CLOSING RANKS: ORGANIZED LABOR AND IMMIGRATION

Carlo Medici*

November 27, 2023

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

Abstract

This paper shows that immigration positively affected the development of organized labor in the United States at the beginning of the 20th century. I digitize archival data to construct the first county-level dataset on historical union membership in the U.S. and use a shift-share instrument to exploit plausibly exogenous variation in immigration between 1900 and 1920. I find that counties that received more immigration experienced an increase in the probability of having any labor union, the share of unionized workers, the number of local union branches, and the average branch size. Exploring the mechanisms driving the effect, I find that the increase occurred only among unions representing skilled workers, particularly in counties more exposed to the immigrants' labor competition, and in places harboring less favorable attitudes towards immigration. Taken together, these results indicate that existing workers formed and joined labor unions due to both economic and social motivations. The findings shed light on a novel driver of unionization in the early 20th-century United States: in the absence of immigration, the average union density of this period would have been 17% lower. They also identify an unexplored consequence of immigration: the development of institutions that aim to protect workers' status in the labor market.

*Northwestern University, Kellogg School of Management. carlo.medici@kellogg.northwestern.edu.

I am indebted to Joel Mokyr, Matthew Notowidigdo, Nancy Qian, Marco Tabellini, and Edoardo Teso for extensive advice and guidance throughout this project. I thank Ivan Canay, Devis Decet, Georgy Egorov, Silvia Farina, Carola Frydman, Walker Hanlon, Leander Heldring, Joris Mueller, Laura Murphy, Nicola Persico, Silvia Vannutelli, seminar participants at Northwestern University, and participants of the Economic History Association 2023 Annual 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 are among the most important labor market institutions in all advanced economies. Throughout the 20th century, they have contributed to reducing inequality (Farber et al., 2021), improving working conditions (Rosenfeld, 2019), and influencing policy through extensive political activities (Ahlquist, 2017). Despite ebbs and flows in their membership, labor unions remain central to today's economy. In the U.S., they recently gained prominent victories for several categories of workers, including autoworkers, UPS drivers, and Hollywood writers.[1] In Europe and Canada, where collective bargaining also boasts a long tradition, organized labor continues to expand to previously unorganized sectors and to shape the policy agenda.[2] Given the long-lasting prominence of labor unions, it is perhaps surprising that we have relatively little evidence on the determinants of their origins. The primary aim of this paper is to address this question with systematic empirical evidence.

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

This paper investigates the second mechanism: the effect of a large and protracted increase in the labor supply on the formation and emergence of labor unions, leveraging the episodes of mass immigration to the U.S. of the early 20th century. The effect is *ex ante* ambiguous because it affects both the incentives of workers to organize and the ability of capital owners to undermine organized labor. On the one hand, the increased competition for jobs can motivate workers to organize in response to economic threats to their employment and wages. On the other hand, a larger labor supply lowers the cost to business owners to replace uncooperative workers and break strikes. Thus, how increased labor supply impacted the emergence of unions 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 grew and industrialized swiftly during this period, and the labor movement experienced its first national expansion (Foner, 1947). Unions represented workers'

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

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

interests in the workplace and, at the same time, advocated for significant pro-labor legislation (Goldin, 1994; Mink, 2019). Second, these years witnessed the creation and growth of several labor unions that remain influential today (Stewart, 1926), despite the legal and judicial frameworks of the time allowing employers to dismiss and replace unionizing workers (Taft, 1964). Third, this context provides a natural experiment to establish causal identification, given by the large and prolonged influx of European immigrants during this period, often referred to as the Age of Mass Migration (Hatton and Williamson, 1998).

Two main challenges are associated with this study. The first is the need for disaggregated data on the presence and membership of labor unions. The second is establishing the causal effect. For example, the presence of unions may deter labor migration. Such reverse causality would result in a negative association between labor and union presence. Alternatively, both unions and labor supply may increase in response to economic growth. Such joint determination would lead to a positive association between unions and labor supply.

To measure unionization, I hand-collect and digitize archival documents on the quantity, location, and membership of labor union branches across the United States. The main sources of these records are the convention proceedings of the state federations of labor, which report detailed information on the number and location of union branches within each state's territory, along with the names of the delegates sent by each branch to the conventions. I collect these data for the years 1900, 1910, and 1920. To calculate the membership of each local branch, I exploit the constitutional rules of these organizations, which specified that local union representation at the conventions be proportional to their membership. I complement these data with proceedings of national unions' annual conventions to improve and validate these measures. The information is then aggregated to the county and year levels, and merged with the historical U.S. Census. These data constitute the first comprehensive dataset measuring historical union 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 (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020). I estimate 2SLS regressions that include county and year fixed effects, in addition to baseline county characteristics which are likely correlated with the initial presence of immigrants and the evolution of unionization, such as the urban share of the population and the labor force participation rate, interacted with year dummies.

The main results of this paper show that immigration positively affected the emergence of organized labor. Counties that received more immigrants as a fraction of the population experienced an increase in the probability of having a branch of any labor union, the share of unionized workers, the number of local union branches, and the average branch size. This novel finding documents empirically a new driver of unionization and highlights an unexplored effect of immigration on the labor market. According to my estimates, a 4 percentage points (1 standard deviation) increase in immigration raised the share of the unionized workforce by one percentage point. In areas receiving high volumes of immigration – such as the large New York County (NY), or the smaller Lake (IN) and Kenosha (WI) counties – immigration increased the fraction of union workers by 50–75%. A back-of-the-envelope calculation reveals that in the absence of immigration, the average union density between 1900 and 1920 would have been 17% lower overall. The estimates are robust to a variety sensitivity checks, such as using an alternative instrument that replaces the actual immigration flows with plausibly exogenous ones (Borusyak et al., 2022) 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 (total and from each European country), the baseline shares of the labor force in major industries and occupations, and measures of income and economic growth.

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

Second, I construct a measure of exposure to the immigrants' labor market competition, whereby a county is more exposed if occupations prevalent among immigrants entering the country in each decade are also predominant among the U.S.-born workers of that county in the previous decade. I then interact this measure with the main regressor of interest, to investigate the hetero-

[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 5.2.

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

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

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

in occupations represented by unions. Therefore, the positive effects on union density cannot be explained mechanically by a decrease in the denominator of this measure.

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

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

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

that led to the emergence and development of modern labor unions during a highly formative period for the American labor movement.

This paper also relates to studies that explore the historical drivers of unionization (Alesina and Glaeser, 2004; Archer, 2010; Asher, 1982; Bernstein, 1954; Briggs, 2001; Burgoon et al., 2010; Brody, 1993; Collomp, 1988; Foner, 1947; Freeman and Medoff, 1984; Griffin et al., 1986; Hannan and Freeman, 1987; Haydu, 1988; 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 (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 important and unexplored driver of unionization. The findings show that immigration contributed to the early emergence of organized labor, and shed light on the mechanisms driving this effect.

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

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

Further, this study relates to the vast literature about the consequences of immigration on domestic workers' employment and wages, which has not reached an agreement on whether immigration has a positive, negative, or null effect (Dustmann et al., 2016). In particular, the findings of this paper are in line with Abramitzky et al. (2023), Card (2001b, 2005, 2009), Foged and Peri (2016), Ottaviano and Peri (2012), and Tabellini (2020), who find negligible or positive impacts on domestic workers. The results of this study suggest that labor unions may play a role in mitigating the possible adverse effects of immigration on domestic workers' wages and employment, and that part of the heterogeneous results in this literature could be explained by differences in unionization. Along these lines, this work also relates to the research on the effect of unions on wage inequality (Ahlquist, 2017; Collins and Niemesh, 2019; Farber et al., 2021).

Finally, this paper is closely related to the recent political economy studies showing that higher

levels of immigration increased the vote share for conservative politicians and support for anti-immigration legislation, both historically (Alsan et al., 2020; Goldin, 1994; Tabellini, 2020) and recently (Barone et al., 2016; Dustmann et al., 2019; Edo et al., 2019; Halla et al., 2017; Mayda et al., 2022; Mendez and Cutillas, 2014; Otto and Steinhardt, 2014). The results of this study identify a novel and unexamined consequence of immigration on the development of institutions that have had – and still have today – vast political influence. Although anecdotal and historical evidence has acknowledged the instrumental role that organized labor played in the introduction of immigration restrictions in the 1920s (Goldin, 1994; Mink, 2019), this paper is the first to empirically estimate a causal and positive effect of immigration on unionization, and document that this was due to both economic and social motivations. Moreover, this paper is related to the work by Alesina and Glaeser (2004), which links the weak labor and socialist movements of the U.S. 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 long-term trends in unionization. Section 8 concludes.

2 Historical Background

2.1 Labor Unions at the Turn of the Twentieth Century

Around the end of the 1880s, a new phase for the American labor movement started, as the American Federation of Labor (AFL) became the largest and most influential group of labor unions.[4] 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,[5] 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.[6] 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

[4]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).

[5]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).

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

unionism. This meant that workers were organized based on their particular occupation (or craft).[7] 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.[8] This industry was dominated by skilled craftsmen, and characterized by small employing units (Taft, 1964). Only few unions organized unskilled laborers in industrial settings. The United Mine Workers of America (UMWA) was the largest of these, along with unions representing dockworkers and workers in the meat-packing and textile industries. These sectors, mining in particular, were dominated by large employers, who owned and operated several plants or mining sites (Beik, 1996), and strongly opposed unionization efforts (Northrup, 1943).

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

Until the mid-1930s, there was no federal law requiring employers to recognize unions or punish retaliatory behavior against union members. This situation promoted an environment where company owners, with the support of the courts, made use of strikebreakers, lockouts, retaliatory firing, and other strategies to oppose unions and prevent their organization (Foner, 1947; Taft, 1964).[9]

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 sending 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 immigration moved to the rest of Europe. In 1850, the 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

[7] 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.

[8] The Bricklayers and the Carpenters' unions were the dominant organizations among building trades.

[9] 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).

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 the rural areas to the cities in the 1880s. Very frequently the newly-arrived immigrants, eager to earn a livelihood in a new country, made their first appearance into the American workforce as strikebreakers, hired by business owners in order to undermine the incumbent workers' bargaining power and unionization efforts (Foner, 1947). Over the years, the political climate grew hostile towards European immigrants, based on concerns about labor market competition and xenophobia toward new arrivals (Goldin, 1994). In response, starting in the late 1890s members of Congress proposed legislation to limit immigration, and in 1917 the U.S. Congress eventually introduced a literacy requirement for all immigrants.[10] Though during WWI immigration temporarily slowed down, after the end of the war flows 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, with the National Origins Act, this restriction was made permanent and more stringent (Abramitzky and Boustan, 2017). The immigration quotas remained in effect for the next forty years, until they were eliminated in 1965 by the Immigration and Nationality Act.

2.3 The Labor Movement and Immigration

Organized labor has always been concerned with the effects of large inflows of new workers on employment and wages, and particularly those caused by immigration (Taft, 1964). This is the main reason why it favored immigration restrictions since its inception. In 1881, in the founding meeting of its precursor organization, the American Federation of Labor adopted a resolution against Chinese laborers, and lobbied Congress to ban Chinese immigration through the Chinese Exclusion Act of 1882 (Foner, 1947). In 1885, the labor movement succeeded again when the Alien Contract Labor Law (also known as Foran Act) was approved, which banned the importation of foreigners to perform labor in the United States.[11] In 1896, in response to the shift of immigration to ethnic and national groups whose schooling levels, skills, and standards of living were substantially below those of previous groups, the AFL endorsed further restrictive measures. It was widely held that Southern and Eastern Europeans lowered wages, dragged down working conditions, were not responsive to the discipline of labor unions, and therefore constituted a threat to the American working man (Mink, 2019; Taft, 1964). The federation vigorously supported fur-

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

[11] 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, 2019).

ther restrictive measures, until it obtained the introduction of the 1921 and 1924 nationality quotas (Goldin, 1994).

Throughout this period, the labor movement used increasingly popular racial and eugenics-based arguments to discuss threats to employment and gain momentum in calling for an outright ban on European immigration.[12] Nativism was triggered by the increased presence of foreign laborers that inundated labor markets and intensified by the mounting pressure of mechanization (Mink, 2019; Yellowitz, 1981). These events added credibility to the fears that machines and the new unskilled workers could substitute skilled unionized labor (Olzak, 1989), and led unions to concentrate on securing job control for skilled workers by organizing the workplace and the work process (Mink, 2019). At the same time, the immigration-induced expansion of the labor supply was deemed responsible for weakening unions' bargaining power, by creating a reservoir of potential strikebreakers and freeing employers from the constraints of a tight, unionizing labor market (Montgomery, 1979).

3 Data

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

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

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, covers a different set of unions in different years, and in a time period that precedes the first significant expansion of the labor movement and the largest waves of immigration. Farber et al. (2021) combine survey data, primarily from Gallup, to compute historical levels of union membership for most of the 20th century. However, their data are not disaggregated below the state level, and only start in 1937,

[12] Statements made by union men expressing hatred on new immigrants abound. In 1884, a labor leader would describe 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).

after immigration restrictions had been in place for over a decade and the first national expansion of the American labor movement had occurred. Similarly, Downes (2023) constructs county-level union membership estimates for selected years in the mid-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 my 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 my dataset on unionization are the convention proceedings of the state federations of labor. The state federations of labor were state-level subordinate bodies of the American Federation of Labor. 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 elects general officers. All affiliated local unions were entitled to representation.[13]

I digitize the proceedings of these conventions every ten years between 1900 and 1920.[14] From these documents, I extract the lists of the 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).[15] 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. Using this information, I construct an estimate of union membership for each local branch. Since the representation rules were often expressed in terms of ranges (e.g., one delegate every 100 members), I use the mid-points of these intervals as the estimates of membership. For example, if the constitution states that a branch is represented by one delegate every 100 members, its membership is estimated to be 50 if one delegate is present at the convention, 150 if two delegates are recorded, and so on. The results are unchanged if membership is estimated using the lower or the upper bound of the intervals instead.

I geocode the location of all the union branches based on their town, village, or city, and retrieve

[13] 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.

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

[15] 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 my sample.

their coordinates. I use the names of each branch's national union to establish which occupations and industries they operated in.[16] 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. Similarly to the state federations, the AFL-affiliated unions met in national conventions to legislate, elect officers, and set the guidelines for the local branches to follow in their bargaining agreements. I digitize the proceedings of these conventions for six of the largest AFL-affiliated unions of this period, every ten years between 1900 and 1920.[17] The members of these six unions accounted for approximately 40% of the over 100 AFL-affiliated unions' total membership between 1900 and 1920 (Wolman, 1924).[18] 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 representation rule listed in the convention proceedings or in the constitution of each of these organizations. Finally, I aggregate the data at the county level.

These data sources complement the records from the state federations in three main ways. First, they provide a validation for the estimates constructed using the main data source. In particular, for the six unions that I observe across both sources, I am able to compare the estimates of union membership and the number of branches. In all cases, the measures display a highly positive correlation (Figure A.3). Nonetheless, it is possible that some branches may appear in only one of the two types of convention documents I digitize. This may be more likely to occur for smaller branches, which may have found it harder to find individuals to send to both conventions (or funds to cover their travel expenses). Similarly, 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 conventions. Another possibility is that some delegates may have been erroneously omitted from the roll calls of the convention proceedings. Any of these occurrences would lead me to underestimate the number of members and/or the number of branches in a given county if I relied on only one of the sources. Unfortunately, there is no way of knowing with certainty if and how many locals fall into these circumstances, since this information is never systematically

[16] 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.)

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

[18] These unions are: the Bricklayers, Masons, and Plasterers International Union of America (BMPIU), the International Association of Machinists (IAM), the International Brotherhood of Teamsters (IBT), the International Typographical Union (ITU), the United Brotherhood of Carpenters and Joiners (UBC), and the United Mine Workers of America (UMWA). These are all the unions among the ten largest that systematically and consistently reported information in their convention proceedings about delegates and the local branches they represented, and whose proceedings are available either in physical or digital copy.

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 my dataset. This is due to the fact that some state federations of labor were constituted (and hence convened for the first time) only after 1900.[19] Relying only on the first data source would lead me to measure no presence of union branches and zero union membership for counties in states and decades before the first federation of labor's convention. Although the lack of an AFL state subordinate body is intrinsically suggestive of a limited presence of organized labor, it is still possible that some national unions may have already been present in at least some counties of these states. The additional information on the branches (and its delegates) of these six large unions operating throughout the whole U.S. territory in 1900, 1910, and 1920, allows me to more accurately measure unionization at the early stages of a state's labor movement.

Combined data sources. To construct my final measures of unionization, I combine the information collected from the two sources described above. I first reduce the number of missing observations and misreportings from each of these sources by linearly interpolating the number of union branches and members for counties that are not reported in the convention proceedings of a certain year, but that have representation both in the previous and in the following decade.[20] Next, for the six 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. In the case in which only one data source reports a positive membership or number of local branches, this is the value I use in my analysis. Finally, I sum the total number of branches and members across all unions at the county-decade level, and obtain the total number of these quantities in each county over time.

In order to construct measures of union density, I divide the number of union members by the size of the labor force, by occupation and total. For example, the measure of union density for carpenters will be the number of members of the carpenters' union divided by the labor force in carpentry occupations. When computing the overall union density for the county, the total number of union members is divided by the total labor force in occupations within the jurisdiction

[19] The following state federations of labor first convened after 1900, the first year of my 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.

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

of the AFL-affiliated unions in existence in the period 1900–1920.[21] Additionally, I construct an indicator for whether a county has any union, the number of union branches within its territory, and their average size, defined as the number of members divided by the number of branches. As a final validation exercise, I compare my measures of union density to those contained in another existing historical dataset. While only aggregated national estimates of union membership exist for the period I study, I ensure that my 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. In Figure F.1, I show that this is the case.

My final dataset contains information on the number of labor union branches and their membership in over 2,400 counties.[22] Throughout the empirical analysis, I will restrict the sample to a balanced panel of 1,675 counties, which represent approximately 65% of the total U.S. labor force during this time period.

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 at the national level, are taken from the decennial U.S. Census of Population. For 1900, 1910, and 1920, I use the full count census datasets, made available by IPUMS (Ruggles et al., 2021). For 1890, I use census datasets aggregated at the county level, made available by ICPSR (Haines, 2010).[23]

Labor market outcomes. I compile data on labor force, occupation, and yearly income from the U.S. Census of Population.[24]

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

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 (1999).

Railroad network. Data on the expansion of the railroad network rely on the database compiled

[21] As in the rest of the paper, the labor force variables are computed restricting the sample to men 16–64. The jurisdiction of each union is taken from Stewart (1926). In case the total number of estimated union members exceeds the total labor force, union density is coded to be 1. This is a rare event, which occurs for the main measure of union density only in 12 out of the 5,025 county-year observations of the main sample. In Section 5.2, I show that the results are unchanged when using alternative definitions of union density, such as dividing union members by the total labor force in occupations organized both by AFL and non-AFL unions, or by the total non-agricultural labor force, and when excluding outliers.

[22] The counties not part of my 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). In Section 5.2, I show that the results are unchanged when extending the analysis to the whole unbalanced sample of counties.

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

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

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.

3.3 Summary Statistics

Figure 6 plots the share of the unionized workforce (union density) in my sample, in 1900, 1910, and 1920. Unionization in 1900 was predominantly concentrated in areas of the Northeast and Midwest. By 1920, unions had spread in many other areas of country as well, including the Western United States and in selected areas of the South. Overall, the maps display substantial variation in union density across counties, both within and across states.

In Table 1, I present the summary statistics for the main variables on demographic characteristics (Panel A), labor market (Panel B), and unionization (Panel C). On average, the share of recent immigrants – defined as the number of European immigrants who entered the U.S. in the previous decade as a fraction of the population (men 16–64) – is 2%. However, this masks substantial heterogeneity across counties, as indicated by the size of the standard deviation. The average labor force participation is 91%; on average, 87% of the labor force is made of U.S.-born workers, and 11% is composed of European immigrants. Turning to the unionization outcomes, the average county membership is slightly short of 255, and the average union density is 4%. Overall, over a quarter of the observations in my sample have positive union membership.

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.[25] Throughout the analysis, standard errors are clustered at the county level, and all variables are harmonized to reflect 1930 county boundaries (Hornbeck, 2010).[26]

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

[26] Since county boundaries change over time, I maintain consistent geographic units by holding county boundaries constant throughout my sample period. I follow the procedure in Hornbeck (2010) and harmonize all the variables used in the analysis to reflect 1930 county boundaries. This procedure uses area-based weights to harmonize county

In my baseline specification, Imm_{ct} refers to the stock of working-age male European immigrants who entered the U.S. during the previous decade, as a share of the total working-age male population. All the labor force variables are similarly computed on the sample of men 16–64. Focusing on this group allows me to more confidently interpret my findings as the consequences of an inflow of new (immigrant) workers into the labor market.[27]

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). 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 U.S.), which I denote $\alpha^{j}_{c,1890}$. The second is the change, or *shift*, in the number of European immigrants from country j entering the U.S. in a given decade, net of those that eventually settled in county c, denoted by O^{j}_{-ct}.[28] 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 $\alpha^{j}_{c,1890}$. I choose 1890 as the base year because it captures many of the important migration networks established in the first part of the Age of Mass Migration, but predates both the peak of immigration flows from Europe and the largest periods of union growth (Figure 1 and Figure 2).[29] Crucially, 1890

boundaries across years. Following alternative border harmonization procedures that use population-based weights, such as the one in Ferrara et al. (2022), yields almost identical results.

[27] 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). Results are very similar when considering all immigrants, regardless of their age or arrival year.

[28] A similar "leave-out" strategy is also used in Tabellini (2020).

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

also predates the large compositional shift in immigration that occurred at the turn of the 20th century (Figure 4). As previous work has argued (Abramitzky et al., 2023; Tabellini, 2020), this period is particularly suited to the use of shift-share instruments, not only because of the changes in the quantity of immigration over time, but also due to the variation in the immigrants' country of origins in each decade. Differently from Tabellini (2020), who employs an analogous identification strategy to predict immigration between 1910 and 1930, this shift-share instrument exploits the additional variation in the composition of immigration that took place between 1890 and 1900.

Table 2 reports first stage coefficients and shows that actual and predicted immigration are highly correlated.

Identification assumption. The key identifying assumption behind the instrument described in equation (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 the immigration patterns from European countries after 1890.[30] 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 World War I (WWI) lowers the concern that the shift-share instrument may be correlated with shocks jointly affecting local condition 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 my 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 my baseline specification by including interactions between year dummies and county characteristics measured at baseline that might have attracted more immigrants (from each sending country) before 1890, and may have had a time-varying effect on unionization across counties. In my preferred specification, such controls include: (i) the share of the urban population living in county c at baseline; and, (ii) the baseline labor force participation rate, defined as the number of individuals in the labor force divided by the total working-age population.[31] The former accounts for the fact that both immigration and labor unions were a predominantly urban phenomenon in this period (Abramitzky and Boustan, 2017; Taft, 1964), and therefore early urbanization levels may have been correlated with both the initial settlement of immigrants and the subsequent evolution of organized labor. Similarly, tighter labor markets likely attracted more immigration early on and affected the growth of labor unions in the beginning of the 20th century.[32] Second, I directly control for the size of

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

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

[32] In Appendix B, I show that the estimates are robust to the inclusion of several other baseline county characteristics with a potential effect on both the 1890 levels of the immigrant population and unionization in the subsequent decades (Table B.4), and also that the results do not depend on the inclusion of any of these controls (Table B.5).

the 1890 European immigrant population, interacted with year dummies. This implies that the effects of immigration are identified exploiting variation only in the ethnic composition of immigrant enclaves across counties, holding constant the size of their foreign born populations. Since 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. Third, I include interactions between year dummies and the share of immigrants from each European country, $a^{j}_{c,1890}$, to assuage concerns that the 1890 settlements of specific European groups across U.S. counties might be correlated with both the long-run trends in unionization and the migration patterns of those specific immigrants groups, in each decade between 1890 and 1920.

Alternative instrument. In addition, I construct an alternative version of the instrument described in equation (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 initial 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 1890 as a fraction of the county population.[33] Then, I re-estimate the 2SLS analysis controlling for fixed effects for the 800+ county pairs interacted with year dummies.

I summarize all the other robustness checks in Section 5.2, after presenting the main results.

5 Main Results

5.1 The Effect of Immigration on Unionization

In Table 3, I investigate the effects of immigration on the formation and the development of organized labor by estimating equation (1). I examine four unionization outcomes: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union

[33] 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 American Federation of Labor became the dominant federation of unions. For this exercise, I use data from Garlock (2009) to measure union presence as of 1890, when the American Federation of Labor was only recently established and did not yet have substantial national presence (Foner, 1947).

members divided by the labor force (column 2);[34] the log number of union branches (column 3);[35] and, the average branch size, defined as the number of members divided by the number of branches (column 4).[36] All regressions include county and year fixed effects, and interactions between year dummies and the baseline urban population share and labor force participation rate (as discussed in Section 4.2). I present OLS estimates in Panel A. Although imprecisely estimated (with the exception of column 3), all coefficients are positive. This suggests that counties that received more immigration were also more likely to display higher levels of unionization.

Panel B and C show the reduced form and the 2SLS estimates. The F-stat for weak instruments, reported at the bottom of the table, is always above the conventional levels, and indicates that the instrument is strong. In all cases, the point estimates are positive and statistically significant at either the 5% or the 1% level. The 2SLS estimates imply that a 4 percentage points (1 standard deviation) increase in the share of recent immigrants causes: a 6.3 percentage point (24% relative to the mean) higher probability that the county has any union (column 1), a higher share of unionized workers by over 1 percentage points (column 2); 70% more union branches (column 3);[37] and, ten more members per branch, or 35% relative to the sample mean (column 3).

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

The difference between OLS and 2SLS estimates indicates that the former are biased downwards, and suggests that European immigrants selected areas where unionization was growing more slowly. This might have happened because, during this period, the vast majority of labor unions actively discriminated against immigrants, precluding them from joining their ranks and the occupations they represented (Asher, 1982). Consistent with the historical evidence, in Table A.2 I show that there is a negative and statistically significant relationship between all my four measures of unionization and immigration flows in the following decade. Another possibility, which may co-occur with the previous one, is that the instrument identifies a local average treatment effect (LATE) for counties that received more European immigrants because of country-of-origin

[34] My preferred definition of union density is the number of union members divided by the total labor force in occupations covered by the AFL-affiliated national unions during this period. This measure has the main advantage of not being influenced by the relative importance of such occupations in the labor force. In Table B.11, I show that the results hold when using different definitions of the dependent variable.

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

[36] To maintain the same sample throughout the table, and for consistency with the other outcomes, I define this variable as zero if the county has no union branch (and, therefore, also no union members). Results are qualitatively similar if restricting only to county-year observations with at least one union branch. See Section 6 for a discussion about the effects on the extensive and intensive margins of unionization.

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

networks, and not because of economic or political characteristics of the destination county. If such immigrants were more likely to generate an increase in unionization – either because of their own preferences, or because of the reactions they would cause among existing workers – this could explain why OLS coefficients are smaller than the 2SLS estimates.

5.2 Summary of Robustness Checks

I perform several exercises to verify the robustness of my 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 for the period 1890–1920 to predict the flows of European immigration (Table B.2).[38] This alternative identification strategy relies on the observation that the validity of shift-share instruments can be achieved from the exogeneity of the shocks (Borusyak et al., 2022).

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

Morevoer, I verify that the results are robust to the inclusion of several county characteristics which 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: an indicator for whether the county was connected to the railroad network, the share of the immigrant (total and European) and Black population, the share of the labor force in the largest industries, the share of the labor force in occupations covered by AFL-affiliated national unions, the average occupational income score, the growth rate of the manufacturing output, the share of land used for farming, and the vote share for the Democratic Party in presidential elections.

Further, I show that the findings are unchanged when: using alternative baseline specifications, such as not controlling for any baseline characteristics or including state by year fixed effects (Table B.5); dropping potential outliers (Table B.6); clustering standard errors at the SEA level or using Conley (1999) standard errors to account for spatial correlation (Table B.7); estimating population-weighted regressions (Table B.8); extending the analysis to an unbalanced sample of counties (Table B.9); excluding the South from the estimation sample (Table B.10); and, using alternative definitions of union density (Table B.11).

I also re-estimate the baseline 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 driving the results by absorbing most of the variation in the data (Goldsmith-Pinkham et al., 2020).[39]

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

[39] 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.

Finally, I check for the absence of pre-trends by regressing the pre-period change in several unionization, population, and economic outcomes on the 1900 to 1920 change in immigration predicted by the instrument (Table B.12). The fact that all these coefficients are never statistically significant suggests that, before 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or in other economic variables.

6 Mechanisms

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

6.1 Heterogeneous Effects by Skill

As described in Section 2, the massive labor supply expansion caused by immigration inundated the urban labor markets with large pools of laborers in search of employment. This, in turn, increased employers' bargaining power, by lowering their cost to break strikes and replace workers willing to unionize (Asher, 1982; Mink, 2019; Olzak, 1989). Such process was facilitated by the political and legal framework of this period, when workers' right to unionize was not granted by law and courts often sided with the employers in disputes over the dismissal of unionizing or striking workers (Foner, 1947; Taft, 1964).

On the other hand, the positive effects presented in Table 3 suggest that immigration provided workers with greater incentives to organize against the economic threats to their employment and wages. Nonetheless, one should expect such a reaction to occur only among those workers who cannot be easily and immediately replaced. For example, a minimum required level of human capital represents a barrier to enter some occupations, and therefore can provide incumbent workers an advantage in forming and sustaining a labor union at their workplace.

I leverage differences in the skills required across occupations to test this hypothesis. As described in Section 2, most AFL-affiliated labor unions in the early 20th century organized skilled workers, especially craftsmen working in relatively small employment units and operating in the construction industry. However, some unions represented unskilled laborers, mostly in industrial plants and company towns. Between 1900 and 1920, such unions organized: coal and metal miners, dockworkers, and laborers in the meat-packing and textile industries (Stewart, 1926).[40] I formally test whether immigration had heterogeneous effects across occupations by examining its impact on the four unionization outcomes of Table 3, separately for skilled and unskilled unions. I report the results in Table 4. The estimates indicate that immigration had a positive impact on

[40]The United Mine Workers of America, which organized coal miners, was one of the largest union in terms of total national membership during this period (Wolman, 1924).

all skilled unionization outcomes. Counties that received larger shares of recently arrived immigrants experienced an increase in the probability of having any union, in the share of the unionized workforce, in the number of union branches, and their size. Instead, immigration had no effect on the expansion of unskilled unions. These results are consistent with the hypothesis that skilled jobs could not be immediately taken on by the immigrants. This likely gave incumbent workers in those occupations an advantage in their quest to establish and maintain a labor union against their employer, who could less credibly threaten to replace the existing workers with the newly arrived immigrants.

Most of the outcomes shown until now, however, do not distinguish between an increase along the extensive or the intensive margin of unionization. In other words, we know that counties that received larger shares of immigrants became more likely to have some union presence. But did immigration also increase the strength of labor unions in already unionized labor markets? I answer this question by restricting the estimation sample to a balanced set of counties that were unionized in every decade between 1900 and 1920. This allows me to rule out the possibility that my results simply reflect a comparison between unionized and non-unionized counties. The coefficients, presented in Table A.1, indicate that immigration positively affected skilled unionization also along the intensive margin, increasing the share of unionized workers in always unionized counties.

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

6.2 Labor Market 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 U.S. in each decade between 1890 and 1920. Both unskilled (e.g., miners) and skilled (e.g., carpenters, machinists) occupations feature among the most frequent ones.

To formally estimate the effect of the immigrant labor market competition on unionization, I interact my main regressor of interest from equation (1) with a time-varying measure of a county's exposure to immigrants' competition for jobs.[41] This measure consists of two terms. The first is given by the number of immigrant workers in each occupation o who entered the U.S. (net of those

[41] 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.

that settled in county c) between $t-10$ and t, as a fraction of the total immigrants in the labor force who entered the U.S. between $t-10$ and t. The second is a weight, represented by the share of U.S.-born workers in county c and occupation o in the previous decade:

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

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

In Table 5, I show the results separately for skilled (Panel A) and unskilled (Panel B) unions. I interact my main regressor of interest with a standardized version of the measure described in equation (3), computed with the respect to the occupations organized by each of the two set of unions. In Panel A, the uninteracted estimates are all positive and statistically significant. Remarkably, also all the coefficients of the interactions (except for column 1) are statistically significant. These findings indicate that counties more exposed to the immigrant labor market competition in skilled occupations experienced larger growth in skilled unionization. On the contrary, the estimates in Panel B show no statistically significant effect on the uninteracted coefficients, while all the estimates of the interaction terms are negative. These results suggest that, among unskilled workers, increased labor market competition may have instead hampered the growth of labor unions.

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

6.3 Social Concerns About Immigration

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

Discrimination against culturally distant immigrants. As described in Section 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 the individuals arriving from Southern and Eastern Europe. Such immigrants 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 not responsive to the discipline of labor unions (Goldin, 1994; Higham, 1955; Taft, 1964). If the increased unionization were caused in part also by xenophobic reactions, the effects should be more prominent in places that received larger shares of more culturally distant immigrants. To test this idea, I estimate

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

where Imm_{ct}^{SE} is the fraction of immigrants from Southern or Eastern Europe, and Imm_{ct}^{NW} the one of immigrants from Northern or Western Europe. 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 sending region. I present the results in Table 6. As expected, larger increases in unionization are caused by the inflow of immigrants from Southern and Eastern Europe.[42]

Heterogeneity by attitudes towards immigration. However, the previous result may conflate economic and cultural concerns, to the extent that immigrants from those areas may have also had lower wages expectations, and made coordination within unions harder due to their higher illiteracy rates and larger linguistic distance than immigrants from Northern and Western Europe. To further explore this channel, I test whether the effects are stronger in counties with worse attitudes towards immigration. In the absence of a direct measure, I use two proxies that likely reflect a county's higher hostility towards immigrants. The first is the historical vote share for the Know Nothing Party, a nativist political party that in the mid-1850s ran on an anti-Catholic and anti-Irish platform (Alsan et al., 2020). The second is a measure of residential segregation between U.S.-born and European immigrants.[43] Since residential segregation usually arises either from collective action to exclude minorities or from individuals from the majority group moving away from ethnically mixed neighborhoods (Boustan, 2013), such characteristic likely reflects higher levels of discrimination against immigrants.

I report the results in Table 7, separately estimating my baseline 2SLS regressions separately for the sample of counties above and below the median of vote share for the Know Nothing party and the measure of residential segregation, respectively. Using either proxy, I find that immigration strengthened organized labor more prominently in counties with higher resentment towards immigration.

Altogether, these results suggest that non-economic motives also contributed to the expansion

[42] The results are unchanged when separately estimating immigration from Protestant and non-Protestant countries (Table A.3).

[43] 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 my instrument, may qualify as a "bad control" (Angrist and Pischke, 2009). For more details on its construction, see Appendix D.

of labor unions. Unionization occurred more prominently in counties that received larger shares of culturally distant immigrants, namely those from Southern and Eastern Europe. Moreover, immigration strengthened the American labor movement more in counties that harbored less favorable attitudes towards immigration.

6.4 Ruling Out Alternative Explanations

Immigrant-Driven Unionization One first alternative explanation for my results is that immigrants may have joined or created labor unions at greater rates than U.S.-born workers. Although detailed information on the origin or ancestry of union members is not available, I can exploit my data on the local union representatives, described in Section 3, to gauge the ethnic composition of these organizations. Union delegates can be considered as 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.[44] 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 only very few from Southern or Eastern Europe.

Although the share of U.S.-born delegates increased – and the one of Europeans decreased – over time at the national level, it may still be the case that counties that received more immigrants experienced an increase in the proportion of European leaders. If, for example, the newly arrived immigrants joined labor unions *en masse*, we would expect to see an increase in the share of European delegates, as the newcomers would likely obtain the voting power to elect them. To test whether this is the case, I use the proportion of leaders with last names prevalent among U.S.-born and Europeans, computed at the county level, as dependent variables in equation (1). The coefficients, plotted in Panel A of Figure A.5, indicate that the inflow of immigrants did not increase the proportion of leaders with immigrant last names. The coefficients on the left, estimated on the whole sample of counties, show that immigration increased the share of U.S.-born leaders more than the one of immigrants. The ones on the right, computed on the counties where I observe at least a delegate in every year – although imprecisely estimated – suggest that immigration caused a redistribution of delegates in favor of the U.S.-born.[45]

These findings confirm the anecdotal and historical evidence that the observed increase in

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

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

unionization was not caused by a larger participation of immigrant laborers, but rather by U.S.-born workers (Mink, 2019; Taft, 1964), who maintained the control of labor unions throughout the first twenty years of the 20th century.

Previous exposure to labor unions. A second possibility is that immigrants coming from European countries that already had labor unions by the end of the 20th century may have brought into the United States their experience of collective bargaining from their home country, and, in turn, contributed to the growth of unionization in their destination counties. This explanation would be in line with existing work arguing that Europeans who migrated to the U.S. 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.[46] The results, shown in Table A.4, rule out this possibility. The coefficients of the share of immigrants from the U.K. and Ireland, the only countries with a strong labor movement at the turn of the 20th century, are never statistically significant; on the contrary, the coefficient for the share of immigrants from the rest of Europe are positive and statistically significant.

Other economic channels. Another possibility is that the growth in unionization may have been caused by a differential economic expansion – or contraction – experienced by counties receiving larger shares of immigrants. Alternatively, the observed effect may have been a result of a decrease in the number of individuals working in occupations represented by AFL unions, which would mechanically increase the measure of union density. In Figure A.6, I show that this is not the case. Immigration had no effect on economic indicators such as the (male) labor force participation rate or total manufacturing output (measured both as a fraction of the manufacturing labor force in the county, or of the total output in the U.S.). Moreover, the effect on the (log of the) total number of workers in occupations covered by skilled unions is actually positive, although imprecisely estimated; if anything, this goes against the estimated effect, as it mechanically reduces union density.

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

7 Economic Implications

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

Effects on U.S.-born workers' outcomes. A question unexplored thus far in the paper is whether

[46] 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.

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

In addition, consistent with existing evidence in both historical settings (Abramitzky et al., 2023; Tabellini, 2020) and recent times (Card, 2001b, 2005, 2009; Foged and Peri, 2016; Ottaviano and Peri, 2012), immigration did not have negative effects on the labor market outcomes of domestic workers, which I measure with the labor force participation rate and the (log) occupational income score (Table A.5).[47] In light of the increased unionization caused by immigration that this paper documents, these results suggest that labor unions may have mitigated the potentially adverse economic consequences induced by an increased labor supply on existing workers.

Unions and inequality. Another central economic question that arises from the findings of this paper is about the consequences of unionization on inequality. Recent evidence (Farber et al., 2021) has documented a causal impact of labor unions in reducing inequality for most of the 20th century, combining national and state-level survey data on unionization from the mid-1930s onwards. I explore the relationship between labor unions and inequality in my sample. More specifically, I use data on wages from the U.S. Census of 1940 – the first year in which such information was collected – to compute measures of wage inequality at the county level, and investigate the correlation between them and my measures of unionization in 1920 – the last year in my sample. Following the literature (Autor et al., 2008), I measure inequality as the log wage differentials for full-time, full-year workers computed at the following percentiles: 90 to 10; 90 to

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

50; and, 50 to 10.[48] I present the results in Table 9. The coefficients in Panel A display a negative correlation between the presence of labor unions in the county and wage inequality. Although not causal, these results are consistent with existing studies documenting labor unions' contribution in reducing inequality (Collins and Niemesh, 2019; Farber et al., 2021), and suggest that unions may have done so also in the first four decades of the last century.

Persistence of unionization. Further, I examine whether the local patterns of unionization that emerged in the early 20th century, and that I document for the first time in this paper, persisted until today. I aggregate the data on union density between 1900 and 1920 at the metropolitan-area level, to make them consistent with the current measures of unionization from Macpherson and Hirsch (2023), and explore their cross-sectional correlation with the average levels of union density over the first two decades of the 21st century. The results are presented in Figure 10. Remarkably, even after controlling for Census division fixed effects, which account for differences in attitudes towards organized labor across areas of the country, past and present unionization are positively correlated. This suggests that the conditions that favored the initial development of labor unions in the early 1900s may have provided the labor movement with a head start that perdures throughout decades.

8 Conclusion

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

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

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

roots organizations and other self-organized institutions, expanding our perspective on its societal impact.

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 U.S. 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^{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^{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.

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.

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.

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.

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

Barth, E., 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.

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.

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.

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.

Downes, H. (2023). 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.

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

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

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

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*.

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.

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

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

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.

Inter-University Consortium for Political and Social Research (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.

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

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.

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*.

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

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

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

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

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

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

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.

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

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. (2021). 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.

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.

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			
Total Membership	5,025	254.58	1,557.79
Union Density	5,025	0.04	0.11
Nr. Branches	5,025	1.62	5.59
Avg. Branch Size	5,025	29.94	71.31
Share of Unionized Counties	5,025	0.26	0.44
Panel B: Demographics			
Share Recent Immigrants	5,025	0.02	0.04
Share Urban Population	5,025	0.18	0.24
Total Population	5,025	33,010.55	102,216.67
Panel C: Labor Market			
LF Participation Rate	5,025	0.91	0.04
Share of U.S.-Born LF	5,025	0.87	0.15
Share of Immigrant LF	5,025	0.11	0.12

Notes: The table presents summary statistics for the over 2,400 counties in the sample described in Section 3, in the years 1900, 1910, and 1920. Share Immigrants is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. Union Density is the number of union members divided by the labor force in occupations represented by AFL unions. Avg. Branch Size is the number of union members divided by the number of branches.

Table 2: First Stage of the Instrumental Variable Estimation

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

Notes: Observations are at the county-decade level. The table reports the first stage of the instrument described in Section 4.2. The dependent variable is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The main regressor of interest is the predicted number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the 1890 male population in the county. All regressions include county and year fixed effects. The following controls, interacted with year dummies, are also included: the 1890 share of urban population (column 2); and, the 1880 male labor force participation rate (column 3). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 3: The Effect of Immigration on Organized Labor

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). Panel A shows OLS estimates, where the regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. Panel B shows reduced form estimates, with the instrument described in Section 4.2. Panel C shows 2SLS estimates. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 4: Heterogeneous Effects by Workers' Skills

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 5: Heterogeneous Effects by Immigrants' Labor Market Competition

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. Competition is a (standardized) measure of the immigrants' labor market competition, based on the prevailing occupations among the immigrants that enter the U.S. in each decade and the ones of the U.S.-born workers in each county in the previous decade, as described in Section 6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 6: Heterogeneous Effects by Origin of Immigrants

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from Southern/Eastern Europe or Northern/Western Europe who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 4.2 and Section 6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table 7: Heterogeneous Effects by Attitudes Towards Immigration

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. In Panel A, the estimation sample is split around the median of the vote share for the Know Nothing party in the 1856 presidential elections. In Panel B, the estimation sample is split around the median of the index of residential segregation calculated in 1880 and described in Section 6 and Appendix D. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

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

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

Notes: Observations are at the county-decade level. The dependent variables are the shares of U.S.-born workers (men 16–64) in the labor force who are in occupations that have positive union membership in the county (columns 1–3), or no union representation in the county (columns 4–6). All (columns 1 and 4) refers to all occupations covered by an AFL-affiliated national union; Skilled (columns 2 and 5) refers to the occupations covered by the ten largest AFL-affiliated national unions that represented skilled workers; Unskilled (columns 3 and 6) refers to the AFL-affiliated national unions that represented unskilled workers. The sample in each column is restricted to counties that have at least one worker in the indicated set of occupations in every decade between 1900–1920. The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

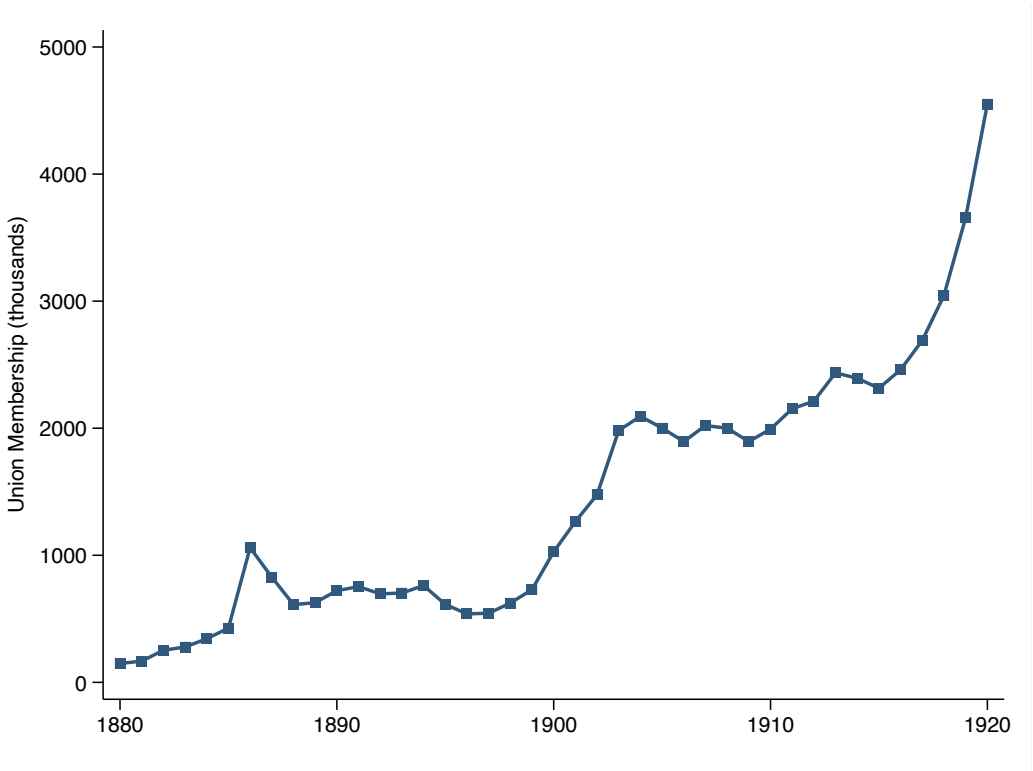
Table 9: Union Density and Wage Inequality

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

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

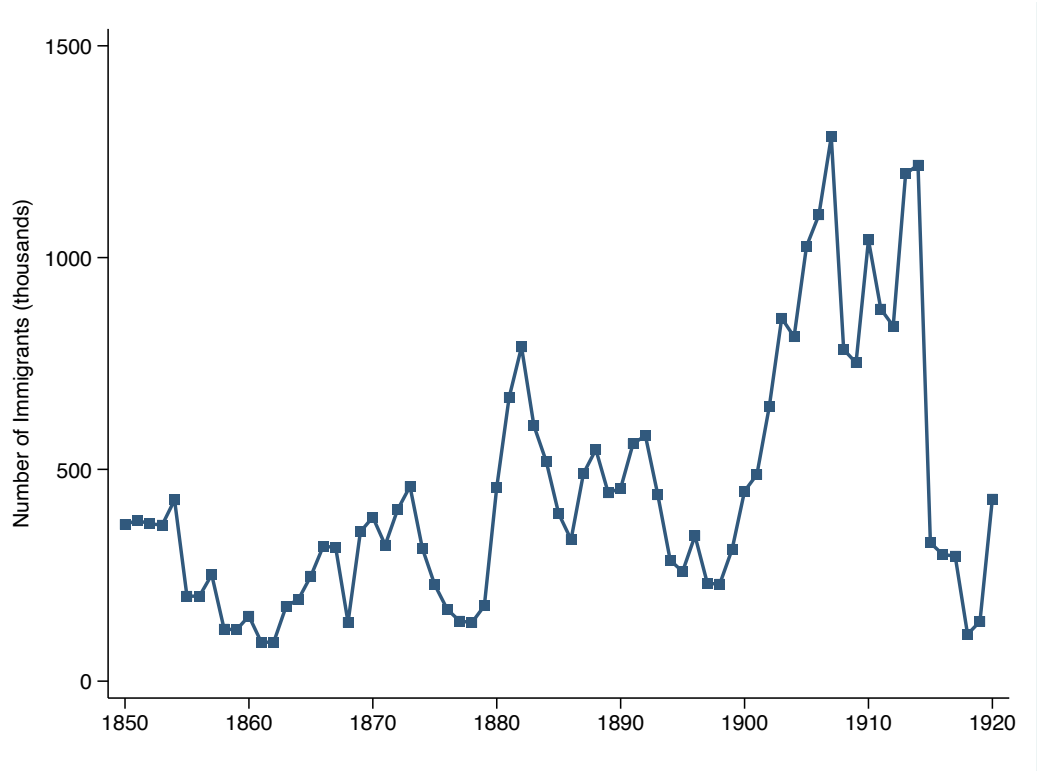
Figures

Figure 1: Estimates of Total Union Membership (1880–1920)



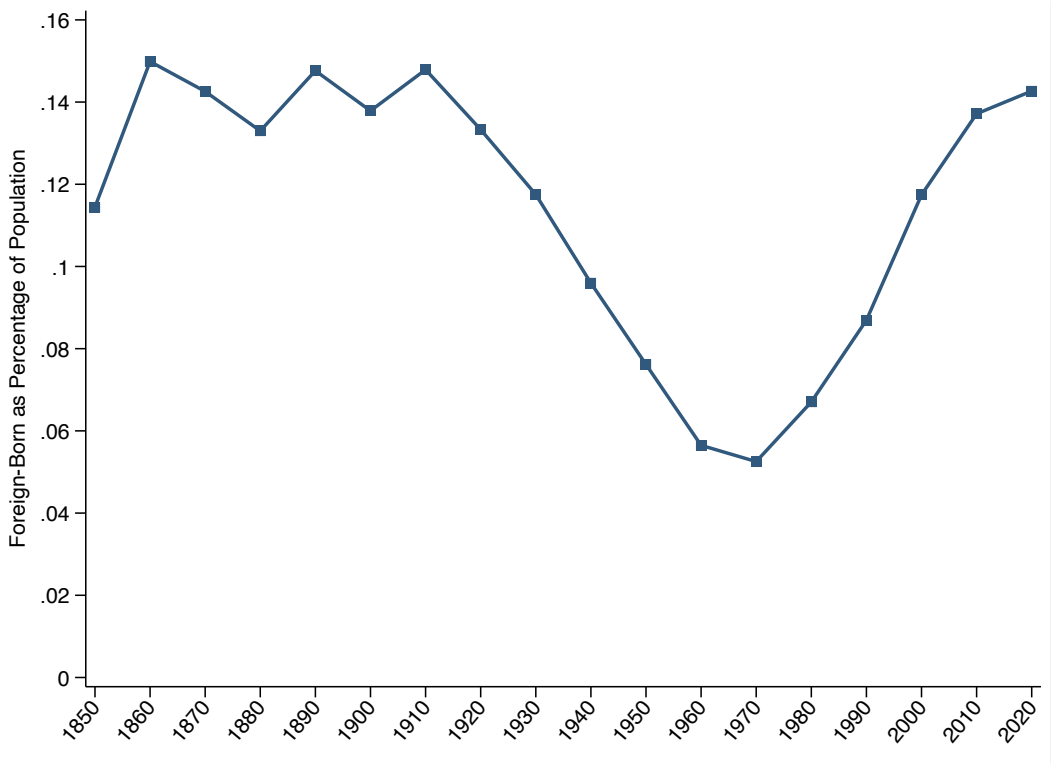
Notes: The figure shows the total number of union members in the U.S., between 1880 and 1920. Source: Freeman (1998).

Figure 2: Annual Inflow of Immigrants (1850–1920)



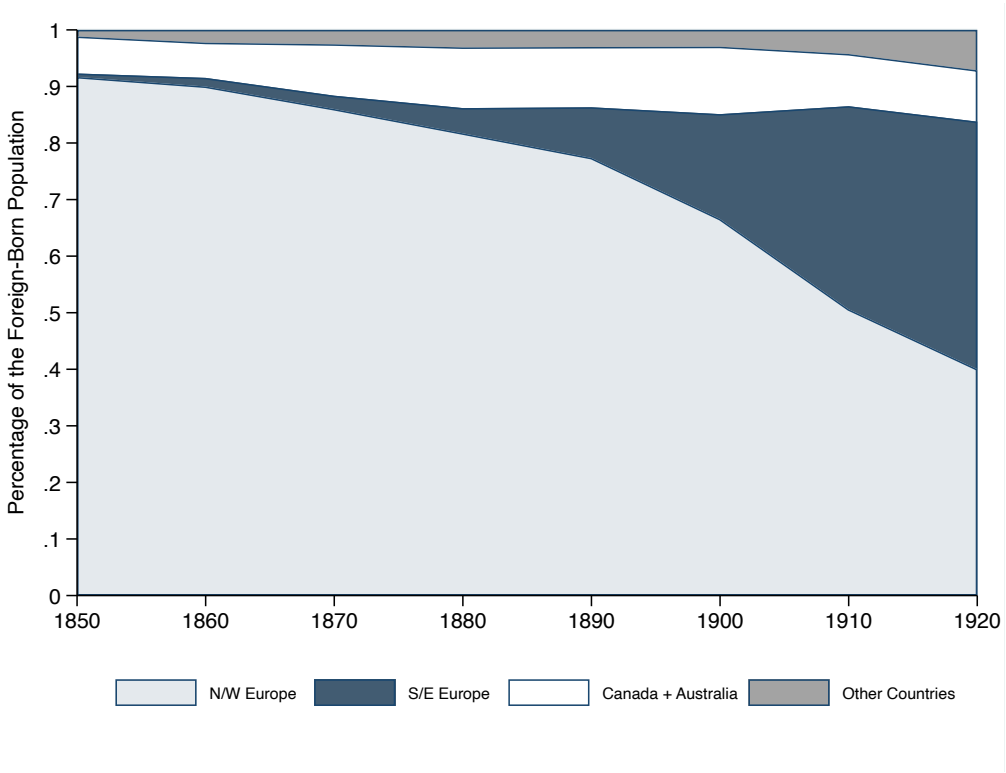
Notes: The figure shows the total number of immigrants to the United States, between 1850 and 1920. Source: Immigration Policy Institute.

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



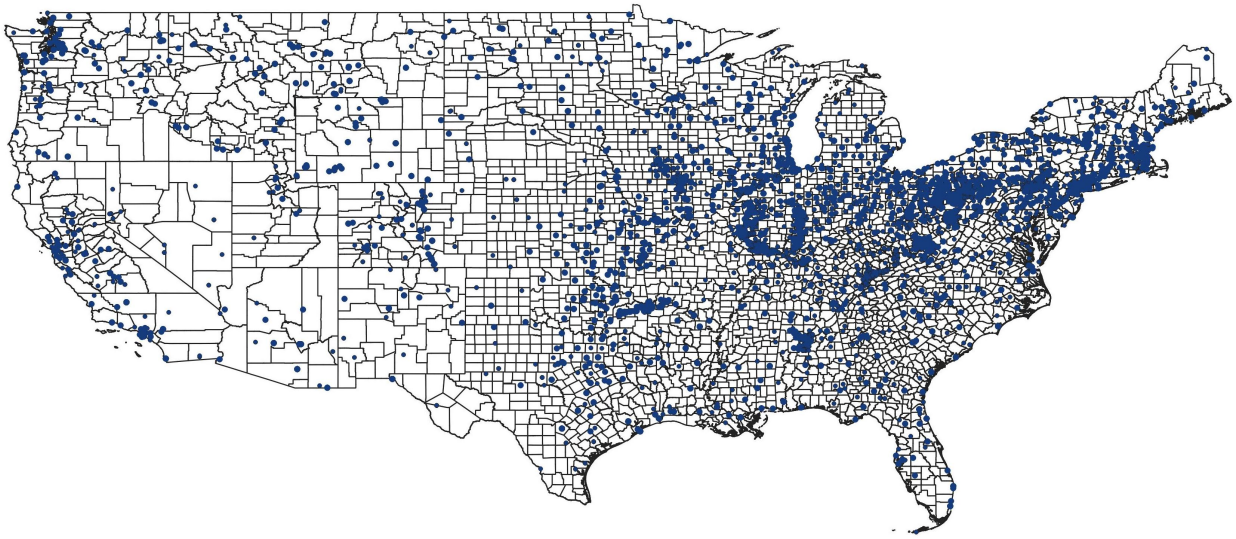
Notes: The figure shows the number of foreign-born individuals as a percent of the U.S. population, between 1850 and 2020. Source: Author's calculations from full count and samples of the U.S. Census of Population, made available by IPUMS (Ruggles et al., 2021) and ICSPR (Haines, 2010).

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



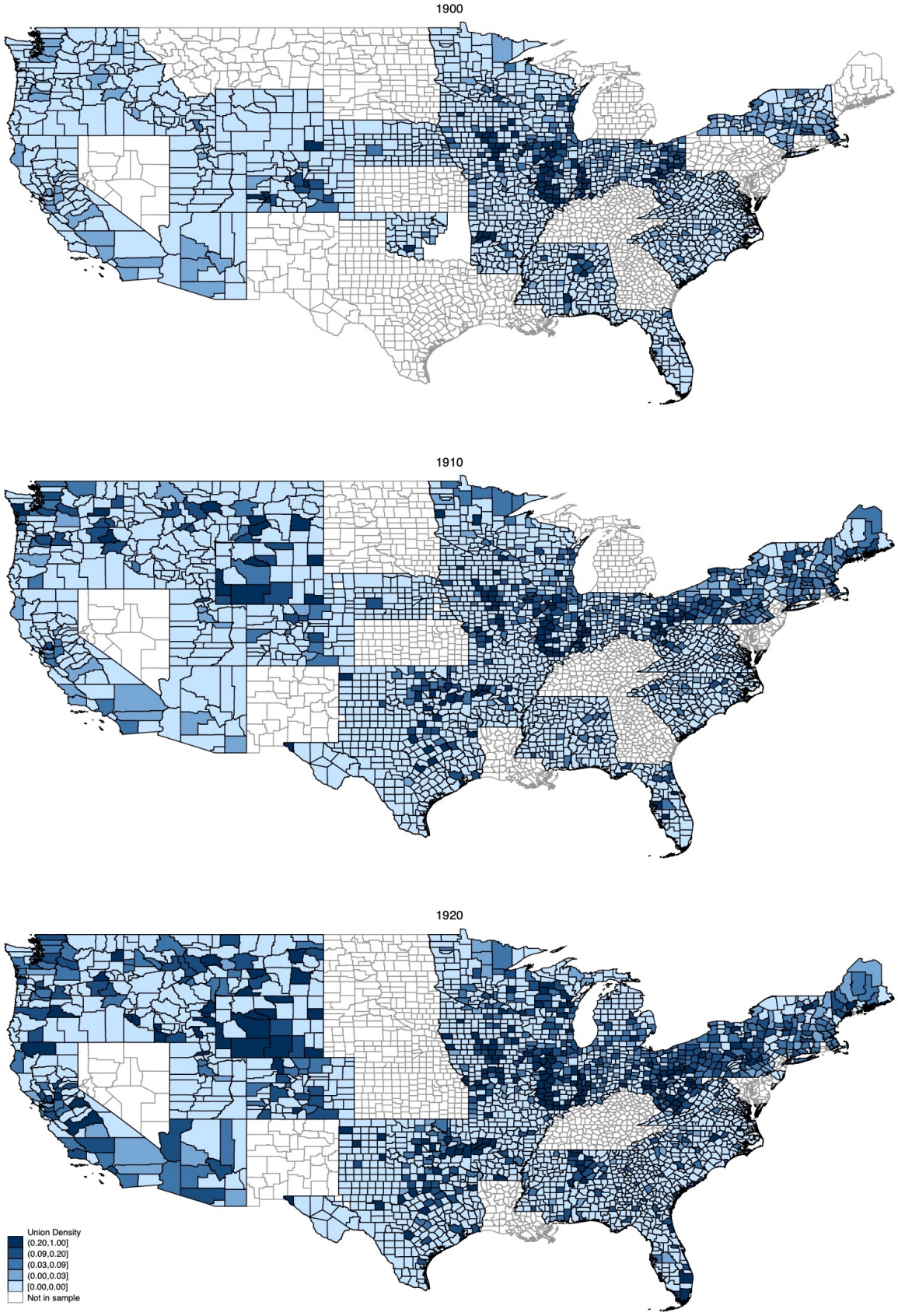
Notes: The figure shows the number of foreign-born individuals by region of origin, as a share of the total foreign-born population, between 1850 and 1920. Source: Author's calculations from full count U.S. Census of Population, made available by IPUMS (Ruggles et al., 2021) and ICSPR (Haines, 2010).

Figure 5: Geographic Distribution of Union Branches 1900–1920



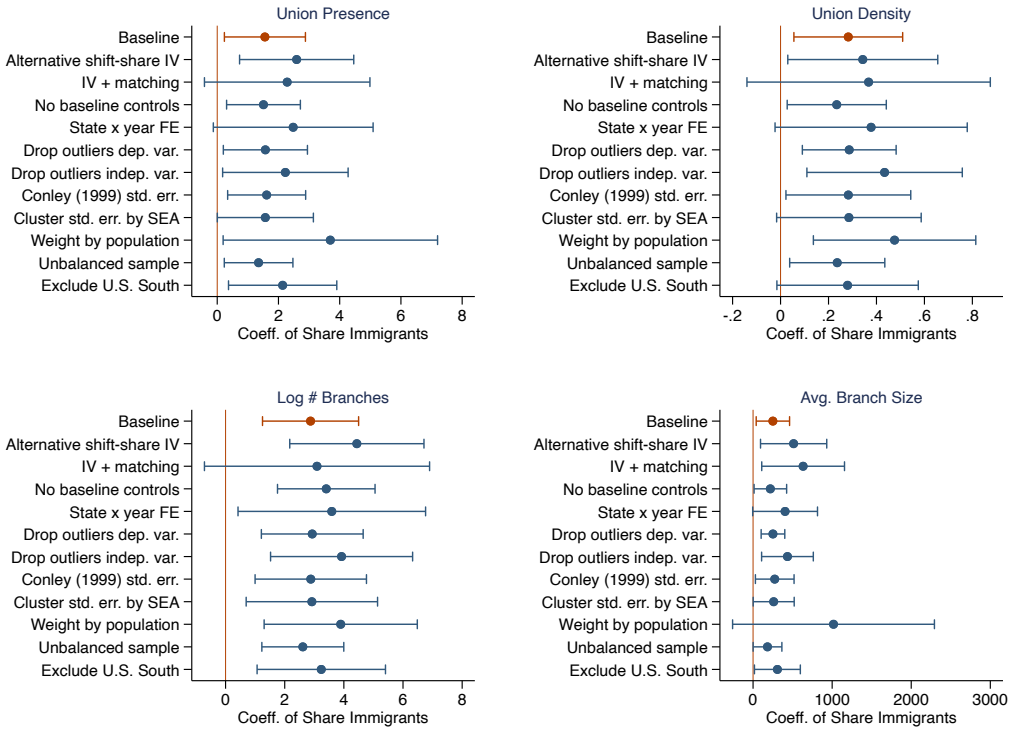
Notes: The map plots the all the union branches recorded and geocoded from the newly digitized labor union records described in Section 3.

Figure 6: Geographic Distribution of Union Density Over Time



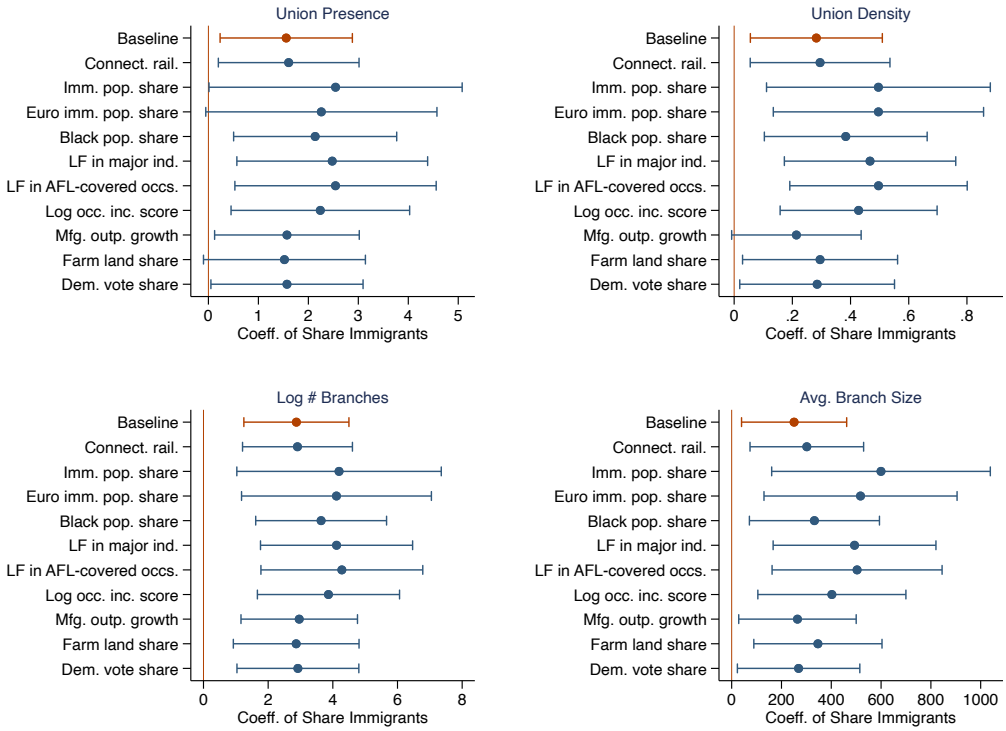
Notes: The maps plot the county-level shares of the union membership rate in 1900, 1910, and 1920. The legend shows the deciles with respect to the 1920 distribution. Source: Author's calculations from union convention proceedings, as described in Section 3.

Figure 7: Summary of Robustness Checks



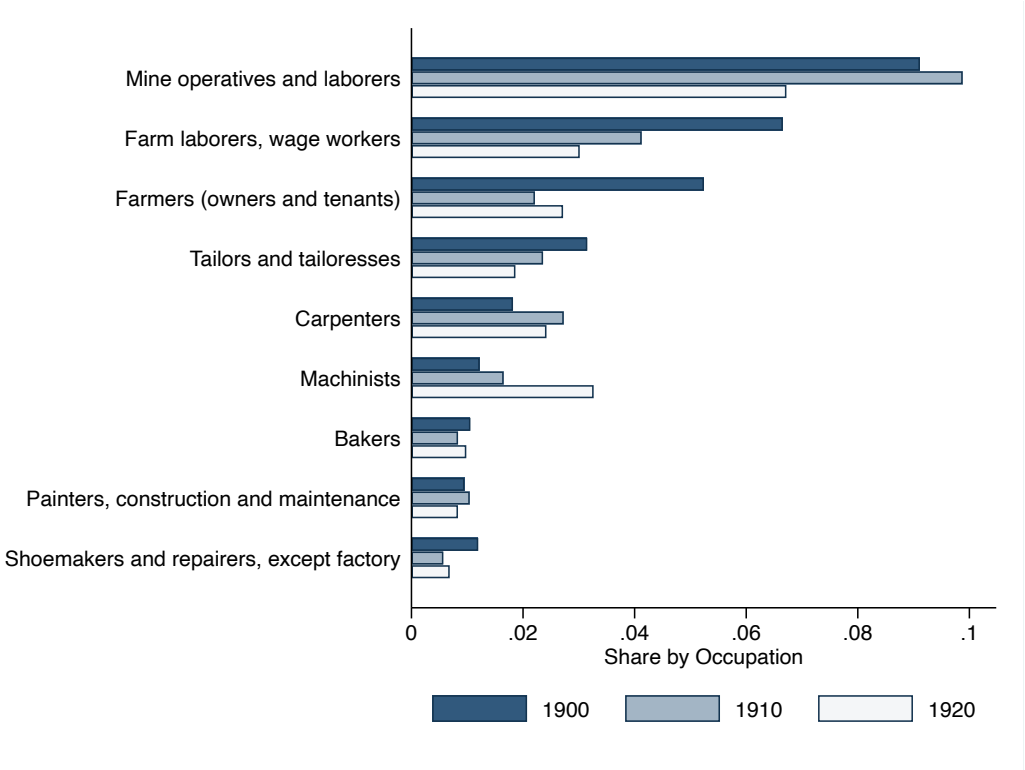
Notes: The figure presents a summary of the main robustness checks described in Section 5.2. The estimates plotted are the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, the main independent variable of equation (1). The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county. For more details and formal estimates, see also Appendix B.

Figure 8: Robustness Check – Controlling for Additional Baseline Characteristics



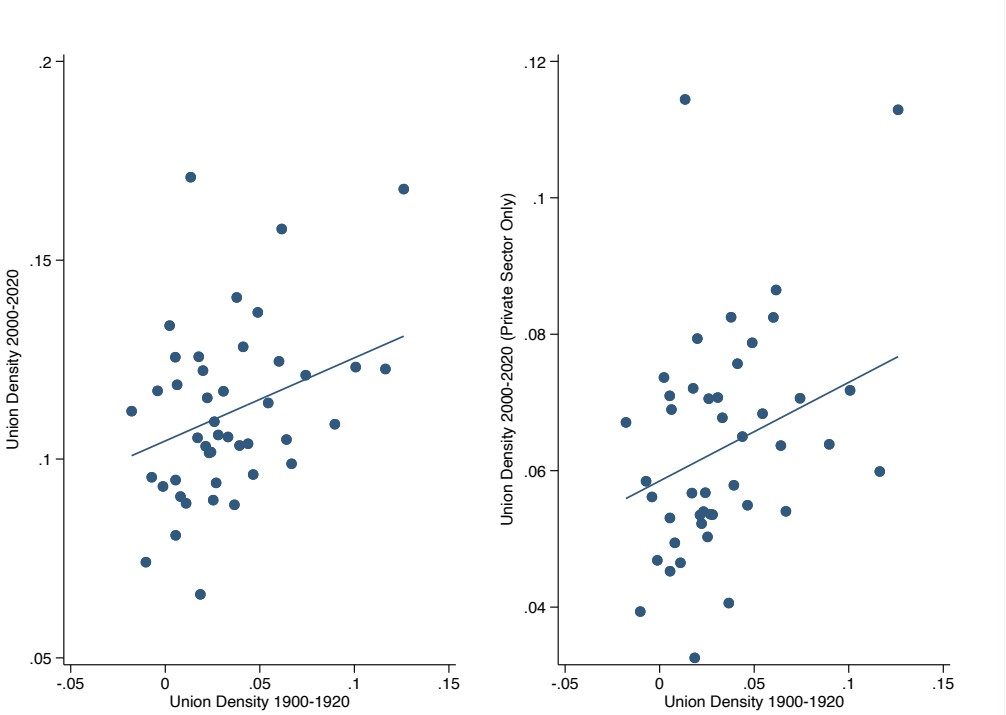
Notes: The figure plots the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, the main independent variable of equation (1), augmenting the specification of Table 3 with the variable(s) indicated in each row, measured at baseline and interacted with year dummies. The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county. For more details, see the description of the robustness checks in Section 5.2 and the formal estimates presented in Appendix B.

Figure 9: Prevailing Occupations Among Immigrants 1900–1920



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

Figure 10: Persistence of Unionization



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

Appendix – Table of Contents

A Additional Tables and Figures

Table A.1: Heterogeneous Effects by Workers' Skills – Intensive Margin

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

Notes: Observations are at the county-decade level. The sample is restricted only to counties that have some union presence in every year they are observed. The dependent variables are: union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 1); the log number of union branches (column 2); or, the average branch size, defined as the number of union members divided by the number of branches (column 3). In Panel A, the dependent variables are computed with respect to the AFL craft unions, which organized skilled workers only. In Panel B, with respect to the AFL industrial unions, which organized predominantly unskilled workers. See Section 6 for more details. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.2: Unionization and Immigration Flows

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

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

Table A.3: Heterogeneous Effects by Religion of Immigrants

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from non-Protestant or Protestant European countries who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 4.2 and Section 6. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.4: Heterogeneous Effects by Strength of Labor Movement in Country of Origin

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressors of interest are the number of immigrants (men 16–64) from European countries with a strong (UK-Ireland) and weak (other countries) labor movements as of 1870 (see Appendix E) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instruments used to predict them are described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. SW F-stat refers to the Sanderson-Windmeijer F-stat of the instruments in the two separate first-stage regressions. Square brackets report standardized coefficients. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.5: Effect on U.S.-Born Workers' Labor Market Outcomes

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

Notes: Observations are at the county-decade level. The dependent variables are the shares of the labor force participation rate among U.S.-born workers, men 16–64 (column 1), or the log of the average occupational income score among U.S.-born workers (column 2). The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.6: Effect on the Composition of Union Leaders

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

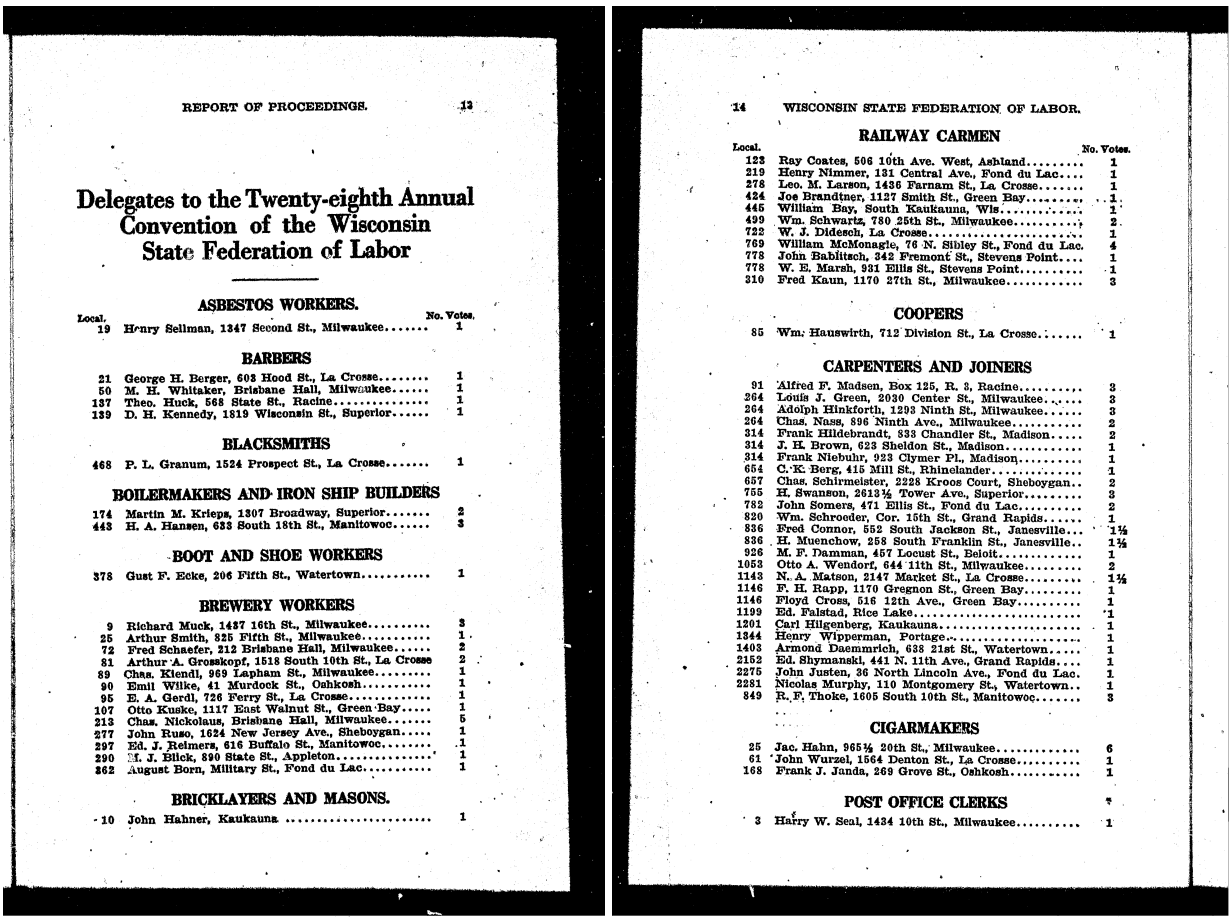
Notes: Observations are at the county-decade level. The dependent variable is the share of union delegates whose last name is of the origin (Panel A) or ancestry (Panel B) indicated in the column headings. The procedure used to infer the origin or the ancestry is described in Section C. The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 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. The regressions in Panel A also include year dummies interacted with the 1890 share of urban population and the 1880 male labor force participation rate. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table A.7: Effect on Local Economic Outcomes

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

Notes: Observations are at the county-decade level. Dependent variables are: the male labor force participation rate (column 1); the log of manufacturing output divided by the manufacturing labor force (column 2); the manufacturing output as a share of the total output in the U.S. in that year (column 3); or, the log of the total male labor force in occupations represented by the AFL craft unions (column 4). The main regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Square brackets report standardized coefficients. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

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. Klendl, 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. Hanswirth, 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. Schrooder, Cor. 15th St., Grand Rapids	1
836	Fred Connor, 552 South Jackson St., Janesville	1½
836	H. Muenchow, 258 South Franklin St., Janesville	1½
926	M. F. Damman, 457 Locust St., Beloit	1
1053	Otto A. Wendorf, 644 11th St., Milwaukee	2
1143	N. A. Matson, 2147 Market St., La Crosse	1½
1146	F. H. Rapp, 1170 Gregnon St., Green Bay	1
1146	Floyd Cross, 516 12th Ave., Green Bay	1
1199	Ed. Falstad, Rice Lake	1
1201	Carl Hilgenberg, Kaukauna	1
1344	Henry Wipperman, Portage	1
1403	Armond Daemmrich, 638 21st St., Watertown	1
2152	Ed. Shymanski, 441 N. 11th Ave., Grand Rapids	1
2275	John Justen, 36 North Lincoln Ave., Fond du Lac	1
2281	Nicolas Murphy, 110 Montgomery St., Watertown	1
849	R. F. Thoke, 1605 South 10th St., Manitowoc	3

CIGARMAKERS

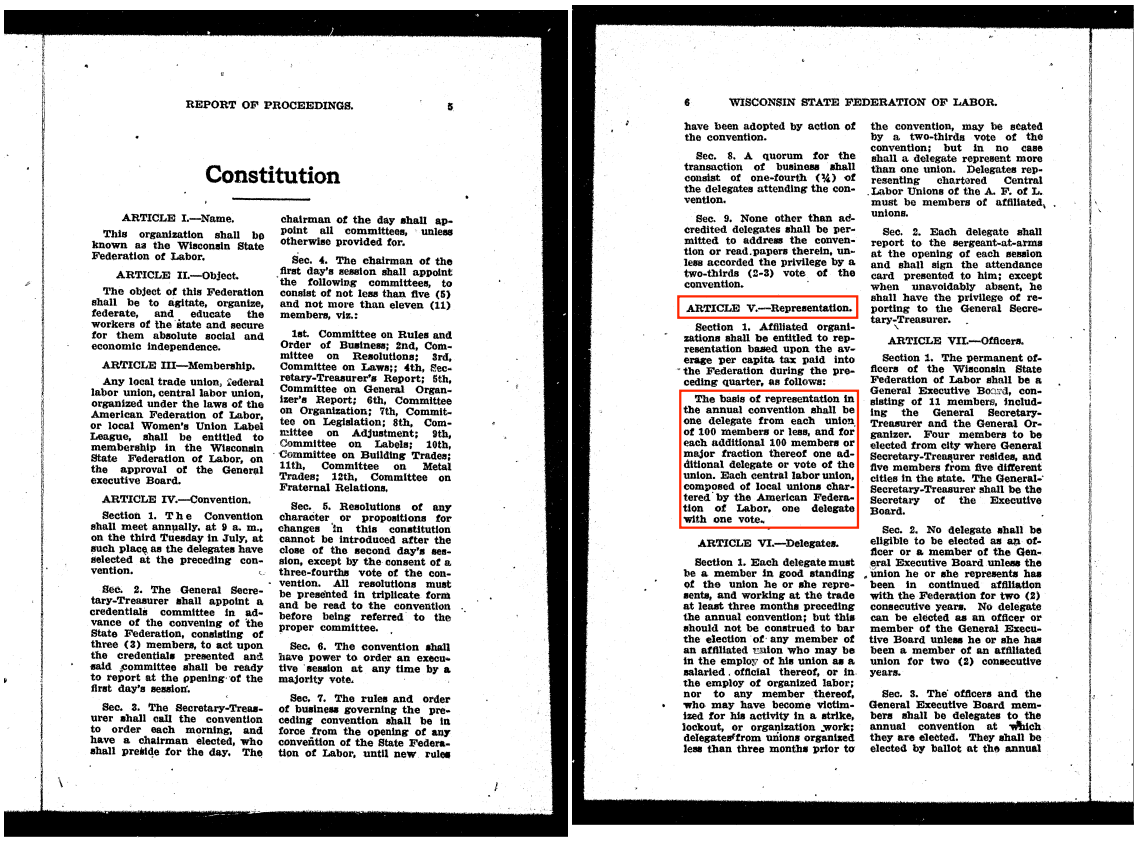
25	Jac. Hahn, 965½ 20th St., Milwaukee	6
61	John Wurzel, 1564 Denton St., La Crosse	1
168	Frank J. Janda, 269 Grove St., Oshkosh	1

POST OFFICE CLERKS

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

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

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



REPORT OF PROCEEDINGS. 5

Constitution

ARTICLE I.—Name.

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

ARTICLE II.—Object.

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

ARTICLE III—Membership.

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

ARTICLE IV.—Convention.

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

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

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

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

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

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

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

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

6 WISCONSIN STATE FEDERATION OF LABOR.

have been adopted by action of the convention.

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

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

ARTICLE V.—Representation.

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

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

ARTICLE VI.—Delegates.

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

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

ARTICLE VII.—Officers.

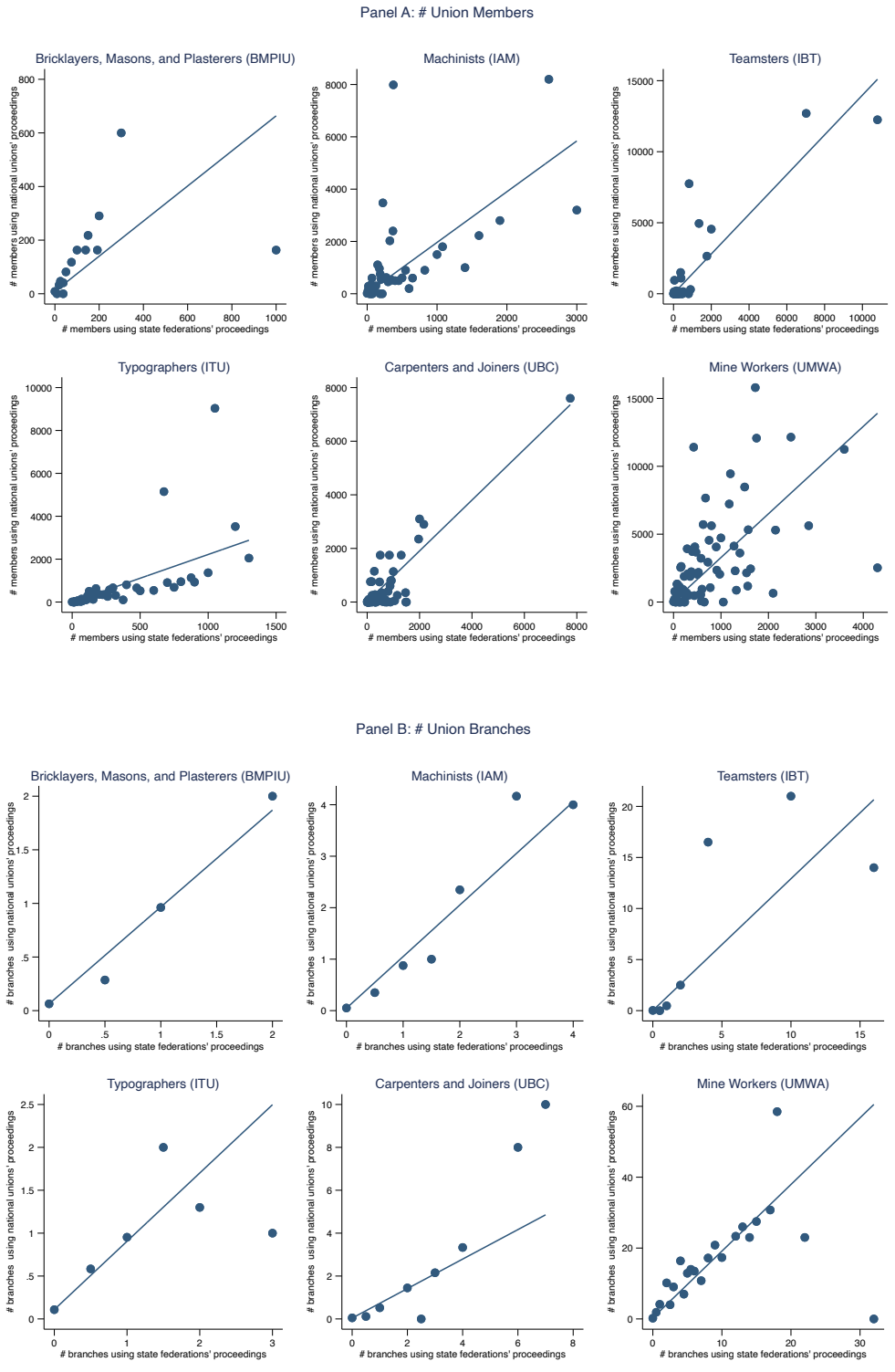
Section 1. The permanent officers of the Wisconsin State Federation of Labor shall be a General Executive Board, consisting of 11 members, including the General Secretary-Treasurer and the General Organizer. Four members to be elected from city where General Secretary-Treasurer resides, and five members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

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

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

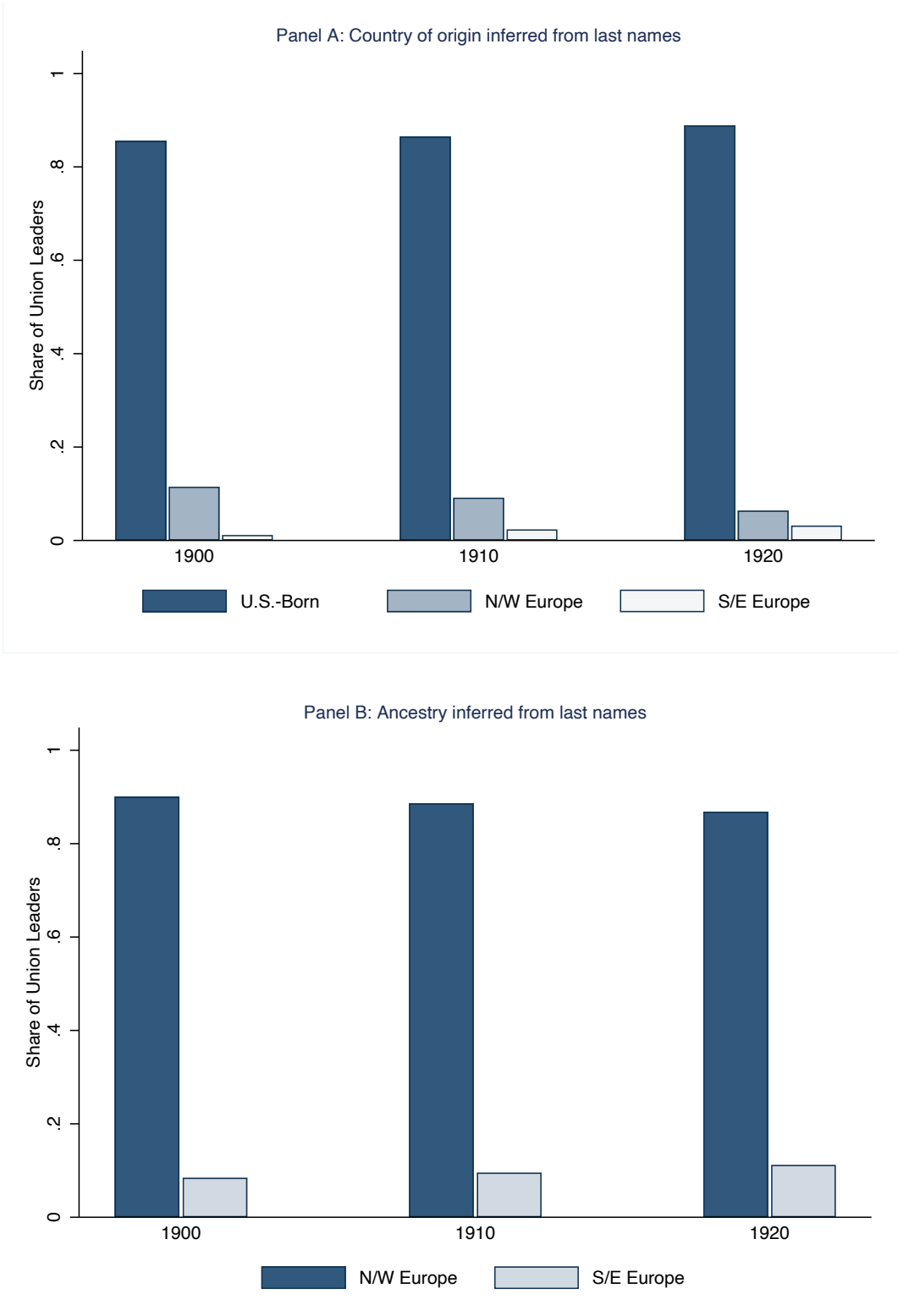
Notes: The figure shows a digitized document from the constitutions of the state federations of labor. The documents contain information on the rules that establish the number of delegates that local branches could send to the conventions. The highlighted paragraph on the page on the right provides an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

Figure A.3: Correlation Between Measures Across Data Sources



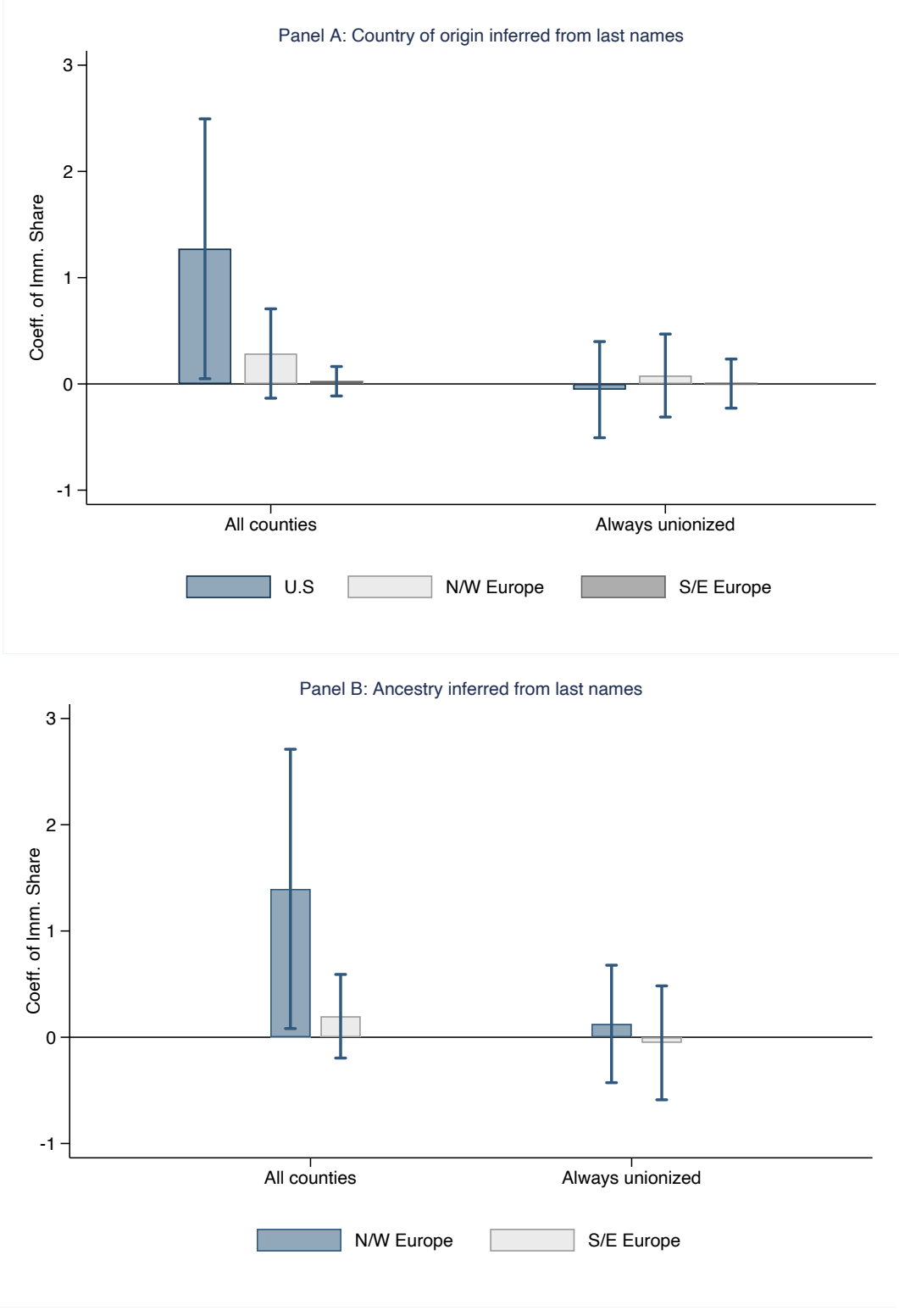
Notes: The figure shows binned scatter plots of the county-level union membership estimates (Panel A) and number of union branches (Panel B), constructed using the main data source (convention proceedings of the state federations of labor, on the x-axis) and the complementary data source (convention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

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



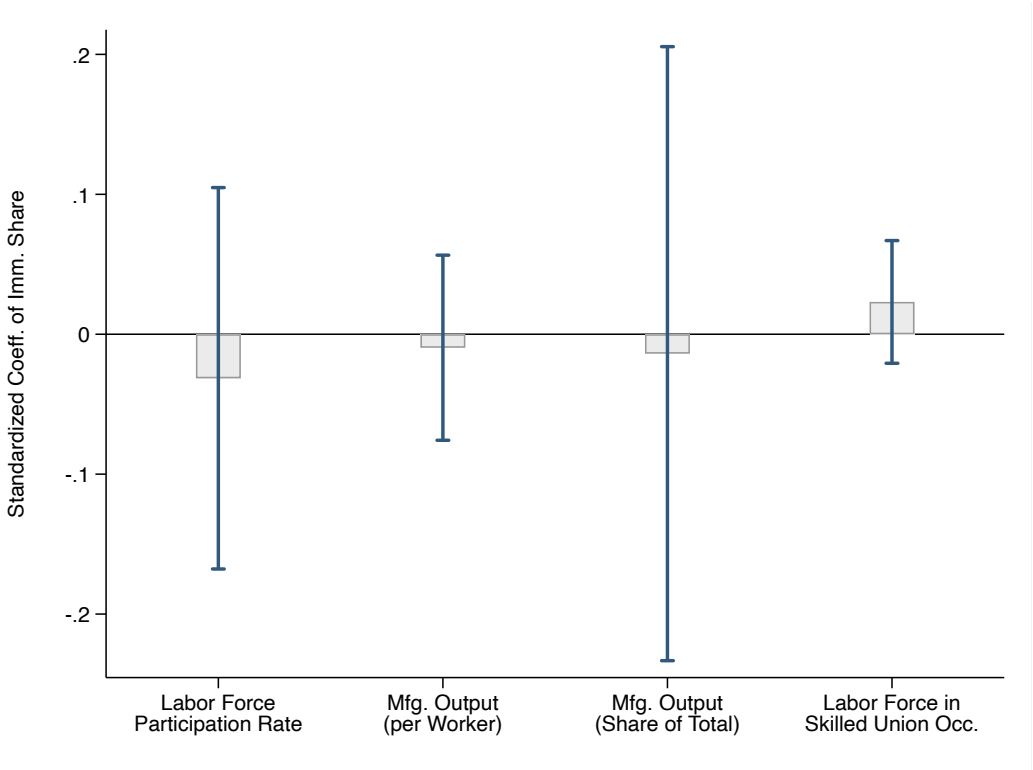
Notes: The figure plots the shares of union leaders of U.S.-born, Northern/Western Europe, and Southern/Eastern Europe origin (Panel A) and of Northern/Western and Southern/Eastern Europe ancestry (Panel B), at the beginning of each decade between 1900 and 1920. Union leaders are the delegates sent by 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.

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



Notes: Bars plot coefficients (with corresponding 95% confidence intervals) of a 2SLS regression of the share of union leaders with inferred country of origin (U.S., N/W Europe, S/E Europe, in Panel A) or ancestry (N/W and S/E Europe, Panel B) on the share of recently arrived immigrants. Union leaders are the delegates sent by the local union branches to the national convention of their union, or to the state conventions of the American Federation of Labor. On the left, the sample includes all counties as in Table 3 (in counties with no unionization, both shares are set to zero); on the right, the sample is restricted only to counties for which a union delegate is observed in every year. The country of origin and the ancestry are inferred from delegates' last names, as described in Appendix C Formal estimates are presented in Table A.6.

Figure A.6: Effects on the Local Economy



Notes: Bars plot standardized coefficients (with corresponding 95% confidence intervals) of a 2SLS regression of outcomes on the share of recently arrived immigrants. The dependent variables are: the male labor force participation rate; the log of manufacturing output divided by the manufacturing labor force; the manufacturing output as a share of the total output in the U.S. in that year; or, the log of the total male labor force in skilled occupations represented by AFL-affiliated national unions. See Table 3 for more details on controls, and sample, and Section 3 for the description of the data. Standard errors are clustered by county. Formal estimates are presented in Table A.7.

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),[49] 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.[50] For this exercise, my sample includes nineteen European countries for which immigration, weather, and crop data are available.[51] In my baseline specification, I consider temperature shocks, but results are unchanged if using precipitations.

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

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

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

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

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 outcomes. Panel A reports the baseline estimates of Table 3 using the main instrument, while Panel B displays the estimates from using the alternative instrument based on weather shocks. In either case, all coefficients are highly statistically significant and positive.

B.2 Matching Exercise

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

I present the results in Table B.3. In Panel A, I re-estimate my baseline specification of Table 3 for the counties that can be included in the county-pair strategy.[52] In Panels B and C, I re-estimate equation (1), replacing the baseline controls with fixed effects for the 800+ county pairs, interacted with year dummies. In Panel B, counties are matched on the number of Knights of Labor branches in 1890. In Panel C, on the one of 1880. The resulting coefficients identify the effect of immigration inflows on unionization for counties with nearly identical levels of union presence at baseline.[53] Despite the very demanding nature of this specification, reassuringly all the point estimates remain positive, large in magnitude, and similar to the baseline coefficients of Panel A.

B.3 Controlling for Additional Baseline Characteristics

In this section, I address the possibility 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 re-estimate my baseline specification by further controlling for several characteristics measured at baseline and interacted with year dummies. This exercise is meant to reduce the concern that factors that are jointly correlated with the 1890 size

[52] 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.

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

of immigration and with the development of labor unions between 1900 and 1920 may bias my estimates.

Connection to the railroad network. Previous work has shown that, between 1860 and 1920, the timing of the connection to the railroad network had a positive effect on both the inflow of immigrants to a county and on its economic growth in the medium- and long-run (Sequeira et al., 2020). Therefore, whether a county was crossed by a railroad or not may bias my estimates. To rule out this possibility, I use data from Atack (2016) to construct an indicator for whether each county in my sample was connected to the railroad network as of 1890, and interact this variable with year dummies (Table B.4, column 1).

Share of immigrant population. I directly control for the size of the 1890 immigrant population (total and European only), interacted with year dummies. This implies that the effects of immigration are identified exploiting variation only in the ethnic composition of immigrant enclaves across counties, holding constant the size of their foreign born populations. Since mechanically the instrument predicts higher immigration to counties with a larger stock of immigrants at baseline, by doing this I also address the concern that a larger 1890 immigrant population may itself have an independent and time-varying effect on unionization. Despite the highly demanding nature of this specification, all estimates remain statistically significant above the conventional levels (Table B.4, columns 2 and 3).

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

Labor force composition. I further control for the shares of the labor force in all major industries (agriculture, manufacturing, transportation, trade, manufacturing, and mining) and the share of the labor force in occupations covered by AFL-affiliated national unions in the period 1900–1920, all measured in 1890. These regressions therefore estimate the effect of immigration among counties with similar initial composition of workers across sectors. The results are all positive and statistically significant, and larger in magnitude (Table B.4, columns 5 and 6).

Average income and economic growth. Similarly, I control for the initial levels of average income (proxied by the occupational income score) and economic growth (measured by the growth rate of manufacturing output), to reduce any concern that counties with different economic conditions may have attracted more immigration earlier on and also witnessed a different growth of labor unions over time. The estimates are robust to the inclusion of these additional controls (Table B.4, columns 7 and 8).

Share of farm land. An additional concern is represented by the structural transformation away from agriculture towards manufacturing that occurred in the U.S. between 1880 and 1920 (Eckert and Peters, 2022). This may have implied larger growth rates for counties that were rural at the

beginning of the time period, with potential implications on the evolution of labor unions too. Although in the baseline specification I already control for the urban share of the population in 1890, I further include interactions between year dummies and the 1890 share of land in farms. The results are almost unchanged (Table B.4, column 9).

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

B.4 Additional Robustness Checks

Alternative baseline specification. Table B.5 reports results from using different specifications. In particular, in columns 1 to 3 I estimate less stringent specifications, by gradually including the two controls that are part of my preferred specification (the 1890 share of the urban population and the 1880 labor force participation rate). In columns 4 to 6, I do the same, while also always including state by year dummies, implying that the coefficients are estimated from changes in the fraction of immigrants within the same county over time, compared to other counties in the same state in a given year. The estimates are quantitatively and qualitatively unchanged.

Drop potential outliers. I verify that the results are robust to omitting counties with very large and very low levels of the dependent and independent variables, which could be potential outliers. In Table B.6, I re-estimate the baseline results dropping counties 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 computations of standard errors. In the baseline specification, standard errors are clustered at the county level. To address potential concerns of spatial correlation, in Table B.7 I verify that the precision of the estimates is unchanged when clustering standard errors at the SEA level (Panel A) and when computing Conley (1999) standard errors (with a 100km bandwidth).

Population-weighted regressions. I also re-estimate the results of Table 3 weighting the observations by total population, measured in the previous decade (Table B.8). By doing so, the estimates will return the effects for the average county. All coefficients remain positive, and if anything, are larger than the ones estimated with unweighted regressions. Except for column 4, whose coefficient is slightly above the conventional levels of significance, all other estimates are significant at either the 5% or 1% level.

Alternative samples. In Table B.9, I relax the restriction of having a balanced sample, and re-estimate the baseline specification on the full sample of counties for which the unions data are available. This yields a larger number of observations: 5,971 against the 5,025 of the baseline specification. All coefficients are remarkably similar to the ones of Table 3, both in terms of magnitude and significance. In Table B.10, I re-estimate the baseline regressions omitting the counties in the South. This exercise is motivated by the fact that this region of the U.S. received limited vol-

umes of immigration between 1900 and 1920, and also saw smaller labor unions' activity. Hence, a possible concern is that Southern counties may be driving the positive relationship between immigration and unionization. Reassuringly, even after dropping such counties, all estimates remain positive, statistically significant, and with magnitude similar to the ones reported in Table 3 (although less precisely estimated in some cases, due to the smaller sample size).

Alternative definitions of union density. The preferred definition of union density used throughout the paper is the number of union members divided by the total labor force in occupations covered by the AFL-affiliated national unions during the period 1900–1920, collected from Stewart (1926). This measure has the main advantage of not being influenced by the relative importance of such occupations in the labor force. In Table B.11, I show that the results are unchanged when using different definitions of the dependent variable. In particular, in column 2, the number of union members is divided by the total labor force in occupations covered by any labor union in existence during this period (regardless of whether it was affiliated with the AFL or not); and, in column 3, by the total labor force in any occupation not in the agricultural industry. As expected, the magnitudes change, but all coefficients remain statistically significant.

Test of pre-trends. The validity of the shift-share instrument defined by equation (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 Table 8 already reduce the concerns about this assumption being invalidated, in Table B.12, I test for pre-trends more directly, regressing the pre-period change (from 1880 to 1890) in several outcomes on unionization, population, and economic growth, on the 1900 to 1920 change in immigration predicted by the instrument. Panel A reports the coefficients from reduced form regressions. Panel B display 2SLS estimates, although conclusions from this second set of cofficients should be taken with caution, given the low F-stat. All regressions control for urbanization and labor force participation rate in 1880, in an analogous way to the specification of Table 3. The choice of the dependent variables is constrained by data availability. Given the absence of data on union membership before the sample period for which I construct my dataset, and the fact that the American Federation of Labor was constituted only in 1886, I measure unionization with the number of branches of the Knights of Labor (Garlock, 2009); for economic outcomes, I rely on the Census of Manufactures, and use information on the number of workers in manufacturing, the number of manufacturing establishments, and the value of manufacturing output, which are available for both 1880 and 1890 (Haines, 2010). Reassuringly, no coefficient of Table B.12 is statistically significant. These results indicate that, before 1890, European immigrants did not settle in counties that were already undergoing changes in union presence or in other economic variables.

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

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

Notes: Observations are at the county-decade level. The table reports the first stage of the alternative instrument described in Appendix B.1. The dependent variable is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The main regressor of interest is the predicted number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the 1890 male population in the county. All regressions include county and year fixed effects. The following controls, interacted with year dummies, are also included: the 1890 share of urban population (column 2); and, the 1880 male labor force participation rate (column 3). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.2: Alternative Shift-Share Instrument Using Predicted Immigration Flows

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

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

Table B.3: Matching Counties with Similar Union Presence at Baseline

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county, year, and county pair by year fixed effects. County pairs are matched within states on the 1890 number of Knights of Labor branches (from Garlock, 2009) divided by county population. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county-pair, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.4: Controlling for Additional Baseline Characteristics

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (Panel A); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (Panel B); the log number of union branches (Panel C); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (Panel D). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with the 1890 share of urban population, the 1880 male labor force participation rate, and the following variables: an indicator whether a county was connected to the railroad network in 1890 (column 1); the immigrant share of the population in 1890 (column 2); the European immigrant share of the population in 1890 (column 3); the Black share of the population in 1890 (column 4); the shares of the male labor force in the mining, manufacturing, construction, trade, transportation, and agricultural industries in 1880 (column 5); the share of the male labor force in occupations covered by AFL-affiliated national unions in 1880 (column 6); the log of the average occupational income score in 1880 (column 7); the growth rate of manufacturing output between 1880 and 1890 (column 8); the share of land used in farming in 1890 (column 9); the average vote share for the Democratic Party in the presidential elections of 1888 and 1892 (column 10). KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.5: Using Alternative Baseline Specifications

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

Notes: Observations are at the county-decade level. Dependent variables are: the number of union members divided by the male labor force in occupations represented by the American Federation of Labor (Panel A); the log number of union branches (Panel B); the number of union members divided by the number of branches, or zero if no union is present (Panel C); or, an indicator for whether the county has a positive union membership in any occupation (Panel D). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects. Columns 2, 3, 5, and 6 include year dummies interacted with the 1890 share of urban population. Columns 3 and 6 include year dummies interacted with the 1880 male labor force participation rate. Columns 4 to 6 include state by year fixed effects. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.6: Dropping Outliers

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. Observations below the 1st or above the 99th percentile of the dependent variable (Panel A), or of the independent variable (Panel B), are excluded from the sample. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.7: Computing Standard Errors with Alternative Procedures

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors are shown in parentheses. In Panel A, standard errors are robust and clustered by State Economic Area (SEA). In Panel B, standard errors are computed with the procedure described by Conley (1999) to account for spatial correlation, with a bandwidth of 100km. *** p<0.01; ** p<0.05; * p<0.1.

Table B.8: Weighting Counties by Population

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. Observations are weighted by the total population in the previous decade. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.9: Using an Unbalanced Sample

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

Notes: Observations are at the county-decade level. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.10: Excluding the South

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

Notes: Observations are at the county-decade level. The estimation sample is restricted to counties in the Northeast, Midwest or West regions. The dependent variables are: an indicator for whether the county has any labor union (column 1); union density, defined as the number of union members divided by the total male labor force in occupations represented by the American Federation of Labor (column 2); the log number of union branches (column 3); or, the average branch size, defined as the number of union members divided by the number of branches or zero if the county has no labor union (column 4). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.11: Alternative Definitions of Union Density

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

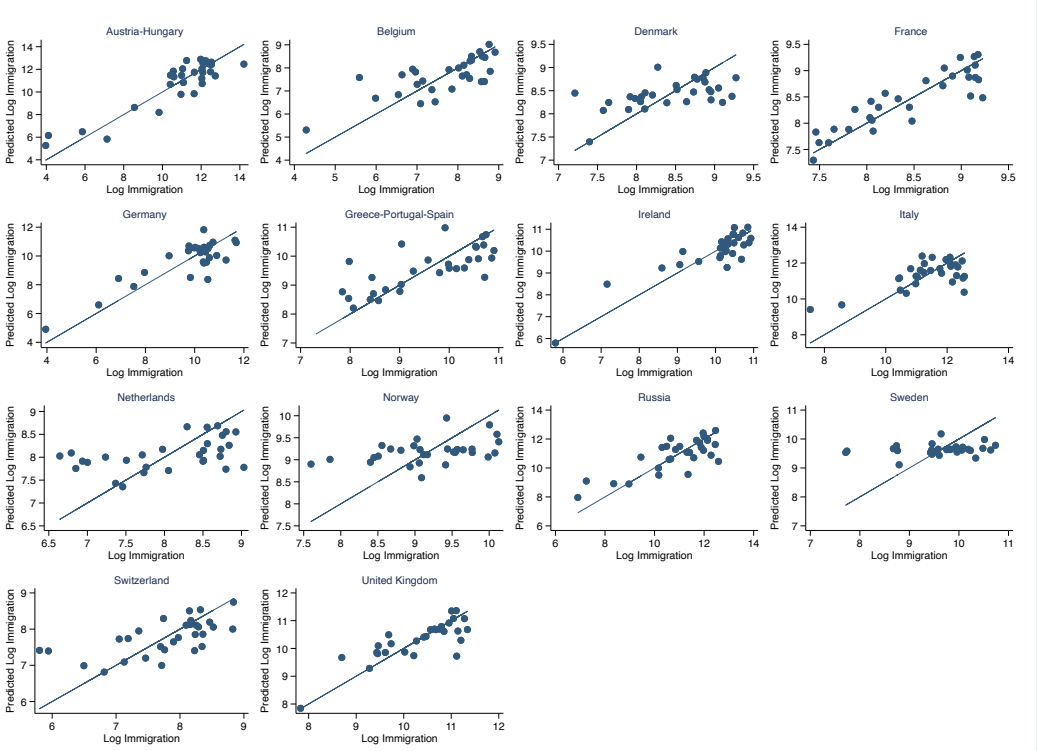
Notes: Observations are at the county-decade level. The dependent variables are the number of union members divided by: the total male labor force in occupations represented by the American Federation of Labor (column 1); the total male labor force in occupations represented by any labor union (column 2); the total male labor force in any non-agricultural occupation (column 3). The regressor of interest is the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county. The instrument used to predict it is described in Section 4.2. All regressions include county and year fixed effects, and year dummies interacted with: the 1890 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Standard errors, robust and clustered by county, are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Table B.12: Test of Pre-Trends in Unionization and Economic Outcomes

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

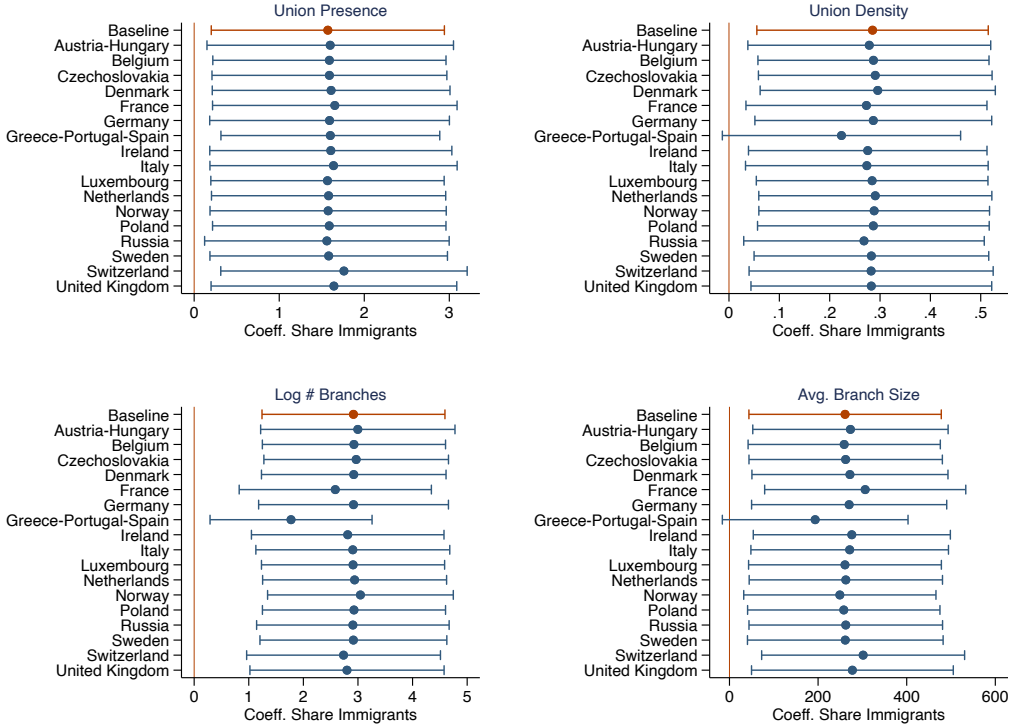
Notes: Observations are at the county level. The dependent variables are the 1880–1890 change in: the log of the number of Knights of Labor branches (column 1); the number of Knights of Labor branches divided by population (column 2); the log of population density (column 3); the share of the population employed in manufacturing (column 4); the log of the number of manufacturing establishments divided by the number of workers in manufacturing (column 5); the log of the manufacturing output divided by the number of manufacturing workers (column 6). The regressor of interest is 1900–1920 change in the number of European immigrants (men 16–64) who entered the U.S. in the previous decade, as a fraction of the male working-age population in the county, as predicted by the instrument described in Section 4.2. Panel A reports reduced form coefficients; Panel B displays 2SLS estimates. All regressions control for the 1880 share of urban population and the 1880 male labor force participation rate. KP F-stat refers to the Kleibergen-Paap F-stat for weak instruments. Robust standard errors are shown in parentheses. *** p<0.01; ** p<0.05; * p<0.1.

Figure 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 the European countries in the sample.

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



Notes: The figures plot the 2SLS coefficients (with corresponding 95% confidence intervals) of Share Immigrants, augmenting the specification reported in Table 3 with the 1890 immigrant share from each sending country (relative to all immigrants from that country in the U.S. in that year), separately. The first coefficient at the top of each figure (in orange) corresponds to that from the baseline specification. Standard errors are robust and clustered by county.

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 U.S. from foreign-born father.

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

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

D Index of Residential Segregation

In Section 6, I explore the heterogeneity of the effects of European immigration on unionization, by splitting counties above and below the sample median of the 1880 index of residential segregation of immigrants. In this section, I briefly described how the measure is constructed.[54]

First, I identify next-door neighbors from full-count U.S. Census data. Then, I follow the procedure described in Logan and Parman (2017), and I construct an indicator variable equal to one if a European immigrant has a next-door neighbor who is U.S.-born (from both U.S.-born parents).[55] 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.

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

[55] The original measure in Logan and Parman (2017) is constructed to compute an index of residential segregation for Black households. In my 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

Data on the development of labor unions in Europe used in Section 6.4 come from Crouch (1993). Estimates on union membership at the country level are available approximately every twenty or thirty years, starting in 1870. In most countries, the right to organize had been gained between 1860 and 1870, and was still often precarious. Similarly to the U.S., organization was limited to the skilled crafts and mining. At the turn of the 20th century, the only countries with an active and strong labor movement where the U.K. and Ireland. In 1900, there had been some, but limited, union activity also in Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, and Switzerland, although most of it had started only in the year 1900 or after (Crouch, 1993).

In Section 6.4, 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 my sample. The idea behind this exercise reflects the fact that individuals emigrating from countries with stronger unions may have been exposed to the experience of collective bargaining by the time they arrived in the U.S., and therefore might have been particularly interested in forming or joining labor unions in their new country. Table E.1 reports union membership at the national level for the years 1870 and 1900.

Table E.1: Union Membership in European Countries

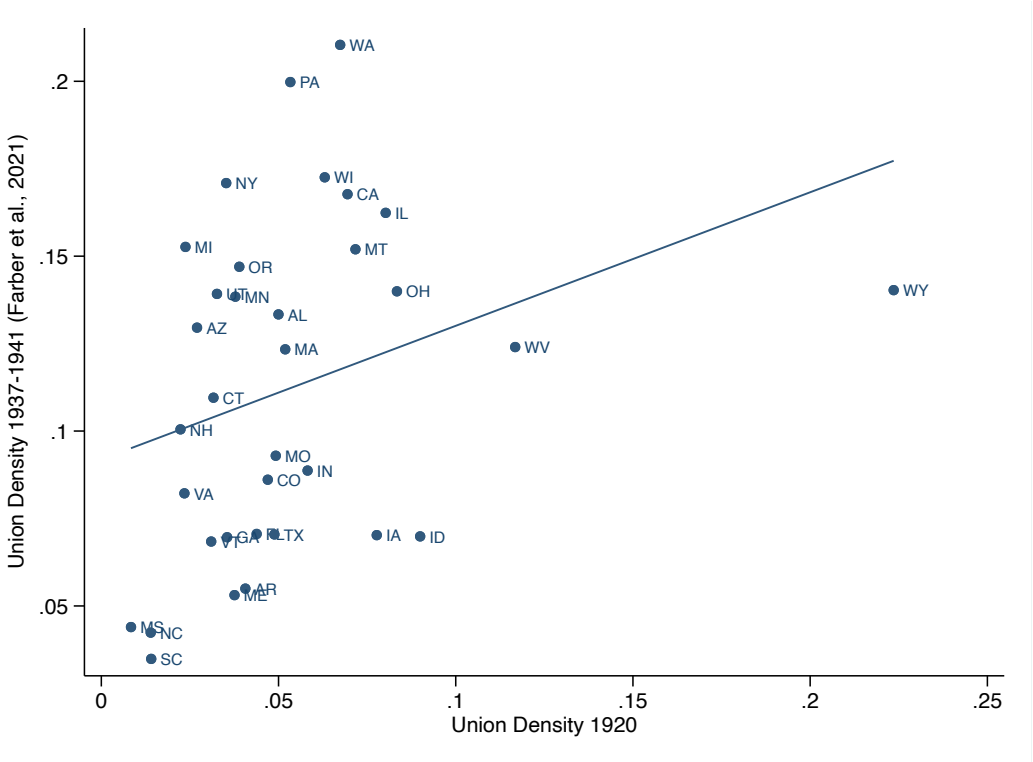
	Members (as % of LF)	
Country	1870	1900
Austria	0.28	1.00
Belgium	2.42	3.29
Denmark	0.54	8.76
France	0.20	2.99
Germany	0.39	3.40
Italy	n.a.	3.07
Norway	n.a.	2.30
Sweden	n.a.	2.53
U.K. and Ireland	8.32	12.50

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

F Dataset on Unionization

I provide a validation of my estimates of union density by investigating their correlation with the only other measures available in a historical period. This comes from Farber et al. (2021), who harmonize household-level survey data from Gallup starting in 1937. In Figure F.1, I show a scatter plot between the two measures. Since the data from Farber et al. (2021) are at the state level, I aggregate union membership in my data at the same unit and, to improve comparability with their measure, I divide it by the total non-agricultural labor force in the state. Unfortunately, the two sources do not overlap in time. Hence, I plot on the x-axis my measure in the last year of observation (Census year 1920) and the measure from Farber et al. (2021) as an average of the first five years of observations (1937–1941). Although the two measures do not agree in levels (and they are not expected to, since by 1937 several industrial unions affiliated with the Congress of Industrial Organizations had been constituted, which represented large masses of workers previously unorganized), the two measures display a positive correlation. The correlation coefficient is over 0.3, and approaches 0.4 once Wyoming (an outlier in the graph) is excluded from the sample.

Figure F.1: Correlation Between My Data and State-Level Gallup Data



Notes: The figure plots a scatter plot for state-level union density measured in 1920 using my 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.