46
Presence of a Railroad	876	0.95	0.22

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 (Haines, 2010), the 1890 Census of Agriculture (Haines et al., 2018), the 1890 Report on Mineral Industries (Day, 1892), and the database on the expansion of the railroad network compiled by Atack (2016).

Table 2: First Stage of the Instrumental Variable Estimation

	Dependent variable: Share of Immigrants				
	(1)	(2)	(3)	(4)	(5)
Predicted Share of Immigrants	0.336***	0.273***	0.274***	0.272***	0.271***
	(0.037)	(0.033)	(0.033)	(0.033)	(0.034)
Observations	2628	2628	2628	2628	2628
Dep. var. mean	0.028	0.028	0.028	0.028	0.028
Indep. var. mean	0.030	0.030	0.030	0.030	0.030
KP F-statistic	80.67	67.89	67.76	67.59	65.27
Share of Farming in 1890	No	Yes	Yes	Yes	Yes
Share of Manufacturing in 1890	No	No	Yes	Yes	Yes
Number of Coal Mines (per 1,000 ppl.) in 1890	No	No	No	Yes	Yes
Presence of Railroad in 1890	No	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 families in farming (from column 2); the share of population in manufacturing (from column 3), the number of coal mines per 1,000 people (from column 4), and an indicator for the presence of a railroad in the county (column 5). 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 3: The Effect of Immigration on Organized Labor

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: OLS				
Share of Immigrants	0.637*	1.780***	0.097*	165.265***
	(0.348)	(0.464)	(0.050)	(56.997)
Panel B: Reduced Form				
Predicted Share of Immigrants	0.684**	1.079***	0.079***	137.174***
	(0.285)	(0.382)	(0.029)	(41.006)
Panel C: 2SLS				
Share of Immigrants	2.526**	3.985***	0.293***	506.451***
	(1.161)	(1.518)	(0.104)	(157.544)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
Indep. var. mean	0.028	0.028	0.028	0.028
KP F-statistic	65.27	65.27	65.27	65.27

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 log of one plus 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. 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 4: The Effect of Immigration on Organized Labor – Intensive and Extensive Margin

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Intensive Margin				
Share of Immigrants		5.768*	0.539**	781.821**
		(3.301)	(0.230)	(317.146)
Observations		825	825	825
Dep. var. mean		7.250	0.087	112.711
Indep. var. mean		0.042	0.042	0.042
KP F-statistic		21.00	21.00	21.00
Panel B: Extensive Margin				
Share of Immigrants	3.349**	3.842**	0.223**	427.873**
	(1.507)	(1.622)	(0.112)	(171.012)
Observations	1,803	1,803	1,803	1,803
Dep. var. mean	0.189	0.456	0.014	18.571
Indep. var. mean	0.021	0.021	0.021	0.021
KP F-statistic	46.60	46.60	46.60	46.60

Notes: The observations are at the county-year level. In Panel A, the sample is restricted only to counties that have some union presence in every year they are observed. In Panel B, the sample is restricted only to counties that do not have union presence in every year they are observed. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus 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 (1) reflects the average number of union branches, not the log-transformed value. The instrument used to predict it is described in Section 4.2. 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. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 5: Heterogeneous Effects by Workers' Skills

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Skilled Workers				
Share of Immigrants	2.466**	3.769***	0.764***	491.027***
	(1.164)	(1.366)	(0.247)	(157.210)
Dep. var. mean	0.440	2.147	0.070	47.588
Indep. var. mean	0.028	0.028	0.028	0.028
Panel B: Unskilled Workers				
Share of Immigrants	1.175	0.785	0.010	239.538*
	(1.066)	(0.866)	(0.042)	(141.528)
Dep. var. mean	0.209	0.403	0.015	27.464
Indep. var. mean	0.028	0.028	0.028	0.028
Observations	2,628	2,628	2,628	2,628
KP F-statistic	65.27	65.27	65.27	65.27

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 log of one plus 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. 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 6: Heterogeneous Effects by Immigrant Labor Competition

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Skilled Workers				
Share of Immigrants	2.173**	3.094***	0.659***	385.534***
	(1.010)	(1.043)	(0.225)	(130.720)
Share of Immigrants x Competition	0.922**	2.302**	0.326**	334.184***
	(0.459)	(0.959)	(0.132)	(120.823)
Observations	2,624	2,624	2,624	2,624
Dep. var. mean	0.441	2.150	0.070	47.598
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	44.86	44.86	44.86	44.86
SW F-statistic (Sh. of Imm.)	79.30	79.30	79.30	79.30
SW F-statistic (Sh. of Imm. x Competition)	45.52	45.52	45.52	45.52
Panel B: Unskilled Workers				
Share of Immigrants	2.436*	1.938*	0.044	378.288**
	(1.303)	(1.061)	(0.048)	(187.052)
Share of Immigrants x Competition	-0.789***	-0.719***	-0.021	-85.828**
	(0.284)	(0.256)	(0.014)	(41.120)
Observations	2,624	2,624	2,624	2,624
Dep. var. mean	0.209	0.404	0.015	27.505
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	38.75	38.75	38.75	38.75
SW F-statistic (Sh. of Imm.)	79.65	79.65	79.65	79.65
SW F-statistic (Sh. of Imm. x Competition)	87.22	87.22	87.22	87.22

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 log of one plus 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. 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 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	3.302	5.653**	0.401**	706.499**
	(2.059)	(2.856)	(0.197)	(300.796)
Standardized coefficient	*[0.208]*	*[0.187]*	*[0.164]*	*[0.300]*
Share of N/W European Immigrants	0.370	-0.646	-0.007	-49.074
	(2.768)	(3.381)	(0.389)	(488.242)
Standardized coefficient	*[0.014]*	*[-0.013]*	*[-0.002]*	*[-0.013]*
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
Indep. var. mean (S/E Europe)	0.031	0.031	0.031	0.031
Indep. var. mean (N/W Europe)	0.019	0.019	0.019	0.019
KP F-statistic	13.16	13.16	13.16	13.16
SW F-statistic (S/E Europe)	27.81	27.81	27.81	27.81
SW F-statistic (N/W Europe)	45.44	45.44	45.44	45.44

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); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.

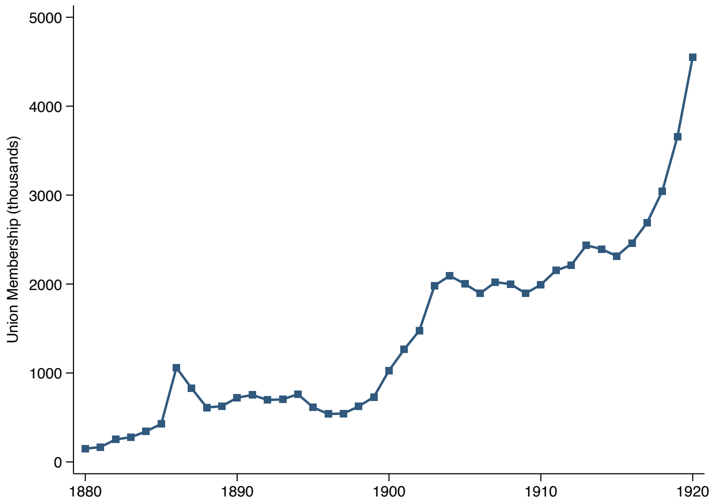
Table 8: Heterogeneous Effects by Attitudes Towards Immigration

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
	Panel A: V = Vote share Know-Nothing party			
Share of Immigrants [1]	1.796	0.268	0.068	408.596
	(1.634)	(1.898)	(0.180)	(303.896)
Share of Immigrants x High V [2]	0.851	3.910**	0.165	149.018
	(1.393)	(1.758)	(0.146)	(244.905)
Observations	2,103	2,103	2,103	2,103
Dep. var. mean	0.457	2.656	0.039	51.473
Indep. var. mean (Share of Immigrants)	0.023	0.023	0.023	0.023
KP F-statistic	23.39	23.39	23.39	23.39
SW F-statistic [1]	39.50	39.50	39.50	39.50
SW F-statistic [2]	87.40	87.40	87.40	87.40
Sum of coefficients: [1] + [2]	*2.647*	*4.177**	*0.234**	*557.614****
	1.673	*2.210*	*0.122*	*209.924*
	Panel B: V = Index of residential segregation			
Share of Immigrants [1]	0.148	0.601	-0.014	9.921
	(1.364)	(1.647)	(0.151)	(184.434)
Share of Immigrants x High V [2]	2.791**	3.896***	0.371**	584.684***
	(1.161)	(1.436)	(0.149)	(166.961)
Observations	2,565	2,565	2,565	2,565
Dep. var. mean	0.448	2.624	0.037	48.789
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	29.91	29.91	29.91	29.91
SW F-statistic [1]	45.91	45.91	45.91	45.91
SW F-statistic [2]	68.95	68.95	68.95	68.95
Sum of coefficients: [1] + [2]	*2.938***	*4.496****	*0.358****	*594.605****
	(1.187)	*(1.599)*	*(0.109)*	*(165.826)*

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); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The 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 log-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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.

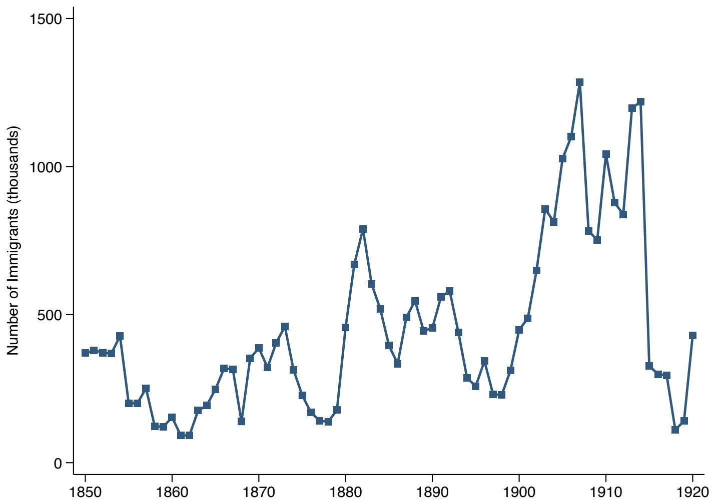
Figures

Figure 1: Estimates of Total Union Membership, 1880–1920



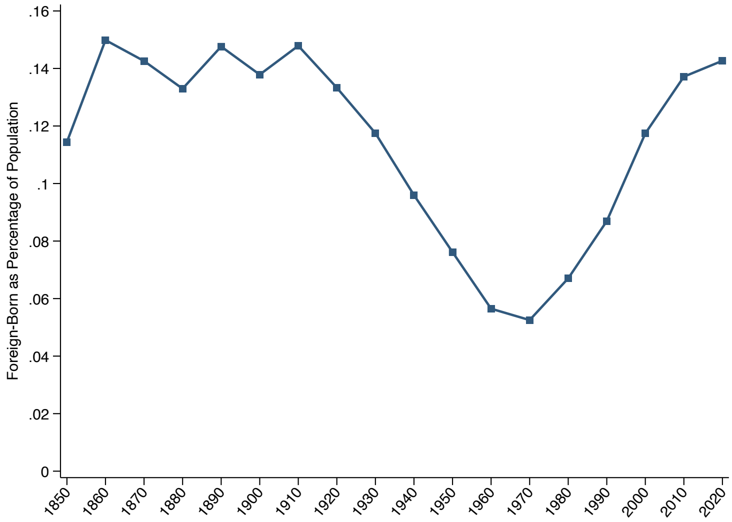
Notes: The figure shows the total number of union members in the U.S., between 1880 and 1920. Source: Freeman (1998).

Figure 2: Annual Inflow of Immigrants, 1850–1920



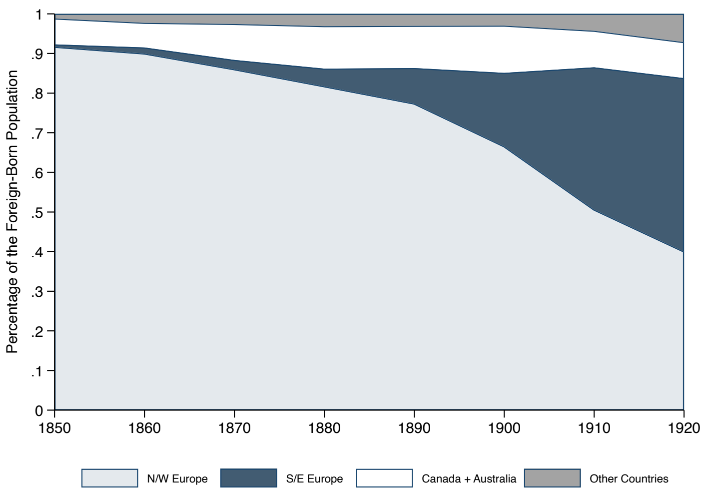
Notes: The figure shows the total number of immigrants to the United States, between 1850 and 1920. Source: Immigration Policy Institute.

Figure 3: Foreign-Born Stock as a Percentage of the U.S. Population, 1850–2020



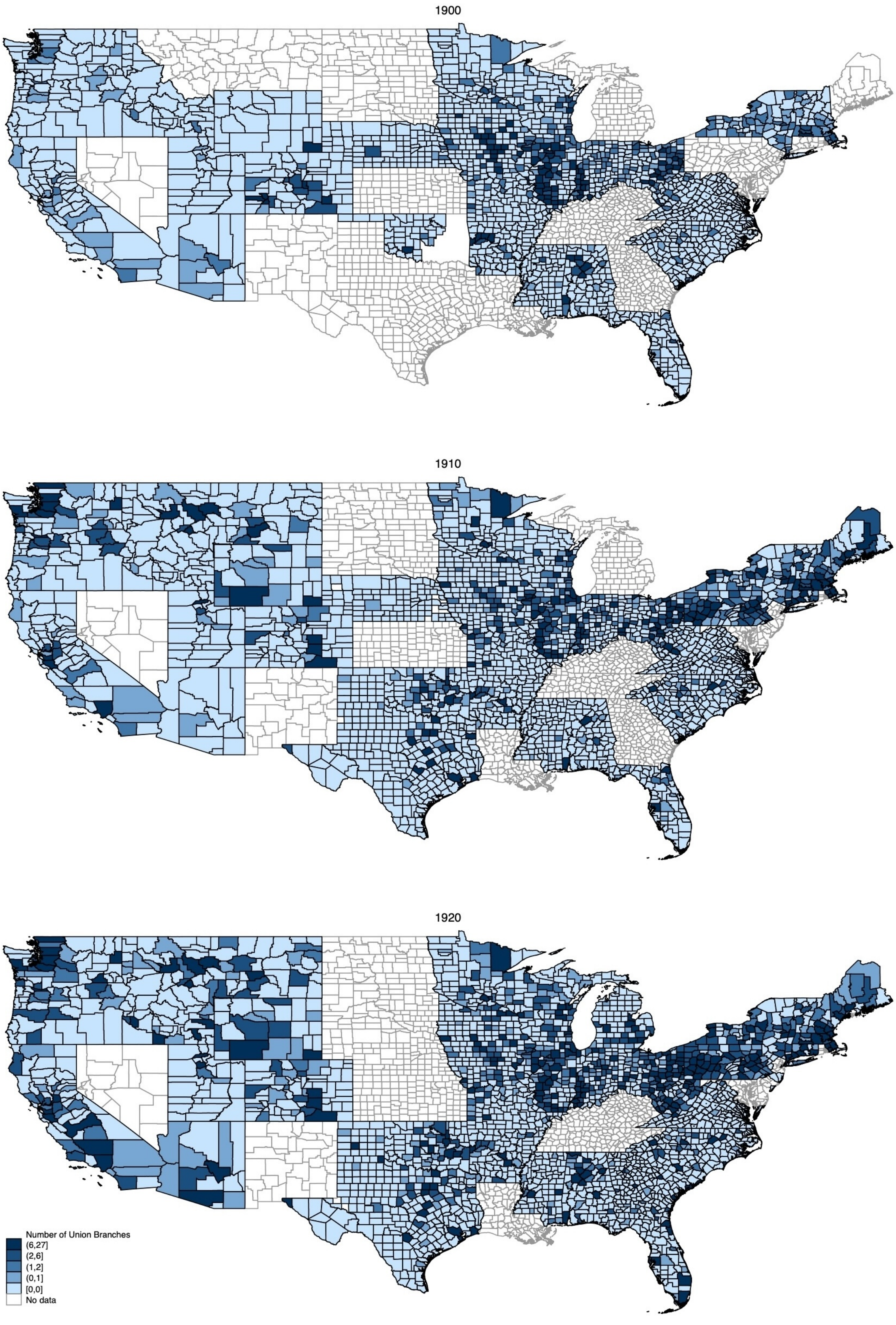
Notes: The figure shows the number of foreign-born individuals as a percent of the U.S. population, between 1850 and 2020. Source: Author's calculations from full count and samples of the U.S. Census of Population, made available by IPUMS (Ruggles et al., 2022) and ICSPR (Haines, 2010).

Figure 4: Sending Regions within the Foreign Born Population, 1850–1920



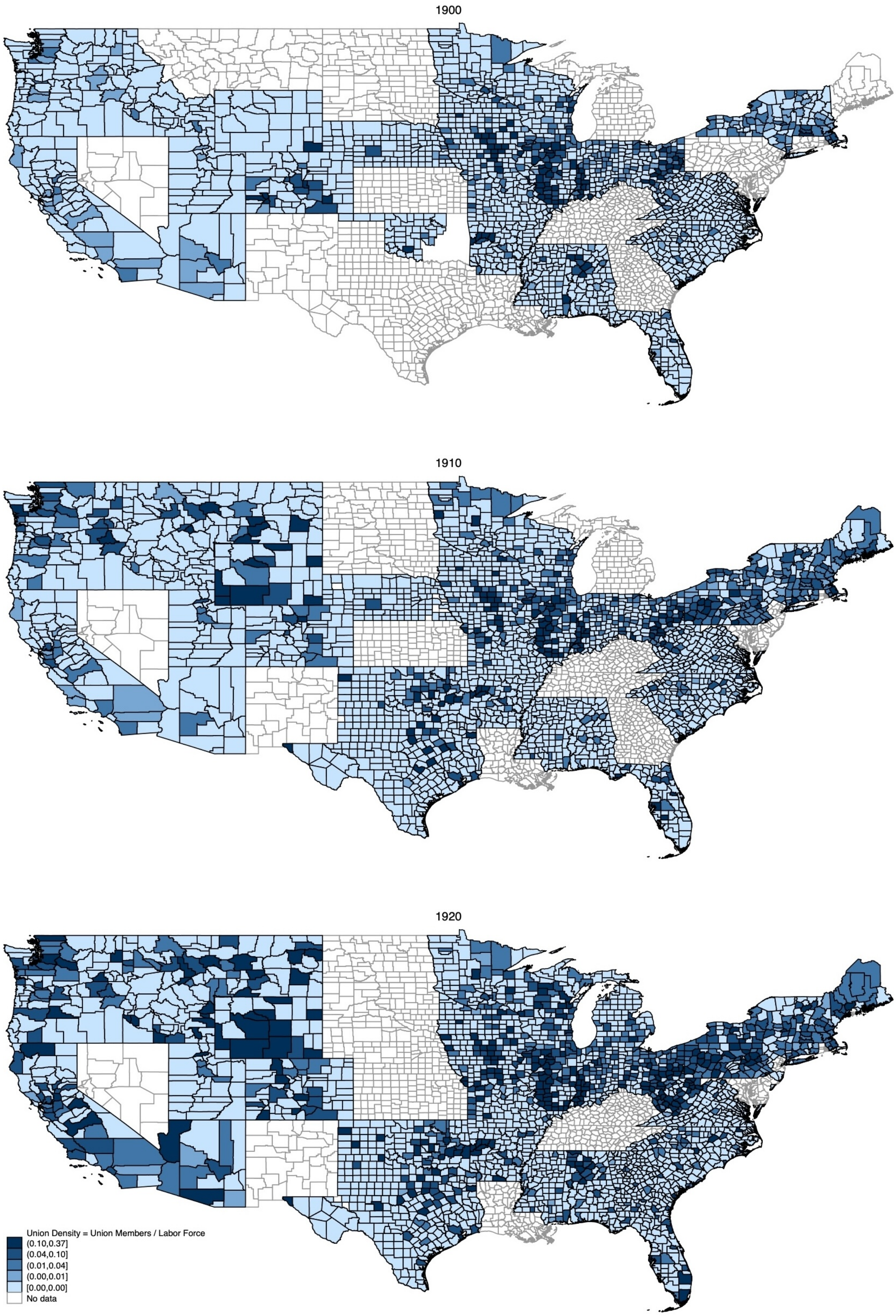
Notes: The figure shows the number of foreign-born individuals by region of origin, as a share of the total foreign-born population, between 1850 and 1920. Source: Author's calculations from full count U.S. Census of Population, made available by IPUMS (Ruggles et al., 2022) and ICSPR (Haines, 2010).

Figure 5: Geographic Distribution of Union Branches, 1900–1920



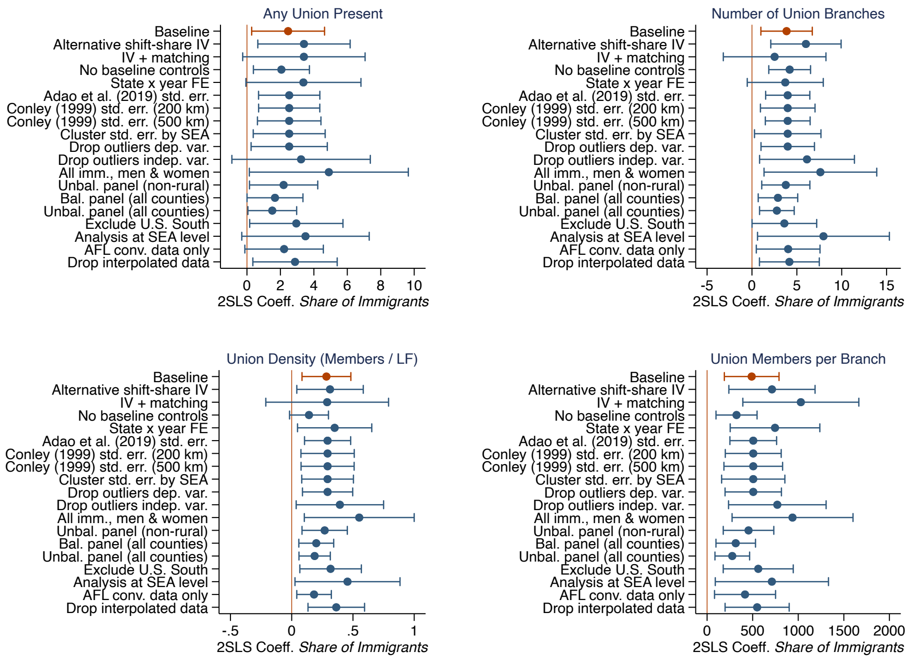
Notes: The maps plot the county-level number of union branches in 1900, 1910, and 1920. The legend shows the deciles of the distribution in 1920. Source: Author's calculations from union convention proceedings, as described in Section 3.

Figure 6: Geographic Distribution of Union Density, 1900–1920



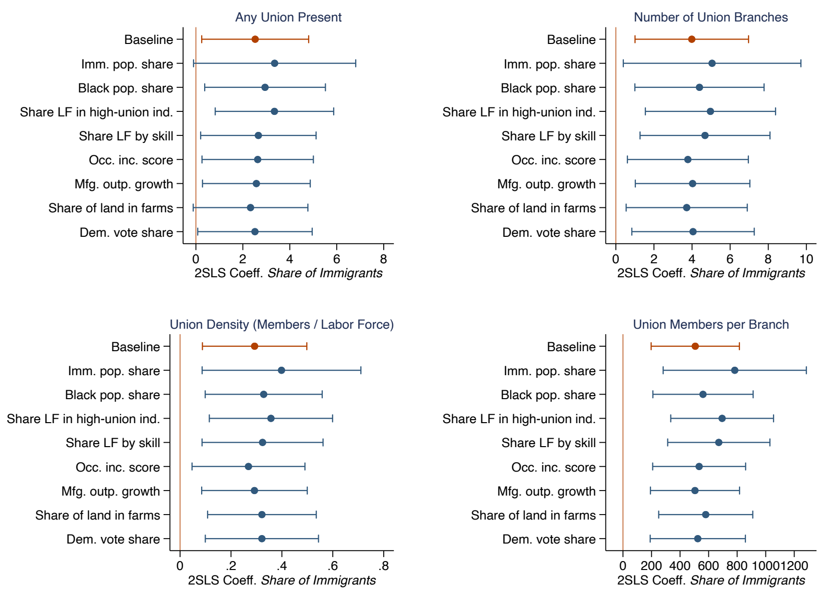
Notes: The maps plot the county-level union density (i.e., the number of union members as a fraction of the labor force, as defined in Section 3.1) in 1900, 1910, and 1920. The legend shows the deciles of the distribution in 1920. Source: Author's calculations from union convention proceedings, as described in Section 3.

Figure 7: Summary of Robustness Checks



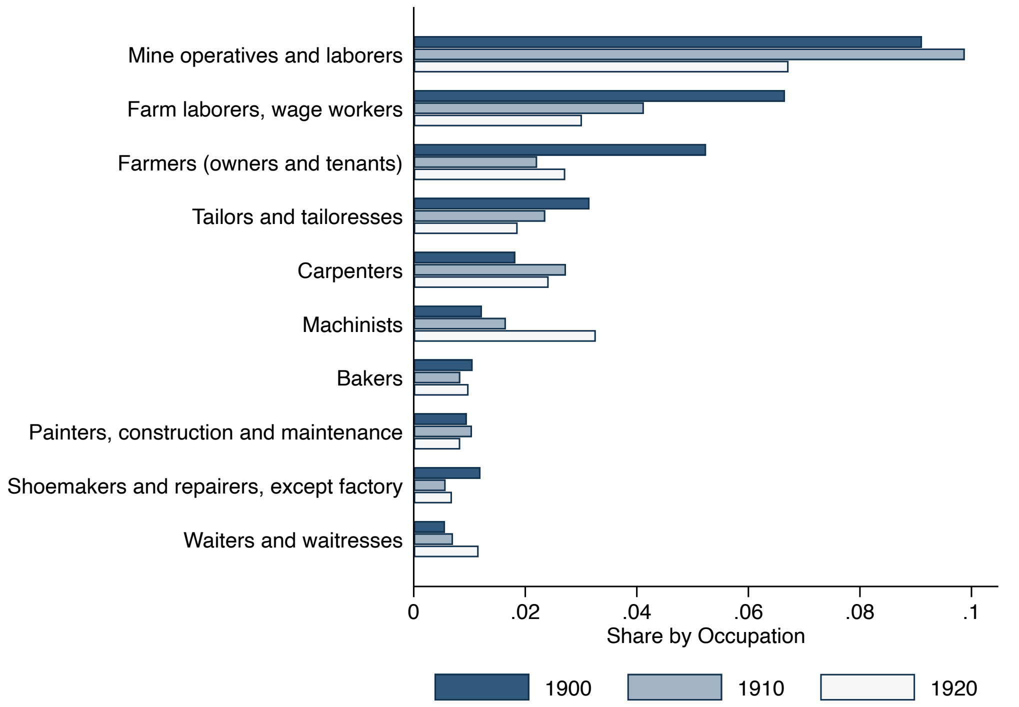
Notes: The figure presents a summary of the main robustness checks described in Section 5.3. The estimates plotted are the 2SLS coefficients (with corresponding 95% confidence intervals) of *Share of Immigrants*, the main independent variable of equation (1). The first coefficient at the top of each figure (in orange) corresponds to that from the preferred specification shown in Table 3. Standard errors are robust and clustered by county, unless indicated otherwise. For more details and formal estimates, see also Appendix B.

Figure 8: Robustness Check – Controlling for Additional Baseline Characteristics



Notes: The figure plots the 2SLS coefficients (with corresponding 95% confidence intervals) of *Share of Immigrants*, the main independent variable of equation (1), augmenting the preferred specification of Table 3 with the variable(s) indicated in each row, measured at baseline and interacted with year dummies. The first coefficient at the top of each figure (in orange) corresponds to that from the preferred specification shown in Table 3. Standard errors are robust and clustered by county. For more details, see the description of the robustness checks in Section 5.3 and the formal estimates presented in Appendix B.

Figure 9: Prevailing Occupations Among Immigrants 1900–1920



Notes: The figure shows the prevailing occupations among the European immigrants who entered the United States between 1891 and 1920. The shares indicate the number of immigrants with the reported occupation as a fraction of the total number of immigrants. Generic or undefined categories of occupations (e.g., "laborers (n.e.c.)") are not reported.

Appendix – Table of Contents

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: European Immigrant Population
Share of Families in Farming	-2.705***
	(0.404)
Share of Population in Manufacturing	15.931***
	(2.139)
Number of Coal Mines per 1,000 people	0.459***
	(0.099)
Presence of Railroad	1.775***
	(0.257)
Observations	876

Notes: The observations are at the county level for the year 1890. The dependent variable is the log of one plus the number of European immigrants living in the county. The independent variables are the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. Robust standard errors are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.3: List of Digitized Proceedings of State Federations of Labor Conventions

		Year	
State	1900	1910	1920
Alabama	-	1911	1920
Arizona	-	-	1920
Arkansas	-	1910	1922
California	-	1910	1920
Colorado	1902	1911	-
Connecticut	-	1909	-
District of Columbia	-	1910	1922
Florida	1902	1908	1920
Georgia	-	-	1920
Idaho	-	-	1920
Illinois	1901	1910	1920
Indiana	1900	1910	1920
Iowa	1902	1911	1919
Maine	-	1912	1920
Massachussetts	1901	1910	1921
Michigan	-	1910	1920
Minnesota	1900	1912	1920
Mississppi	-	1918	-
Missouri	-	1909	1921
Montana	-	1910	1921
Nebraska	-	1910	-
New Hampshire	-	1908	1920
New Jersey	-	-	1920
New York	-	1910	1920
North Carolina	-	-	1919
Ohio	1902	1910	1920
Oklahoma	-	1909	1920
Oregon	1902	1912	1919
Pennsylvania	1902	1912	1921
South Carolina	-	-	1921
South Dakota	-	-	1920
Texas	-	1910	1922
Utah	-	-	1920
Vermont	-	-	1922
Virginia	1900	1911	-
Washington	-	1910	1920
West Virginia	-	-	1920
Wisconsin	1900	1910	1920
Wyoming	-	1910	1922

Notes: The table lists the state federations of labor convention proceedings digitized for this paper. Only documents containing all the required information (i.e., delegate names or numbers, branch locations, and union names or occupations) and whose information has been used in the analysis are included. Missing years are due to incomplete records, unavailable documents in either physical or scanned version, or the state federation of labor being established at a later date. Additional details on data construction, requirements, and sources are provided in Section 3.1.

Table A.4: Heterogeneous Effects by Workers' Skills – Intensive Margin

	Dependent variable:		
	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)
Panel A: Skilled Workers			
Share of Immigrants	6.191**	1.441***	759.067**
	(3.117)	(0.520)	(311.867)
Observations	819	819	819
Dep. var. mean	6.041	0.163	112.043
Indep. var. mean	0.042	0.042	0.042
KP F-statistic	21.03	21.03	21.03
Panel B: Unskilled Workers			
Share of Immigrants	0.199	0.034	228.491
	(2.493)	(0.105)	(314.680)
Observations	276	276	276
Dep. var. mean	2.514	0.099	151.673
Indep. var. mean	0.040	0.040	0.040
KP F-statistic	26.34	26.34	26.34

Notes: The observations are at the county-year level. The sample is restricted only to counties that have some union presence in every year they are observed. The dependent variables are: the log of one plus 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 log-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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.5: Heterogeneous Effects by Workers' Skills – Extensive Margin

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Skilled Workers				
Share of Immigrants	3.309**	3.334**	0.549**	419.454**
	(1.517)	(1.436)	(0.278)	(170.855)
Observations	1,809	1,809	1,809	1,809
Dep. var. mean	0.187	0.385	0.028	18.407
Indep. var. mean	0.021	0.021	0.021	0.021
KP F-statistic	46.58	46.58	46.58	46.58
Panel B: Unskilled Workers				
Share of Immigrants	1.217	0.850	0.004	246.005*
	(1.169)	(0.902)	(0.044)	(149.382)
Observations	2,352	2,352	2,352	2,352
Dep. var. mean	0.116	0.156	0.005	12.888
Indep. var. mean	0.026	0.026	0.026	0.026
KP F-statistic	64.36	64.36	64.36	64.36

Notes: The observations are at the county-year level. The sample is restricted only to counties that do not have union presence in every year they are observed. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus 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. 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.6: Unionization and Immigration Flows

	Dependent variable: Share of Immigrants (x 100)			
	(1)	(2)	(3)	(4)
Any Union Present (t-10)	-0.645**			
	(0.264)			
Number of Union Branches (t-10)		-0.469**		
		(0.208)		
Union Density (t-10)			-3.995*	
			(2.382)	
Union Members per Branch (t-10)				-0.005**
				(0.002)
Observations	2,626	2,626	2,626	2,626
Dep. var. mean	2.274	2.274	2.274	2.274
Indep. var. mean	0.444	2.590	0.037	48.108

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 log of one plus 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 log-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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.7: Heterogeneous Effects by Workers' Skills – Based on Apprenticeship Rules

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Skilled Workers				
Share of Immigrants	2.452**	3.343***	0.402***	417.654***
	(1.096)	(1.237)	(0.142)	(129.391)
Dep. var. mean	0.364	1.683	0.030	33.204
Indep. var. mean	0.028	0.028	0.028	0.028
Panel B: Unskilled Workers				
Share of Immigrants	1.666	1.551	0.061	283.684*
	(1.112)	(1.152)	(0.093)	(166.481)
Dep. var. mean	0.221	0.795	0.034	29.537
Indep. var. mean	0.028	0.028	0.028	0.028
Observations	2,628	2,628	2,628	2,628
KP F-statistic	65.27	65.27	65.27	65.27

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 log of one plus 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. In Panel A, the dependent variables refer to skilled workers; in Panel B, to unskilled workers (based on whether the national labor union representing their occupation had established apprenticeship terms (Stewart, 1926)). 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 Immigrant Labor Competition

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants	3.167***	4.316***	0.318***	511.991***
	(1.222)	(1.389)	(0.113)	(165.052)
Share of Immigrants x Skilled Comp.	0.656	2.488**	0.149**	316.044***
	(0.523)	(1.034)	(0.069)	(119.805)
Share of Immigrants x Unskilled Comp.	-0.530**	-0.670*	-0.045	-65.491
	(0.257)	(0.358)	(0.035)	(44.727)
Observations	2,624	2,624	2,624	2,624
Dep. var. mean	0.444	2.592	0.037	48.135
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	31.99	31.99	31.99	31.99
SW F-statistic (Share of Immigrants)	83.59	83.59	83.59	83.59
SW F-statistic (Sh. of Imm. x Sk. Comp.)	42.91	42.91	42.91	42.91
SW F-statistic (Sh. of Imm. x Unsk. Comp.)	51.16	51.16	51.16	51.16

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 log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.9: Heterogeneous Effects by Religion 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 non-Protestant Immigrants	3.234	5.668**	0.395**	692.297**
	(2.034)	(2.815)	(0.191)	(291.670)
Standardized coefficient	*[0.213]*	*[0.197]*	*[0.169]*	*[0.308]*
Share of Protestant Immigrants	0.498	-0.835	0.001	-25.793
	(2.777)	(3.356)	(0.392)	(493.444)
Standardized coefficient	*[0.018]*	*[-0.016]*	*[0.000]*	*[-0.006]*
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
Indep. var. mean (non-Protestant)	0.017	0.017	0.017	0.017
Indep. var. mean (Protestant)	0.010	0.010	0.010	0.010
KP F-statistic	12.78	12.78	12.78	12.78
SW F-statistic (non-Protestant)	26.89	26.89	26.89	26.89
SW F-statistic (Protestant)	43.36	43.36	43.36	43.36

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 log of one plus 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 non-Protestant or Protestant European countries 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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 Terciles of Attitudes Towards Immigration

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
	Panel A: V = Vote share Know-Nothing party			
Share of Immigrants x Low V	1.722	0.108	0.057	429.845
	(1.637)	(1.963)	(0.182)	(311.061)
Share of Immigrants x Medium V	2.322	3.474	0.184	651.331**
	(1.738)	(2.874)	(0.124)	(282.101)
Share of Immigrants x High V	2.852	4.621*	0.265*	498.510**
	(2.002)	(2.410)	(0.142)	(208.050)
Observations	2,103	2,103	2,103	2,103
Dep. var. mean	0.457	2.656	0.039	51.473
Indep. var. mean (Share of Immigrants)	0.023	0.023	0.023	0.023
KP F-statistic	14.57	14.57	14.57	14.57
SW F-statistic (Low V)	38.48	38.48	38.48	38.48
SW F-statistic (Medium V)	41.67	41.67	41.67	41.67
SW F-statistic (High V)	41.30	41.30	41.30	41.30
	Panel B: V = Index of residential segregation			
Share of Immigrants x Low V	0.166	0.961	0.006	24.271
	(1.359)	(1.678)	(0.150)	(185.783)
Share of Immigrants x Medium V	3.072**	7.141***	0.501***	699.907***
	(1.277)	(2.094)	(0.136)	(208.550)
Share of Immigrants x High V	2.824**	2.227	0.235**	504.284***
	(1.393)	(1.410)	(0.100)	(168.767)
Observations	2,565	2,565	2,565	2,565
Dep. var. mean	0.448	2.624	0.037	48.789
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	20.09	20.09	20.09	20.09
SW F-statistic (Low V)	46.37	46.37	46.37	46.37
SW F-statistic (Medium V)	60.48	60.48	60.48	60.48
SW F-statistic (High V)	48.02	48.02	48.02	48.02

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); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The mean of the dependent variable in column (2) reflects the average number of union branches, not the log-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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.11: Heterogeneous Effects by Origin of Immigrants – Skilled Workers

	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	3.491*	5.178**	0.989**	691.989**
	(2.084)	(2.559)	(0.447)	(300.006)
Standardized coefficient	*[0.220]*	*[0.184]*	*[0.205]*	*[0.296]*
Share of N/W European Immigrants	-0.382	-0.143	0.138	-67.036
	(2.832)	(3.263)	(0.796)	(489.045)
Standardized coefficient	*[-0.015]*	*[-0.003]*	*[0.017]*	*[-0.018]*
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.440	2.147	0.070	47.588
Indep. var. mean (S/E Europe)	0.031	0.031	0.031	0.031
Indep. var. mean (N/W Europe)	0.019	0.019	0.019	0.019
KP F-statistic	13.16	13.16	13.16	13.16
SW F-statistic (S/E Europe)	27.81	27.81	27.81	27.81
SW F-statistic (N/W Europe)	45.44	45.44	45.44	45.44

Notes: The observations are at the county-year level. The dependent variables, measured for skilled workers only (see Section 6.1 for more details), are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.12: Heterogeneous Effects by Attitudes Towards Immigration – Skilled Workers

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
	Panel A: V = Vote share Know-Nothing party			
Share of Immigrants [1]	1.198	-0.384	0.348	359.035
	(1.691)	(2.077)	(0.452)	(302.657)
Share of Immigrants x High V [2]	1.370	4.314**	0.230	181.124
	(1.448)	(1.884)	(0.360)	(244.010)
Observations	2,103	2,103	2,103	2,103
Dep. var. mean	0.453	2.186	0.073	50.884
Indep. var. mean (Share of Immigrants)	0.023	0.023	0.023	0.023
KP F-statistic	23.39	23.39	23.39	23.39
SW F-statistic [1]	39.50	39.50	39.50	39.50
SW F-statistic [2]	87.40	87.40	87.40	87.40
Sum of coefficients: [1] + [2]	*2.568*	*3.930**	*0.579***	*540.159***
	(1.680)	*(2.047)*	*(0.283)*	*(210.457)*
	Panel B: V = Index of residential segregation			
Share of Immigrants [1]	0.012	1.065	0.112	-6.162
	(1.355)	(1.561)	(0.333)	(182.562)
Share of Immigrants x High V [2]	2.818**	3.017**	0.780**	583.109***
	(1.150)	(1.368)	(0.345)	(161.965)
Observations	2,565	2,565	2,565	2,565
Dep. var. mean	0.445	2.176	0.071	48.237
Indep. var. mean (Share of Immigrants)	0.028	0.028	0.028	0.028
KP F-statistic	29.91	29.91	29.91	29.91
SW F-statistic [1]	45.91	45.91	45.91	45.91
SW F-statistic [2]	68.95	68.95	68.95	68.95
Sum of coefficients: [1] + [2]	*2.830***	*4.082****	*0.892****	*576.948****
	(1.192)	*(1.440)*	*(0.264)*	*(165.009)*

Notes: The observations are at the county-year level. The dependent variables, measured for skilled workers only (see Section 6.1 for more details), are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The 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 log-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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.13: 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	2.339**	0.534	0.065	0.141	-0.136	0.038
	(1.053)	(0.369)	(0.133)	(0.280)	(0.234)	(0.175)
Dep. var. mean	0.343	0.031	0.009	0.870	0.089	0.024
Panel B: Ancestry						
Share of Immigrants		2.521**	0.505		0.029	0.108
		(1.148)	(0.373)		(0.394)	(0.384)
Dep. var. mean		0.344	0.038		0.881	0.101
Observations	2,628	2,628	2,628	582	582	582
Indep. var. mean	0.028	0.028	0.028	0.046	0.046	0.046
KP F-statistic	65.27	65.27	65.27	13.76	13.76	13.76

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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.14: 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	-3.551	-6.298	-0.997	-3,217.642
	(13.035)	(18.655)	(1.476)	(2,239.034)
Standardized coefficient	*[-0.037]*	*[-0.034]*	*[-0.067]*	*[-0.224]*
Share of Immigrants from Other Countries	2.737*	4.342**	0.338**	635.483***
	(1.540)	(2.072)	(0.148)	(238.661)
Standardized coefficient	*[0.221]*	*[0.185]*	*[0.177]*	*[0.347]*
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
Indep. var. mean (UK-Ireland)	0.003	0.003	0.003	0.003
Indep. var. mean (Other countries)	0.025	0.025	0.025	0.025
KP F-statistic	10.67	10.67	10.67	10.67
SW F-statistic (UK-Ireland)	23.23	23.23	23.23	23.23
SW F-statistic (Other Countries)	39.61	39.61	39.61	39.61

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 log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.15: 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	1.425	-0.407	0.099	-158.555
	(2.538)	(3.384)	(0.274)	(391.905)
Standardized coefficient	[0.062]	[-0.009]	[0.028]	[-0.047]
Share of Immigrants from Countries w/ Low Socialist Vote Share	3.194	6.653**	0.411*	910.279***
	(2.685)	(3.274)	(0.215)	(329.971)
Standardized coefficient	[0.169]	[0.185]	[0.141]	[0.325]
Indep. var. mean (High Socialist Vote Share)	0.003	0.003	0.003	0.003
Indep. var. mean (Low Socialist Vote Share)	0.025	0.025	0.025	0.025
KP F-statistic	14.55	14.55	14.55	14.55
SW F-statistic (High Socialist Vote Share)	27.41	27.41	27.41	27.41
SW F-statistic (Low Socialist Vote Share)	37.16	37.16	37.16	37.16
Panel B: High Vote Share = Above 10% during 1890-1919				
Share of Immigrants from Countries w/ High Socialist Vote Share	-1.522	-3.573	-0.126	-278.393
	(1.919)	(2.645)	(0.162)	(286.817)
Standardized coefficient	[-0.109]	[-0.134]	[-0.058]	[-0.134]
Share of Immigrants from Countries w/ Low Socialist Vote Share	24.177*	44.408***	2.531***	4,703.714***
	(13.878)	(16.385)	(0.842)	(1,534.743)
Standardized coefficient	[0.597]	[0.577]	[0.405]	[0.783]
Indep. var. mean (High Socialist Vote Share)	0.003	0.003	0.003	0.003
Indep. var. mean (Low Socialist Vote Share)	0.025	0.025	0.025	0.025
KP F-statistic	3.89	3.89	3.89	3.89
SW F-statistic (High Socialist Vote Share)	12.42	12.42	12.42	12.42
SW F-statistic (Low Socialist Vote Share)	7.88	7.88	7.88	7.88
Dep. var. mean	0.444	2.589	0.037	48.124
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 log of one plus 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 European countries with high or low support for socialist parties. 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), 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.16: 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	-0.031	-0.667	-0.003	0.081
	(0.100)	(1.175)	(0.015)	(0.108)
Standardized coefficient	*[-0.035]*	*[-0.055]*	*[-0.033]*	*[0.038]*
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	65.27	64.58	64.58	65.27

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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.

Table A.17: Changes to U.S.-Born Workers' Occupations

	Dependent variable: Share of U.S.-Born LF in AFL-Covered Occupations					
	With local union branch			Without local union branch		
	All	Skilled	Unskilled	All	Skilled	Unskilled
	(1)	(2)	(3)	(4)	(5)	(6)
Share Immigrants	0.941**	0.148*	0.029	-0.836**	-0.023	-0.049
	(0.446)	(0.082)	(0.037)	(0.379)	(0.080)	(0.032)
Observations	2,628	2,628	2,628	2,628	2,628	2,628
Dep. var. mean	0.149	0.021	0.007	0.110	0.076	0.021
Indep. var. mean	0.028	0.028	0.028	0.028	0.028	0.028
KP F-statistic	65.27	65.27	65.27	65.27	65.27	65.27

Notes: The observations are at the county-year level. The dependent variables are: the shares of U.S.-born workers (men ages 16–64) in the labor force who are in occupations that have positive union membership in the county (columns 1–3) or no union representation in the county (columns 4–6). All (columns 1 and 4) refers to all occupations covered by an AFL-affiliated national union; Skilled (columns 2 and 5) refers to the occupations covered by the ten largest AFL-affiliated national unions that represented skilled workers; Unskilled (columns 3 and 6) refers to the AFL-affiliated national unions that represented unskilled workers. The 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.18: Effect on U.S.-Born Workers' Labor Market Outcomes

	Dependent variable:	
	Labor Force Participation Rate	Occupational Income Score
	(1)	(2)
Share of Immigrants	-0.011	0.219
	(0.109)	(0.181)
Observations	2,628	2,628
Dep. var. mean	0.903	20.412
Indep. var. mean	0.028	0.028
KP F-statistic	65.27	65.27

Notes: The observations are at the county-year level. The dependent variables are the labor force participation rate among U.S.-born workers, men ages 16–64 (column 1), and the log of the average occupational income score among U.S.-born workers (column 2). The main regressor of interest is the number of European immigrants (men 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.19: Persistent Effect of Immigration on Unionization

	Dependent variable:		
	Union Density in Private Sector (avg. 2000–2020)		
	Total	Construction	Manufacturing
	(1)	(2)	(3)
Share of Immigrants	0.724***	1.878***	0.269
(avg. 1900–1920)	(0.123)	(0.354)	(0.265)
Observations	163	163	163
Dep. var. mean	0.066	0.157	0.102
Indep. var. mean	0.036	0.036	0.036
KP F-statistic	67.62	67.62	67.62

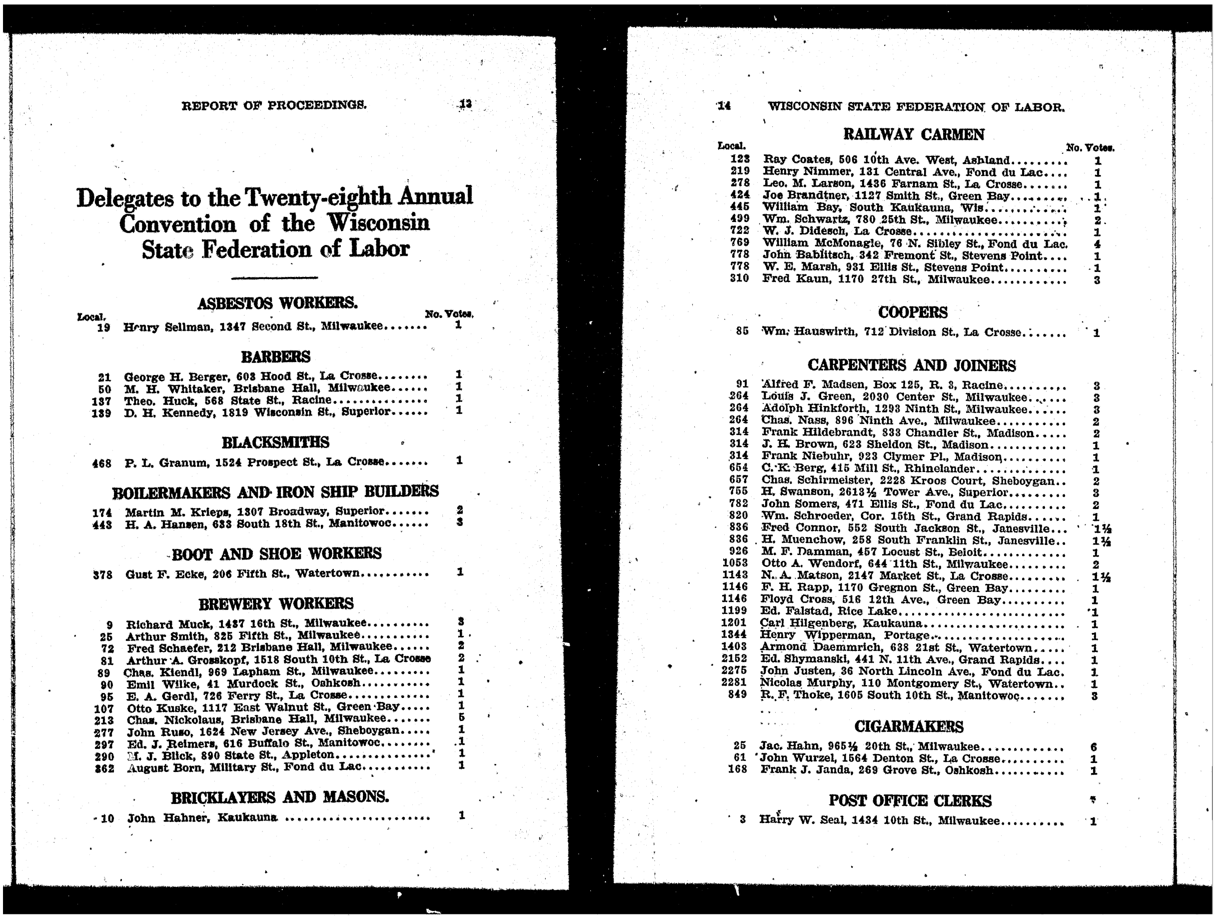
Notes: The observations are at the metropolitan area (MSA) level. The dependent variables are the average union density between 2000 and 2020 in: the whole private sector (column 1), the private construction sector (column 2), and the private manufacturing sector (column 3). The main regressor of interest is the 1900–1920 average 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 metropolitan area. The instrument used to predict it is the average between 1900 and 1920 of the one described in Section 4.2, after aggregating the data at the metropolitan area level. All regressions include the following controls, measured in 1890: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the metropolitan area. 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.

Table A.20: Labor Unions and Wage Inequality

	Dependent variable:		
	(1)	(2)	(3)
Panel A: Overall Wage Inequality			
	Log(wages at x pctile / wages at y pctile)		
	90 pctile/10 pctile	90 pctile/50 pctile	50 pctile/10 pctile
Any Union Present	-0.058***	-0.053***	-0.005
	(0.020)	(0.012)	(0.011)
Panel B: Wage Inequality Across Groups			
	Log(U.S.-born's wages/immigrants' wages) at z pctile		
	90 pctile	50 pctile	10 pctile
Any Union Present	0.026	0.091***	0.018
	(0.027)	(0.027)	(0.045)
Observations	837	837	837

Notes: The observations are at the county level. The dependent variables are measures of wage inequality in 1940 for full-time, full-year workers. In Panel A, they are log wage differentials computed at the following percentiles: 90 to 10 (column 1); 90 to 50 (column 2); and 50 to 10 (column 3). In Panel B, they are log wage differentials between U.S.-born and immigrant workers, computed at the following percentiles: 90 (column 1); 50 (column 2); and 10 (column 3). The main regressor of interest is an indicator for whether the county has any labor union in 1920. All regressions include the following controls, measured in 1890: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. Robust standard errors are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

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



REPORT OF PROCEEDINGS. 13

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

ASBESTOS WORKERS.

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

BARBERS

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

BLACKSMITHS

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

BOILERMAKERS AND IRON SHIP BUILDERS

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

BOOT AND SHOE WORKERS

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

BREWERY WORKERS

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

BRICKLAYERS AND MASONS.

10	John Hahner, Kaukauna	1

14 WISCONSIN STATE FEDERATION OF LABOR.

RAILWAY CARMEN

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

COOPERS

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

CARPENTERS AND JOINERS

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

CIGARMAKERS

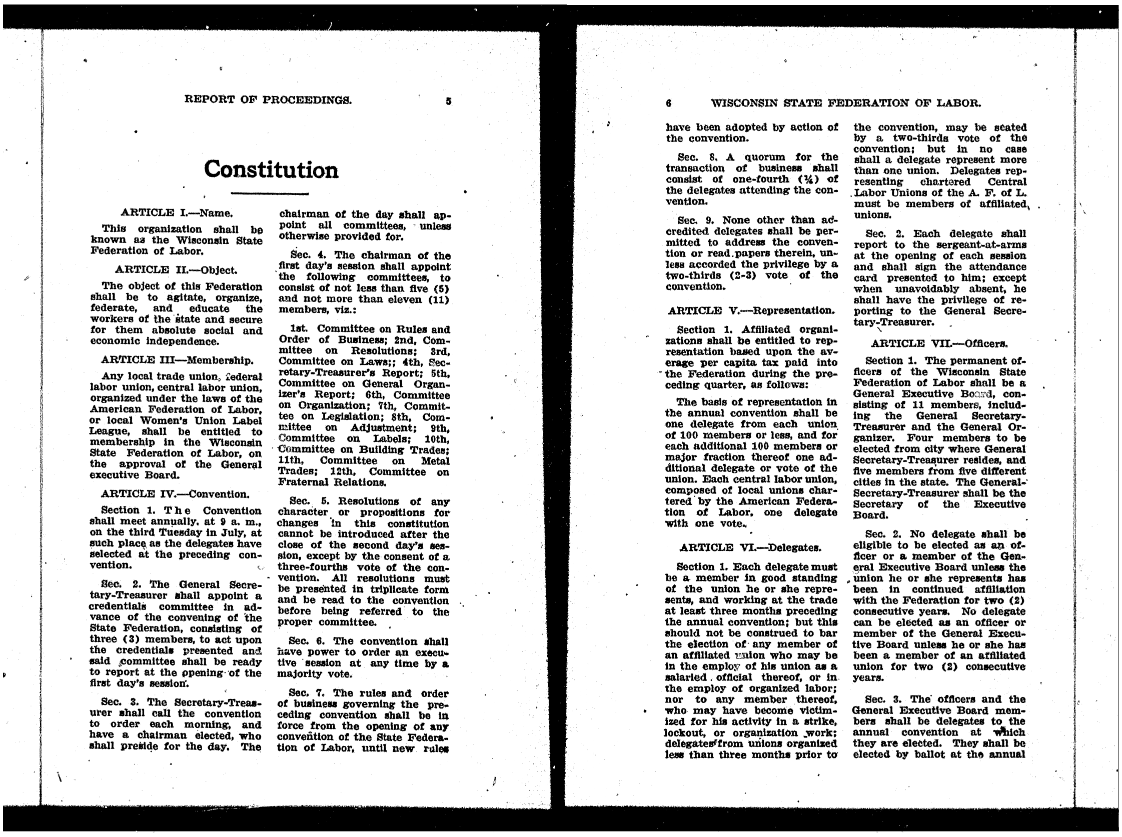
25	Jac. Hahn, 965½ 20th St., Milwaukee	6
61	John Wurzel, 1564 Denton St., La Crosse	1
168	Frank J. Janda, 269 Grove St., Oshkosh	1

POST OFFICE CLERKS

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

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

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



REPORT OF PROCEEDINGS. 5

Constitution

ARTICLE I.—Name.

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

ARTICLE II.—Object.

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

ARTICLE III—Membership.

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

ARTICLE IV.—Convention.

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

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

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

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

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

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

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

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

6 WISCONSIN STATE FEDERATION OF LABOR.

have been adopted by action of the convention.

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

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

ARTICLE V.—Representation.

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

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

ARTICLE VI.—Delegates.

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

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

ARTICLE VII.—Officers.

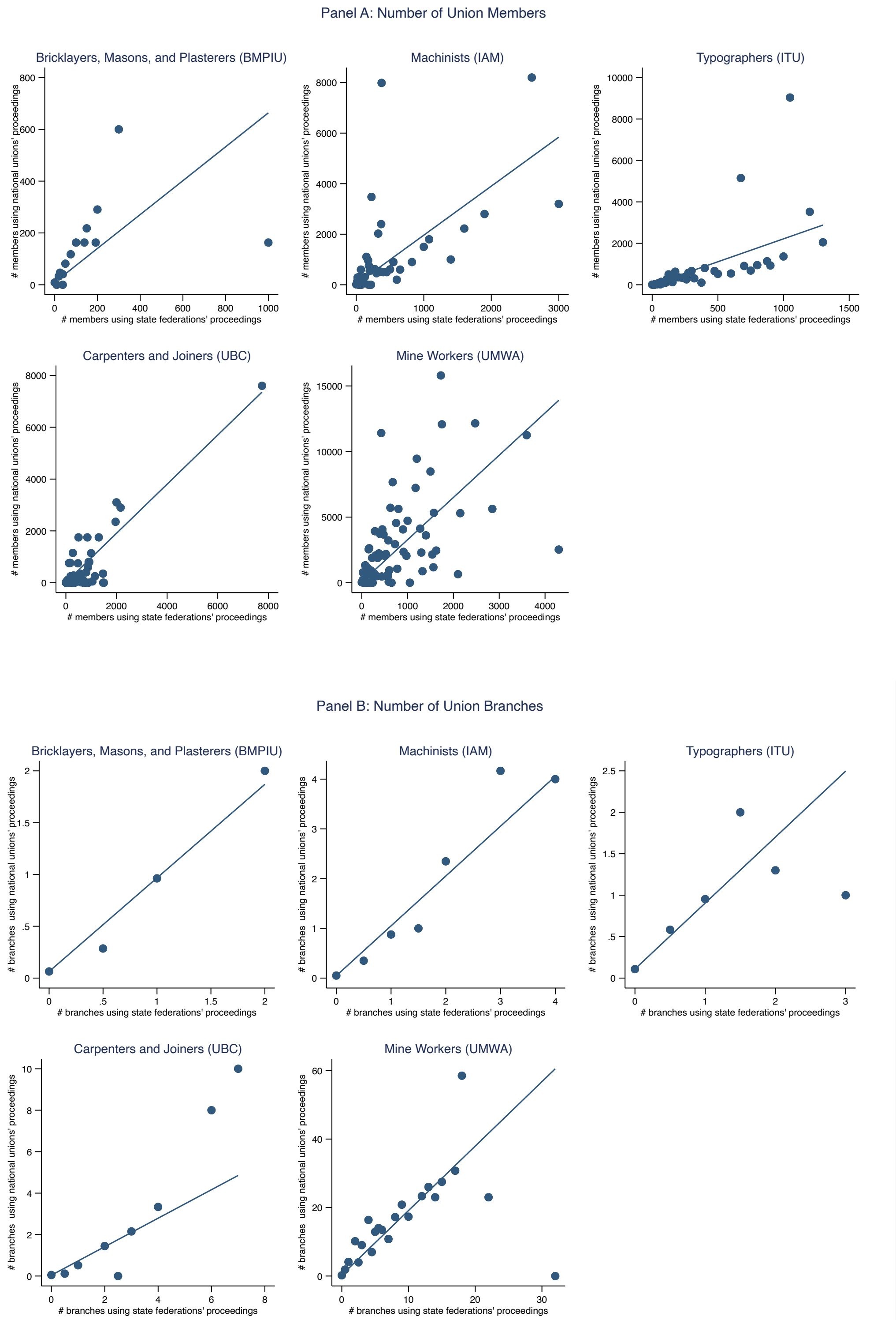
Section 1. The permanent officers of the Wisconsin State Federation of Labor shall be a General Executive Board, consisting of 11 members, including the General Secretary-Treasurer and the General Organizer. Four members to be elected from city where General Secretary-Treasurer resides, and five members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

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

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

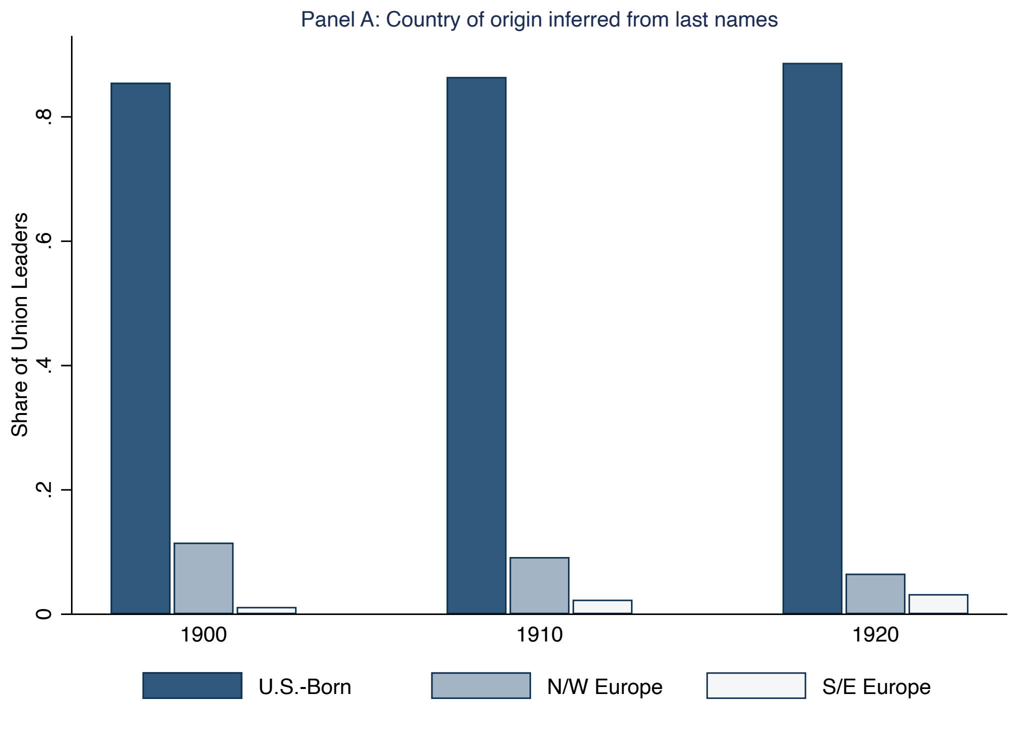
Notes: The figure shows a digitized document from the constitutions of the state federations of labor. The documents contain information on the rules that establish the number of delegates that local branches could send to the conventions. The highlighted paragraph on the page on the right provides an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

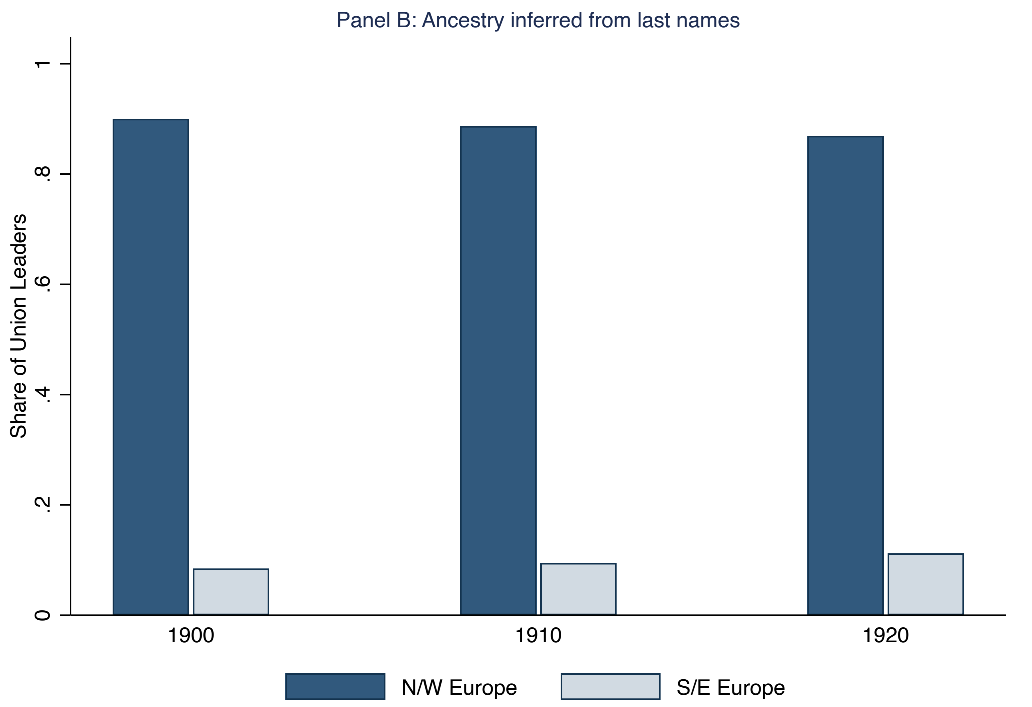
Figure A.3: Correlation Between Measures Across Data Sources



Notes: The figure shows binned scatter plots of the county-level union membership estimates (Panel A) and number of union branches (Panel B), constructed using the main data source (convention proceedings of the state federations of labor, on the x-axis) and the complementary data source (convention proceedings of the AFL-affiliated national unions, on the y-axis). Each graph shows the correlation between the two measures for each of the five national unions that are observed in both sources. See Section 3 for more details.

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





Notes: The figure plots the shares of union leaders of U.S.-born, Northern/Western Europe, and Southern/Eastern Europe origin (Panel A) and of Northern/Western and Southern/Eastern Europe ancestry (Panel B), at the beginning of each decade between 1900 and 1920. Union leaders are the delegates sent by the local union branches to the national convention of their union, or to the state conventions of the American Federation of Labor. The country of origin and the ancestry are inferred from delegates' last names, as described in Appendix 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),[65] 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.[66] For this exercise, the sample includes nineteen European countries for which immigration, weather, and crop data are available.[67] In the baseline specification, I consider temperature shocks, but results are unchanged if using precipitations.

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

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

[67] 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:

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

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. Table B.2 shows the main results on the effect of immigration on the four unionization measures. Panel A reports the baseline estimates of Table 3 using the main instrument, while Panel B displays the estimates with the alternative instrument based on weather shocks. In either case, all coefficients are positive and statistically significant at the 1% or 5% level.

B.2 Matching Exercise

Similar to Bazzi et al. (2023), I combine the main empirical strategy based on a shift-share instrumental variable with a matching exercise. I identify county pairs within the same state that have the closest number of Knights of Labor branches as a fraction of the county population, in 1880 and in 1890. In the absence of comprehensive information on unions affiliated with the American Federation of Labor before 1890 (the AFL was only established in 1886), or of complete data on the union membership of the Knights of Labor, this is one way to measure unionization at the county level before the time period analyzed in the paper.

I present the results in Table B.3. In Panel A, I re-estimate the baseline specification of Table 3 for the counties that can be included in the county-pair strategy.[68] In Panels B and C, I re-estimate equation (1), replacing the baseline controls with fixed effects for the county pairs, interacted with year dummies. In Panel B, counties are matched on the number of Knights of Labor branches per capita in 1880. In Panel C, on the one of 1890. The resulting coefficients identify the effect of immigration inflows on unionization for counties with nearly identical levels of union presence at baseline.[69] Despite the very de-

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

[69]In case of equal values of the matching variable, I further match counties on these additional variables, in the following order: total number of Knights of Labor branches in the county, share of population in manufacturing, share of population in farming, share of population in mining (for the year 1880; for the year 1890, due to data availability, this variable is replaced by the number of coal mines per population), and an indicator for the presence of a railroad in the county. 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.

manding nature of this specification, reassuringly all the point estimates remain positive, in some cases are precisely estimated, and are similar to the baseline coefficients of Panel A.

B.3 Controlling for Additional Baseline Characteristics

In this section, I address 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 unionization growth, for either economic or political reasons. In Table B.4, I augment the preferred specification 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 reported in column 1 are repeated from Table 3 for comparison.

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

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 (column 3).

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 (columns 4 and 5).

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 (columns 6 and 7).

Share of land in farms. An additional concern is represented by the structural transformation away from agriculture towards manufacturing that occurred in the U.S. between 1880 and 1920 (Eckert and Peters, 2022). This may have implied larger growth rates for counties that were more rural at the beginning of the time period, with potential implications on the evolution of labor unions too. Although in the baseline specification I already control for the share of the population in farming in 1890, I further include interactions between year dummies and the 1890 share of land in farms. The results are almost unchanged (column 8).

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 (column 9).

B.4 Additional Robustness Checks

Alternative baseline specification. Table B.5 reports results from using different specifications. In particular, in columns 1 to 5 I estimate less stringent specifications, by gradually including the controls, measured in 1890 and interacted with year dummies, that are part of the preferred specification: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. In columns 6 to 10, I do the same, while also always including state by year fixed effects, implying that the coefficients are estimated from changes in the fraction of immigrants within the same county over time, compared to other counties in the same state in a given year. The estimates are quantitatively and qualitatively unchanged.

Alternative standard errors. All the results in the paper report standard errors clustered at the county level. In Table B.6, I report standard errors using four alternative procedures: the adjustement proposed by Adao et al. (2019) for shift-share instrumental variables, the approach of Conley (1999) to account for spatial correlation (using a 200km- or 500km-bandwidth), and clustering by State Economic Area, i.e., single counties or groups of contiguous counties within the same state identified by the Census Bureau as having similar economic characteristics (Bogue, 1951). All coefficients reamain 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. In Table B.7, I re-estimate the baseline results dropping observations with measures of unionization (Panel A) and immigration (Panel B) below the 1st and above the 99th percentile. Reassuringly, in all cases the coefficients are in line with those reported in Table 3.

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 ten years as a fraction of the male working-age population. In Table B.8, I show that the results are robust when using alternative definitions of the main independent variable. In Table A, *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 Panel B, 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 Panel C, it is the number of all working-age European immigrants as a fraction of the total working-age population. In Panel D, the independent variable is the total number of European immigrants as a fraction of the total population. In all four panels, the coefficients are positive and statistically significant for all four unionization measures.

Alternative definitions of *Union Density*. The preferred definition of union density used throughout the paper is the number of union members as a fraction of the total male labor force, except farmers and farm laborers, managers and proprietors, and those working in private household service. In Table B.9, I show that the results are unchanged when using different definitions of this unionization measure. In column 1, the number of union members is divided by the total male labor force, except farmers and farm laborers. In column 2, the denominator is the total male labor force. In column 3, it is the total labor force. As expected, the size of the coefficient decreases as the denominator increases from column 1 to column 3, but all three estimates are statistically significant at the 1% or 5% level.

Alternative samples. In Table B.10, I re-estimate the preferred specification of Table 3 using alternative samples. In Panel A to C, I 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. More specifically, in Panel A, the sample is a balanced panel of both urban and rural counties. In Panel B, the sample is an unbalanced panel of only urban or mining counties. In Panel C, the sample is composed of all county-year observations for which the unionization data are available. Finally, in Panel D, 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 panels, the coefficients are positive and statistically significant.

Analysis at the SEA level. In Table B.11, I re-estimate the main results from Table 3 using data aggregated at the State Economic Area (SEA) level, a reasonable proxy for integrated labor markets at the time (similar to today's commuting zones). Despite the lower number of observations and consequently a lower F-statistic, the effects of immigration on all four measures of unionization remain positive and statistically significant at the 5% or 10% level.

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. In Table B.12, I ensure that the results are not driven by any interpolated value of unionization (in Panel A I omit such observations; in Panel B I control for an indicator equal to one if the observation is interpolated). In Table B.13, I re-estimate the analysis only relying on the onvention proceedings of the state federations of labor (withouth combining them with the proceedings from some of the largest AFL-affiliated national unions). In both tables, the results are very similar to those reported in Table 3 and always 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 and Goldsmith-Pinkham et al., 2020). Although the results of Figure 8 already reduce the concerns about this assumption being invalidated, in Table B.14, 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 1900 to 1920 average immigration flows predicted by the instrument. Similarly to the preferred specification used in the rest of the paper, all regressions include the following county controls, measured in 1880: the share of population in farming, the share of population in manufacturing, the share of population in mining, and an indicator for the presence of a railroad in the county.[70] The choice of the dependent variables is constrained by data availability. Given the absence of local data on union membership before the sample period analyzed in the paper, and the fact that the American Federation of Labor was constituted only in 1886, I measure unionization with an indicator for the presence of any branch of the Knights of Labor (column 1); for population and economic measures, I examine the share of urban population (column 2) and three measures of from the Census of Manufacturing: the number of establishments (column 3), the value of manufacturing output (column 4), and the value of output per manufacturing worker (column 5).[71] Reassuringly, no coefficient of Table B.14 is statistically significant.

[70] For the analysis reported in this table I include controls measured in 1880. The same variables measured in 1890 would constitute *bad controls* (Angrist and Pischke, 2009), since the dependent variables are changes between 1880 and 1890.

[71] The dependent variables in columns 3 to 5 are log-transformed after adding one to their value to allow

These results indicate that, before 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or in other economic variables.

for zeroes.

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

	Dependent variable: Share of Immigrants				
	(1)	(2)	(3)	(4)	(5)
Predicted Share of Immigrants	0.211***	0.167***	0.166***	0.166***	0.164***
	(0.042)	(0.035)	(0.035)	(0.035)	(0.035)
Observations	2,628	2,628	2,628	2,628	2,628
Dep. var. mean	0.028	0.028	0.028	0.028	0.028
Indep. var. mean	0.094	0.094	0.094	0.094	0.094
KP F-statistic	25.30	23.27	23.12	23.07	22.47
Share of Farming in 1890	No	Yes	Yes	Yes	Yes
Share of Manufacturing in 1890	No	No	Yes	Yes	Yes
Number of Coal Mines (per 1,000 ppl.) in 1890	No	No	No	Yes	Yes
Presence of Railroad in 1890	No	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 families in farming (from column 2); the share of population in manufacturing (from column 3), the number of coal mines per 1,000 people (from column 4), and an indicator for the presence of a railroad in the county (column 5). 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: Alternative Shift-Share Instrument Using Predicted Immigration Flows

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Main instrument				
Share of Immigrants	2.526**	3.985***	0.293***	506.451***
	(1.161)	(1.518)	(0.104)	(157.544)
KP F-statistic	65.27	65.27	65.27	65.27
Panel B: Alternative instrument				
Share of Immigrants	3.410**	6.014***	0.313**	711.691***
	(1.406)	(1.997)	(0.138)	(240.875)
KP F-statistic	22.47	22.47	22.47	22.47
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
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 log of one plus 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 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. In Panel A, the instrument used to predict immigration is the one described in Section 4.2. In Panel B, the instrument is the one that uses predicted rather than actual immigration flows (predicted using weather shocks in each European country, following Sequeira et al., 2020), as described in Appendix B.1. All regressions include county and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.3: Matching Counties with Similar Union Presence at Baseline

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Baseline (matching sample)				
Share of Immigrants	2.502**	3.926***	0.288***	500.549***
	(1.158)	(1.511)	(0.105)	(157.293)
KP F-statistic	67.79	67.79	67.79	67.79
Panel B: Matching on 1880 union presence				
Share of Immigrants	3.404*	2.525	0.290	1,028.357***
	(1.862)	(2.909)	(0.255)	(323.755)
KP F-statistic	33.17	33.17	33.17	33.17
Panel C: Matching on 1890 union presence				
Share of Immigrants	1.901	3.752	0.203	579.740*
	(2.086)	(3.205)	(0.205)	(315.750)
KP F-statistic	12.15	12.15	12.15	12.15
Observations	2,598	2,598	2,598	2,598
Dep. var. mean	0.448	2.609	0.037	48.598
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 log of one plus 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 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. In Panel A, the following controls, measured in 1890 and interacted with year dummies, are included: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. In Panel B and C, the regressions include county-pair by year fixed effects. County pairs are matched within states on the 1880 (Panel B) or 1890 (Panel C) number of Knights of Labor branches divided by county population. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by county (Panel A) or county-pair (Panel B and Panel C), are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.4: Controlling for Additional Baseline Characteristics

	Baseline	*Control: Year Dummies Interacted with Baseline Value of*							
		Immigrant Pop. Share	Black Pop. Share	Share of LF in High-Union Ind.	Share of LF by Skill Level	Occ. Income Score	Mfg. Output Growth	Share of Land in Farms	Dem. Vote Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A - Dependent variable: Any Union Present									
Share of Immigrants	2.526**	3.354*	2.946**	3.346***	2.661**	2.632**	2.576**	2.328*	2.515**
	(1.161)	(1.760)	(1.310)	(1.286)	(1.252)	(1.209)	(1.169)	(1.244)	(1.242)
Panel B - Dependent variable: Number of Union Branches									
Share of Immigrants	3.985***	5.052**	4.389**	4.965***	4.680***	3.780**	4.028***	3.719**	4.051**
	(1.518)	(2.375)	(1.728)	(1.741)	(1.738)	(1.617)	(1.531)	(1.620)	(1.637)
Panel C - Dependent variable: Union Density (Members / Labor Force)									
Share of Immigrants	0.293***	0.398**	0.329***	0.357***	0.324***	0.269**	0.292***	0.322***	0.322***
	(0.104)	(0.159)	(0.117)	(0.123)	(0.121)	(0.113)	(0.106)	(0.109)	(0.113)
Panel D - Dependent variable: Members per Branch									
Share of Immigrants	506.451***	783.186***	560.617***	694.683***	671.271***	533.596***	504.483***	579.758***	524.107***
	(157.544)	(255.644)	(178.960)	(183.510)	(182.506)	(165.861)	(159.052)	(167.935)	(169.963)
Observations	2,628	2,628	2,628	2,616	2,616	2,616	2,592	2,628	2,571
KP F-statistic	65.27	42.42	58.93	57.10	56.59	59.85	64.77	60.51	60.69

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (Panel A); the log of one plus the number of union branches (Panel B); union density, defined as the number of union members divided by the total male nonfarm labor force (Panel C); and the number of members per branch, or zero if the county has no union branch (Panel D). 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. In addition, the following controls, interacted with year dummies, are included: the immigrant share of the population in 1890 (column 2); the Black share of the population in 1890 (column 3); the shares of the male labor force in the mining, construction, and transportation industries in 1880 (column 4); the share of the male labor force in skilled occupations in 1880 (column 5); the log of the average occupational income score in 1880 (column 6); the growth rate of manufacturing output between 1880 and 1890 (column 7); the share of land in farms in 1890 (column 8); and the average vote share for the Democratic Party in the presidential elections of 1888 and 1892 (column 9). 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.5: Alternative Baseline Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A - Dependent variable: Any Union Present									
Share of Immigrants	2.060**	2.495**	2.505**	2.570**	2.526**	2.576*	3.266*	3.417*	3.416*	3.376*
	(0.853)	(1.145)	(1.138)	(1.137)	(1.161)	(1.328)	(1.743)	(1.759)	(1.764)	(1.753)
	Panel B - Dependent variable: Number of Union Branches									
Share of Immigrants	4.201***	3.716**	3.900***	4.010***	3.985***	4.021**	3.620*	3.631*	3.605*	3.711*
	(1.185)	(1.490)	(1.502)	(1.502)	(1.518)	(1.683)	(2.107)	(2.153)	(2.151)	(2.159)
	Panel C - Dependent variable: Union Density (Members / Labor Force)									
Share of Immigrants	0.141*	0.244**	0.263***	0.283***	0.293***	0.202*	0.341**	0.350**	0.345**	0.351**
	(0.081)	(0.101)	(0.101)	(0.100)	(0.104)	(0.116)	(0.149)	(0.151)	(0.151)	(0.154)
	Panel D - Dependent variable: Members per Branch									
Share of Immigrants	322.985***	456.123***	474.620***	474.860***	506.451***	491.773***	699.741***	736.856***	732.901***	745.228***
	(114.712)	(149.469)	(150.124)	(151.028)	(157.544)	(184.463)	(242.184)	(247.350)	(247.423)	(250.502)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Year FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Share of Farming in 1890	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Share of Manufacturing in 1890	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Number of Coal Mines (per 1,000 ppl.) in 1890	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Presence of Railroad in 1890	No	No	No	No	Yes	No	No	No	No	Yes
Observations	2,628	2,628	2,628	2,628	2,628	2,628	2,628	2,628	2,628	2,628
KP F-statistic	80.67	67.89	67.76	67.59	65.27	37.23	30.55	30.43	30.34	31.31

Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (Panel A); the log of one plus the number of union branches (Panel B); union density, defined as the number of union members divided by the total male nonfarm labor force (Panel C); and the number of members per branch, or zero if the county has no union branch (Panel D). 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 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. Columns 6 to 10 also include state by year fixed effects. In addition, the following controls, measured in 1890 and interacted with year dummies, are included: the share of families in farming (columns 2 to 5 and 7 to 10), the share of population in manufacturing (columns 3 to 5 and 8 to 10), the number of coal mines per 1,000 people (columns 4 and 5, and 9 and 10), and an indicator for the presence of a railroad in the county (columns 5 and 10). 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.6: Alternative Standard Errors

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants	2.526	3.985	0.293	506.451
Standard errors:				
Adao et al. (2019) adjustement	(0.935)	(1.257)	(0.096)	(130.697)
Conley (1999), 200km bandwidth	(0.933)	(1.552)	(0.111)	(156.421)
Conley (1999), 500km bandwidth	(0.968)	(1.272)	(0.111)	(164.133)
Clustered by State Economic Area	(1.092)	(1.880)	(0.107)	(176.208)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
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 log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. The following standard errors are shown in parentheses: standard errors based on the Conley (1999) approach to account for spatial correlation, with a bandwidth of 200km or 500km; standard errors using the correction proposed by Adao et al. (2019) for shift-share estimators; and standard errors clustered by State Economic Area (SEA).

Table B.7: Dropping Outliers

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Outliers of dependent variable				
Share of Immigrants	2.526**	3.985***	0.293***	506.451***
	(1.161)	(1.518)	(0.104)	(157.544)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.444	2.589	0.037	48.124
Indep. var. mean	0.028	0.028	0.028	0.028
KP F-stat	65.27	65.27	65.27	65.27
Panel B: Outliers of independent variable				
Share of Immigrants	3.238	6.133**	0.393**	770.402***
	(2.109)	(2.692)	(0.182)	(272.663)
Observations	2,592	2,592	2,592	2,592
Dep. var. mean	0.441	2.544	0.037	47.806
Indep. var. mean	0.025	0.025	0.025	0.025
KP F-statistic	52.75	52.75	52.75	52.75

Notes: The observations are at the county-year level. The observations below the 1st or above the 99th percentile of the dependent variable (Panel A), or of the independent variable (Panel B), are excluded from the sample. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.8: Alternative Definitions of *Share of Immigrants*

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Share of Immigrants =				
# Working-age immigrants / Working-age population, men only				
Share of Immigrants	4.186*	6.604**	0.485**	839.202***
	(2.212)	(2.895)	(0.189)	(306.470)
Indep. var. mean	0.113	0.113	0.113	0.113
KP F-statistic	21.06	21.06	21.06	21.06
Panel B: Share of Immigrants =				
# Working-age recent (< 10 years) immigrants / Working-age population, men and women				
Share of Immigrants	2.937**	4.574**	0.331**	562.335***
	(1.355)	(1.805)	(0.132)	(190.046)
Indep. var. mean	0.023	0.023	0.023	0.023
KP F-statistic	53.89	53.89	53.89	53.89
Panel C: Share of Immigrants =				
# Working-age immigrants / Working-age population, men and women				
Share of Immigrants	4.357*	6.786**	0.491**	834.281**
	(2.300)	(3.055)	(0.213)	(324.286)
Indep. var. mean	0.102	0.102	0.102	0.102
KP F-statistic	20.37	20.37	20.37	20.37
Panel D: Share of Immigrants =				
# Immigrants / Total population, men and women				
Share of Immigrants	4.895**	7.624**	0.551**	937.356***
	(2.422)	(3.199)	(0.228)	(338.008)
Indep. var. mean	0.078	0.078	0.078	0.078
KP F-statistic	31.57	31.57	31.57	31.57
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 log of one plus 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 regressor of interest is the share of European immigrants, defined in the title of each panel. Working-age refers to individuals ages 16–64. 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.9: Alternative Definitions of *Union Density*

	Dependent variable: Union Density = # Members as a Fraction of		
	Total Male Labor Force, Except Farming	Total Male Labor Force	Total Labor Force
	(1)	(2)	(3)
Share of Immigrants	0.289**	0.211***	0.180***
	(0.118)	(0.081)	(0.069)
Observations	2,628	2,628	2,628
Dep. var. mean	0.033	0.022	0.018
Indep. var. mean	0.028	0.028	0.028
KP F-statistic	65.27	65.27	65.27

Notes: The observations are at the county-year level. The dependent variables consist of three alternative definitions of union density, different from the one used in the rest of the paper (i.e., the number of union members as a fraction of the total male labor force, except farmers and farm laborers, managers and proprietors, and those working in private household services): the number of union members divided by the total male labor force, except farmers and farm laborers (column 1); the number of union members divided by the total male labor force (column 2); and the number of union members divided by the total (male and female) labor force (column 3). 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.10: Alternative Samples

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Balanced panel, urban and rural counties				
Share of Immigrants	1.676**	2.896***	0.200***	314.372***
	(0.852)	(1.122)	(0.073)	(110.516)
Observations	4,958	4,958	4,958	4,958
KP F-statistic	30.44	30.44	30.44	30.44
Panel B: Unbalanced panel, urban counties				
Share of Immigrants	2.193**	3.772***	0.269***	454.612***
	(1.040)	(1.367)	(0.094)	(141.243)
Observations	2,908	2,908	2,908	2,908
KP F-statistic	75.17	75.17	75.17	75.17
Panel C: Unbalanced panel, urban and rural counties				
Share of Immigrants	1.514**	2.775***	0.187***	276.303***
	(0.743)	(0.985)	(0.065)	(96.863)
Observations	5,738	5,738	5,738	5,738
KP F-statistic	38.25	38.25	38.25	38.25
Panel D: Excluding the South				
Share of Immigrants	2.951**	3.609**	0.318**	561.132***
	(1.420)	(1.834)	(0.128)	(195.605)
Observations	1,953	1,953	1,953	1,953
KP F-statistic	53.33	53.33	53.33	53.33

Notes: The observations are at the county-year level. In Panel A, the sample is restricted to a balanced panel of both urban and rural counties; in Panel B, to an unbalanced panel of urban counties, i.e., with a non-zero urban population in 1890; in Panel C, to an unbalanced panel of both urban and rural counties; and in Panel D the states in the South are excluded from the sample. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.11: Analysis at the State Economic Area (SEA) Level

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants	3.500*	7.967**	0.456**	710.853**
	(1.935)	(3.731)	(0.218)	(315.250)
Observations	765	765	765	765
Dep. var. mean	0.754	8.663	0.040	87.513
Indep. var. mean	0.035	0.035	0.035	0.035
KP F-statistic	18.82	18.82	18.82	18.82

Notes: The observations are at the State Economic Area (SEA)-year level. The dependent variables are: an indicator for whether the SEA has any labor union (column 1); the log of one plus 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 SEA 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 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 SEA. The instrument used to predict it is described in Section 4.2. All regressions include SEA and year fixed effects, and the following controls, measured in 1890 and interacted with year dummies: the share of families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the SEA. KP F-statistic refers to the Kleibergen-Paap F-statistic for weak instruments. Standard errors, robust and clustered by SEA, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.12: Omitting or Controlling for Interpolated Observations

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Panel A: Omit observations with interpolated union data				
Share of Immigrants	2.874**	4.170**	0.364***	549.105***
	(1.285)	(1.699)	(0.118)	(179.077)
KP F-statistic	57.07	57.07	57.07	57.07
Observations	2,426	2,426	2,426	2,426
Panel B: Control for whether observation has interpolated union data				
Share of Immigrants	2.077*	2.980**	0.223**	443.409***
	(1.149)	(1.502)	(0.105)	(155.956)
KP F-statistic	65.77	65.77	65.77	65.77
Observations	2,628	2,628	2,628	2,628

Notes: The observations are at the county-year level. In Panel A, the observations where the union data are interpolated (as described in Section 3.1) are excluded from the sample; in Panel B, the sample is the same as in Table 3 and an indicator for whether the observation is interpolated is added among the controls. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus 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 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.13: Using Data Only From the State Convention Proceedings of the AFL

	Dependent variable:			
	Any Union Present	Number of Union Branches	Union Density (Members / LF)	Union Members per Branch
	(1)	(2)	(3)	(4)
Share of Immigrants	2.221*	4.033**	0.182**	416.133**
	(1.196)	(1.809)	(0.072)	(170.581)
Observations	2,628	2,628	2,628	2,628
Dep. var. mean	0.274	0.425	0.014	28.617
Indep. var. mean	0.028	0.028	0.028	0.028
KP F-statistic	65.27	65.27	65.27	65.27

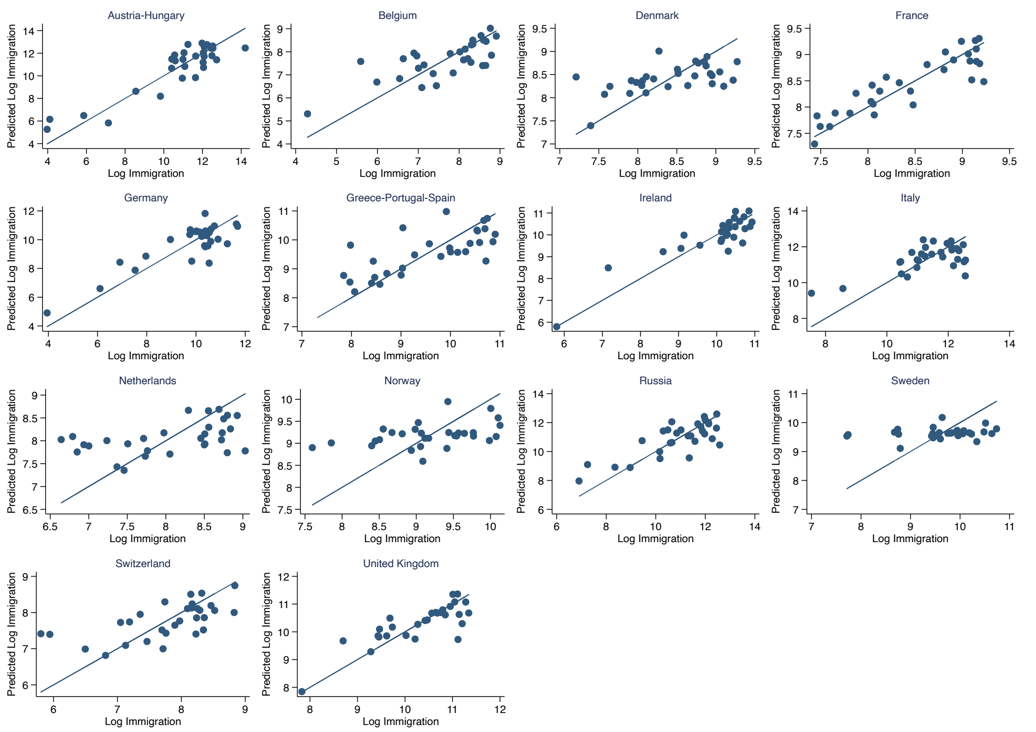
Notes: The observations are at the county-year level. The dependent variables are: an indicator for whether the county has any labor union (column 1); the log of one plus the number of union branches (column 2); union density, defined as the number of union members divided by the total male nonfarm labor 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 dependent variables are constructed using data from the convention proceedings of the state federations of labor (the state-level subordinate bodies of the AFL) only, without combining them with the convention proceedings of the AFL-affiliated national unions as in the rest of the paper (see Section 3.1 for details on the data sources and construction). 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 families in farming, the share of population in manufacturing, the number of coal mines per 1,000 people, and an indicator for the presence of a railroad in the county. 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.14: Test of Pre-Trends

	Dependent variable (1880-1890 difference):				
	Any Union Present	Share of Urban Pop.	Nr. of Mfg. Establishments	Total Mfg. Output	Mfg. Output per Worker
	(1)	(3)	(4)	(5)	(6)
Share of Immigrants	-0.582	0.083	0.822	-0.524	-1.920
(average 1900–1920)	(0.909)	(0.319)	(1.683)	(3.016)	(1.306)
Observations	871	871	871	871	871
KP F-statistic	111.11	111.11	111.11	111.11	111.11

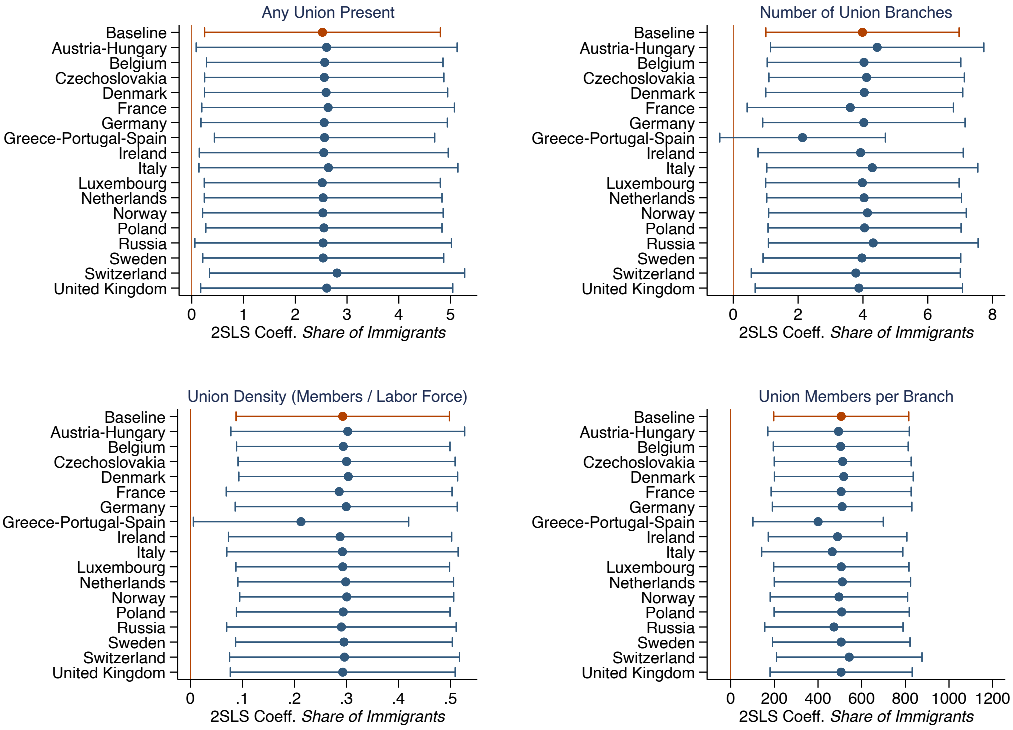
Notes: The observations are at the county level. The dependent variables are the 1890–1880 difference in: the presence of any branch of the Knights of Labor (column 1); the share of urban population (column 2); the log of one plus the number of manufacturing establishments (column 3); the log of one plus the value of manufacturing output (column 4); and the log of one plus the value of manufacturing output 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. All regressions include the following controls, measured in 1880: the share of population in farming, the share of population in manufacturing, the share of population in mining, and an indicator for the presence of a railroad in the county. 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*, 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 Mapping Delegates' Last Names to Origins and Ancestry

In Section 6, I use union delegates' last names to infer their ethnic origins. In this section, I describe how this mapping is constructed.

I start with de-anonymized full count U.S. Census data between 1900 and 1920, which contain information on names and birthplaces of the whole U.S. population. I then restrict the sample to the male population, and classify individuals depending on their country of birth and their ancestry, defined as their country of birth if born abroad, or the country of birth of the father if born in the 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.[72]

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.[73] 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).[74] 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.

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

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

[74] 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.15, 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.[75]

[75]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.[76]

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". Although, as discussed in Section 4 and Section 5.3, the identification strategy of this paper is more in line with assumptions on the exogeneity of the shocks (Borusyak et al., 2022) rather than the shares,[77] 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.[78] 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 correlation between the weights and the immigration flows is 0.917, which implies that the variation in the shocks explains approximately 84% ($0.917^2 = 0.841$) of the variation in the Rotemberg weights. This confirms that the identifying variation of the overall shift-share instrument is primarily driven by the plausibly exogenous shocks (immigration flows) rather than the shares (Borusyak et al., 2022). 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

[76] 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.

[77] See also Table B.2, which reports results using an alternative instrument built on purely exogenous, weather-predicted, immigration flows from European countries to the United States.

[78] 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.

Table 7 and Table A.9). It is also worth noting that none of these weights are particularly large, with the top weight being only 27%.[79] 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 almost the entire identifying variation (92.3%), and the unweighted mean of the estimates are remarkably similar across positively and negatively weighted $\widehat{\beta}_j$'s.

G.2 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 larger Rotemberg weight also display a larger F-statistic, while the complete opposite is true for most of the ones with a very low (or even negative) weight.

G.3 Test of Pre-Trends for Top Rotemberg Weight Instruments

In Table G.2, I conduct a test of pre-trends in the spirit of what recommended by Goldsmith-Pinkham et al. (2020). More specifically, I re-estimate the test reported in Table B.14, but predicing the share of immigrants only using each of the top five Rotemberg weight instruments (one per panel). The results show, with only a hanfdul of exceptions, that the coefficients are not statistically significant. This indicates that, even just exploiting the variation coming from the countries with the highest Rotemberg weights, counties that received more immigration between 1900 and 1920 were not trending differently between 1880 and 1890. Moreover, it is reassuring to see that none of the panels report pre-trends in column 1, where the dependent variable is the change in union presence (proxied by the presence of branches of the Knights of Labor), and that the largest evidence of pre-trends in the economic meaures (columns 4 to 6) is visible only when using the instrument with the lowest Rotemberg weight among the top five (Sweden).

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

Table G.1: Summary of Rotemberg Weights

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.063	-0.016	0.056
Positive	1.063	0.082	0.944

Panel B: Correlations

	$\widehat{\alpha}_j$	g_j	$\widehat{\beta}_j$	$\widehat{F}_j$	var(z_j)
α_k	1				
g_k	0.917	1			
β_k	0.125	-0.070	1		
F_k	0.107	0.115	0.070	1	
Var(z_k)	-0.222	-0.391	0.461	-0.384	1

Panel C: Variation across years in α_k

	Sum	Mean
1900	0.012	0.001
1910	1.702	0.100
1920	-0.714	-0.042

Panel D: Top five Rotemberg weight countries of origin

	$\widehat{\alpha}_j$	g_j	$\widehat{\beta}_j$	95% CI
Austria-Hungary	0.270	7.55e+05	0.341	(-0.900,2.000)
Italy	0.257	7.38e+05	-1.396	(-3.600,-0.400)
Russia	0.223	6.64e+05	4.585	(-0.200,2.800)
Greece-Portugal-Spain	0.100	1.06e+05	12.617	(-2.000,4.100)
Sweden	0.065	1.20e+05	1.245	(-0.700,2.200)

Panel E: Estimates of β_j for positive and negative weights

	α-weighted sum	Share of overall β	Mean
Negative	0.180	0.070	2.008
Positive	2.375	0.930	1.574

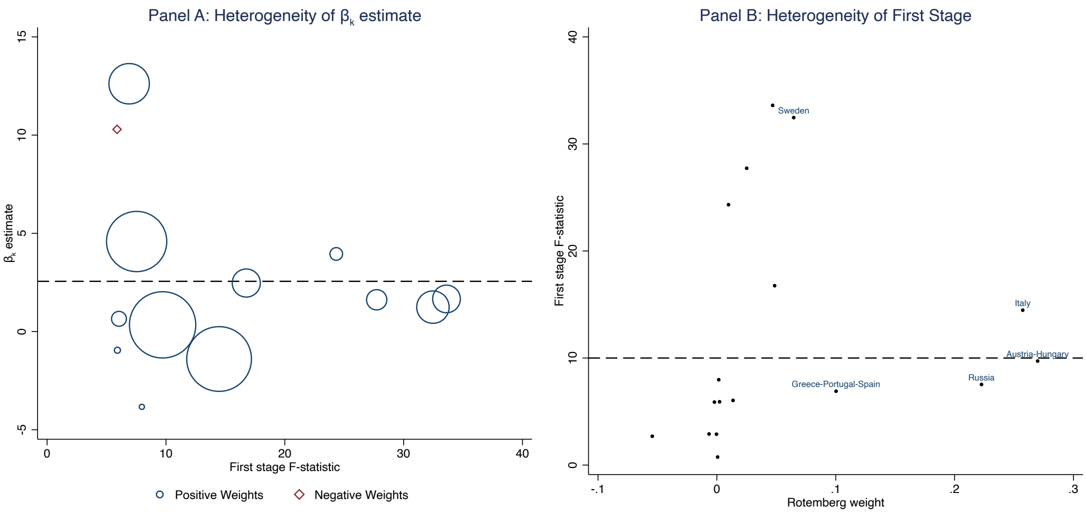
Notes: The table reports statistics about the Rotemberg weights (Goldsmith-Pinkham et al., 2020). In all cases, the reported statistics are about aggregated weights for a given country (or group of countries) of origin across years. Panel A reports the share and sum of negative weights. Panel B reports 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 variation in the weights across years. Panel D 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 E reports statistics about how the values of β_j vary with the positive and negative Rotemberg weights.

Table G.2: Test of Pre-Trends – Top Rotemberg Weight Instruments

	Dependent variable (1880-1890 difference):				
	Any Union Present (1)	Share of Urban Pop. (3)	Nr. of Mfg. Establishments (4)	Total Mfg. Output (5)	Mfg. Output per Worker (6)
Panel A: Austria-Hungary					
Share of Immigrants (average 1900–1920)	0.370	0.135	4.202	7.370*	1.255
	(1.574)	(0.664)	(2.560)	(4.026)	(2.584)
KP F-statistic	14.47	14.47	14.47	14.47	14.47
Panel B: Italy					
Share of Immigrants (average 1900–1920)	0.768	-0.493	-3.504	-3.960	0.039
	(1.888)	(0.683)	(2.865)	(4.332)	(2.494)
KP F-statistic	12.53	12.53	12.53	12.53	12.53
Panel C: Russia					
Share of Immigrants (average 1900–1920)	-0.039	1.174*	3.276	-0.473	0.481
	(1.377)	(0.626)	(3.061)	(9.040)	(2.342)
KP F-statistic	12.81	12.81	12.81	12.81	12.81
Panel D: Greece-Portugal-Spain					
Share of Immigrants (average 1900–1920)	-3.588	-0.799	-6.922	-8.747	-5.859
	(2.809)	(0.891)	(5.137)	(7.972)	(4.459)
KP F-statistic	58.52	58.52	58.52	58.52	58.52
Panel E: Sweden					
Share of Immigrants (average 1900–1920)	1.064	0.701	5.557**	6.619*	-4.341**
	(1.505)	(0.536)	(2.694)	(3.508)	(1.695)
KP F-statistic	61.59	61.59	61.59	61.59	61.59
Observations	871	871	871	871	871

Notes: The observations are at the county level. The dependent variables are the 1890–1880 difference in: the presence of any branch of the Knights of Labor (column 1); the share of urban population (column 2); the log of one plus the number of manufacturing establishments (column 3); the log of one plus the value of manufacturing output (column 4); and the log of one plus the value of manufacturing output 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 using exclusively the country-of-origin share indicated in the heading of each panel. All regressions include the following controls, measured in 1880: the share of population in farming, the share of population in manufacturing, the share of population in mining, and an indicator for the presence of a railroad in the county. 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 G.1: Heterogeneity by Rotemberg Weights

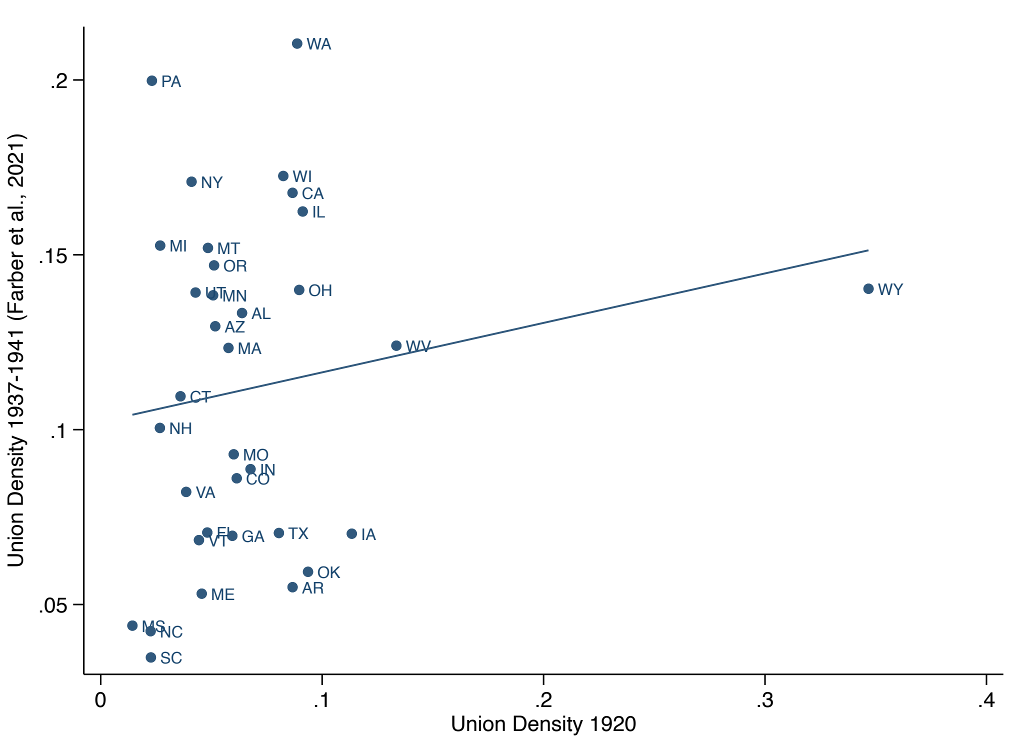


Notes: Panel A plots the relationship between each instrument's $\widehat{\beta}_j$, first-stage F-statistics and Rotemberg weights. Each point is a separate instrument's estimates (country of origin share). The figure plots the estimated $\widehat{\beta}_j$ for each instrument on the y-axis and the 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 Unionization Data

In addition to the steps illustrated in Section 3.1 to ensure the accuracy and completeness of the data collected, in this Section I provide a validation of the estimates of union density. In particular, I explore their correlation with the only other measures available in a historical period. Such data come from Farber et al. (2021), who harmonize household-level survey data from Gallup starting in 1937. In Figure H.1, I show a scatter plot between the two measures, where the data from this paper are aggregated at the state level to match the unit of observation from Farber et al. (2021). Unfortunately, the two sources do not overlap in time. Therefore, the figure plots on the x-axis the union density from this paper in 1920 and on the y-axis the measure of union density from Farber et al. (2021), calculated as an average of the first five years of observations (1937–1941). Although the two measures do not agree in levels (and they are not expected to, since by 1937 several new industrial unions had been constituted, which represented large masses of workers previously unorganized), they display a positive correlation (which becomes stronger once Wyoming, an outlier in the figure, is excluded from the sample).

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



Notes: The figure plots a scatter plot for state-level union density measured in 1920 using the newly collected archival data (x-axis) and average union density between 1937–1941 measured using Gallup data as in Farber et al. (2021). See Section 3 for more details on the dataset on labor unions I assemble for the period 1900–1920.