The Union Threat[*]

Mathieu Taschereau-Dumouchel[†]
Cornell University

March 4, 2020

Abstract

This paper develops a search theory of labor unions in which the *possibility* of unionization distorts the behavior of nonunion firms. In the model, unions arise endogenously through a majority election within firms. As union wages are set through a collective bargaining process, unionization compresses wages and lowers profits. To prevent unionization, nonunion firms over-hire high-skill workers—who vote against the union—and under-hire low-skill workers—who vote in its favor. As a consequence of this distortion in hiring, firms that are threatened by unionization hire fewer workers, produce less and pay a more concentrated distribution of wages. In the calibrated economy, the threat of unionization has a significant negative impact on aggregate output, but it also reduces wage inequality.

JEL Classifications: J51, E24

[*] I thank the editor Michèle Tertilt and three anonymous referees for valuable suggestions. I am grateful to Nobuhiro Kiyotaki, Esteban Rossi-Hansberg and Oleg Itskhoki for their advice and support. I also thank Andy Abel, Lukasz Drozd, Zvi Eckstein, Pablo Fajgelbaum, Henry Farber, João Gomes, Lawrence Kahn, Morris Kleiner, Thomas Lemieux, Douglas Miller, Edouard Schaal, Aleh Tsyvinski, as well as seminar participants at Booth, Boston University, Chicago, Chicago Fed, Cornell/PSU Macro Workshop, CREI, HEC Lausanne, IIES, McGill, Minneapolis Fed, Penn, Princeton, SED Ghent, St. Louis Fed, UCLA, Université de Montréal, UQAM, Wharton and Yale for helpful discussions.

[†] Email: mt763@cornell.edu; Address: 480 Uris Hall, Cornell University, Ithaca NY 14853

1 Introduction

As unions are now covering only about 7% of private sector jobs in the United States, many observers have argued that their impact on the aggregate economy must be small. In opposition to this view, this paper investigates how unions can nonetheless have a sizable impact on the macroeconomy through their influence on *nonunion* firms. Indeed, if unionization lowers profits, like many studies find, vulnerable nonunion firms might distort their behavior to prevent their own unionization. Through that channel, unions may influence employment, wages and output in many nonunion firms and, therefore, have a larger impact on macroeconomic aggregates than the unionization rate alone would suggest.

To analyze this mechanism, this paper proposes a novel general equilibrium theory of endogenous union formation in which each firm hires multiple workers who differ in their productivity. In the model, unionization is simply a way for the workers to force the firm into a different wage setting mechanism. If a simple majority of the workers vote in favor of unionization, a union is created and wages are bargained collectively between the firm and its employees. If, instead, the vote fails to gather enough support, the firm remains union-free and wages are bargained individually between each worker and the firm.

By changing how the surplus from production is split, unionization generates a conflict between the firm and its employees. Indeed, since collective bargaining allows the workers to extract a higher share of the surplus, creating a union increases the average wage and lowers profits. But unionization also creates a second conflict, this time between the workers themselves. As collective bargaining compresses the distribution of wages, high-productivity workers tend to vote against the creation of the union, while low-productivity workers tend to vote in its favor. To avoid unionization, the firm can therefore hire more high-skill workers and fewer low-skill workers to increase the employees' opposition to the union and push the outcome of the vote in its favor.

This change in hiring in response to the threat of unionization is not motivated by production efficiency and leads to a higher marginal cost of production. As a consequence, threatened firms hire fewer workers, produce less and, because of decreasing returns to labor, pay higher wages. The threat also affects the variance of wages through the change in hiring. Since the firm overhires high-productivity workers, their marginal product goes down as do their wages. The opposite happens to low-productivity workers, and nonunion firms therefore pay a narrower range of wages in response to the threat of unionization.

In the model, the labor market is subject to search frictions so that it takes time for workers to be matched with vacancies. The unemployment rate is also affected by the union threat. In general equilibrium, as threatened firms hire fewer workers, the unemployment rate goes up and it takes more time for workers to find jobs. Since unemployment becomes less attractive, firms are able to extract a higher share of the production surplus which also pushes wages down.

The model provides a microfounded bargaining theory of unionization that is able to replicate important empirical facts associated with unions: i) union wages have a smaller variance and are on average higher than nonunion wages (Card et al., 2004), ii) the preference for unionization and the difference between union and nonunion wages decrease with skill (Farber and Saks, 1980), and iii) unionized firms are on average less profitable than their nonunion counterpart (Hirsch, 2004).

To quantify the impact of the union threat, I estimate the model using data from the private sector of the United States in 2005, and I use the parametrized economy to conduct three experiments in general equilibrium. In the first experiment, the formation of new unions is prohibited. As a result, nonunion firms no longer need to take action to prevent unionization and this first experiment therefore captures the impact of the threat of unionization alone, as the union status of the firms remain unchanged. In the new general equilibrium, output and the variance of log wages go up by about 1.2%, while the unemployment rate decreases by about 1.5 percentage points. If, in addition to removing the threat, all union firms are forced to become union free, the variance of log wages goes up by an additional 6.5%, but output and unemployment are not further affected. This second experiment therefore suggests that the threat of unionization on its own, more than the fact that some firms are actually unionized, might be a key channel through which unions affect output and unemployment in the U.S. economy. Finally, in the third experiment all firms are forced to be unionized. Comparing this new equilibrium to the calibrated economy, the variance of log wages goes down substantially while output and employment increase more than in the experiment in which unions were banned.

The paper also shows that often-used reduced-form estimators tend to underestimate the full impact of labor unions on wage inequality. For instance, the classical Freeman (1980) estimator finds that, in the calibrated economy, unions reduce the variance of log wages by 3.63% while their true impact is of 7.73%. More sophisticated estimators that take into account the heterogeneity between workers do worse by suggesting that unions lower the variance of log wages by only 0.72%. These large differences between reduced-form and model-based estimators can be partly explained by the threat of unionization, as it induces nonunion firms to pay a more equal distribution of wages. Standard estimators do not capture this channel.

The theory also provides an explicit mechanism to explain why some regression discontinuity studies, such as DiNardo and Lee (2004), find little impact of unionization on firms. These studies compare firms before and after unionization. But according to the theory firms before unionization are actively distorting their behavior in response to the threat. As a result, regression discontinuity estimators only capture part of the full impact that unions have on firm behavior.[1]

Finally, I provide supporting evidence for the mechanisms of the model by using the passage of right-to-work laws by some U.S. states as a source of variation in union strength. These laws prevent labor unions and employers from signing contracts that requires that workers pay union

[1] DiNardo and Lee (2004) also discuss how the union threat may contribute to their results.

membership fees as a condition of employment. As a result, under these laws unions have access to fewer resources, which leads to a weaker threat of unionization. Through a series of regressions, I find that the passage of a right-to-work law is associated with lower earnings for nonunion workers, suggesting that nonunion firms no longer feel the need to pay high wages to prevent unionization.

1.1 Literature review

Rosen (1969) was perhaps the first to mention that the threat of unionization could affect nonunion firms. Dickens (1986) considers the impact of the union threat on a firm's employment and wage level in a static environment in which workers can form coalitions to force the firms into specific work contracts. In contrast, the current paper proposes a dynamic, general equilibrium framework with heterogenous workers to evaluate the impact of the union threat on wage inequality, output and unemployment. Corneo and Lucifora (1997) also consider a model in which firms preemptively increase wages if they believe a union will force costly negotiations.

This paper is also part of a literature that includes labor unions in search models. Pissarides (1986) finds that introducing a monopoly union with control over the wage in a search framework might lead to efficiency. Alvarez and Veracierto (2000) study the impact of many labor market policies in a search model and find that unions who control hiring have adverse effects on unemployment and welfare. Ebell and Haefke (2006) and Delacroix (2006) investigate the interaction between union formation and product market regulations. Boeri and Burda (2009) look into the impact of an endogenous bargaining regime on economic activity. Açikgöz and Kaymak (2014) estimate the impact of a rising skill premium on the decline of union membership in the United States. Krusell and Rudanko (2016) have studied the dynamic problem of a monopoly union that sets wages with or without commitment. None of these papers investigate the impact of the threat of unionization on decision makers and the macroeconomy.

Several empirical papers find reduced-form evidence that the threat of unionization affects the behavior of firms. Part of that literature uses the passage of right-to-work laws across U.S. states as a source of variation in union strength. Farber (2005) finds that nonunion wages fell by 4.2% after the passage of a right-to-work law in Idaho in 1981. More recent work has shown that the threat remains active today. For instance, Manzo and Bruno (2017) investigate the impact of right-to-work laws that were enacted between 2012 and 2015 in Indiana, Michigan and Wisconsin. Controlling for a variety of factors, they find a decline of 2.3% in nonunion wages after the legislation passed.[2] Overall, this literature suggests that nonunion firms respond to the threat of unionization by raising wages, a finding consistent with the model presented in this paper and with up-to-date estimates of the impact of right-to-work laws provided in Section 5.2.

Other studies have used union densities as measures of the importance of the union threat.

[2] The impact of the right-to-work laws on nonunion wages is not reported in Manzo and Bruno (2017) but was communicated to me via private correspondence.

Hirsch and Neufeld (1987) find a strong positive relationship between union density and nonunion wages. Dickens and Katz (1987) use a principal component analysis to study interindustry wage differences and also find a positive relationship between union coverage and nonunion wages. In contrast, Neumark and Wachter (1995) find that an increase in union coverage is linked to lower nonunion wages at the industry level. They, however, find a positive relationship at the city level. In terms of wage dispersion, Kahn and Curme (1987) find a lower nonunion wage dispersion in more heavily unionized industries. Foulkes (1980) documents from survey data that, like in the model, large nonunion firms increase wages and working conditions preemptively to incentivize workers to vote against the formation of a union.

A literature also documents the negative impact of unionization on firm profitability. Lee and Mas (2012) use a regression discontinuity approach to show that, on average, unionization leads to a decline in the firm's equity value of $40,500 per unionized worker, which translates into a 10% decline in cumulative abnormal stock return. Such an important loss in firm value is indicative of the strong pressure on management to prevent unionization.[3]

Several studies documents that firms employ a wide variety of techniques, legal and illegal, and expand a lot of resources to prevent their own unionization (Dickens, 1983; Freeman and Kleiner, 1990; Bronfenbrenner, 1994). Plenty of anecdotal evidence also show the extent to which some firms are willing to go to avoid unionization. Wal-Mart, the largest employer in the U.S., has been known for its anti-union stance, providing a large amount of support to store managers for that purpose and going as far as shutting down stores after a successful union vote (Vieira, 2014). Recently, several private universities have improved graduate student salaries and benefits substantially in anticipation of a decision by the National Labor Relations Board allowing them to unionize (Elejalde-Ruiz, 2016; Flaherty, 2016).

The next section introduces the model. An explanation of how firms respond to the union threat follows. The model is then calibrated to the U.S. economy and experiments are conducted to evaluate the impact of the union threat. The following section discusses how the model relates to reduced-form estimators commonly used in the literature. The last section concludes.

2 Model

This section describes the model. Here is an overview of the main ingredients. The economy is populated by heterogeneous workers and heterogenous firms that meet through frictional labor markets to produce consumption goods. Before production takes place, workers have the option to unionize through a majority election. If the workers unionize, the surplus generated by production is split through a collective bargaining process between the firm and the workers. If instead the union

[3] Lee and Mas (2012) find that this abnormal return is larger in the later part of their sample, from 1984 to 1999, which suggests that the pressure to avoid unionization remained important even under lower unionization rates.

election fails, workers bargain individually with the firm. Because of the difference in bargaining protocol, unionization leads to different wages for the workers and profit levels for the firm.

2.1 Preferences and technology

Each worker is endowed with a skill level $s \in \mathcal{S} = \{1, \ldots, S\}$ which remains constant over time. The mass of workers of each skill is given by a vector $\boldsymbol{N} = \{N_s\}_{s \in \mathcal{S}}$, with $N_s > 0$ for all s. Workers live forever, are risk-neutral and discount future consumption at a rate $0 < \gamma < 1$.[4]

Each firm is endowed with a technology $j \in \mathcal{J} = \{1, \ldots, J\}$ that converts the labor provided by a vector of workers $\boldsymbol{g} = \{g_s\}_{s \in \mathcal{S}}$ into consumption goods according to the production function

$$F_j(\boldsymbol{g}) = A_j \left(\sum_{s \in \mathcal{S}} z_{j,s} g_s^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}\alpha_j} \tag{1}$$

where $A_j > 0$ is total factor productivity and $\sigma > 0$ is the elasticity of substitution between skills. The vector $\boldsymbol{z_j} = \{z_{j,s}\}_{s \in \mathcal{S}}$, with $z_{j,s} > 0$, determines the relative skill intensity in firm j and is normalized to sum to one. The parameter $0 < \alpha_j < 1$ describes the returns to scale of the production function. To keep the exposition simple, labor is the only factor of production in the benchmark model but it is straightforward to also add capital as an additional input, as is done in the quantitative model of Section 4. To avoid cluttering the notation, the subscript j is often omitted when this creates no confusion.

The technology that a firm operates, and in particular the returns to scale parameter α_j and the skill intensity vector $\boldsymbol{z_j}$, will determine its union status in equilibrium. Consider first the role played by α_j. Since firms operate decreasing returns to scale technologies, production involves a fixed factor whose returns, governed by α_j, accrue to the firm owner. When negotiations with the union break down, the firm remains idle and the returns to that fixed factor are lost. Changes in α_j therefore influence the bargaining strength of the union in negotiations with the firm. Another important determinant of the firm's union status is its skill intensity vector $\boldsymbol{z_j}$, which influences how many workers of each skill group the firm hires. Since low-skill workers will tend to vote in favor of unionization while high-skill workers will tend to oppose it, $\boldsymbol{z_j}$ directly influences the outcome of the union vote and how costly it is for the firm to prevent unionization.

2.2 Labor markets

The labor market is divided into S submarkets, one for each skill $s \in \mathcal{S}$, in which unemployed workers search for jobs and firms post vacancies. Workers only search in the labor market corresponding to their skill but firms are free to post multiple vacancies in multiple markets, at a unit

[4] Through the paper bold typeface is used to denote vectors.

cost κ. This segmentation of the labor markets by skill groups allows the firm to control precisely the skill composition of its workforce and, through this channel, to influence the union vote.[5]

In a submarket where u unemployed workers are searching and v vacancies are posted, $m(u,v)$ matches are created in a period. The matching function m is assumed to be strictly concave, strictly increasing and homogenous of degree one. By defining the labor market tightness as $\theta = v/u$, the probability that a vacancy is filled can be written as $q(\theta) = m(u,v)/v = m(1/\theta, 1)$, and the probability that an unemployed worker finds a job can be written as $p(\theta) = m(u,v)/u = m(1,\theta)$. Since search requires no effort, all unemployed workers are searching. At the end of each period, a fraction $\delta > 0$ of jobs are exogenously destroyed.

2.3 Firms

A firm that previously employed a vector $\boldsymbol{g_{-1}}$ of workers enters the current period with the $(1-\delta)\boldsymbol{g_{-1}}$ workers whose jobs were not randomly destroyed. It can then post a vector of vacancies $\boldsymbol{v} \geq 0$ to maximize its expected discounted profits. Since the firm is posting a continuum of vacancies in each labor market, a law of large numbers implies that the mass of successful matches is deterministic.

By defining the current-period profit as $\pi(\boldsymbol{g}) = F(\boldsymbol{g}) - \sum_{s\in\mathcal{S}} w_s(\boldsymbol{g})\, g_s$, where $w_s(\boldsymbol{g})$ is the wage of the g_s workers of skill s, we can write the recursive problem of a firm as

$$J(\boldsymbol{g_{-1}}) = \max_{\boldsymbol{v}\geq 0} \pi(\boldsymbol{g}) - \kappa \sum_{s\in\mathcal{S}} v_s + \gamma J(\boldsymbol{g}), \tag{2}$$

subject, for each j, to the law of motion for employment

$$g_s = (1-\delta)\, g_{-1,s} + v_s q(\theta_s),$$

so that current workers were either with the firm last period or are newly hired.

At a steady state, we can simplify this problem substantially. In this case, the firm has a fraction $1-\delta$ of its optimal employment at the beginning of a period and, because of the linear hiring costs, it immediately hires back to that optimal level. The constraint $\boldsymbol{v} \geq 0$ is therefore never binding and we have the following lemma.

Lemma 1. *In a steady-state equilibrium, the firm's dynamic problem is equivalent to*

$$\max_{\boldsymbol{g}} \pi(\boldsymbol{g}) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + \kappa(1-\delta)\gamma \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)}. \tag{3}$$

[5] As long as a firm has some control over the type of workers it hires, the threat of unionization will influence its decision. As such, the assumption of perfectly segmented markets is not necessary for the main mechanisms to operate.

Proof. All the proofs are in the appendix. □

This equation states that a firm sets its employment $\boldsymbol{g}$ to maximize its present-period profit (first term) net of some vacancy posting costs (second term), and taking into account that the $(1-\delta)\,\boldsymbol{g}$ workers that remains with the firm next period are lowering future hiring costs (last term).

2.4 Workers

In each period, a worker is either employed or unemployed. Employed workers lose their jobs with probability δ, in which case they become unemployed. The lifetime discounted expected utility of a worker of type s who is matched with a firm of type j and who is currently earning a wage w is therefore

$$V_{j,s}^{E}(w) = w + \gamma\left[\delta V_{s}^{U} + (1-\delta)\, V_{j,s}^{E}(w_{j,s})\right], \tag{4}$$

where V_s^U is the lifetime utility of being unemployed and $w_{j,s}$ is the equilibrium wage that the worker expects to receive next period if there is no job separation. Since wages are bargained every period, the negotiations with the firm are over the current wage w only. Both parties take the future equilibrium wage $w_{j,s}$ as given.

At the beginning of a period, an unemployed worker finds a job with probability $p(\theta_s)$. The expected value of this job is $\mathbb{E}\left(V_{j,s}^{E}\right)$, where the expectation is taken over all the vacancies, posted by different types of firms, in submarket s. If no job is found, the worker receives home production b_s, which is assumed to be increasing in s. The lifetime discounted utility of an unemployed worker is therefore

$$V_s^U = p(\theta_s)\,\mathbb{E}\left(V_{j,s}^{E}\right) + (1-p(\theta_s))\left[b_s + \gamma V_s^U\right]. \tag{5}$$

By combining the last two equations we can write the gain in utility provided by employment at wage w as

$$V_{j,s}^{E}(w) - b_s - \gamma V_s^U = w - c_{j,s}, \tag{6}$$

where

$$c_{j,s} = b_s + \gamma(1-\delta)\,\frac{(1-\gamma)\,V_s^U - w_{j,s}}{1-\gamma(1-\delta)} \tag{7}$$

is the net outside option of a worker s who is bargaining with a firm j. This convenient notation makes explicit the fact that the worker loses all potential future wages $w_{j,s}$ if the bargaining breaks down.

2.5 Wages

In the United States, the typical unionization process starts when a group of workers petition the National Labor Relation Board (NLRB) for a union recognition. If there is sufficient interest from employees, the NLRB makes a ruling on whether the workers that would be covered by the union share a “community of interest”. In practice, the coverage of the union is often at the enterprise level (Traxler, 1994; Nickell and Layard, 1999).[6] Then, the NLRB organizes a vote at the work site and a simple majority is required for the union to be certified as the exclusive bargaining agent of the workers. All work-related negotiations between the workers and the firm must then be conducted by the union.

The model incorporates these features of the institutional environment. The sequence of events that occurs once a firm has hired its new workers is shown in Figure 1. First, the workers vote to decide whether to form a union or not. Then, if the union vote is successful, wages are bargained *collectively*. The outcome of this bargaining is a wage schedule $\boldsymbol{w^{u}}\left(\boldsymbol{g}\right)$ and a profit function $\pi^{u}\left(\boldsymbol{g}\right)$. Instead, if the union vote fails to gather enough support, wages are bargained *individually*, which leads to a wage schedule $\boldsymbol{w^{n}}\left(\boldsymbol{g}\right)$ and a profit function $\pi^{n}\left(\boldsymbol{g}\right)$. Unionization is therefore a way for the workers to force the firm into a different wage setting mechanism.

Both individual and collective bargaining are modeled using Nash bargaining, but the surplus that is bargained over is different. In a union firm, the workers and the firm bargain collectively over the *total* surplus generated by all the workers. If an agreement on wages cannot be reached, the whole workforce leaves the firm and no production takes place. In a nonunion firm, each worker bargains individually with the firm over the *marginal* surplus he or she alone generates. If the bargaining fails, this specific worker goes to unemployment but the firm can still produce with the remaining workers. As we will see, this asymmetry between collective and individual bargaining interacts with the decreasing returns of the production function and has important consequences for profits and wages. It is the only difference between a union and a nonunion firm in the model.

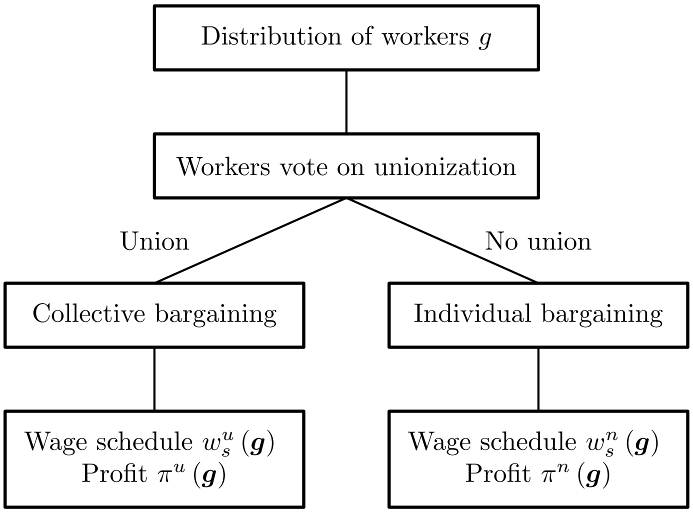

Distribution of workers g
Workers vote on unionization
Union
No union
Collective bargaining
Individual bargaining
Wage schedule $w_s^u\left(\boldsymbol{g}\right)$
Profit $\pi^u\left(\boldsymbol{g}\right)$
Wage schedule $w_s^n\left(\boldsymbol{g}\right)$
Profit $\pi^n\left(\boldsymbol{g}\right)$

Figure 1: Sequence of events after hiring

[6] The bulk of the literature models unions at the level of the production function. For recent examples, see Ebell and Haefke (2006) and Dinlersoz and Greenwood (2016).

Collective bargaining

Collective bargaining is modeled as an n-player Nash (1950) bargaining between the firm and all its workers.[7] If an agreement on a wage schedule $\boldsymbol{w}$ is reached, a worker s receives $V_s^E(w_s)$, otherwise he or she receives home production b_s today and starts the next period as unemployed, which has value γV_s^U. The net benefit of an agreement to a worker is therefore $V_s^E(\boldsymbol{w}) - b_s - \gamma V_s^U$. On the firm side, if an agreement is reached production takes place and wages are paid. Otherwise, the firm loses all its workers and needs to hire extensively next period to get back to its optimal size.

The following lemma formalizes this collective bargaining problem.

Lemma 2. *If all the workers have the same bargaining power, and the firm has bargaining power* $1-\beta_u$*, the collective Nash bargaining problem can be written as*

$$\max_{\boldsymbol{w}} \left[\prod_{s\in\mathcal{S}} \left(V_s^e(\boldsymbol{w}) - b_s - \gamma V_s^U\right)^{\frac{g_s}{n}}\right]^{\beta_u} \left[F(\boldsymbol{g}) - \sum_{s\in\mathcal{S}} w_s g_s + (1-\delta)\kappa\gamma \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)}\right]^{1-\beta_u} \tag{8}$$

where $n = \sum_{s\in\mathcal{S}} g_s$ *is the total mass of employed workers. Furthermore, the wage schedule*

$$w_s^u(\boldsymbol{g}) - c_s = \frac{\beta_u}{n}\left(F(\boldsymbol{g}) - \sum_{k\in\mathcal{S}} c_k g_k + \gamma(1-\delta)\kappa \sum_{k\in\mathcal{S}} \frac{g_k}{q(\theta_k)}\right) \tag{9}$$

solves this bargaining problem.

This collective bargaining problem is very similar to the usual 2-player bargaining. The first term between brackets in (8) can be interpreted as the surplus of the union; it takes the simple form of a geometric average of all the workers' individual surpluses. The second term between brackets is the surplus of the firm. Its interpretation is straightforward: if negotiations break down, the firm loses the current-period profit and pays a higher hiring cost tomorrow to compensate for the loss of the fraction $1-\delta$ of its current workforce that would have remained next period if negotiations had been successful.

From (9), it is straightforward to compute the current-period profit of a union firm employing workers $\boldsymbol{g}$ as

$$\pi^u(\boldsymbol{g}) = (1-\beta_u)F(\boldsymbol{g}) - (1-\beta_u)\sum_{s\in\mathcal{S}} c_s g_s - \beta_u(1-\delta)\kappa\gamma \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)}, \tag{10}$$

where c_s is given by (7).

Other works in the literature also rely on some form of Nash-bargaining to model wage setting

[7] Nash bargaining with more than two players is microfounded in axiomatic bargaining theory (Roth, 1979) and in game theory (Krishna and Serrano, 1996).

in union firms (Bauer and Lingens, 2010; Açikgöz and Kaymak, 2014). In general, the literature assumes that the firm enters a 2-player bargaining problem with some separate organization referred to as a union. Importantly, to properly define the bargaining problem the union must be endowed with its own preferences. In contrast, in the current model a union is simply the collective of the workers who are entering the n-person Nash bargaining with the firm. Each agent uses his or her own preferences and there is no need to model a union as a separate middleman between the workers and the firm. As a result, here, the "preferences of the union" are microfounded directly from the preferences of the individual workers.[8]

Individual bargaining

If, instead, the union fails to gather a majority of the votes, each worker bargains individually with the firm. In this case, the firm compares producing with and without that worker. Importantly, it understands that if that worker leaves, the marginal product of the remaining workers might change. In this case, these workers may want to reopen negotiations with the firm.[9]

In this context, the firm's marginal gain from employing an extra worker of type s is

$$\Delta_s^n(\boldsymbol{w}) = \frac{\partial F(\boldsymbol{g})}{\partial g_s} - w_s(\boldsymbol{g}) - \sum_{k \in \mathcal{S}} g_k \frac{\partial w_k(\boldsymbol{g})}{\partial g_s} + \gamma(1-\delta)\frac{\kappa}{q(\theta_s)}.$$

The first term is the extra output produced by the worker. The next term is simply the wage paid to the worker. The third term is the marginal effect of this worker on the wage of the other members of the workforce. Finally, the last term is the expected vacancy costs saved from retaining, with probability $1-\delta$, this worker into the next period.

Defining $0 < \beta_n < 1$ as the bargaining power of a nonunion worker, Nash bargaining implies that the nonunion wage vector $\boldsymbol{w}$ must solve the system of partial differential equations

$$\beta_n \Delta_s^n(\boldsymbol{w}) = (1-\beta_n)\left(V_s^E(\boldsymbol{w}) - b_s - \gamma V_s^U\right) \tag{11}$$

for all $s \in \mathcal{S}$ with the standard boundary conditions $\lim_{g_s \to 0} w_s^n(\boldsymbol{g})\, g_s = 0$ for all $s \in \mathcal{S}$.

The solution to this system is characterized in the following lemma.

[8] The literature generally assumes that the union maximizes the sum of the surplus going to the workers. With risk-neutral heterogenous workers, this assumption only pins down the total share of the surplus going to the workers, not how it is divided among them. In contrast, the current modeling assumption pins down what each worker receives. Another advantage of this approach is that it is robust in the sense that other bargaining environments yield the same outcome. For instance, Appendix E.2 shows that introducing a union organization between the workers and the firms leaves the wage equation (9) unchanged if the workers bargain collectively with the union.

[9] Stole and Zwiebel (1996) and Brügemann et al. (2019) provide theoretical foundations for this type of bargaining. Bertola and Garibaldi (2001) show that this bargaining procedure is broadly consistent with the empirical "relationship between employer size, the mean and variance of employees' wages, and the character of gross job creation and destruction." See also Cahuc and Wasmer (2001), Elsby and Michaels (2013) and Acemoglu and Hawkins (2014) for search models using this bargaining protocol. Appendix E.1 shows that the key mechanisms are preserved in an alternative model in which firms can pick nonunion wages unilaterally instead of through individual bargaining.

Lemma 3. *The wage schedule*

$$w_s^n(\boldsymbol{g}) - c_s = \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\partial F(\boldsymbol{g})}{\partial g_s} - \beta_n c_s + \beta_n \gamma (1-\delta)\frac{\kappa}{q(\theta_s)} \tag{12}$$

solves the individual bargaining problem (11)*.*

It follows directly that the current-period profit of a nonunion firm is

$$\pi^n(\boldsymbol{g}) = \frac{1-\beta_n}{1-(1-\alpha)\beta_n}F(\boldsymbol{g}) - (1-\beta_n)\sum_{s\in\mathcal{S}} c_s g_s - \beta_n(1-\delta)\kappa\gamma\sum_{s\in\mathcal{S}}\frac{g_s}{q(\theta_s)}. \tag{13}$$

Comparing collective and individual bargaining

The wage equations (9) and (12) derived from the collective and the individual bargaining problems have remarkably similar structures. They each consist of three terms that relate to production, the outside option of the workers and the hiring costs paid by the firm. They, however, differ in how these quantities influence wages. Indeed, the union wage (9) is mostly a function of the *average* characteristics of the workers, while the nonunion wage (12) is a function of the *individual* characteristics of each worker. In particular, the union wage of a worker of skill s depends on the average product $F(\boldsymbol{g})/n$ of all the workers in the firm while their nonunion wage is a function of the marginal product $\partial F(\boldsymbol{g})/\partial g_s$ of that worker alone.

This asymmetry has two important consequences. First, the presence of a union influences wage inequality within the firm, with union wages being naturally compressed. Second, the possibility of unionization creates a conflict between workers of different skills. Workers with valuable characteristics, for instance high marginal products, would rather bargain individually with the firm than share their high productivity with the other employees. As a result, high-skill workers are more likely to be against unionization than low-skill ones.[10]

The following proposition shows that unionization also creates a second conflict, this time between the firm and the workers.

Proposition 1. *If the bargaining powers are equal ($\beta = \beta_n = \beta_u$), the difference between the average nonunion and union wage is*

$$\mathbb{E}_s(\boldsymbol{w^n}(\boldsymbol{g})) - \mathbb{E}_s(\boldsymbol{w^u}(\boldsymbol{g})) = -\frac{\beta(1-\beta)(1-\alpha)}{1-(1-\alpha)\beta}\frac{F(\boldsymbol{g})}{n} < 0,$$

where $\mathbb{E}_s$ is the expectation across skills. It follows that the difference between nonunion and union profits is

$$\pi^n(\boldsymbol{g}) - \pi^u(\boldsymbol{g}) = \frac{\beta(1-\beta)(1-\alpha)}{1-(1-\alpha)\beta}F(\boldsymbol{g}) > 0.$$

[10] Verna (2005) discusses the literature on the relationship between worker productivity and pay in union firms. Consistent with the theory, pay is more correlated with ability and performance in nonunion firms.

This proposition shows that, for any set of workers $\boldsymbol{g}$, a firm prefers to bargain individually, while the workers, on average, would rather bargain collectively. This conflict between the workers and the firm is a direct consequence of the decreasing returns to scale in production. Indeed, as $\alpha \to 1$ the differences in profits and in average wages go to zero. To understand why, consider that when bargaining individually, the firm contemplates producing with or without the *marginal* worker. Because of diminishing returns to labor, this marginal worker has a relatively small impact on total production, limiting their possibility to bargain. The firm can then extract a large share of the total surplus. On the other hand, when the firm bargains with the union, the surplus is a function of the *total* production, which includes the relatively high output generated by the inframarginal workers. By forming a union, the workers can thus extract a bigger share of these high marginal products, which lowers the firm's profit.[11]

While Proposition 1 compares wages and profits when β_u and β_n are equal, the conflict between the workers and the firm naturally becomes more severe as β_u increases or β_n falls. In these cases, the workers are more strongly in favor of unionization while the firm's preference for remaining union free grows. As a result, workers are more likely to vote in favor of unionization and it becomes harder for the firm to incentives them otherwise. Together, the parameters α, β_u and β_n are therefore key determinants of the strength of the threat of unionization that the firm is facing.

2.6 Voting procedure

When the union vote takes place, workers know the wages they will get after either outcome of the vote, and the difference between these wages is the key driver of how they will vote. In addition, workers might favor or oppose unions for reasons unrelated to wages. For instance, workers with different political views might vote differently even if they face the same union and nonunion wages. To capture these additional voting motives, I assume that a worker of skill s votes for the union if and only if $w_s^u(\boldsymbol{g}) - w_s^n(\boldsymbol{g}) > \varepsilon$, where ε is a random variable with mean 0 that is drawn independently across workers in each period and that has a CDF ϕ with $\phi(0) = 1/2$.[12]

Since the firm employs a continuum of workers of each skill, a law of large numbers applies and a deterministic fraction $\phi(w_s^u(\boldsymbol{g}) - w_s^n(\boldsymbol{g}))$ of workers of type s votes for unionization. Denoting by

$$\Lambda(\boldsymbol{g}) = \sum_{s \in \mathcal{S}} g_s \phi(w_s^u(\boldsymbol{g}) - w_s^n(\boldsymbol{g})) - \frac{1}{2}n \tag{14}$$

the excess number of workers in favor of unionization, a firm is unionized if and only if $\Lambda(\boldsymbol{g}) > 0$.

[11] Proposition 1 is consistent with evidence from Kleiner (2001) showing that firms generally oppose unions. Freeman and Kleiner (1990) and Bronfenbrenner (1994) also detail various tactics used by firms to prevent unionization. Hirsch (2004) summarizes the literature on union and profitability and concludes that union firms are in general less profitable than firms that are not unionized.

[12] This random disutility term is not necessary for the results but helps to convexify the firm's optimization problem so that standard algorithms can be used to solve that problem easily. An earlier version of the model assumed no random preferences for unionization and found similar quantitative results.

Notice that the outcome of the vote is deterministic and that the firm therefore knows whether a union will be formed when it decides which workers to hire. As a result, the firm is effectively deciding its union status.

2.7 Steady-state equilibrium

In a steady state, the flows in and out of unemployment in each labor market must be equal. In submarket s, this implies the following relationship between the mass u_s of job searchers and the labor market tightness θ_s:

$$u_s = \frac{\delta \left(1 - p\left(\theta_s\right)\right)}{\delta + p\left(\theta_s\right)\left(1 - \delta\right)}. \tag{15}$$

We can now define a steady-state equilibrium in this economy.

Definition 1. A steady-state equilibrium is, a set of value functions $\left\{V_{j,s}^E, V_s^U\right\}_{s\in\mathcal{S},j\in\mathcal{J}}$, labor market tightnesses $\{\theta_s\}_{s\in\mathcal{S}}$, employment vectors $\left\{g_s^j\right\}_{s\in\mathcal{S},j\in\mathcal{J}}$ and wage schedules $\left\{w_s^j\right\}_{s\in\mathcal{S},j\in\mathcal{J}}$ such that,

1. the workers' value functions (4) and (5) are satisfied;
2. the employment vectors $\left\{g_s^j\right\}_{s\in\mathcal{S},j\in\mathcal{J}}$ solve the optimization problem of the firms given by (3) and where $\boldsymbol{w}\left(\boldsymbol{g}\right)$ is given by (12) if $\Lambda_j\left(\boldsymbol{g}\right) \leq 0$ or (9) otherwise;
3. the labor market tightnesses $\{\theta_s\}_{s\in\mathcal{S}}$ are consistent with stationarity (15) and with the hiring decisions of the firms, i.e. total vacancy posting in each submarket $s \in \mathcal{S}$ is $v_s = \sum_{j\in\mathcal{J}} \delta g_s^j / q\left(\theta_s\right)$.

3 Economic forces at work

We now investigate how the union threat influences the hiring decisions of a firm. As shown in Lemma 1, at a steady state, a firm's optimal employment decision solves

$$\max_{\boldsymbol{g}} \Pi\left(\boldsymbol{g}, \boldsymbol{w}\left(\boldsymbol{g}\right)\right), \tag{16}$$

where the objective function Π is defined as

$$\Pi\left(\boldsymbol{g}, \boldsymbol{w}\left(\boldsymbol{g}\right)\right) = F\left(\boldsymbol{g}\right) - \sum_{s\in\mathcal{S}} g_s w_s\left(\boldsymbol{g}\right) - \kappa\left(1 - \left(1 - \delta\right)\gamma\right)\sum_{s\in\mathcal{S}} \frac{g_s}{q\left(\theta_s\right)},$$

and where the wage schedule $\boldsymbol{w}$ is given by

$$\boldsymbol{w}\left(\boldsymbol{g}\right) = \begin{cases} \boldsymbol{w^u}\left(\boldsymbol{g}\right) & \text{if } \Lambda\left(\boldsymbol{g}\right) > 0 \quad \text{(the union vote succeeds)} \\ \boldsymbol{w^n}\left(\boldsymbol{g}\right) & \text{if } \Lambda\left(\boldsymbol{g}\right) \leq 0. \quad \text{(the union vote fails)} \end{cases}$$

In an economy in which unions are weak, perhaps because of a low bargaining power β_u, firms do not have to worry about unionization. They simply hire to maximize discounted profits under the nonunion wage schedule $\boldsymbol{w^n}(\boldsymbol{g})$. Denote this optimal hiring decision by $\boldsymbol{g^{n*}} = \text{argmax}_{\boldsymbol{g}} \Pi(\boldsymbol{g}, w^n(\boldsymbol{g}))$.

As the strength of unions increases, unionization becomes more attractive to the workers. If the firm keeps hiring according to $\boldsymbol{g^{n*}}$, there comes a point at which a majority of the workers will vote for unionization. The firm is then *constrained* by the unionization vote, and hiring according to $\boldsymbol{g^{n*}}$ is no longer optimal. In that situation, the firm modifies its hiring so that the workers reject the union. This new hiring decision, denoted by $\boldsymbol{g^n}$, maximizes $\Pi(\boldsymbol{g}, \boldsymbol{w^n}(\boldsymbol{g}))$ subject to the constraint that workers vote against the union, i.e. $\Lambda(\boldsymbol{g}) \leq 0$. Through this additional constraint, the hiring decisions of firms that are union free in equilibrium, as well as the wages they pay, are affected by the workers' option to unionize.

If the strength of unions increases even more, the firm contemplates letting its workers unionize. In this case, its profits would be $\Pi(\boldsymbol{g^{u*}}, \boldsymbol{w^u}(\boldsymbol{g^{u*}}))$, where $\boldsymbol{g^{u*}}$ is the optimal employment vector under collective bargaining. If preventing unionization is so costly that $\Pi(\boldsymbol{g^{u*}}, \boldsymbol{w^u}(\boldsymbol{g^{u*}})) > \Pi(\boldsymbol{g^n}, \boldsymbol{w^n}(\boldsymbol{g^n}))$, the firm chooses to be unionized as an optimal reaction to the threat of unionization.[13]

To understand how the threat affects decisions, it is useful to first consider an equilibrium in which the union status of the firms is given exogenously, such that no union vote takes place. In this case, we can characterize how a firm hires, the wages it pays as well as the workers' preference for unionization. We then take a step back and consider the full problem of a firm when unionization is endogenous—where the union threat matters. To focus the analysis on the empirically relevant cases, assume from now on that the value of unemployment W_s^u and the labor market tightness θ_s are increasing in s. These assumptions are satisfied in the calibrated economy presented in the next section.

3.1 Exogenous union status

We first consider the problem of a firm whose union status is exogenously given, such that the union threat has no impact on its behavior. This firm maximizes profits $\Pi\left(\boldsymbol{g}, w^i(\boldsymbol{g})\right)$ where the superscript i indicates whether the firm is unionized ($i = u$) or not ($i = n$). By defining

$$\text{MC}_s^i = (1 - \beta_i)\, c_s + (1 - \gamma(1 - \delta)(1 - \beta_i)) \frac{\kappa}{q(\theta_s)} \tag{17}$$

[13] It is possible to build examples in which $\Lambda(\boldsymbol{g^{u*}}) < 0$ but this usually requires extreme parameters.

as the marginal cost paid to hire a worker of skill s, the firm's optimal hiring decision $\boldsymbol{g^{i*}}$ solves

$$\mathrm{MC}_s^i = B_i \frac{\partial F(\boldsymbol{g})}{\partial g_s}, \tag{18}$$

where

$$B_i = \begin{cases} 1-\beta_u & \text{if } i = u \\ \frac{1-\beta_n}{1-(1-\alpha)\beta_n} & \text{if } i = n \end{cases}$$

is the share of revenues retained by the firm after bargaining. Equation (18) simply states that at the optimum the marginal revenue generated by an extra worker of type s is equal to the marginal cost of hiring that worker.

Solving (18), the optimal hiring decision g^{i*} is

$$g_s^{i*} = (\alpha A B_i)^{\frac{1}{1-\alpha}} \left(\frac{z_s}{\mathrm{MC}_s^i}\right)^{\sigma} \left(\sum_{k\in\mathcal{S}} z_k \left(\frac{z_k}{\mathrm{MC}_k^i}\right)^{\sigma-1}\right)^{\frac{1-\sigma(1-\alpha)}{(\sigma-1)(1-\alpha)}} \tag{19}$$

which, together with (17), shows that workers who search in tight labor markets (θ_s large) or who have attractive outside options (c_s large) are expensive to hire (MC_s^i large), and that the firm therefore relies less on them for production (g_s^{i*} small). When bargaining powers are equal, nonunion firms are also larger than union firms (since $B_n > B_u$) as they tend to hire more workers to lower their marginal products and thus pay lower wages.

The following proposition characterizes the wages paid by the firms.

Proposition 2. *The equilibrium wage schedules $w_s^u(\boldsymbol{g^{u*}})$ and $w_s^n(\boldsymbol{g^{n*}})$ are increasing in skill s, and the union wage gap $w_s^u(\boldsymbol{g^{u*}}) - w_s^n(\boldsymbol{g^{n*}})$ is decreasing in s.*

This proposition is consistent with a large empirical literature that finds that the union wage gap in the U.S. declines with income (Card, 1996; Card et al., 2004). It characterizes the *observed* wages that are paid in equilibrium, but not the workers' preferences about unionization. For those, we need to consider the counterfactual wages that the workers would receive if the union status of their firm were to change.

Proposition 3. *The counterfactual union wage gap $w_s^u(\boldsymbol{g^{i*}}) - w_s^n(\boldsymbol{g^{i*}})$ is decreasing in skill s for both firm union status $i \in \{u, n\}$.*

Proposition 3 characterizes the counterfactual, unobserved union wage gap that workers consider when casting their vote. It is consistent with work by Farber and Saks (1980), who show that the desire of a worker to be unionized goes down with her position in the intra-firm earnings distribution.

Propositions 2 and 3 are direct consequences of the individual and collective bargaining protocol outlined earlier. As individually bargained wages depend more on a worker's own characteristics, they tend to favor high-skill workers at the expense of low-skill workers.

3.2 Preventing unionization

We now consider the problem of a firm whose union status is endogenously determined by the vote of its workers. To maximize profits, the firm compares the optimal employment decision under which the workers unionize, $\boldsymbol{g^{u*}}$, to the optimal employment decision under which the workers reject the union, $\boldsymbol{g^n}$. In the latter case, the firm takes the voting constraint $\Lambda(\boldsymbol{g}) \leq 0$ into consideration such that $\boldsymbol{g^n}$ solves a modified version of the first-order conditions (18) that takes into account the impact of the marginal worker on the union vote. The new conditions are, for all $s \in \mathcal{S}$,

$$\mathrm{MC}_s^n + \lambda \frac{\partial \Lambda(\boldsymbol{g^n})}{\partial g_s} = B_n \frac{\partial F(\boldsymbol{g^n})}{\partial g_s}, \tag{20}$$

where $\lambda \geq 0$ is the Lagrange multiplier associated with the voting constraint $\Lambda(\boldsymbol{g}) \leq 0$. We see that this constraint effectively increases the marginal cost of hiring a worker who votes for the union.

We can expand the derivative of the voting constraint as follows:

$$\frac{\partial \Lambda(\boldsymbol{g})}{\partial g_s} = \underbrace{\phi(\Delta_s(\boldsymbol{g})) - \frac{1}{2}}_{\text{(a)}} + \underbrace{\sum_{s' \in \mathcal{S}} g_{s'} \frac{\partial \Delta_{s'}(\boldsymbol{g})}{\partial g_s} \frac{\partial \phi(\Delta_{s'}(\boldsymbol{g}))}{\partial \Delta_{s'}(\boldsymbol{g})}}_{\text{(b)}}, \tag{21}$$

where $\Delta_s(\boldsymbol{g}) = w_s^u(\boldsymbol{g}) - w_s^n(\boldsymbol{g})$ is the counterfactual union wage gap and where $\phi(\Delta_s)$ is, as before, the fraction of workers of skill s who vote for the union when the wage gap that they face is Δ_s. The terms (a) and (b) in (21) highlight two mechanisms through which hiring a worker of skill s influences the union vote.

(a) Direct impact on voting As a fraction $\phi(\Delta_s)$ of workers of skill s vote in favor of the union, adding an extra worker of this type directly increases the excess number of voters in favor of unionization by $\phi(\Delta_s) - 1/2$.

(b) Indirect effect through wages An extra hire of skill s also affects the union wage gap of the other workers in the firm ($\partial \Delta_{s'} / \partial g_s$) which, in turn, influences how they vote ($\partial \phi_{s'} / \partial \Delta_{s'}$). For instance, hiring an additional worker of skill s lowers the marginal product of all skill s workers. As a result, their nonunion wage also declines and, through this channel, these workers are more likely to vote for the union. An extra hire also changes union wages throughout the firm. Since union wages are driven by the average product, hiring a high-s worker shifts the union wage schedule upward which helps the union. If instead the firm hires many low-skill workers, their relatively low marginal product pushes the union wage schedule downward, which increases the number of votes against unionization.

To prevent unionization, the firm takes both of these channels into account. Channel (a), as it directly affects the union vote, is particularly important and is used by firms to their advantage. It pushes them to hire more high-skill workers, who vote against the union, and fewer low-skill workers, who vote in its favor. These changes in hiring then affect the wages paid by the firms.

Simplified Economy

In general, the problem of a firm constrained by the union vote must be solved numerically. We can however derive some analytical results in a static ($\gamma = 0$) economy in which there are only two types of workers (high-skill h and low-skill l) and in which workers have no random disutility from unionization ($\varepsilon = 0$). For tractability, assume also that the firm combines labor inputs using a Cobb-Douglas technology ($\sigma = 1$) and that there is no home production ($b_s = 0$ for all $s \in \mathcal{S}$).[14]

For the union threat to have an impact on the behavior of the firms, I also assume that the following assumption holds throughout this section.

Assumption 1. *$0 < B_u < B_n < 1$, $z_h q(\theta_h) < z_l q(\theta_l)$ and $z_h \geq z_l$.*

The first part of Assumption 1, $B_n > B_u$, implies that workers, as a group, prefer to be unionized. The second part, $z_h q(\theta_h) < z_l q(\theta_l)$, guarantees that the firm would hire more low-skill than high-skill workers in an environment without the voting constraint. The third part of the assumption, $z_h \geq z_l$, implies that, all else equal, high-skill workers are more productive than low-skill workers.

The following proposition establishes that the union threat influences the behavior of nonunion firms in this environment.

Proposition 4. *The union threat is binding for nonunion firms, i.e $\Lambda(\boldsymbol{g}) = 0$.*

Under the conditions of Assumption 1, low-skill workers have the majority of the votes in the union election. As these workers vote in favor of the union, the firm must distort its hiring decision, from $\boldsymbol{g^{n*}}$ to $\boldsymbol{g^{n}}$, to prevent unionization. The firm does so by over-hiring high-skill workers and under-hiring low-skill workers. The following proposition highlights the impact of this change in hiring on the size and profits of the firm.

Proposition 5. *The union threat lowers the profits, employment and output of nonunion firms.*

Intuitively, the voting constraint $\Lambda(\boldsymbol{g}) \leq 0$ distracts the firm from maximizing the production surplus and pushes it to use an inefficient mix of workers. As a result, the unit cost of production increases, which leads the firm to hire fewer workers to take advantage of the steeper part of the production function, and output declines.

The threat also affects wages, as the following proposition shows.

[14] Online Appendix A provides a numerical example of the behavior of a firm under threat in an economy with a full distribution of skills.

Proposition 6. *The union threat increases the average wage and decreases wage inequality, defined as the ratio of high-skill to low-skill wages, in nonunion firms.*

As the firm reduces its size in response to the threat, the average marginal product of the workers increases, which pushes wages higher. In addition, since the firm now hires a higher ratio of high-skill to low-skill workers, the marginal product—and therefore the wage—of high-skill workers falls relative to that of low-skill workers. As a result, wage inequality decreases when a firm is subject to the threat.

We can also use this simplified model to shed light on which type of firms are more likely to be unionized. For that purpose, it is useful to define $\Delta\Pi = \Pi\left(\boldsymbol{g^n}\right)/\Pi\left(\boldsymbol{g^{u*}}\right)$ as the ratio of a firm's profits if it successfully prevents unionization, $\Pi\left(\boldsymbol{g^n}\right)$, to its profits if its workers unionize, $\Pi\left(\boldsymbol{g^{u*}}\right)$. $\Delta\Pi$ therefore provides a natural measure of the gain in profitability associated with preventing unionization. In particular, if $\Delta\Pi < 1$ the firm optimally decides to become unionized.

The following proposition highlights how firms with different skill intensity vectors respond to the threat.

Proposition 7. $\Delta\Pi$ *is maximized when worker heterogeneity is minimal, in the sense that* $z_l q\left(\theta_l\right) = z_h q\left(\theta_h\right)$.

The intuition for this result is straightforward. The term $z_s q\left(\theta_s\right)$—the product of the intensity z_s of skill s in production with the inverse of a measure of how expensive these workers are to hire—captures how many workers of type s a firm wants to hire when it is unconstrained. If $z_l q(\theta_l)$ and $z_h q(\theta_h)$ are close to each other, the firm prefers to hire a similar number of workers of each skill. Under these circumstances, if the firm becomes subject to the voting constraint, it only needs to hire a few additional high-skill workers and a few less low-skill workers to prevent unionization. The distortion created by the constraint is therefore small and so is the loss in profits from preventing unionization. As the gap between $z_l q(\theta_l)$ and $z_h q(\theta_h)$ widens, the firm must depart more substantially from its optimal unconstrained skill mix and it becomes more costly to win the union vote.

A firm's labor intensity α also matters for its union status, as the following proposition shows.

Proposition 8. *There is a threshold* $\bar{\alpha} \in [0, 1]$ *such that* $\Delta\Pi > 1$ *for* $\alpha < \bar{\alpha}$ *and* $\Delta\Pi < 1$ *for* $\alpha > \bar{\alpha}$. *In addition, there is a threshold* $\hat{\alpha} \in [0, 1]$ *such that the firm cannot prevent unionization if* $\alpha < \hat{\alpha}$.

This result comes directly from the difference in bargaining protocol, and is reminiscent of Proposition 1. As the labor intensity α gets smaller, individual bargaining becomes increasingly attractive to the firm and, at some point, preventing unionization becomes profitable. The second part of Proposition 8 shows, however, that the firm might not be able to prevent unionization. For labor intensities below $\hat{\alpha}$, the gains in wages from unionization are so large that the workers vote for the union no matter what.[15]

Together, Propositions 4 and 8 show that, in this simple economy, firms with labor intensities in the interval $\alpha \in [\hat{\alpha}, \bar{\alpha}]$ would have their decisions affected by the threat of unionization while remaining union free. If the distribution of labor intensities has full support, a strictly positive mass of nonunion firms would therefore be affected. This last result shows that the threat of unionization can affect not only firms that are at the margin of being unionized, but also those for which unionization might seem a more distant prospect.

In addition to α and $\boldsymbol{z}$, the bargaining powers matter for whether the firm manages to prevent unionization, as the following proposition shows.

Proposition 9. *There are thresholds $\bar{\beta}_u \in [0,1]$ and $\bar{\beta}_n \in [0,1]$ such that the firm cannot prevent unionization if $\beta_u > \bar{\beta}_u$ or if $\beta_n < \bar{\beta}_n$.*

The bargaining powers β_n and β_u are key determinants of the strength of the union threat and, as such, influence whether the firm can prevent unionization. For instance, if β_u is large enough, union wages are so high that the firm cannot incentive the workers to vote against the union no matter what. Similarly, if β_n is small enough, nonunion wages are low and unionization is too attractive a prospect for the workers to vote against it.

3.3 Impact of the threat on social welfare

The threat of unionization also affects welfare by distorting the type of workers that the firm hires. To highlight the mechanisms at work, it is useful to consider the problem of a social planner that seeks to maximize steady-state welfare in this economy. To keep the analysis tractable I assume here that all jobs are destroyed at the end of a period ($\delta = 1$).

The planner's problem involves choosing the employment of all skill levels $s \in \mathcal{S}$ in all firms $j \in \mathcal{J}$ to maximize the social welfare function

$$\sum_{j \in \mathcal{J}} \left[F_j \left(\boldsymbol{g_j} \right) + \sum_{s \in \mathcal{S}} \left(N_s - g_{j,s} \right) b_s - \sum_{s \in \mathcal{S}} g_{j,s} \frac{\kappa}{q \left(\theta_s \right)} \right], \tag{22}$$

where N_s denotes, as before, the total mass of agents s in the economy.[16,17] The first term is the total amount of output produced, the second term is the home production of the unemployed, and

[15] The firm prevents unionization by hiring more high-skill workers and fewer low-skill workers. If $\alpha < \hat{\alpha}$ the firm will reach a point at which the nonunion wage of the high skill workers $w_s^n(g^n)$ is equal to their counterfactual union wage $w_s^u(g^n)$ while the low-skill workers still have a majority of the vote $g_l^n > g_h^n$. In this case, adding an extra high-skill worker would push their nonunion wage $w_s^n(g^n)$ under $w_s^u(g^n)$ and they would vote for the union. Removing a low-skill worker would push the union wage of the high skill workers $w_s^u(g^n)$ above their nonunion wage $w_s^n(g^n)$ and they would also vote for the union.

[16] Since all agents are risk-neutral, the planner can simply maximize total production in that economy and then use lump-sum transfers to redistribute that production among agents.

[17] Lemma A1 in the Appendix shows that we can normalize the mass of each firm to 1 by adjusting their total factor productivity A_j.

the last term corresponds to the costs that must be paid to hire the workers. The planner maximizes (22) subject to the constraint that the flows in and out of the labor market must equal each other at a steady state, i.e. (15) must be satisfied. Taking first-order conditions and simplifying, one can show that the planner's chosen allocation must satisfy

$$\frac{\partial F_j}{\partial g_{j,s}} - b_s - \frac{\kappa}{p'(\theta_s)} = 0 \tag{23}$$

for all $s \in \mathcal{S}$ and all $j \in \mathcal{J}$. Notice that the union status of the firm does not show up in (23). Indeed, unionization, on its own, only affects how the surplus from production is split between the workers and the firm. It does not affect how big that surplus is, which is what the planner cares about.

We can compare this equation to its equivalent in the decentralized equilibrium. Under our assumption that $\delta = 1$, the first-order condition (20) for hiring workers of skill s in a nonunion firm j becomes

$$\frac{1-\beta_n}{1-(1-\alpha)\beta_n}\frac{\partial F_j}{\partial g_{j,s}} - (1-\beta_n) b_s - \frac{\kappa}{q(\theta_s)} - \lambda_j \frac{\partial \Lambda(g_j)}{\partial g_{j,s}} = 0, \tag{24}$$

where, as before, $\lambda_j > 0$ is the Lagrange multiplier associated with the union vote constraint $\Lambda(\boldsymbol{g}) \leq 0$. Similarly, the first-order condition for a union firm j hiring a worker s is

$$(1-\beta_u)\frac{\partial F_j}{\partial g_{j,s}} - (1-\beta_u) b_s - \frac{\kappa}{q(\theta_s)} = 0. \tag{25}$$

Contrasting the equilibrium conditions (24) and (25) with the planner's allocation (23) reveals three sources of inefficiency in this economy. First, by comparing (23) and (25), we see that the employment decisions of a union firm coincide with those of the social planner when $\beta_u = -\theta q'(\theta) / q(\theta)$, which is the standard Hosios (1990) condition for efficiency in random search models. Under that condition, a fully unionized economy would be efficient. The second source of inefficiency comes from the individual bargaining protocol. By comparing (23) and (24), we see that even if the threat of unionization is inactive ($\lambda_j = 0$), there is no condition like Hosios' that would make these two equations coincide. Under individual bargaining, the firm seeks to lower wages by overhiring workers compared to what is socially optimal (Stole and Zwiebel, 1996). Finally, the threat of unionization ($\lambda_j > 0$) creates a third source of inefficiency. As we can see by comparing (23) and (24), it creates an additional wedge that leads to misallocation of employment across skill groups. Since high-skill workers vote against unionization, the threatened firms hire too many of them compared to what is socially optimal, and vice versa for low-skill workers.

4 Quantitative exploration

In this section, I first estimate the theoretical model using United States data. For this exercise, I assume that there is a distribution of heterogenous firms that differ in their technologies. Some of these firms will endogenously be unionized in equilibrium while others will be affected by the threat of unionization. With the estimated model in hand, I conduct several experiments to evaluate the impact of unions on the economy. Finally, I use the calibrated model to shed light on the rapid decline in unionization rates that the United States experienced over the last decades.

4.1 Specification and estimation

As we have seen in Section 3, the overall curvature of the production function matters for the impact of the threat on the decisions of the firms. To better capture this curvature in the data, I augment the model with capital and assume that firms now operate the technology

$$\tilde{A}_j \left[K_j^{1-\gamma_j} \left(\sum_{s \in \mathcal{S}} z_{j,s} g_s^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}\gamma_j} \right]^{\omega_j}, \tag{26}$$

where $\tilde{A}_j$ is total factor productivity, and γ_j and ω_j are the labor intensity and returns to scale parameters. I also assume that firms can access capital K_j frictionlessly at a constant interest rate $r > 0$.

Appendix B shows that once the firm has optimized over its capital input, the new production function (26) takes the same form as the one in the simpler model of Section 2. As a result, this change in production function is innocuous and all the mechanisms explored in the previous sections remain unchanged when capital is included as a factor of production.

Data

I use data on the private sector of the United States economy in 2005. Data on wages and the union status of workers come from the Merged Outgoing Rotation Groups of the Current Population Survey (CPS).[18] I combine these data with the Bureau of Economic Analysis (BEA) Annual Industry Accounts to calculate the labor share in the union and nonunion sectors of the economy, which will be targets in the estimation.[19]

[18] I clean the sample by removing agricultural, public sector and workers who are out of the labor force. I also remove individuals with an hourly wage higher than \$100 or lower than \$5, and individuals younger than 16 or older than 65.

[19] For each industry in the BEA dataset, I compute the labor share by dividing total workers' compensation by value added. I then associate each worker in the CPS sample with the labor share of the industry in which they are currently working. By averaging this variable separately over all union and nonunion workers, I find a labor share of 0.597 for union firms and of 0.613 for nonunion firms.

To build a skill index for each worker, I follow Card (1998) and regress log monthly *nonunion* wages on a set of worker characteristics. Denoting by w_i the monthly wage of a worker i who is working in industry j, the regression is

$$\log w_i = \Lambda X_i^1 + \Psi X_{i,j}^2 + \varepsilon_i,$$

where X^1 contains indicator variables that reflect characteristics that are intrinsic to the worker (age, education, occupation, race and sex) while X^2 contains indicator variables that are related to the job in which the individual currently works (industry and U.S. state). I then construct the skill index $\hat{s}_i$ of worker i as the predicted values associated with the intrinsic variables X^1, so that $\hat{s}_i = \exp\left(\hat{\Lambda} X_i^1\right)$. This way of constructing the index isolates the impact of variables intrinsically related to the individual, and therefore more associated with a notion of skill, from match-related factors that could also influence the wage. Notice that even though the regression is run on nonunion workers only, the predicted values $\hat{s}_i$ are computed for *all* members of the labor force.[20] The support of the distribution is then split into $S = 6$ bins of equal size, which is enough to observe the impact of union policies across skills while keeping the computational complexity at a reasonable level.

Parameters calibrated directly

Several parameters, mostly about preferences and the workings of the labor market, are taken directly from the literature. The remaining parameters—the bargaining powers and technology parameters—are key determinants of the strength of the threat and are estimated directly from the data using a method of simulated moments.

To reflect the typical duration of labor contracts the time period is set to one year. All monetary amounts are measured in thousands of 2005 dollars. The discount rate is set to $\gamma = 0.95$ and the job destruction rate is set to $\delta = 0.113$ to match the job destruction rate in the data (Davis et al., 1998). For the matching function, I use the functional form of den Haan et al. (2000) along with the parametrization of Krusell and Rudanko (2016) so that $m(u, v) = uv/(u + v)$. The elasticity of substitution between skills is set to $\sigma = 1.5$ as in Johnson (1997) who summarizes estimates from the literature. For the cost of posting a vacancy, I follow Silva and Toledo (2009) who document that training and vacancy posting costs amount to 69% of quarterly wages. This translates to $\kappa = 1.8$ in the model.[21] The value of these parameters is listed in Table 1. Finally, I use a linear

[20] This approach implicitly assigns to unemployed workers the average occupation and the average industry, in terms of their contribution to skill. An alternative regression that does not include occupation and industry yields a similar skill distribution.

[21] Alternative calibrations with a higher job destruction probability of $\delta = 0.4$ and lower vacancy costs, equivalent to 14% of quarterly wages, find a similar impact of labor unions on the economy. The benchmark parameters offer the best fit of the model to the data.

function to approximate the CDF ϕ that describes the workers' preference for unionization.[22] The parameters taken directly from the literature are listed in Table 1.

Definition	Parameter	Value	Source/target
Discount factor	γ	0.95	5% annual interest rate
Job destruction probability	δ	0.113	Davis et al. (1998)
Skill elasticity of substitution	σ	1.5	Johnson (1997)
Cost of posting a vacancy	κ	1.8	Silva and Toledo (2009)
Number of skills	S	6	See text

Table 1: Parameters calibrated directly

In the model, the value of non-work activities $\boldsymbol{b}$ takes into account unemployment insurance, home production and the value of the extra leisure provided by unemployment. Krueger and Meyer (2002) describe unemployment benefits in major U.S. states and find that the average replacement rate is 54% up to an annual maximum of about \$19,280 in 2005 dollars, when averaged across states. I include these benefits in $\boldsymbol{b}$. To capture the components associated with home production and leisure, I also include a second term in $\boldsymbol{b}$ that scales linearly with the average wage of the worker. I set the slope of this term so that the average value of non-work activities across workers amounts to 85% of the average wage as in Hall (2009).

Technologies

There is a distribution of firms, each using a technology indexed by $j \in [0,1]$.[23] These technologies differ in terms of the curvature α_j of their production functions, their skill intensity vectors $\boldsymbol{z_j}$, and their total factor productivities A_j. As a normalization, I order firms such that a larger j indicates a larger α_j. In addition, I assume that $\alpha_0 = 0.55$ and $\alpha_1 = 0.95$ so as to cover a broad range of curvatures.[24] Finally, as Lemma A1 in the appendix shows, we can normalize the mass of each type of firm j to one and let A_j determine the importance of that technology in the economy.

The estimation will determine the technologies used by the firms, but I impose some functional forms to limit the dimensionality of the parameter space to explore. The skill intensities $\{\boldsymbol{z}_j\}_{j\in[0,1]}$ are modeled as probability density functions of truncated log-normal distributions with mean parameters $\{\mu_j\}_{j\in[0,1]}$ and variance parameters $\{\xi_j\}_{j\in[0,1]}$ that vary linearly with j, such that

[22] This linear approximation has several advantages. First, it greatly simplifies the numerical computations. Second, because of the nature of the voting constraint, there is no need to specify the slope of the function, and there is therefore one less parameter to estimate (see Lemma A2 in the Appendix). An earlier version of the calibration used micro data from the 1970's about workers' preference for unionization to parametrize a logistic CDF for ϕ. That calibration found a similar impact of unions on the economy but had to rely on older data.

[23] The problem of each firm must be solved numerically so that the set of firms must be discretized for the computations. For that purpose, I assume that there are 20 different types of firms—a good compromise between precision and computation time.

[24] Since capital is now used as a production factor, α captures the whole curvature of the production function after the firm has optimized on its capital input. See Appendix B for details.

$\mu_j = a^\mu + b^\mu j$ and $\xi_j = a^\xi + b^\xi j$. The parameters of these linear relationships are part of the estimation. Similarly, I assume that A_j is a piece-wise linear function of j with one potential break point. That is, there is a $j^* \in [0,1]$ such that $A_j = a_1^A + b_1^A j$ for $j \in [0, j^*]$ and $A_j = a_2^A + b_2^A j$ for $j \in (j^*, 1]$, and $a_1^A + b_1^A j^* = a_2^A + b_2^A j^*$. The parameters $\left\{a_1^A, b_1^A, a_2^A, b_2^A, j^*\right\}$ are also part of the estimation. This simple specification has the advantage of having few parameters to estimate and to fit the data well.

Calibrated economy

I use a method of simulated moments to estimate the parameters of the model. In addition to the technology parameters listed above, the estimation also includes the bargaining powers β_u and β_n. The estimation seeks to bring a set of model quantities as close as possible to their data counterparts. The targeted quantities are the union and nonunion wage and employment of each skill group, as well as the labor shares in the union and nonunion sectors.[25],[26]

The parameters are jointly estimated but some intuition can be provided about the moments of the data that matter the most in determining their values. Broadly speaking, the average wage in the union and nonunion sectors identifies the bargaining powers β_u and β_n. The parameters that determine the skill intensities $\{\boldsymbol{z_j}\}_{j\in[0,1]}$ are identified from the observed skill distributions in the union and nonunion sectors. Finally, the parameters that govern $\{A_j\}_{j\in[0,1]}$ are identified by the overall employment levels and the labor shares in the union and nonunion sectors.

Table 2 shows the parameter values that best fit the data.[27] In the calibrated economy, firms with technology indexes between 0 and 0.1 are unionized, while those between 0.1 and 0.55 are affected, to varying degrees, by the union threat. The remaining firms are union free and their decisions are not directly affected by the possibility of unionization.

Importantly, the estimation finds that the union threat distorts the decisions of a sizable mass of firms. The key features of the data that push for that conclusion are the union and nonunion wage schedules. Since union wages are relatively high in the data, the estimated value of β_u is high. As a result, nonunion workers know that they would have high wages if they were to unionize which pushes them to vote in favor of the union. Nonunion firms must then react to prevent unionization.

[25] To compute the labor share, I assume that the returns to scale parameter ω_j is 0.8, in the range estimated by Burnside et al. (1995).

[26] Adding the outcome of the Freeman (1980) estimator as an additional targeted moment does not affect the results much. Note also that the unemployment rate is implicitly targeted since the total number of workers (employed plus unemployed) is taken directly from the data and that the number of employed workers is a target of the estimation.

[27] The estimation finds β_n to be larger than β_u. A few unmodeled features of the data may explain this difference. First, there are costs to unionization: workers may have to pay dues or spend time organizing the union (Voos, 1983). The estimation captures these costs by lowering β_u. Second, consistent with evidence from Farber (1987), union workers might want the firm to hire more workers even if it leads to lower wages. In the model, since an increase in the bargaining power leads to higher wages and to lower employment, this preference would also be captured by a lower β_u. Finally, Bronfenbrenner (1994) and Freeman and Kleiner (1990) detail various tactics, some legal and some illegal, used by firms to prevent unionization. These tactics make it easier for firms to stay union free and they would also be captured by a low β_u.

In the calibrated economy, 9.0% of the population works in a union firm, 24% works in a nonunion firm that is subject to the threat of unionization and 67% works in a nonunion firm that is not directly affected by the threat. The unemployment rate is 6%.[28] Figure 2 shows how the estimated model fits the wage and employment schedules in the union and nonunion sectors.

4.2 The impact of unions on the economy

I now evaluate the impact of labor unions on the calibrated economy by considering three experiments in general equilibrium. First, I investigate the role played by the union threat alone ("no threat" experiment). To do so, I assume that workers in nonunion firms cannot form a union anymore. As a result, these firms no longer distort their behavior to prevent unionization. In the second experiment, I assume that unions are simply forbidden ("no unions" experiment). In this case, not only does the union threat disappear, but all firms that were previously unionized become union free. This experiment therefore captures the overall impact of labor unions on the economy. Finally, in the third experiment all firms are unionized ("all unions" experiment). Notice that the union threat is inactive in all three experiments.

The results are presented in Figure 3 and Table 3. Figure 3 shows how the experiments influence wages and unemployment rates across the skill distribution. Table 3 shows how aggregate output, unemployment, welfare and wages react to the experiments. The rest of this section describes how the economic forces at work in the model generate these results.[29]

Impact of the union threat

We begin by considering the first experiment: the removal of the union threat. Figure 4 shows how nonunion wages and employment levels react. To better highlight the various mechanisms at work, the solid lines represent the partial equilibrium changes (when all aggregate quantities are kept unchanged) and the dashed lines show the overall impact of the threat removal in general equilibrium.[30]

Let us consider the partial equilibrium reaction of the firms first. From Panel (a), we see that, once the threat is gone, firms hire substantially more, as predicted by Proposition 5. Indeed, when the threat disappears firms are no longer distort their hiring and the marginal cost of production goes down. As a result, firms increase their size to reach the flatter part of their production function.

[28] The mean and variance of log wages are 3.57 and 0.16, respectively.

[29] Figures 8 and 9 in Appendix C.1 show how union and nonunion employment change in response to the policy exercises. To evaluate the robustness of the experiments, I also include an additional exercise in Appendix C.2 in which the mass of firms in the economy adjusts through a free-entry condition. The union threat has a substantial impact on the economy in that environment as well.

[30] To be precise, the partial equilibrium exercises keep the labor market tightness $\boldsymbol{\theta}$ and the value of non-work activities $\boldsymbol{b}$ fixed at their calibrated values. Since some firms have curvature α_j close to unity, their employment reacts substantially to the removal of the threat in partial equilibrium, as seen in Panel (a). General equilibrium forces push back these changes in hiring to more modest levels.

Definition	Parameter	Calibrated value
Bargaining powers of workers		
Individual bargaining	β_n	0.46
Collective bargaining	β_u	0.33
Skill intensity vectors $\{\boldsymbol{z_j}\}_{j\in[0,1]}$		
Intercept of the mean	a^μ	1.19
Slope of the mean	b^μ	0.50
Intercept of the variance	a^ξ	0.56
Slope of the variance	b^ξ	0.38
Total factor productivities $\{A_j\}_{j\in[0,1]}$		
Intercept of the first segment	a_1^A	36.3
Slope of the first segment	b_1^A	66.6
Intercept of the second segment	a_2^A	-22.0
Slope of the second segment	b_2^A	223.6
Break point	j^*	0.35

Table 2: Estimated parameters

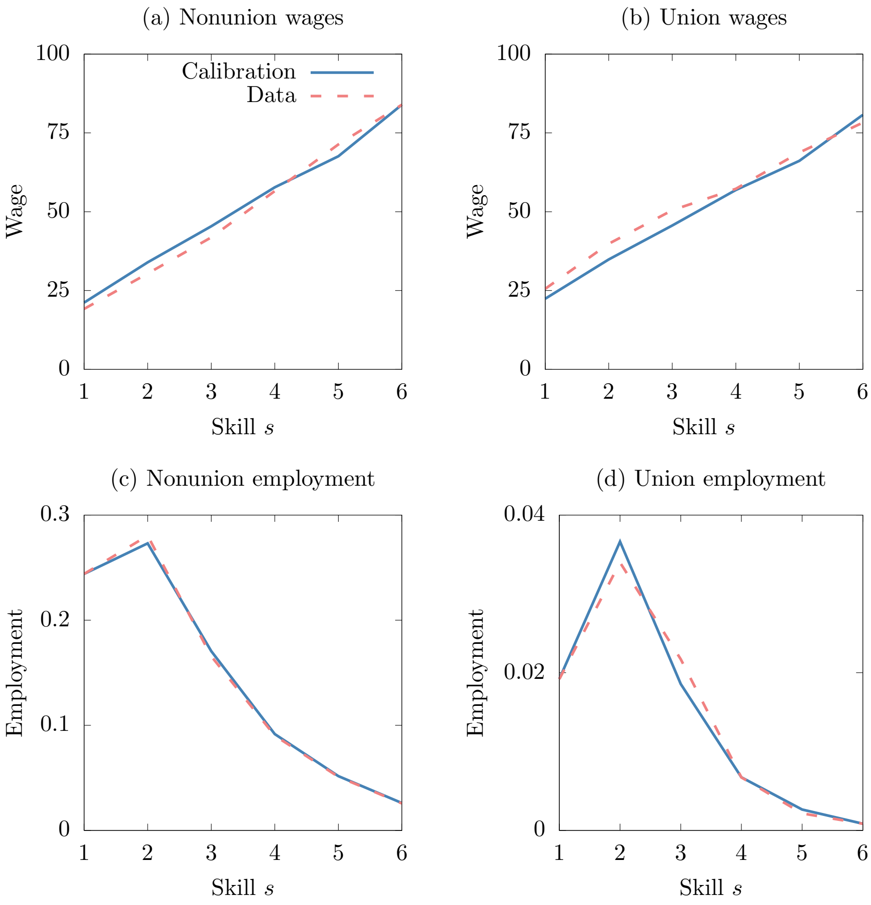


Figure 2: Fit of the calibrated model

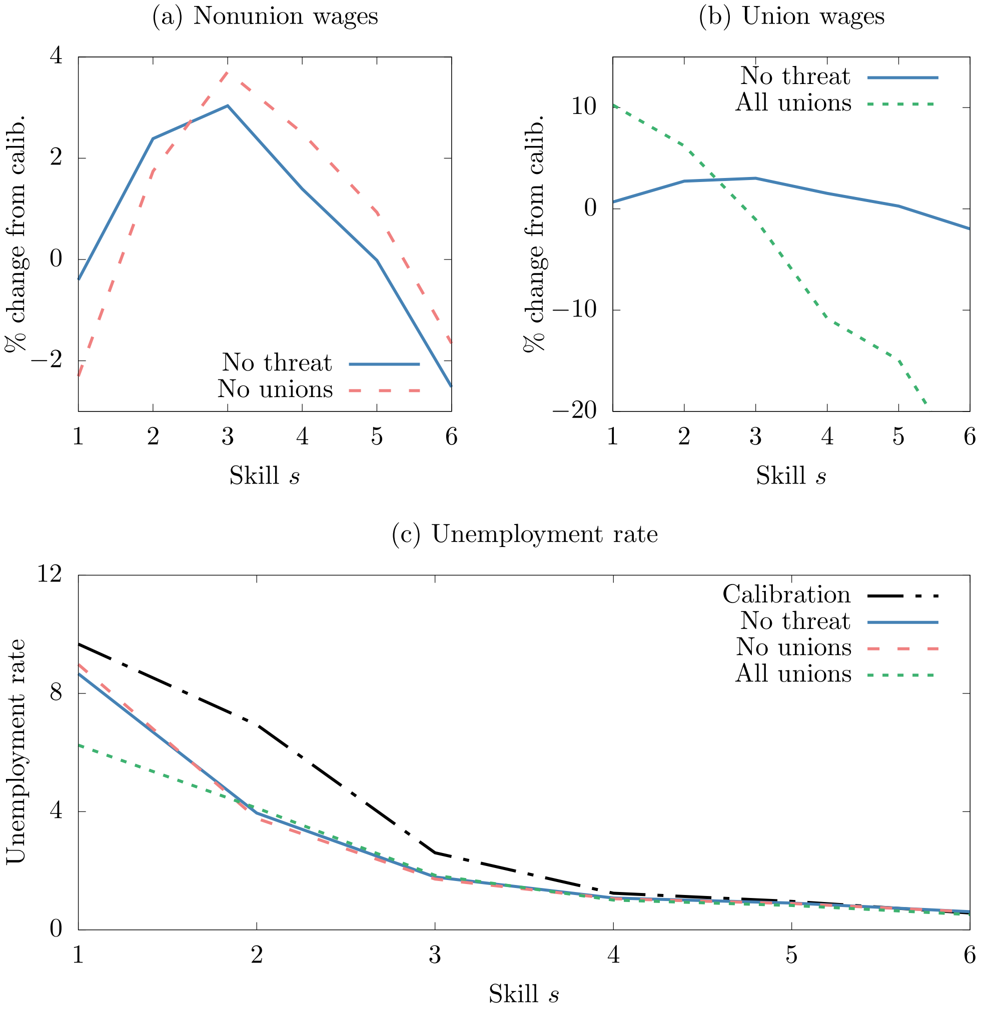


Figure 3: General equilibrium impact of experiments

While this increase in hiring affects all workers, the impact is particularly important at the bottom of the skill distribution. When the threat was active, firms were biased against hiring these workers since they voted in favor of unionization. In contrast, high-skill workers were favored since they voted against the union. The removal of the threat therefore leads to a more modest increase in hiring at the top of the skill distribution than at its bottom.

These changes in employment affect wages in partial equilibrium, as shown by the solid line in Panel (b) of Figure 4. Since firms now hire more, the marginal product of the workers decline which, through individual bargaining, adversely affects wages. Notice that, in partial equilibrium, the disappearance of the threat leads firms to pay a broader range of wages, which pushes for an increase in wage inequality, as predicted by Proposition 6. Indeed, we can see from Figure 4 that high-s wages remain essentially unchanged while low-s wages decline.

In general equilibrium, the increase in hiring pushes unemployment down for all skill groups

	Changes from calibrated economy		
	No threat	No unions	All unions
Output	+1.18%	+1.20%	+1.50%
Unemployment rate	−1.48pp	−1.47pp	−2.11pp
Welfare	+0.22%	+0.22%	0.28%
Wages			
Mean	+0.35%	+0.18%	+0.99%
Variance	+1.19%	+7.73%	−47.4%

Notes: All numbers are percentage changes except for the unemployment rates which are differences in percentage points. Wages are measured in logs. Output refers to value added.

Table 3: Impact of experiments

(bottom of Figure 3) for an overall decline in the unemployment rate of 1.48 percentage points. These lower unemployment levels benefit the bargaining position of the workers, since they can now find other jobs quickly if negotiations break down, which leads to higher wages. In turn, this increase in wages is strong enough to undo the wage decline that was observed in partial equilibrium such that, in general equilibrium, the threat removal leads to higher wages for most workers. Finally, these higher wages tamper the initial increase in employment, so that the increase in hiring generated by the removal of the threat is much smaller in general than in partial equilibrium (Panel (a) of Figure 4).

Overall, the removal of the threat benefits nonunion workers above the median skill level ($s = 2$) more than those below it, as Panel (b) of Figure 4 shows. In general equilibrium, the average wage of workers below the median falls by about 0.4% while that of workers above the median increases by 1.8%.[31] As a result, the removal of the threat increases the variance of log wages by 1.19%. The removal also benefits production, with the increase in hiring that follows the disappearance of the threat pushing output up by 1.18%. Welfare also benefits from the threat removal. Since firms no longer distort their skill mix to avoid unionization, welfare goes up by 0.22%.[32] This increase is smaller than the increase in output since, as many unemployed workers find employment once the threat is gone, the extra value of leisure $\boldsymbol{b}$ that unemployment brought is lost.

Finally, we can consider the impact of the removal of the threat on the distribution of wages across firms. The black line in Figure 5 shows the average wage paid by each firm j in the calibrated economy. The blue line provides the same information but once the threat is gone. We see that the removal of the threat leads to lower wages for workers in previously threatened firms. Once the threat is gone, these firms hire more which decreases the marginal product of the workers and leads to lower wages (see discussion around Proposition 6). The change in wages is also more pronounced

[31] The bulk of the skill distribution is at skill level $s = 3$ and below, as shown in Figure 2. Workers in skill groups 5 and 6 actually suffer from the removal of threat but there are so few workers there that their impact on overall wage inequality is minimal.

[32] In contrast, in the social planner's preferred allocation welfare increases by 1.5%.

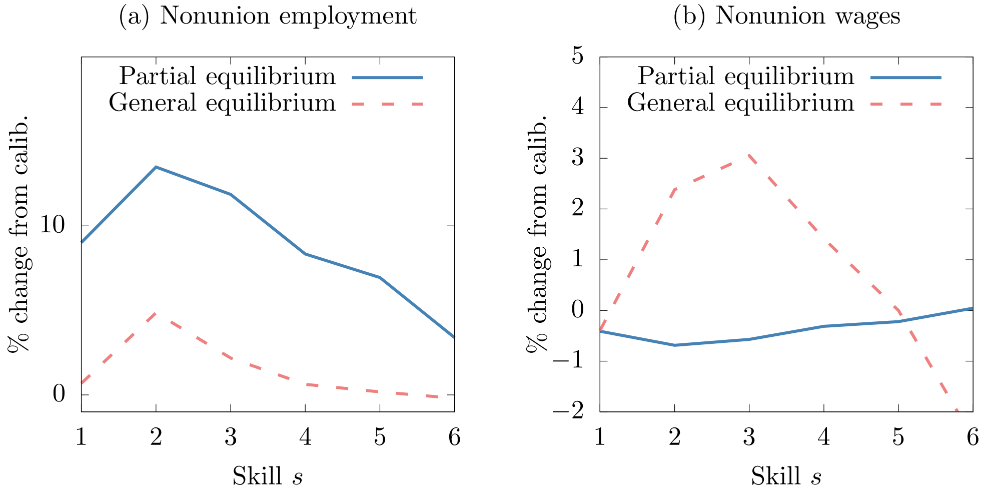


Figure 4: Removing the union threat: partial and general equilibrium impact

for firms facing the strongest threat, around $j = 0.1$.

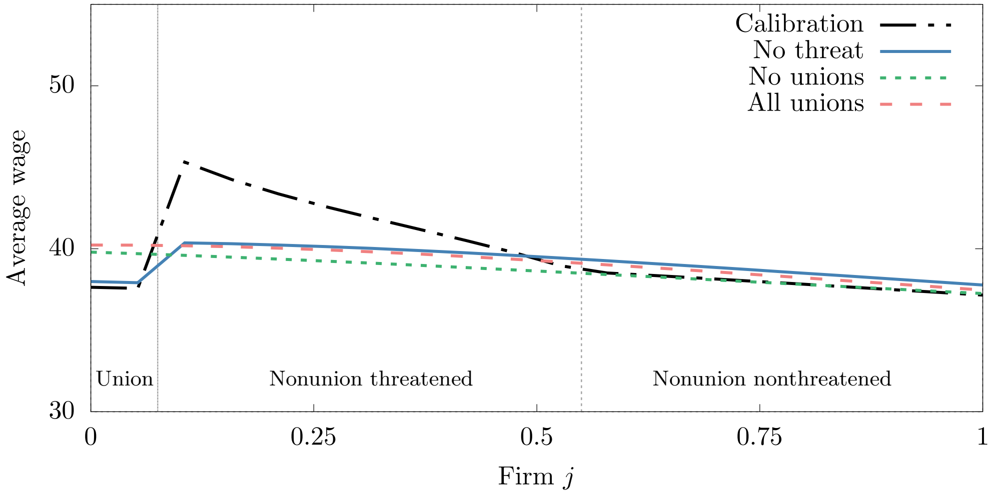


Figure 5: Changes in firm-level average wage

Mandating or prohibiting unions

Figure 3 also shows the impact of the two other experiments: prohibiting all unions or, to the opposite, forcing all firms to be unionized. We see from Panel (a) that removing all unions leads to a substantial increase in wages for workers in the middle of the distribution and to a decline for those in its tails (all workers are nonunion workers in this experiment). As very few workers are actually in the right tail of the distribution, this experiment leads to a substantial increase in wage inequality. Since all bargaining is now done individually, low-skill wages no longer benefit from the high productivity of the high-skill workers. The average worker below the median skill level see her wage fall by 2.3% while the average workers above it gets a wage increase of 2.6%. As a result, the variance of log wages increases by 7.73% from its calibrated value, as shown in Table 3. Output and welfare do not react much more than under the "no threat" experiments.

In contrast, forcing all firms to be unionized leads to a large decline in wage inequality, as can be seen in Panel (b) of Figure 3 (all workers are union workers in this experiment). Since now all wages are bargained collectively, the high-skill workers do not directly benefit from their high productivity and their wages fall substantially. In contrast, workers at the low-end of the skill distribution see massive wage gains from the inclusion of the high-skill workers in the collective bargaining. Overall, the variance of log wages declines by about 48%. Output and welfare are higher under the "All unions" experiment compared to the "No unions" scenario. Unemployment is also at its lowest. Here, the inefficient over-hiring that occurs under individual bargaining is responsible for the differences (see Section 3.3).[33]

Perhaps surprisingly, these experiments show that the threat, on its own, has a larger impact on output and welfare than whether firms are actually unionized or not. To understand why, remember that unionization, by itself, is simply a different way to share the surplus generated by production. Without the threat, firms still seek to maximize that surplus regardless of their union status. As a result, the decisions that matter for the allocation of resources, such as hiring, are relatively unaffected by unionization. In contrast, when the union threat is active, the additional constraint on the firm's problem distracts from surplus maximization, which leads to a decline in output, welfare and employment.[34]

Policy and the non-monotone relationship between unionization and welfare

The experiments of the last section highlight a non-monotone relationship between the unionization rate and aggregate welfare. Indeed, welfare is higher when the economy is fully unionized, or when there are no unions at all, than under an intermediate situation (the calibrated economy) in which the union threat distorts firms' decisions (see Table 3). Figure 6 emphasizes this point by showing how welfare changes with the union bargaining power β_u. We see that for low values of β_u the economy features a relatively high level of welfare and a low unionization rate, while for high β_u's welfare is still elevated but now the unionization rate is also high. In contrast, for intermediate values of β_u the unionization rate is moderate while welfare in relatively depressed.

Two forces work in opposite directions to create this non-monotonicity. First, keeping the union

[33] Figure 5 shows that since all firms have the same union status, wages are more similar across firms in the last two experiments than in the calibrated economy.

[34] To focus on the threat of unionization, the model abstracts from many union-related mechanisms studied in the literature (Freeman and Medoff, 1984). For instance, if union employees could restrict hiring to increase their wage, as in the insider-outsider literature (Lindbeck et al., 1989), the fully unionized economy could feature a higher unemployment rate and a lower welfare level.

status of each firm constant, an increase in β_u makes the threat worse for nonunion firms. Since these firms' workers now anticipate higher union wages, the firms must distort their skill mix more heavily to prevent unionization, which exacerbates the inefficiencies. Through this first mechanism, an increase in β_u therefore leads to a decline in welfare. There is however a second mechanism that operates through changes in the union status of the firms. As β_u increases, there comes a point at which it is so costly to prevent unionization that a firm prefers to let its workers unionize. In this case, there is no longer any reason to distort hiring, which is beneficial for welfare. These two forces compete to generate Figure 6. Increasing β_u, starting from the calibrated economy (black dot), initially leads to a decline in welfare. For a small increase in β_u, not many firms change their union status but the threat becomes more important and threatened firms distort hiring more heavily. As β_u keeps increasing, there comes a point at which welfare begins to increase. At these high levels of union bargain powers, firms simply decide to let the workers unionize and the threat no longer distorts their decisions. This point is reached around $\beta_u \approx 0.39$ in Figure 6. Further increases in β_u after this point lead to large changes in the unionization rate as firms unionize massively.

The mechanisms at work in Figure 6 have important consequences for policy design. In particular, any policy that slightly strengthens the bargaining position of unions from its calibrated value—for instance the repeal of a right-to-work law—is welfare decreasing as it increases the distortion created by the threat. In contrast, increasing β_u by a large amount, say to 0.45, would be welfare improving as the threat would then affect fewer firms. In practice, the optimal design of a policy should weigh the negative impact of increasing the threat (stronger distortion for remaining nonunion firms) against its positive impact (fewer firms are subject to it).

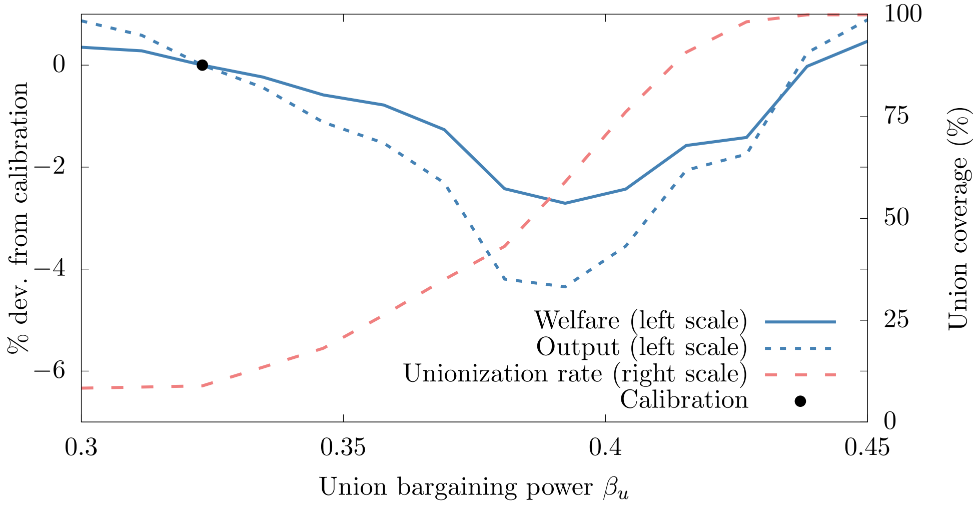


Figure 6: Non-monotone relationship between unionization rate and welfare

The decline of unions in the United States

Over the last few decades the unionization rate has declined significantly in the United States, from 19.7% in 1983 to 9.0% in 2005.[35] Over the same period, heavily unionized sectors such as manufacturing, transportation and utilities have been slowly declining as well, suggesting that a change in industrial composition was one driver behind the decline in unionization.[36] In this subsection, I modify the calibrated model to replicate the industrial composition of the United States in 1983 and then use the model to evaluate how the impact of unions on the economy has changed over the last decades.

In the data, the heavily unionized sectors of the economy feature lower labor shares, so that they correspond to technologies with lower indexes j in the calibration. To match the 1983 economy, I therefore adjust the mass of firms using each technology $j \in [0, 1]$ to match the union and nonunion employment vectors.[37] Once this is done, I then compute the general equilibrium in this new economy.

Table 4 presents key moments of the 1983 and 2005 economies side by side to highlight their differences. We see that the change in industrial composition leads to a lower output, higher unemployment and a higher unionization rate in 1983. In addition, employment in industries with low labor shares was higher, and a much larger fraction of workers where in firms whose decisions were constrained by the threat of unionization.[38]

	1983	2005
Output	41.4	44.5
Welfare	885	919
Unemployment rate	10.9%	5.8%
Unionization rate	19.7%	9.0%
Workers in threatened firms	52%	24%
Workers in low labor share firms	31.0%	26.5%

Notes: Firms under threat are those for which the voting constraint is binding. Low labor share firms are those with technologies $0 < j < 0.5$.

Table 4: Comparing the 1983 and the 2005 economies

We can decompose the changes in Table 4 as the sum of a change in the impact of the union threat and a residual that captures the direct effect of the change in industrial composition on the economy. I measure the impact of the union threat by implementing the "no union threat"

[35] 1983 is the earliest year in the sample with consistent CPS data.

[36] Acemoglu et al. (2001), Açikgöz and Kaymak (2014) and Dinlersoz and Greenwood (2016) investigate the link between technological changes and labor unions. Dinlersoz et al. (2017) documents which firms are targeted by unions.

[37] As Lemma A1 shows, this is equivalent to adjusting the TFPs $\{A_j\}_{j\in[0,1]}$. I therefore pick the slopes b_1^A and b_2^A to match the union and nonunion employment vectors.

[38] The set of firms constrained by the threat is $[0.1, 0.7]$ is 1983 and $[0.1, 0.55]$ in 2005.

policy experiment described in Section 4.2 in both economies and by taking the differences between outcomes. The results are presented in Table 5. We see that the union threat is responsible for a good fraction of the overall changes experienced by the U.S. economy between 1983 and 2005. For instance, the decline in the intensity of the threat was responsible for an increase of 2.6% in output and 0.8% in welfare. Unsurprisingly, the change in industrial composition itself was a key driver of the changes over that period.[39]

	Changes from 1983 to 2005 accounted for by		
	Direct impact of new industrial composition	Union threat	Total
Output	+4.8%	+2.6%	+7.4%
Unemployment rate	−2.7pp	−2.4pp	−5.1pp
Welfare	+3.1%	+0.8%	+3.9%

Notes: Differences from the calibrated economy. All numbers refer to percentage changes except for the unemployment rate number which show the difference in percentage point. Output is measured as value added.

Table 5: The union threat between 1983 and 2005

5 Reduced-form estimates

The empirical literature on unions frequently relies on reduced-form estimators to make predictions about the impact of unions on wages. In this section, I discuss how these estimators do not take into account the threat of unionization and how this can lead to incorrect predictions. I also provide reduced-form estimates of the impact of the union threat on wages by taking advantage of changes in legislation in certain U.S. states. This last exercise provides model-free supporting evidence for the quantitative analysis of the previous section.

5.1 Comparison with common reduced-form estimators

Estimating the impact of unions on wage inequality

In the calibrated economy, the true impact of unions on wage inequality differs from what common reduced-form estimators suggest. To show this, I first consider the well-known estimator introduced by Freeman (1980). That estimator seeks to compute the variance of wages in a counterfactual economy with no unions. It proceed by assigning to each union worker a counterfactual nonunion wage drawn from the observed nonunion wage distribution. This estimator can be written as

$$V - V^n = U\Delta_v + U(1-U)\Delta_w^2,$$

[39] Table 5 isolates the impact of the union threat but the numbers barely change if we isolate the total impact of unionization (threat plus change in bargaining protocol) instead.

where V is the observed variance of log wages, V^n is the variance of log wages without unions in the economy, U is the unionization rate, Δ_v is the difference in the variance of log union and nonunion wages and Δ_w is the difference between the mean log of union and nonunion wages. When used on the calibrated economy, that estimator suggests that unions are responsible for lowering the variance of log wages by 3.63%. In contrast, in the "no union" experiment above unions are responsible for lowering wage inequality by 7.73%, more than twice as much.

More sophisticated estimators also take into account the fact that union and nonunion workers differ in terms of observable characteristics such as education, age, etc. (see for instance Dinardo and Lemieux (1997), Card (2001) and Card et al. (2004)). The idea is to attribute to every union worker a draw from the non-union wage distribution of workers with the same observable characteristics. Taking this heterogeneity into account, these estimators predict that unions are responsible for lowering the variance of log wages by 0.72%, or only about 9% of their true impact on wage inequality.[40,41]

The key reason for these large discrepancies between reduced-form and model-based estimates is that the former assume that the union and nonunion wage schedules themselves are unaffected by the disappearance of labor unions from the economy. In the model, however, these schedules react to the disappearance of unions for multiple reasons. First, in the calibrated economy the union threat distorts the wages that nonunion firms pay. When unionization is no longer an option, the threat disappears and the nonunion wage schedule becomes steeper as a result. Second, union and nonunion firms in the model use different technologies. Indeed, their union status differ precisely because they use different technologies. Therefore, when previously unionized firms become union free they pay wages that differ from those paid by the previously union-free firms. This results in a change in the nonunion wage schedule, which is now coming from a richer mix of technologies. Finally, the reduced-form estimators abstract completely from general equilibrium mechanisms. In particular, when unions disappear firms tend to hire more which, through the increase in the outside option of workers in the labor market, pushes all wages upward. As this channel affects workers with different skills differently it leads to asymmetrical changes in the wage schedules. Putting all these mechanisms together explains the differences between the reduced-form and model-based estimates of the impact of unions on wage inequality.

[40] The Freeman estimator predicts a larger impact of unions on wages inequality than the estimators that control for heterogeneity since the union skill distribution is more concentrated than the nonunion skill distribution (see Figure 2). Since the estimator assumes that a union worker would get a random draw from the nonunion wage distribution if she or he were not unionized, this leads to an overestimation of the true impact of unions on wage inequality.

[41] These estimators can also be thought of as non-targeted moments for the estimation. In the raw data, the Freeman estimator finds that unions lower the variance of log wages by 2.6% while the corresponding number for the estimator that takes into account worker heterogeneity is 1.2%. The equivalent numbers in the calibrated economy are 3.63% and 0.72%, respectively. Both estimators therefore take similar values in the data and in the calibrated economy.

Regression discontinuity estimators

The model can also shed light on why regression discontinuity estimators tend to find a small impact of unionization on firm-level outcomes. For instance, DiNardo and Lee (2004) compare firms that barely win a union election to firms that barely lose an election and find essentially no significant impact of unionization. They mention that a union threat effect would tend to bias the estimates against finding a strong impact of unionization. The idea is that if nonunion firms preemptively change their behavior to prevent unionization, the control group—the firms that barely win the election—is also affected by union policies and the estimator therefore misses the full impact of unions.

We can use the model to evaluate the magnitude of the bias introduced by the threat. To do so, I consider, in the calibrated economy, a threatened nonunion firm that faces a bargaining power β_n such that, if it were to unionize, its total employment would not change, as in DiNardo and Lee (2004).[42] I then compare the impact of unionization on this firm under two different scenarios. In the first scenario, the firm is initially threatened by unionization, as in the calibrated economy. In the second scenario, the firm is initially unaffected by the threat. The differences between these two scenarios is indicative of the impact of union policies that is not captured by the regression discontinuity estimator. Table 6 shows the results of the exercise. We see that for employment, output and the wage bill, the threatened firm reacts less to unionization than its non-threatened counterpart. These results suggest that the full impact of union policies can be larger than implied by regression discontinuity estimators.

	Initially threatened?		Difference
	Yes	No	
Employment	+0.0%	−7.2%	7.2%
Output	−5.2%	−8.8%	3.6%
Wage bill	−1.8%	−8.6%	7.8%

Table 6: Impact of unionization on threatened and non-threatened firm

5.2 Right-to-work laws and the threat of unionization

In this subsection, I provide supporting evidence for the quantitative exercise of Section 4. To do so, I use the passage of right-to-work legislations by U.S. states as a source of variation in union power and rely on reduced-form methods to measure the impact of the threat on the earnings of nonunion workers.

[42] Since the firm does the minimum needed to avoid unionization, the outcome of a union election would be close to 50% and, assuming that some random shock pushes the workers to barely vote in favor of unionization, a regression discontinuity estimator would find no impact of unionization on employment. I assume that the firm has the technology $j = 0.1$, the nonunion firms closest to unionization.

Several states in the U.S. have passed right-to-work (RTW) legislations since World War II. These laws prohibit contracts between labor unions and employers that mandate that workers pay union membership fees as a condition of employment. As a result, under these laws unions have access to fewer resources, which limits their ability to organize and leads to a weaker threat of unionization. By estimating the impact of these laws on nonunion earnings, we can therefore evaluate the impact of the union threat.[43]

The data come from the Merged Outgoing Rotation Groups of the Current Population Survey. The sample covers the period from January 1989 to December 2018 and includes the passage of right-to-work laws in Indiana (2012), Kentucky (2017), Michigan (2012), Oklahoma (2001), Texas (1993), West Virginia (2016) and Wisconsin (2015).[44]

Table 7 shows the outcome of ordinary least-square regressions of log weekly earnings on a right-to-work law indicator variable that equals one if the individual resides in a state with a right-to-work law and zero otherwise.[45] We see from Column 1 that the passage of a RTW law is associated with a significant decline in the earnings of all workers. Breaking down workers by their union status, Column 2 shows that the earnings of *nonunion* workers decline by about 3% after the passage of a RTW law. This decline is consistent with a weaker threat of unionization after the passage of right-to-work laws. Since nonunion firms are less worried about unionization, they no longer have to keep wages high to influence a union vote. The same mechanism operates in the model (see Proposition 6). These results are also consistent with the partial equilibrium reaction of the firms in the calibrated economy, as shown in Figure 4. Finally, the third column of Table 7 suggests that, unsurprisingly, RTW laws have a negative impact on the earnings of union workers, although the estimate is not statistically significant.[46]

[43] This exercise complements the previous literature, discussed in the introduction, in two ways. First, since several RTW laws have been passed in recent years, an up-to-date exercise provides a more current estimate of the impact of the union threat. Second, I consider the impact of the laws on union and nonunion earnings separately, something that few studies do and that allows me to explicitly evaluate the impact of the threat on nonunion firms. One exception is Farber (2005), who studies the impact of RTW laws in Idaho and Oklahoma.

[44] These data come from IPUMS (Flood et al., 2015). January 1989 is the first month with consistent weekly earnings data. Missouri passed a RWT in 2017 but the law was repealed before it could take effect. I restrict the sample to the adult civilian population and I remove government/military employees, part-time workers and individuals in management occupations.

[45] The standard errors are clustered at the state level. The significance levels are the same with two-way clustering at the state-time level instead. With robust standard errors, all coefficients are significant at the 1% level.

[46] One worry about these regressions is that the passage of right-to-work legislations might not be exogenous. In particular, lawmakers might pass these laws in bad economic times in an attempt to sustain the economy. To control for the business cycle, I therefore include the state-level unemployment rate in the regressions. The results are also robust to including as regressors the state-level minimum wage and unemployment benefits extensions, as shown in Appendix D.1. Similar regressions for the skill premium and the skill mix hired by firms are reported in Appendix D.2.

	(1)	(2)	(3)
Dependent variable	Earnings	Earnings	Earnings
Workers in the sample	All	Nonunion	Union
Right-to-work law	-0.038**	-0.029**	-0.035
	(0.019)	(0.013)	(0.033)
State & time fixed effects	yes	yes	yes
Individual & state controls	yes	yes	yes

Notes: Ordinary least-square regressions. The dependent variable is the log of weekly earnings for all workers (1), nonunion workers (2) and union workers (3). Standard errors, clustered at the state level, are in parenthesis. Individual controls are industry, occupation, age, sex and education. The state control is the unemployment rate. The data covers the adult civilian population in the Current Population Survey Merged Outgoing Rotation Groups between January 1989 and December 2018. I remove from the sample government/military employees, part-time workers and individuals in management occupations. Union workers include union members and workers covered by a union. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Impact of right-to-work laws on weekly earnings

6 Conclusion

This paper proposes a general equilibrium theory of endogenous union formation to study the impact of unions on the economy. Unions are created by a majority vote within each firm. If a union is created, wages are bargained collectively otherwise each worker bargains his or her wage individually with the firm. This asymmetry in wage setting mechanisms causes unions to compress the wage distribution inside a firm and to lower its profit. A key mechanism in the theory is that, to prevent their own unionization, nonunion firms distort their hiring decisions in a way that also compresses the range of wages and reduces employment and output. The main predictions of the theory are in line with findings from the empirical literature. Experiments using an estimated version of the model show that removing the threat of unionization increases the variance of wages while also raising output and welfare and lowering unemployment. The model therefore provides guidance about the outcome of policy interventions that would aim to weaken unions, such as the passage of right-to-work legislations by state legislatures, or make them stronger.

This paper also emphasizes the importance of off-equilibrium paths for on-equilibrium quantities. When economic agents try to avoid utility/profits-reducing situations, they can take actions that affect observed aggregates, even though the unwanted situation is never actually observed. In that spirit, firms threatening to outsource production abroad might lead to lower wages even though no outsourcing actually takes place. A similar mechanism could also operate for firms that threaten to use industrial robots to save on labor costs.

References

Açikgöz, O. T. and B. Kaymak (2014). The rising skill premium and deunionization. *Journal of Monetary Economics 63*(0), 37 – 50.

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

Acemoglu, D. and W. B. Hawkins (2014). Search with multi-worker firms. *Theoretical Economics 9*(3), 583–628.

Alvarez, F. and M. Veracierto (2000). Labor-market policies in an equilibrium search model. In *NBER Macroeconomics Annual 1999, Volume 14*, NBER Chapters, pp. 265–316. National Bureau of Economic Research, Inc.

Bauer, C. J. and J. Lingens (2010). Individual vs. collective bargaining in the large firm search model. *University of Munich Discussion Paper 11315*.

Bertola, G. and P. Garibaldi (2001). Wages and the size of firms in dynamic matching models. *Review of Economic Dynamics 4*(2), 335 – 368.

Boeri, T. and M. C. Burda (2009). Preferences for collective versus individualised wage setting. *Economic Journal 119*, 1440–1463.

Bronfenbrenner, K. (1994). Employer behavior in certification elections and first-contract campaigns: Implications for labor law reform. In S. Friedman, R. Hurd, R. Oswald, and R. Seeber (Eds.), *Restoring the promise of American labor law*, pp. 75–89. IRL Press.

Brügemann, B., P. Gautier, and G. Menzio (2019). Intra firm bargaining and shapley values. *The Review of Economic Studies 86*(2), 564–592.

Burnside, C., M. Eichenbaum, and S. Rebelo (1995). *Capital Utilization and Returns to Scale*, pp. 67–124. MIT Press.

Cahuc, P., F. Marque, and E. Wasmer (2008). A theory of wages and labor demand with intrafirm bargaining and matching frictions. *International Economic Review 48*(3), 943–972.

Cahuc, P. and E. Wasmer (2001). Labor market efficiency, wages and employment when search frictions interact with intrafirm bargaining. *IZA Discussion Papers* (304).

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica 64*(4), 957–979.

Card, D. (1998). Falling union membership and rising wage inequality: What's the connection? *National Bureau of Economic Research Working Paper Series No. 6520.*

Card, D. (2001). The effect of unions on wage inequality in the u.s. labor market. *Industrial and Labor Relations Review 54*(2), pp. 296–315.

Card, D., T. Lemieux, and W. Riddell (2004). Unions and wage inequality. *Journal of Labor Research 25*(4), 519–559.

Chodorow-Reich, G., J. Coglianese, and L. Karabarbounis (2018). The Macro Effects of Unemployment Benefit Extensions: a Measurement Error Approach*. *The Quarterly Journal of Economics 134*(1), 227–279.

Corneo, G. and C. Lucifora (1997). Wage formation under union threat effects: Theory and empirical evidence. *Labour Economics 4*(3), 265 – 292.

Davis, S. J., J. C. Haltiwanger, S. Schuh, et al. (1998). Job creation and destruction. *MIT Press Books 1.*

Delacroix, A. (2006). A multisectorial matching model of unions. *Journal of Monetary Economics 53*(3), 573–596.

den Haan, W. J., G. Ramey, and J. Watson (2000). Job destruction and propagation of shocks. *American Economic Review 90*(3), 482–498.

Dickens, W. T. (1983). The effect of company campaigns on certification elections: Law and reality once again. *Industrial and Labor Relations Review 36*(4), pp. 560–575.

Dickens, W. T. (1986). Wages, employment and the threat of collective action by workers. *National Bureau of Economic Research Working Paper Series No. 1856.*

Dickens, W. T. and L. F. Katz (1987). *Interindustry Wage Differences and Industry Characteristics*, pp. 48–89. Blackwell.

DiNardo, J. and D. S. Lee (2004). Economic impacts of new unionization on private sector employers: 1984-2001. *Quarterly Journal of Economics 119*(4), 1383–1441.

Dinardo, J. and T. Lemieux (1997). Diverging male wage inequality in the united states and canada, 1981-1988: Do institutions explain the difference? *Industrial and Labor Relations Review 50*(4), pp. 629–651.

Dinlersoz, E. and J. Greenwood (2016). The rise and fall of unions in the united states. *Journal of Monetary Economics 83*, 129–146.

Dinlersoz, E., J. Greenwood, and H. Hyatt (2017). What businesses attract unions? unionization over the life cycle of u.s. establishments. *ILR Review 70*(3), 733–766.

Ebell, M. and C. Haefke (2006). Product market regulation and endogenous union formation. IZA Discussion Papers 2222, Institute for the Study of Labor (IZA).

Elejalde-Ruiz, A. (2016). Nlrb rules that grad students are employees, opens door to unionization. *Chicago Tribune.*

Elsby, M. W. L. and R. Michaels (2013). Marginal jobs, heterogeneous firms, and unemployment flows. *American Economic Journal: Macroeconomics 5*(1), 1–48.

Farber, H. S. (1987). The analysis of union behavior. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 2 of *Handbook of Labor Economics*, Chapter 18, pp. 1039–1089. Elsevier.

Farber, H. S. (2005). Nonunion wage rates and the threat of unionization. *Industrial and Labor Relations Review 58*(3), 335–352.

Farber, H. S. and D. H. Saks (1980). Why workers want unions: The role of relative wages and job characteristics. *The Journal of Political Economy 88*(2), 349–369.

Flaherty, C. (2016). Nlrb: Graduate students at private universities may unionize. *Inside Higher Ed*.

Flood, S., M. King, S. Ruggles, and J. R. Warren (2015). Integrated public use microdata series, current population survey: Version 4.0. [dataset]. *Minneapolis: University of Minnesota*.

Foulkes, F. (1980). *Personnel policies in large nonunion companies.* Prentice-Hall.

Freeman, R. B. (1980). Unionism and the dispersion of wages. *Industrial and Labor Relations Review 34*(1), 3–23.

Freeman, R. B. and M. M. Kleiner (1990). Employer behavior in the face of union organizing drives. *Industrial and Labor Relations Review 43*(4), pp. 351–365.

Freeman, R. B. and J. L. Medoff (1984). *What Do Unions Do?* New York: Basic Books.

Hall, R. E. (2009). Reconciling cyclical movements in the marginal value of time and the marginal product of labor. *Journal of Political Economy 117*(2), 281–323.

Hirsch, B. T. (2004). What do unions do for economic performance? *Journal of Labor Research 25*(3), 415–456.

Hirsch, B. T. and J. L. Neufeld (1987). Nominal and real union wage differentials and the effects of industry and smsa density: 1973-83. *The Journal of Human Resources 22*(1), 138–148.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies 57*(2), 279–298.

Johnson, G. E. (1997). Changes in earnings inequality: the role of demand shifts. *Journal of Economic Perspectives 11*(2), 41–54.

Kahn, L. M. and M. Curme (1987). Unions and nonunion wage dispersion. *The Review of Economics and Statistics 69*(4), 600–607.

Kleiner, M. M. (2001). Intensity of management resistance: understanding the decline of unionization in the private sector. *Journal of Labor Research 22*(3), 519–540.

Krishna, V. and R. Serrano (1996). Multilateral bargaining. *Review of Economic Studies 63*(1), 61–80.

Krueger, A. B. and B. D. Meyer (2002). Labor supply effects of social insurance. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume Volume 4, pp. 2327–2392. Elsevier.

Krusell, P. and L. Rudanko (2016). Unions in a frictional labor market. *Journal of Monetary Economics 80*, 35–50.

Lee, D. S. and A. Mas (2012). Long-run impacts of unions on firms: New evidence from financial markets, 1961-1999. *The Quarterly Journal of Economics 127*(1), 333–378.

Lindbeck, A., D. J. Snower, et al. (1989). The insider-outsider theory of employment and unemployment. *MIT Press Books 1*.

Manzo, F. and R. Bruno (2017). The impact of "right-to-work" laws on labor market outcomes in three midwest states: Evidence from indiana, michigan, and wisconsin (2010-2016). *PMCR Reports*.

Nash, J. F. (1950). The bargaining problem. *Econometrica 18*(2), 155–162.

Neumark, D. and M. L. Wachter (1995). Union effects on nonunion wages: Evidence from panel data on industries and cities. *Industrial and Labor Relations Review 49*(1), pp. 20–38.

Nickell, S. and R. Layard (1999). Chapter 46 labor market institutions and economic performance. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economic*, Volume Volume 3, Part 3, pp. 3029–3084. Elsevier.

Pissarides, C. A. (1986). Trade unions and the efficiency of the natural rate of unemployment. *Journal of Labor Economics 4*(4), pp. 582–595.

Rosen, S. (1969). Trade union power, threat effects and the extent of organization. *The Review of Economic Studies 36*(2), 185–196.

Roth, A. E. (1979). *Axiomatic Models of Bargaining*. Springer-Verlag.

Silva, J. I. and M. Toledo (2009). Labor Turnover Costs And The Cyclical Behavior Of Vacancies And Unemployment. *Macroeconomic Dynamics 13*(S1), 76–96.

Stole, L. A. and J. Zwiebel (1996). Intra-Firm bargaining under Non-Binding contracts. *Review of Economic Studies 63*(3), 375–410.

Traxler, F. (1994). Collective bargaining: Levels and coverage. In *Employment Outlook 1994*, pp. 167–194. OECD.

Vaghul, K. and B. Zipperer (2016). Historical state and sub-state minimum wage data. *Washington Center for Equitable Growth*.

Verna, A. (2005). What do unions do to the workplace? union impact on management and hrm policies. *Journal of Labor Research 26*(3), 415–449.

Vieira, P. (2014). Canadian supreme court rules against wal-mart over store closing. *The Wall Street Journal*.

Voos, P. B. (1983). Union organizing: Costs and benefits. *Industrial and Labor Relations Review 36*(4), pp. 576–591.

Online Appendices

A Numerical example

The simple economy of Section 3.2 is useful to derive some analytical results, but to get a full sense of the behavior of a firm under threat it helps to consider a richer environment in which there is a large number of skill groups. For that purpose, I consider a firm evolving in the economy described in Table 8.

Definition	Parameter	Value
Number of skills groups	S	20
Probability of job destruction	δ	0.05
Discount rate	γ	0.95
Cost of posting a vacancy	κ	3
Nonunion bargaining power	β_n	0.5
Union bargaining power	β_u	0.5
Outside option of workers	c_s	Linear from 1 to 5
Labor market tightness	θ_s	Linear from 1 to 10
Shape of voting preference	ϕ	$[1 + \exp\{-50(w^u - w^n)\}]^{-1}$
Firm's total factor productivity	A	1000
Firm's return to scale parameter	α	0.7
Skill intensity	z_s	$1/S$
Elasticity of substitution	σ	1

Table 8: Parameters for the simulations

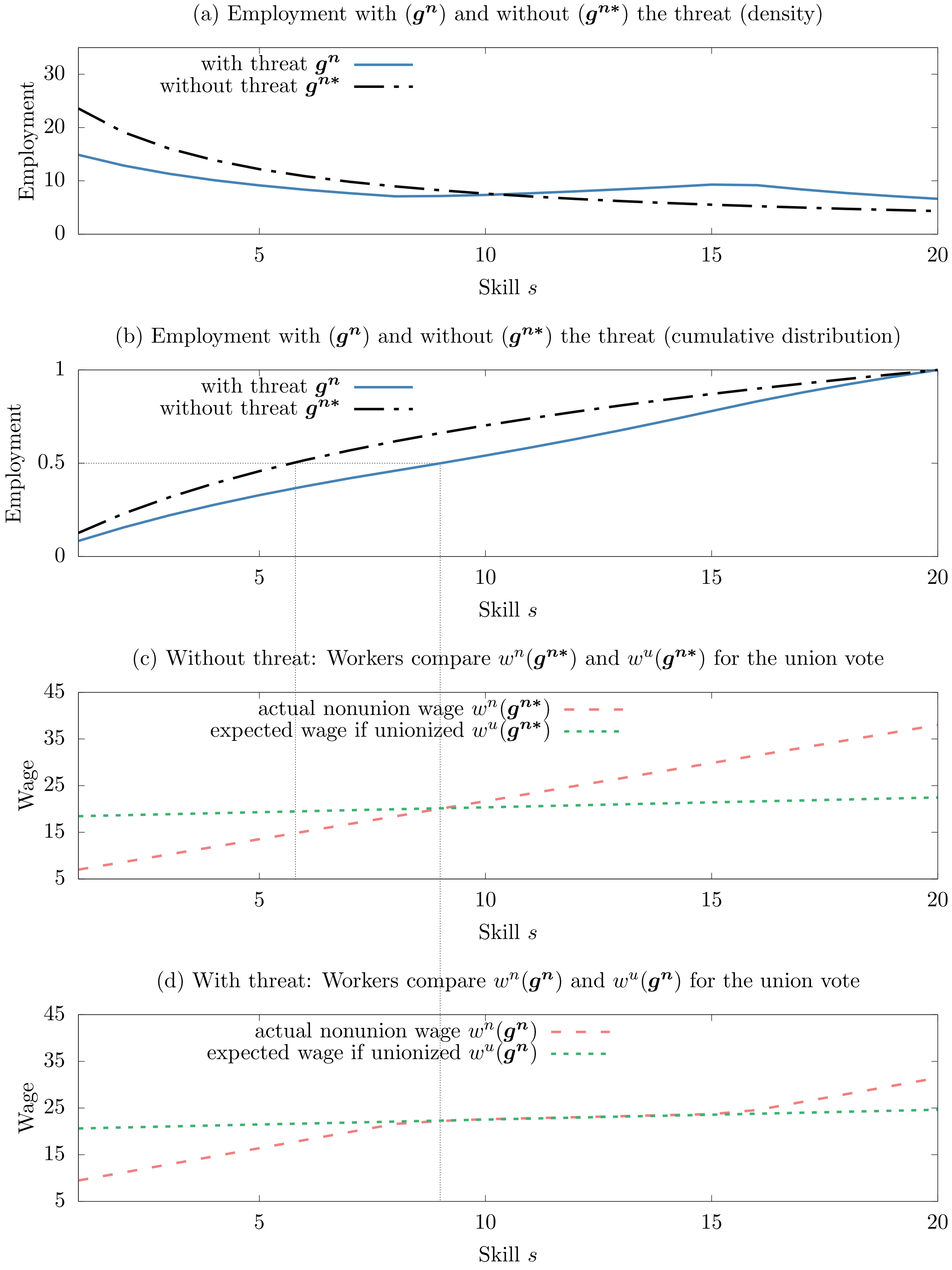
Figure 7 shows how the firm behaves in this environment, with and without the threat. The first two panels show the distribution of workers that the firm employs. Panel (a) shows the density of that distribution, while Panel (b) shows its cumulative density, which makes it easy to identify the median voter within the firm. Panel (c) shows the wages that each worker expects to receive under both union statuses, when the firm is not subject to the union threat. Panel (d) provides the same information but, this time, when the firm is subject to the threat.

To better understand the figure, it is useful to first suppose that this firm operates in an economy in which workers cannot form a union. That firm therefore hires according to (19), and its optimal hiring decision $\boldsymbol{g^{n*}}$ is represented by dashed lines in Panels (a) and (b). This firm also pays wages $\boldsymbol{w^n}(\boldsymbol{g^{n*}})$ shown by the dashed red line in Panel (c). Now, suppose that a change in legislation allows for the creation of a union through a majority election. Hiring according to $\boldsymbol{g^{n*}}$ might no longer be optimal for the firm if it leads to unionization. To see whether that is the case, we need to see how the median worker, identified as $s = 6$ in Panel (b), votes. Comparing that worker's wage with and without a union in Panel (c), we see that he or she will vote to unionize. As a result, if the firm hires according to $\boldsymbol{g^{n*}}$ a union will be created and the firm's profits will suffer.

As a result, the firm attempts to prevent unionization. It does so by over-hiring high-skill workers (who vote against the union) and under-hiring low-skill (who vote in favor of the union) until the outcome of the union election is reversed. This new hiring decision, $\boldsymbol{g^n}$, which solves (20), is shown by the solid line in Panels (a) and (b). Once again, when deciding how to vote, the workers compare their nonunion and union wages, $\boldsymbol{w^n}(\boldsymbol{g^n})$ and $\boldsymbol{w^u}(\boldsymbol{g^n})$, both shown in Panel (d). From Panel (b) we see that the new median worker has skill $s = 9$ and, from Panel (d), we know that this worker is slightly better off without the union, and so the firm remains union free.

While the firm's hiring decisions change to prevent unionization, so do wages. In particular, hiring more high-skill workers lowers their marginal product which lowers their nonunion wage $\boldsymbol{w^n}(\boldsymbol{g^n})$, in Panel (d), compared to the unconstrained wage schedule $\boldsymbol{w^n}(\boldsymbol{g^{n*}})$, seen in Panel (c). In contrast, low-skill wages increase because of the opposite mechanism. As a result, the mere possibility of unionization shrinks the range of wages paid by the firm, even though the firm itself remains union free. Through that channel, the threat of unionization can lower wage inequality.[47]

[47] In Panel (d), $\boldsymbol{w^u}(\boldsymbol{g^n})$ and $\boldsymbol{w^n}(\boldsymbol{g^n})$ are very close to each other for skills in the middle of the distribution. The firm would like to hire more of these workers—they vote against the union—but doing so would lower their marginal product and push their nonunion wage under their union wage. If this were to happen, these workers would change their vote to support the union, which the firm wants to avoid.



(a) Employment with ($\boldsymbol{g^n}$) and without ($\boldsymbol{g^{n*}}$) the threat (density)

(b) Employment with ($\boldsymbol{g^n}$) and without ($\boldsymbol{g^{n*}}$) the threat (cumulative distribution)

(c) Without threat: Workers compare $w^n(\boldsymbol{g^{n*}})$ and $w^u(\boldsymbol{g^{n*}})$ for the union vote

(d) With threat: Workers compare $w^n(\boldsymbol{g^n})$ and $w^u(\boldsymbol{g^n})$ for the union vote

Figure 7: Employment and wages with and without the union threat

B Capital as a factor of production

This section shows that it is straightforward to include capital as an additional input into production. Consider a firm j with a production function

$$\tilde{F}_j\left(K_j, L_j\right) = \tilde{A}_j\left(K_j^{1-\gamma_j} L_j^{\gamma_j}\right)^{\omega_j}$$

where γ_j denotes the labor intensity of the firm, ω_j denotes its total returns to scale and $L_j = \left(\sum_{s\in\mathcal{S}} z_{j,s} g_{j,s}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ is the same aggregated labor input as in the production function (1) in Section 2. Assuming that the firm has access to capital at a constant rental rate r, we can take the first-order conditions with respect to K_j and use the optimal capital decision $K_j^*\left(L_j\right)$ to rewrite the production function as

$$\tilde{F}_j\left(K_j^*\left(L_j\right), L_j\right) = \left(1-\left(1-\gamma_j\right)\omega_j\right)\tilde{A}^{\frac{1}{1-\left(1-\gamma_j\right)\omega_j}}\left(\frac{\left(1-\gamma_j\right)\omega_j}{r}\right)^{\frac{\omega_j\left(1-\gamma_j\right)}{1-\left(1-\gamma_j\right)\omega_j}} L_j^{\frac{\gamma_j\omega_j}{1-\left(1-\gamma_j\right)\omega_j}}.$$

We can see that this equation has the same form as the original production function 1 if we redefine A_j and α_j as

$$A_j = \left(1-\left(1-\gamma_j\right)\omega_j\right)\tilde{A}^{\frac{1}{1-\left(1-\gamma_j\right)\omega_j}}\left(\frac{\left(1-\gamma_j\right)\omega_j}{r}\right)^{\frac{\omega_j\left(1-\gamma_j\right)}{1-\left(1-\gamma_j\right)\omega_j}} \text{ and } \alpha_j = \frac{\gamma_j\omega_j}{1-\left(1-\gamma_j\right)\omega_j}.$$

As a result, including capital as an additional factor of production is equivalent to simply relabeling the parameters of the production function (1).

C Additional results about the quantitative exercises

C.1 Additional figures for the benchmark calibration

Figure 8 shows how employment for union and nonunion workers change in response to the policy exercises of Section 4.2. As explained in the discussion surrounding Figure 4, the disappearance of the threat ("no threat" policy) pushes nonunion firms to hire more workers, particularly low-skill ones since they are the ones that were previously voting in favor of unionization. As a result of the increase demand for workers by the nonunion firms, the union firms shrink down slightly.

When unions are prohibited from the economy ("no union" policy), all workers are now nonunion workers which leads to the large increase in nonunion employment seen in Panel (a). Similarly, when unions are mandatory ("all union" policy), all workers are union workers which explains the large increase in union employment in Panel (b). Notice that these increases in employment are not symmetrical across skill groups since the union and nonunion firms have different skill intensity schedules $\boldsymbol{z}$.

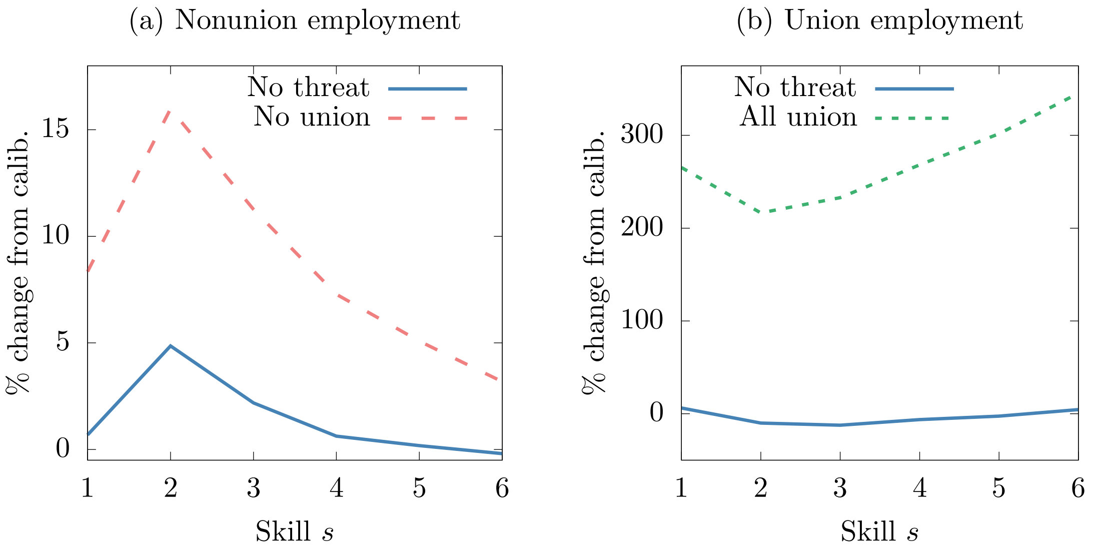


Figure 8: Impact of the policy changes on union and nonunion workers

Figure 9 also shows how employment reacts to the policy experiments but instead of grouping firms by union status it instead groups them by technology. Firms that are *initially* union free in the calibrated economy are in Panel (a) and firms that are *initially* unionized are in Panel (b).

The interpretation of the impact of the threat removal ("no threat" policy) is essentially the same as in Section 4.2. To understand the effect of the two other policies, consider first the impact of the "all union" policy on firms that were initially union free, in Panel (a). As the policy is enacted, these firms begin to bargain collectively with their workers and, in particular, now keep a smaller fraction of the joint surplus (since the estimation finds $\beta_u > \beta_n$). This change in effective bargaining power increases the slope of the marginal cost schedule given by equation (17). As a result, these firms hire relatively fewer high-skill workers and relatively more low-skill workers. Firms that were initially unionized are not directly affected by the change in policy since they keep bargaining with bargaining power β_u. They however change their hiring behavior since the initially union free firms increase their demand of low-skill workers and reduce their demand of high-skill workers. Because of general equilibrium forces (changes in labor market tightness and outside option of the workers), the union firms take the opposite position and become relatively more high-skill intensive.

The situation is reversed when we consider the "no union" policy. In this case, the initially unionized firms are the ones whose bargaining power is changing from β_u to β_n and are therefore the ones driving the changes in hiring behavior. Because of the change in bargaining power, the marginal cost schedule that they face becomes less steep — thereby favoring the hiring of high-skill workers relative to low-skill workers. This change in hiring is visible in the dash curve in Panel (b).

Firms that were initially union free, in Panel (a), react to these changes by hiring more low-skill workers, who are now more attractive, and fewer high-skill workers.

Notice that while the skill composition of employment in each firm changes in response to the different union policies, the overall employment level in each firm remains relatively stable.

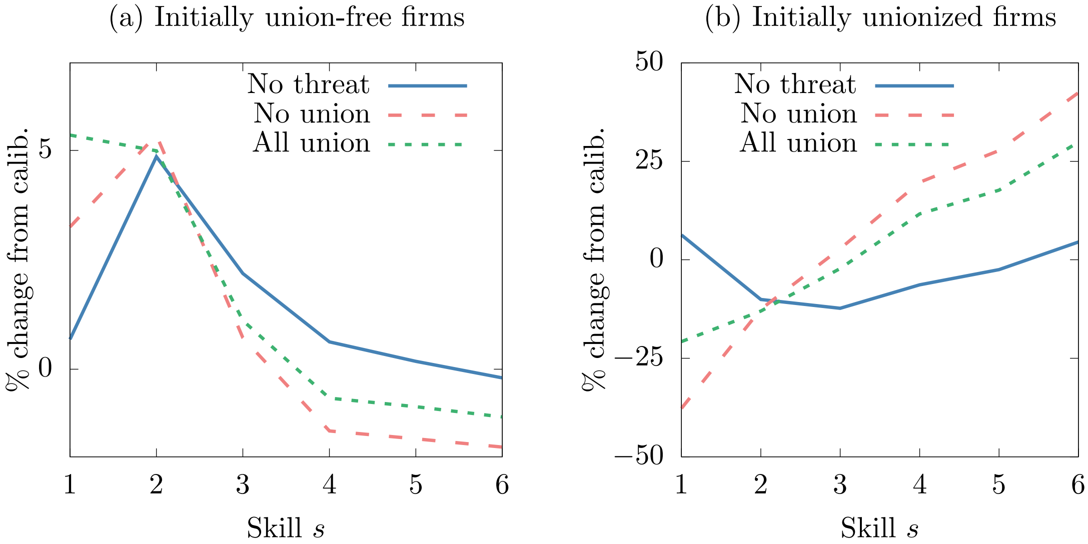


Figure 9: Impact of the experiments on employment as a function of the initial union status of each firm

C.2 Quantitative model with free-entry of firms

This section replicates the exercises of Section 4 in an economy in which firms are free to enter the economy. This change to the environment requires a few additional modeling assumptions. First of all, potential firms are free to enter the economy by paying a fixed cost $f > 0$. After the fixed cost has been paid, each firm is randomly assigned a technology $j \in [0, 1]$. The probability of being assigned to a given type is proportional to the mass of firms of this type in the calibrated economy.[48] Finally, the job destruction shock δ is now assumed to be firm specific and to destroy the firm. In this environment, firms enter until the cost of entering equals the expected discounted profits of entering.

The estimated parameters are identical to those of the benchmark exercise of Section 4. Table 9 shows the outcome of the experiments under free entry. Overall the union threat has a larger impact on output and welfare while its impact on the unemployment rate is smaller. The intuition behind this result is that the removal of the union threat triggers the entry of new firms. The number of workers per firm also increases but less so than in the benchmark economy. As a result,

[48] In practice, I keep the mass of firm of each type fixed and adjust the total factor productivities $\{A_j\}_{j\in[0,1]}$. See Lemma A1 for the equivalence result that validates this approach.

the labor productivity of each firm increases, and output and welfare increase by more than in the benchmark economy even though the unemployment rate declines by a smaller amount.

	Changes from calibrated economy		
	No threat	No unions	All unions
Output	5.30%	5.00%	2.84%
Unemployment rate	-0.94 pp	-1.01 pp	-2.07 pp
Welfare	3.98%	3.59%	0.77%
Wages			
Mean	-0.12%	-0.22%	0.95%
Variance	-0.27%	5.54%	-47.3%

Notes: Differences from the calibrated economy. All numbers are percentage changes except for the unemployment rate number which is a difference in percentage points. Wages are measured in logs. Output refers to value added.

Table 9: Impact of policy experiments under free-entry of firms

D Additional results about the right-to-work regressions

D.1 Robustness results for the earnings regressions

One potential issue with the regressions of Table 7 is that right-to-work legislations might be enacted at the same time as other policies that would affect labor earnings. To alleviate this issue, I include control variables for two additional policies: the effective minimum wage (the maximum of the federal and state minimum wages) and the number of weeks of unemployment benefit extensions in each state.[49] The regression coefficients are in Table 10 below. Interestingly the point estimates are larger and the standard errors are smaller than in the baseline specifications. This last point is all the more surprising since, given that the data on benefit extensions only begins in 1996, there are fewer observations in the sample. Overall, the impact of the unionization threat on the nonunion earnings—the main coefficient of interest—remains strong and statistically significant.

D.2 Impact of right-to-work laws on the skill composition and the skill premium

The model also has predictions for the impact of right-to-work laws on the skill premium and the skill mix hired by the firms. As the threat of unionization gets weaker 1) nonunion firms should

[49] Monthly, state-level minimum wage data from May 1974 up to July 2016 has been put together by Vaghul and Zipperer (2016) and is available online at `https://github.com/equitablegrowth/VZ_historicalminwage/releases`. For data from August 2016 to December 2018, I use annual, state-level data from the U.S. department of labor available online at `https://www.dol.gov/whd/state/stateMinWageHis.htm`. Data about the unemployment benefits extensions has been put together for Chodorow-Reich et al. (2018) and is available online at `https://sites.google.com/site/loukaskarabarbounis/research`. That data only covers the period from 1996 to 2015. I combine these data using data about extension from 2015 to 2018 from the Department of Labor available online at `https://oui.doleta.gov/unemploy/trigger/`. Unfortunately, there is no data available for the period before 1996.

	(1)	(2)	(3)
Dependent variable	Earnings	Earnings	Earnings
Workers in the sample	All	Nonunion	Union
Right-to-work law	-0.039**	-0.031**	-0.037
	0.016	0.012	0.028
State, time fixed effects	yes	yes	yes
Individual & state controls	yes	yes	yes

Notes: Same as Table 7 in the main text, except that the state controls now include the log of the effective minimum wage and the number of weeks of unemployment benefit extensions. Given the data restrictions, the sample now covers the period from January 1996 to December 2018.

Table 10: Impact of right-to-work laws on weekly earnings

hire relatively fewer high-skill workers, and 2) high-skill wages on nonunion firms should go up compared to low-skill wages. This increase in the skill premium pushes, in turn, for an increase in the variance of earnings. I have tested these predictions in the data. To do so, I define high-skill workers as those with a bachelor or advanced degree. Table 11 shows the results of the regressions.[50] We see from columns (1) and (3) that the main predictions of the model are verified in the data. Nonunion firms hire fewer high-skill workers and pay them more. The results of columns (2) and (4) also make sense in light of the theory. If nonunion firms hire relatively fewer high-skill workers, these workers must go somewhere else and we observe the opposite movement in skill composition in the union sector, which explains column (2). We see from column (4) that high-skill workers are also earning relatively more in union firms after the passage of right-to-work legislation. This is not surprising since these workers could move to nonunion firms where earnings are now higher. As a result, high-skill union workers can extract a larger share of the surplus in bargaining and their earnings go up as well.

	(1)	(2)	(3)	(4)
Dependent variable	Share of high skill workers		Skill earnings premium	
Workers in the sample	Nonunion	Union	Nonunion	Union
Right-to-work law	-0.124***	0.229***	0.053***	0.012
	0.027	0.053	0.016	0.036
State, time and industry f.e.	yes	yes	yes	yes
Standard error clustering	state	state	state	state

Notes: Same as Table 7 in the main text. ‘Share of high skill workers’ is defined as the log of the fraction of high-skill workers. The ‘Skill earnings premium’ is defined as the log of high-skill earnings minus the log of low-skill earnings.

Table 11: Impact of right-to-work laws on skill composition and skill premium

[50] These results are robust to including the state-level unemployment rate, the effective minimum wage and unemployment benefit extensions as covariates.

E Alternative wage-setting protocols

This section considers how alternative ways of setting wages interact with the mechanisms of the model.

E.1 Unrestricted nonunion wages

In the benchmark model, the threat of unionization pushes a firm to hire more high-skill workers, relative to their low-skill counterparts, and to compress the range of wages that it pays. To show the robustness of these results, I consider in this section an economy in which firms can pick any nonunion wage schedule. Union wages are still set though collective bargaining. To keep the analysis tractable, I make the same assumptions as in the simple economy of Section 3.2. Recall that, in this environment, the union wage of high-skill and low-skill workers is simply

$$w^u\left(g\right) = \frac{\beta_u}{g_l + g_h} F\left(\boldsymbol{g}\right).$$

Consider first an environment without any possibility of unionization. The firm optimally sets nonunion wages to the worker's reservation value, which is 0 in this simplified environment. In that case, the firm only has to pay the vacancy costs and its problem becomes

$$\left(g_l^{z_l} g_h^{z_h}\right)^{\alpha} - \kappa\left(\frac{g_h}{q(\theta_h)} + \frac{g_l}{q(\theta_l)}\right)$$

Taking first-order conditions yields the optimal employment levels

$$g_s^{n*} = \left(\frac{\alpha}{\kappa}\right)^{\frac{1}{1-\alpha}} \left[\left(z_l q\left(\theta_l\right)\right)^{z_l} \left(z_h q\left(\theta_h\right)\right)^{z_h}\right]^{\frac{\alpha}{1-\alpha}} z_s q\left(\theta_s\right)$$

for $s \in \{l, h\}$.

Suppose now that the workers have the option to unionize, in which case their wages would be determined through collective bargaining. Since all workers would earn more in that case, they would vote to unionize under $\boldsymbol{g^{n*}}$ and this employment vector is no longer optimal—the firm is able to do better if it hires to prevent unionization. In order to do so, one can show that the optimal strategy is to pay high-skill workers a tiny amount over their union wage and to hire enough of them to win the election ($g_h = g_l = g$). In this case, the firm's problem becomes

$$\max_{g} g^{\alpha} - \frac{\beta_u}{2} g^{\alpha} - \kappa\left(\frac{g}{q(\theta_h)} + \frac{g}{q(\theta_l)}\right)$$

and the first-order condition yields

$$g^n = \left(1 - \frac{\beta_u}{2}\right)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{\kappa}\right)^{\frac{1}{1-\alpha}} \left(\frac{1}{q(\theta_h)} + \frac{1}{q(\theta_l)}\right)^{-\frac{1}{1-\alpha}}.$$

Steps similar to those of the proofs of Propositions 5 and 6 show that the firm now hires relatively more high-skill workers than without the threat ($g_h^{n*}/g_l^{n*} < g_h^n/g_l^n$) and that $F\left(\boldsymbol{g^{n*}}\right) > F\left(\boldsymbol{g^n}\right)$. In addition, the average wage paid by the firm obviously increases as a result of the threat. Note finally that the marginal products of the high-skill and low-skill workers gets closer together because of the threat

$$\frac{\mathrm{MP}_h^n}{\mathrm{MP}_l^n} = \frac{z_h}{z_l} < \frac{q\left(\theta_l\right)}{q\left(\theta_h\right)} = \frac{\mathrm{MP}_h^{n*}}{\mathrm{MP}_l^{n*}}.$$

This last result would, in a richer model in which wages reflect marginal products, contribute to pushing wages closer together.

Overall, these predictions are the same as in the benchmark model, as shown in Section 3.2. As a result, the main mechanisms of the model do not hinge on the precise bargaining protocol but are robust to alternative wage-setting mechanisms.

E.2 Alternative ways of setting union wages

This appendix considers alternative bargaining procedures for union wages. In particular, it investigates how the mechanisms of the paper would change if a union organization with its own preferences acted as an intermediary between the workers and the firm. Notice that this is different from the benchmark model in which union wages are determined by an n-player bargaining between the firm and the workers. I keep the framework as simple as possible to make the exposition transparent. In particular, I assume that all jobs are destroyed at the end of the period and that all reservation values are zero. It is straightforward to add them back.

The bargaining now takes place in two stages. In the second stage, a union bargains with the firm on how to split the surplus generated by production. In the first stage, the workers decide on how to split their share of the surplus they will received in the second stage.

Suppose that, in this second stage, the firm bargains with a union with linear preferences. The bargaining problem is

$$\max_T T^\beta \left(F(\boldsymbol{g}) - T\right)^{1-\beta}$$

where T is the transfer between the firm and the union. The solution is the standard outcome of Nash-bargaining: the union keeps a fraction β of the joint surplus $T(g) = \beta F(g)$. Then, the workers have to split this surplus among them. This can be modeled in two ways. We can assume that each worker enters a one-on-one negotiation with the union, or we can assume that there is collective bargaining between the workers and the union. I explore both of these cases below.

One-on-one negotiation between the workers and the union

In this first case, both parties know that if the worker walks away the union will extract a smaller amount from the firm in the second stage (production will happen with one less worker). Let us assume that all workers have the same bargaining power and that they bargain with a union leader who captures what's left of the surplus. The surplus of a worker from agreeing to stay in the union is, under our assumptions, simply w_s. The surplus of the union leader is

$$\frac{\partial T\left(\boldsymbol{g}\right)}{\partial g_s} - w_s\left(\boldsymbol{g}\right) - \sum_k g_k \frac{\partial w_k\left(\boldsymbol{g}\right)}{\partial g_s}$$

where we see that the union internalizes the fact that, if this worker walks away, all the other negotiated wages may change. This is the Stole and Zwiebel (1996) bargaining. This problem is very similar to the one encountered with the nonunion individual bargaining in the main text. The outcome of the bargaining solves the system of equations

$$\frac{\partial T\left(\boldsymbol{g}\right)}{\partial g_s} - w_s\left(\boldsymbol{g}\right) - \sum_k g_k \frac{\partial w_k\left(\boldsymbol{g}\right)}{\partial g_s} = \frac{1-\chi}{\chi} w_s$$

where χ is the bargaining power of the workers in their negotiations with the union. The solution is

$$w_s = \frac{\chi}{1-\chi(1-\alpha)} \frac{\beta\alpha z_s}{g_s^{1/\sigma}} A \left(\sum_k z_k g_k^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}}.$$

We see that, under these assumptions, this new *union* wage schedule is essentially the same as the *nonunion* wage equation (12) in the main text. Therefore, this model does not seem appropriate to explain the union-generated wage compression observed in the data. Furthermore, if we assume that the union leader gets a negligible share of the surplus ($\chi \to 1$), all workers would always vote for the union.

Collective bargaining between all workers and the union

We now consider the other way of splitting the surplus extracted from the firm. Here, the workers and the union enter a single collective bargaining process. Under our assumptions, this problem is

$$\max_{\boldsymbol{w}} \left(\prod_s \left(w_s\right)^{\frac{g_s}{n}}\right)^{\chi} \left(T(g) - \sum_s w_s g_s\right)^{1-\chi}$$

where again $0 < \chi < 1$ denotes the bargaining power of the workers. This problem is similar to the union bargaining problem (8) from the main text except that it features the union surplus $T\left(\boldsymbol{g}\right)$ instead of the production function $F\left(\boldsymbol{g}\right)$. But, since the second stage bargaining yields $T\left(\boldsymbol{g}\right) = \beta F\left(\boldsymbol{g}\right)$, this difference is inconsequential. In fact, we find that the wage with this procedure is $w_s(\boldsymbol{g}) = \chi\beta F(\boldsymbol{g})/n$. If the share receiver by the union leader goes to zero ($\chi \to 1$), we find the union wage equation (9) from the main text (if we impose the same simplifying assumptions there too). This exercise shows that introducing an actual union organization as an intermediary between the workers and the firm might have no impact on wages.

F Proofs

Lemma 1. *In a steady-state equilibrium, the firm's dynamic problem is equivalent to*

$$\max_{\boldsymbol{g}} \pi(\boldsymbol{g}) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + \kappa(1-\delta)\gamma \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)}.$$

Proof. First, the constraint $v_s \geq 0$ are never binding in a steady-state equilibrium. To see why, suppose that in such an equilibrium a firm's optimal measure of workers is given by g_s^*. Two events might move the firm away from g_s^*. First, every period, it loses a fraction δ of its workers. Second, if one of the wage bargaining sessions breaks down without an agreement, the firm loses additional workers.[51] In both of these cases, the firm has to hire a positive number of workers in the next period to replace those that have been lost. Therefore, $v_s > 0$ in all markets s such that $g_s^* > 0$ and $v_s = 0$ elsewhere. From equation (2), the firm's problem is

$$J(\boldsymbol{g_{-1}}) = (1-\delta)\kappa \sum_{s\in\mathcal{S}} \frac{g_{-1,s}}{q(\theta_s)} + \max_{\boldsymbol{g}} \left\{ \pi(\boldsymbol{g}) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + \gamma J(\boldsymbol{g}) \right\}.$$

The term that is maximized is constant with respect to $\boldsymbol{g_{-1}}$. Denote that constant by B. Then, in particular

$$J(\boldsymbol{g}) = (1-\delta)\kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + B.$$

and the firm solves

$$\max_{\boldsymbol{g}} \pi(\boldsymbol{g}) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + \gamma \left((1-\delta)\kappa \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} + B \right)$$

which is the result. □

Lemma 2. *If all the workers have the same bargaining power, and the firm has bargaining power $1-\beta_u$, the collective Nash bargaining problem can be written as*

$$\max_{\boldsymbol{w}} \left[\prod_{s\in\mathcal{S}} \left(V_s^E(\boldsymbol{w}) - b_s - \gamma V_s^U \right)^{\frac{g_s}{n}} \right]^{\beta_u} \left[F(\boldsymbol{g}) - \sum_{s\in\mathcal{S}} w_s g_s + (1-\delta)\kappa\gamma \sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)} \right]^{1-\beta_u}$$

[51] This does not happen in equilibrium but the value function needs to be defined along these paths to correctly specify the bargaining problems.

where $n = \sum g_s$ is the total mass of employed workers. Furthermore, the wage schedule

$$w_s^u\left(\boldsymbol{g}\right) - c_s = \frac{\beta_u}{n}\left(F\left(\boldsymbol{g}\right) - \sum_{k\in\mathcal{S}} c_k g_k + \gamma\left(1-\delta\right)\kappa \sum_{k\in\mathcal{S}} \frac{g_k}{q\left(\theta_k\right)}\right)$$

solves this bargaining problem.

Proof. Axiomatic bargaining theory (Roth, 1979) (see also Krishna and Serrano (1996)) tells us that the solution to an n-players bargaining problem is the payoff that maximizes the geometric average of the n surpluses and where the average weights can be interpreted as bargaining powers. We therefore look at the surpluses of each player and then compute this average.

Consider the firm's surplus from agreeing on a wage schedule w. At this point, the measure $\boldsymbol{g}$ is fixed and the hiring cost is sunk. In a steady state, the difference in discounted profits for the firm, denoted by $\Delta^u(w)$, is

$$\Delta^u\left(\boldsymbol{w}\right) = \left[\pi\left(\boldsymbol{g},\boldsymbol{w}\right) + \gamma J\left(\boldsymbol{g}\right)\right] - \left[\pi\left(\boldsymbol{0}\right) + \gamma J\left(\boldsymbol{0}\right)\right] \tag{27}$$

where the first term between brackets is discounted profits if an agreement is reached and $\pi(0) + \gamma J(0)$ is the firm's discounted profit if negotiations break down. In such a case, the firm has no workers, it produces nothing and pays no wages. Therefore, $\pi(0) = 0$. $J(0)$ is the value function of a firm that starts the period with no workers. Because of risk-neutrality and the linear vacancy cost, it hires back to its steady state optimal level g^* right away (we have seen in the proof of lemma 1 that, at a steady state, g does not depend on g_{-1}). Therefore,

$$J(\boldsymbol{0}) = \pi\left(\boldsymbol{g^*},\boldsymbol{w^*}\right) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s^*}{q(\theta_s)} + \gamma J\left(\boldsymbol{g^*}\right)$$

where w^* is the equilibrium wage schedule for this firm. We can therefore rewrite equation (27) as

$$\Delta^u\left(\boldsymbol{w}\right) = \pi\left(\boldsymbol{g},\boldsymbol{w}\right) + \gamma J\left(\boldsymbol{g}\right) - \gamma\left(\pi\left(\boldsymbol{g^*},\boldsymbol{w^*}\right) - \kappa\sum_{s\in\mathcal{S}} \frac{g_s^*}{q(\theta_s)} + \gamma J\left(\boldsymbol{g^*}\right)\right).$$

But the firm's value function is

$$J\left(\boldsymbol{g}\right) = \pi\left(\boldsymbol{g^*},\boldsymbol{w^*}\right) - \kappa\sum_{s\in\mathcal{S}} \frac{g_s^* - (1-\delta)g_s}{q(\theta_s)} + \gamma J\left(\boldsymbol{g^*}\right)$$

and therefore the firm's surplus from agreeing on a wage w is

$$\Delta^u\left(\boldsymbol{w}\right) = \pi\left(\boldsymbol{g},\boldsymbol{w}\right) + (1-\delta)\gamma\kappa\sum_{s\in\mathcal{S}} \frac{g_s}{q(\theta_s)}.$$

On the workers' side, the net benefit of an agreement is $V_s^e(w) - b_s - \gamma V_s^u$. Assume now that all workers have the same bargaining power and consider the discrete case in which there are $h_s \in \mathbb{N}$ workers of type s who all have mass $\chi > 0$ such that $h_s \times \chi \to g_s$ as we move to the continuum. The bargaining problem with *equal* bargaining power is

$$\left(V_1^E - b_1 - \gamma V_1^U\right)^{h_1} \times \cdots \times \left(V_i^E - b_i - \gamma V_i^U\right)^{h_i} \times \cdots \times \left(V_S^E - b_S - \gamma V_S^U\right)^{h_S} \times \Delta^u.$$

Since the bargaining power of the firm is $1 - \beta_u$ and that bargaining powers must sum to 1 we get

$$\left(V_1^E - b_1 - \gamma V_1^U\right)^{\frac{\beta_u \chi h_1}{\chi H}} \times \cdots \times \left(V_S^E - b_S - \gamma V_S^U\right)^{\frac{\beta_u \chi h_S}{\chi H}} \times \left(\Delta^u\right)^{1-\beta_u}$$

where $H = \sum_s h_s$. Taking the limit to the continuum, $\frac{\chi h_i}{\chi H} \to \frac{g_i}{n}$ for all i and we find equation (8).

When the surplus from the match is positive, the bargaining problem is defined on a convex set and is strictly concave. First-order conditions are therefore necessary and sufficient and yield the wage equation (9).

□

Lemma 3. *The wage schedule*

$$w_s^n(\boldsymbol{g}) - c_s = \frac{\beta_n}{1-(1-\alpha)\beta_n} \frac{\partial F(\boldsymbol{g})}{\partial g_s} - \beta_n c_s + \beta_n \gamma (1-\delta) \frac{\kappa}{q(\theta_s)}$$

solves the individual bargaining problem (11).

Proof. Substituting (6) in (11), a solution to the individual bargaining problem must solve the following system of partial differential equations

$$\frac{\partial F(\boldsymbol{g})}{\partial g_s} - \sum_{k\in\mathcal{S}} g_k \frac{\partial w_k(\boldsymbol{g})}{\partial g_s} - w_s(\boldsymbol{g}) + \gamma(1-\delta)\frac{\kappa}{q(\theta_s)} = \frac{1-\beta_n}{\beta_n}\left(w_s(\boldsymbol{g}) - c_s\right)$$

for all $s \in \mathcal{S}$. One can verify that general solutions to this system are of the form

$$\begin{aligned} w_s^n(\boldsymbol{g}) - c_s = {} & \frac{\beta_n}{1-\beta_n(1-\alpha)} \frac{\alpha z_s}{g_s^{\frac{1}{\sigma}}} A \left(\sum_{k\in\mathcal{S}} z_k g_k^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}} \\ & - \beta_n c_s + \beta_n \gamma(1-\delta)\frac{\kappa}{q(\theta_s)} + C_s g_s^{-\frac{1}{\beta_n}} \end{aligned}$$

where C_s is a constant term that could depend on $\{g_j\}_{j\neq s}$. The boundary conditions $\lim_{g_s\to 0} w_s^n(\boldsymbol{g})\, g_s = 0$ for all $s \in \mathcal{S}$ guarantees that $C_s = 0$ for all s and the wage equation therefore becomes (12).[52] □

[52] Cahuc et al. (2008) study a similar bargaining problem with more general production functions.

Proposition 1. *If the bargaining powers are equal ($\beta = \beta_n = \beta_u$), the difference between the average nonunion and union wage is*

$$\mathbb{E}_s\left(\boldsymbol{w^n}\left(\boldsymbol{g}\right)\right) - \mathbb{E}_s\left(\boldsymbol{w^u}\left(\boldsymbol{g}\right)\right) = -\frac{\beta\left(1-\beta\right)\left(1-\alpha\right)}{1-\left(1-\alpha\right)\beta}\frac{F\left(\boldsymbol{g}\right)}{n} < 0,$$

where $\mathbb{E}_g$ is the expectation across skills. It follows that the difference between nonunion and union profits is

$$\pi^n\left(\boldsymbol{g}\right) - \pi^u\left(\boldsymbol{g}\right) = \frac{\beta\left(1-\beta\right)\left(1-\alpha\right)}{1-\left(1-\alpha\right)\beta}F\left(\boldsymbol{g}\right) > 0.$$

Proof. From equations (9) and (12):

$$\frac{\sum_{s\in\mathcal{S}} w^n(g_s)g_s}{\sum_{s\in\mathcal{S}} g_s} = \frac{1}{n}\left[\frac{\beta}{1-(1-\alpha)\beta}\alpha F\left(\boldsymbol{g}\right) + (1-\beta)\sum_{s\in\mathcal{S}} c_s g_s + \beta\gamma(1-\delta)\kappa\sum_{s\in\mathcal{S}}\frac{g_s}{q(\theta_s)}\right]$$

and

$$\frac{\sum_{s\in\mathcal{S}} w^u(g_s)g_s}{\sum_{s\in\mathcal{S}} g_s} = \frac{1}{n}\left(\beta F\left(\boldsymbol{g}\right) + (1-\beta)\sum_{s\in\mathcal{S}} c_s g_s + \beta\gamma(1-\delta)\kappa\sum_{s\in\mathcal{S}}\frac{g_s}{q(\theta_s)}\right)$$

Taking the difference yields the first result. Subtracting (10) from (13) yields the second result. □

Proposition 2. *The equilibrium wage schedules $w_s^u\left(\boldsymbol{g^{u*}}\right)$ and $w_s^n\left(\boldsymbol{g^{n*}}\right)$ are increasing in skill s, and the union wage gap $w_s^u\left(\boldsymbol{g^{u*}}\right) - w_s^n\left(\boldsymbol{g^{n*}}\right)$ is decreasing in s.*

Proof. We first start with the union wage. From equation (9), we can write $w_s^u\left(\boldsymbol{g^{u*}}\right) = c_s^u + D$ where D is a constant that does not depend on s. Combining with equation (7), we get $w_s^u\left(\boldsymbol{g^{u*}}\right) = \left(1-\gamma(1-\delta)\right)\left(b_s + D\right) + \gamma\left(1-\gamma\right)\left(1-\delta\right)V_s^U$. Since V_s^U and b_s are increasing in s, so is $w_s^u\left(\boldsymbol{g^{u*}}\right)$.

For the nonunion wage, by combining equations (12) and (18), we find

$$w_s^n\left(\boldsymbol{g^{n*}}\right) = c_s^n + \frac{\beta_n}{1-\beta_n}\frac{\kappa}{q(\theta_s)}.$$

Using equation (7) once again yields

$$w_s^n\left(\boldsymbol{g^{n*}}\right) = \left(1-\gamma(1-\delta)\right)\left(b_s + \frac{\beta_n}{1-\beta_n}\frac{\kappa}{q\left(\theta_s\right)}\right) + \gamma\left(1-\gamma\right)\left(1-\delta\right)V_s^U.$$

Since V_s^U and θ_s are increasing in s so is $w_s^n\left(\boldsymbol{g^{n*}}\right)$.

For the union wage gap, notice that

$$w_s^u\left(\boldsymbol{g^{u*}}\right) - w_s^n\left(\boldsymbol{g^{n*}}\right) = \left(1-\gamma\left(1-\delta\right)\right)\left(D - \frac{\beta_n}{1-\beta_n}\frac{\kappa}{q\left(\theta_s\right)}\right).$$

Since θ_s is increasing in s, the union wage gap is decreasing in s. □

Proposition 3. *The counterfactual union wage gap $w_s^u\left(\boldsymbol{g^{i*}}\right)-w_s^n\left(\boldsymbol{g^{i*}}\right)$ is decreasing in skill s for both firm union status $i\in\{u,n\}$.*

Proof. We first start with the unionized firm. This firm hires according to g_s^{u*}. From lemma 2, we know that $w_s^u\left(\boldsymbol{g^{u*}}\right)=c_s^u+D$ and that c_s^u is increasing in s. Consider now the off-equilibrium nonunion wage that the union workers *would* get if they voted against the union. From equation (12) together with the first-order condition of an unconstrained firm we have

$$w_s^n\left(\boldsymbol{g^{u*}}\right)=(1-\beta_n)c_s^u+\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\mathrm{MC}_s^u}{1-\beta_u}+\beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta_s)}.$$

Using the definition of MC^u, it is straightforward to show that

$$w_s^n\left(\boldsymbol{g^{u*}}\right)=c_s^u+c_s^u\frac{\beta_n^2(1-\alpha)}{1-(1-\alpha)\beta_n}+\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\kappa}{q(\theta_s)}\underbrace{\left(\frac{1}{1-\beta_u}-(1-\alpha)\beta_n\gamma(1-\delta)\right)}_{>0}.$$

Since W_s^u and θ_s are increasing, both $w_s^u\left(\boldsymbol{g^{u*}}\right)$ and $w_s^n\left(\boldsymbol{g^{u*}}\right)$ are increasing in s and

$$\begin{aligned}w_s^u\left(\boldsymbol{g^{u*}}\right)-w_s^n\left(\boldsymbol{g^{u*}}\right)=D-\Bigg(&c_s^u\frac{\beta_n^2(1-\alpha)}{1-(1-\alpha)\beta_n}\\&+\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\kappa}{q(\theta_s)}\left(\frac{1}{1-\beta_u}-(1-\alpha)\beta_n\gamma(1-\delta)\right)\Bigg).\end{aligned}$$

so that the union wage gap $w_s^u\left(\boldsymbol{g^{u*}}\right)-w_s^n\left(\boldsymbol{g^{u*}}\right)$ is decreasing in s.

The nonunion firm hires according to g_s^{n*}. From the proof of the previous proposition, we know that

$$w_s^n\left(\boldsymbol{g^{n*}}\right)=c_s^n+\frac{\beta_n}{1-\beta_n}\frac{\kappa}{q(\theta_s)}.$$

Furthermore, from equation (9), we have that $w_s^u\left(\boldsymbol{g^{n*}}\right)=c_s^n+D'$ where D' is a constant that does not depend on s. Therefore, the union wage gap is

$$w_s^u\left(\boldsymbol{g^{n*}}\right)-w_s^n\left(\boldsymbol{g^{n*}}\right)=D'-\frac{\beta_n}{1-\beta_n}\frac{\kappa}{q(\theta_s)}.$$

Since θ_s is increasing, the union wage gap decreases with s. □

F.1 Simplified Economy

This section contains the proofs of Propositions 4 to 9 established in the simplified economy described in Section 3.2. The assumptions on the structure of the economy imply that $c_s=0$, $\mathrm{MC}_s^n=\kappa/q\left(\theta_s\right)$, $F\left(\boldsymbol{g}\right)=\left(g_l^{z_l}g_h^{z_h}\right)^\alpha$ and $\partial F\left(\boldsymbol{g}\right)/\partial g_s=\alpha F\left(\boldsymbol{g}\right)z_s/g_s$. In addition, the propositions assume that Assumption 1 holds which implies that

$$\frac{\alpha \beta_n}{1-(1-\alpha)\beta_n} < \beta_u \tag{28}$$

and

$$q(\theta_l) > q(\theta_h) . \tag{29}$$

Preliminary results

It is useful to establish some preliminary results that hold throughout this section. First, we can characterize the wage schedule in this environment. For an arbitrary employment vector $\boldsymbol{g}$ the nonunion wage schedule is

$$w_s^n(\boldsymbol{g}) = \frac{\beta_n}{1-(1-\alpha)\beta_n} \frac{\partial F(\boldsymbol{g})}{\partial g_s} = \frac{\beta_n}{1-(1-\alpha)\beta_n} \alpha F(\boldsymbol{g}) \frac{z_s}{g_s}, \tag{30}$$

and the union wage schedule is

$$w_s^u(\boldsymbol{g}) = \beta_u \frac{F(\boldsymbol{g})}{n} = \frac{\beta_u}{g_l + g_h} \left(g_l^{z_l} g_h^{z_h}\right)^{\alpha} . \tag{31}$$

We can also write the profits of a firm under some arbitrary employment vector $\boldsymbol{g}$ as

$$\begin{aligned} \Pi^i(\boldsymbol{g}) &= F(\boldsymbol{g}) - \sum_{s \in \mathcal{S}} g_s w_s^i(\boldsymbol{g}) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)} \\ &= B_i F(\boldsymbol{g}) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)} \end{aligned}$$

where $i \in \{u, n\}$ indicates the union status of the firm. Finally, we can solve the problem of an unconstrained firm in this environment. Notice that the problem is convex. The first-order conditions are

$$g_s^{i*} = \frac{\alpha B_i}{\kappa} F(\boldsymbol{g}) z_s q(\theta_s) \tag{32}$$

for $s \in \{l, h\}$ such that output is

$$F\left(\boldsymbol{g^{i*}}\right) = \left(\frac{\alpha B_i}{\kappa}\right)^{\frac{\alpha}{1-\alpha}} \left((z_l q(\theta_l))^{z_l} (z_h q(\theta_h))^{z_h}\right)^{\frac{\alpha}{1-\alpha}} \tag{33}$$

and profits are

$$\begin{aligned} \Pi^i\left(\boldsymbol{g^{i*}}\right) &= (1-\alpha) B_i F(\boldsymbol{g}) \\ &= (1-\alpha) B_i \left(\frac{\alpha B_i}{\kappa}\right)^{\frac{\alpha}{1-\alpha}} \left((z_l q(\theta_l))^{z_l} (z_h q(\theta_h))^{z_h}\right)^{\frac{\alpha}{1-\alpha}} . \end{aligned} \tag{34}$$

We can also compute total employment as

$$g_h^{i*} + g_l^{i*} = \left(\frac{\alpha B_i}{\kappa}\right)^{\frac{1}{1-\alpha}} \left(\left(z_l q\left(\theta_l\right)\right)^{z_l} \left(z_h q\left(\theta_h\right)\right)^{z_h}\right)^{\frac{\alpha}{1-\alpha}} \left(z_h q\left(\theta_h\right) + z_l q\left(\theta_l\right)\right). \tag{35}$$

With these results in hand, we now turn to the proofs of the propositions.

Proposition 4. *The union threat is binding for nonunion firms, i.e* $\Lambda\left(\boldsymbol{g}\right) = 0$.

Proof. Suppose, by contradiction, that there is an unconstrained nonunion firm. Combining the FOC (32) for $s = h$ and $s = l$, we find that $g_l > g_h$. Low-skill workers therefore have the majority of the vote in the union election. Low-skill workers also vote in favor of the union. Indeed, from (30) and (31) we find that low-skill workers vote for unionization if

$$w_l^u \geq w_l^n \Leftrightarrow \frac{\beta_u}{g_l + g_h} F\left(\boldsymbol{g}\right) \geq \frac{\beta_n}{1 - \left(1 - \alpha\right)\beta_n} \alpha F\left(\boldsymbol{g}\right) \frac{z_l}{g_l}$$

simplifying we find

$$\beta_u \geq \frac{\alpha \beta_n}{1 - \left(1 - \alpha\right)\beta_n} z_l \left(1 + \frac{g_h}{g_l}\right). \tag{36}$$

Since $g_l > g_h$ and $z_l \leq 0.5$, we have $z_l\left(1 + g_h/g_l\right) < 1$. Combining with (28), (36) is satisfied and low-skill workers vote in favor of the union. Since they have the majority of the vote we have a contradiction. As a result, any nonunion firm must be constrained. □

Proposition 5. *The union threat lowers the profits, employment and output of nonunion firms.*

Proof. A firm is union free if (1) $g_h \geq g_l$ and $w_h^n\left(\boldsymbol{g}\right) \geq w_h^u\left(\boldsymbol{g}\right)$, or (2) $g_l \geq g_h$ and $w_l^n\left(\boldsymbol{g}\right) \geq w_l^u\left(\boldsymbol{g}\right)$. One can show that option (2) is infeasible or unprofitable for the firm so we focus on option (1). The problem of the firm is therefore

$$\max_{g_l, g_h} F\left(\boldsymbol{g}\right) - \sum_{s \in \mathcal{S}} g_s w_s^n\left(\boldsymbol{g}\right) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)}$$

subject to $g_h \geq g_l$ and $w_h^n\left(\boldsymbol{g}\right) \geq w_h^u\left(\boldsymbol{g}\right)$. Taking advantage of (30) and (31), we can write the Lagrangian as

$$\mathcal{L} = \frac{1 - \beta_n}{1 - \left(1 - \alpha\right)\beta_n} F\left(\boldsymbol{g}\right) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s}{q\left(\theta_s\right)} - \lambda_1\left(g_l - g_h\right) - \lambda_2\left(\beta_u g_h - \frac{\beta_n}{1 - \left(1 - \alpha\right)\beta_n} \alpha z_h\left(g_l + g_h\right)\right)$$

where λ_1 and λ_2 are the Lagrange multipliers on the first and second constraints, respectively. Notice that this is a convex optimization problem so that the first-order conditions are sufficient to characterize a solution.

The first-order condition with respect to g_l is

$$\frac{1-\beta_n}{1-(1-\alpha)\beta_n}\alpha F(\boldsymbol{g})\frac{z_l}{g_l}-\frac{\kappa}{q(\theta_l)}-\lambda_1+\lambda_2\frac{\beta_n}{1-(1-\alpha)\beta_n}\alpha z_h=0$$

and the first-order condition with respect to g_h is

$$\frac{1-\beta_n}{1-(1-\alpha)\beta_n}\alpha F(\boldsymbol{g})\frac{z_h}{g_h}-\frac{\kappa}{q(\theta_h)}+\lambda_1-\lambda_2\beta_u+\lambda_2\frac{\beta_n}{1-(1-\alpha)\beta_n}\alpha z_h=0.$$

We know from Proposition 4 that any nonunion firm is constrained. Therefore, at least one of the two constraints binds. We will show that $g_h \geq g_l$ must binds. Suppose not. Then $w_h^n(\boldsymbol{g}) = w_h^u(\boldsymbol{g})$, $\lambda_2 > 0$ and $\lambda_1 = 0$. Multiply each first-order condition by their respective g_s and add them up, we find

$$\frac{1-\beta_n}{1-(1-\alpha)\beta_n}\alpha F(\boldsymbol{g})=\frac{\kappa}{q(\theta_l)}g_l+\frac{\kappa}{q(\theta_h)}g_h.$$

Similarly, taking their difference we get

$$\frac{1-\beta_n}{1-(1-\alpha)\beta_n}\alpha F(\boldsymbol{g})\left(\frac{z_l}{g_l}-\frac{z_h}{g_h}\right)+\left(\frac{\kappa}{q(\theta_h)}-\frac{\kappa}{q(\theta_l)}\right)+\lambda_2\beta_u=0.$$

Combining these two equations to remove $F(\boldsymbol{g})$, yields

$$\frac{\kappa z_l}{q(\theta_h)}\left(1+\frac{g_h}{g_l}\right)-\frac{\kappa z_h}{q(\theta_l)}\left(1+\frac{g_l}{g_h}\right)+\lambda_2\beta_u=0. \tag{37}$$

The difference between the first two terms is positive if

$$z_l q(\theta_l)\left(1+\frac{g_h}{g_l}\right)>z_h q(\theta_h)\left(1+\frac{g_l}{g_h}\right)$$

which is true because $g_h > g_l$ and $z_l q(\theta_l) > z_h q(\theta_h)$ by Assumption 1. Since $\lambda_2 > 0$, equation (37) cannot be true. As a result, it must be that $g_h = g_l$.

We now proceed to solving the nonunion constrained problem under $g_h = g_l = g$. Solving the first-order condition, it is straightforward to show that the optimal constrained nonunion employment decision is

$$g^n=\left(B_n\left(\frac{\kappa}{q(\theta_l)}+\frac{\kappa}{q(\theta_h)}\right)^{-1}\alpha\right)^{\frac{1}{1-\alpha}}. \tag{38}$$

Using the production function, we can write output as

$$F(\boldsymbol{g}^n)=(\alpha B_n)^{\frac{\alpha}{1-\alpha}}\left(\kappa q(\theta_l)^{-1}+\kappa q(\theta_h)^{-1}\right)^{\frac{\alpha}{\alpha-1}}. \tag{39}$$

Comparing this quantity with its unconstrained equivalent from (33) we find

$$F\left(\boldsymbol{g^{n*}}\right) \quad \geq \quad F\left(\boldsymbol{g^{n}}\right) \Leftrightarrow \left(z_l^{-1}\frac{\kappa}{q(\theta_l)}\right)^{z_l}\left(z_h^{-1}\frac{\kappa}{q(\theta_h)}\right)^{z_h} \leq \left(\frac{\kappa}{q(\theta_l)}+\frac{\kappa}{q(\theta_h)}\right)$$

which is true since the weighted arithmetic mean is larger than the weighted geometric mean (the proper weights here are z_l and z_h). The same argument directly applies to show that $\Pi^{n*} \geq \Pi^n$. We now turn to total employment. Using (35) together with (38), we compare

$$g_h^{n*}+g_l^{n*} \geq g_h^n+g_l^n \Leftrightarrow \frac{z_h q\left(\theta_h\right)+z_l q\left(\theta_l\right)}{\left(q\left(\theta_l\right)^{-1}+q\left(\theta_h\right)^{-1}\right)^{-\frac{1}{1-\alpha}}} \geq 2\left(\left(z_l^{-1}q\left(\theta_l\right)^{-1}\right)^{z_l}\left(z_h^{-1}q\left(\theta_h\right)^{-1}\right)^{z_h}\right)^{\frac{\alpha}{1-\alpha}}. \quad (40)$$

Using the means inequality again, we know that

$$\left(z_l^{-1}q\left(\theta_l\right)^{-1}\right)^{z_l}\left(z_h^{-1}q\left(\theta_h\right)^{-1}\right)^{z_h} \leq q\left(\theta_l\right)^{-1}+q\left(\theta_h\right)^{-1}.$$

So a sufficient condition for (40) to be true is

$$\left(z_h q\left(\theta_h\right)+z_l q\left(\theta_l\right)\right)\left(\frac{1}{q(\theta_l)}+\frac{1}{q(\theta_h)}\right)^{\frac{1}{1-\alpha}} \geq 2\left(q\left(\theta_l\right)^{-1}+q\left(\theta_h\right)^{-1}\right)^{\frac{\alpha}{1-\alpha}}$$

which after simplification becomes

$$\left(z_h q\left(\theta_h\right)+z_l q\left(\theta_l\right)\right)\left(\frac{1}{q(\theta_l)}+\frac{1}{q(\theta_h)}\right) \geq 2.$$

Multiplying the parenthesis on the left-hand side and simplifying this sufficient condition becomes

$$z_h\frac{q\left(\theta_h\right)}{q\left(\theta_l\right)}+z_l\frac{q\left(\theta_l\right)}{q\left(\theta_h\right)} \geq 1.$$

For the lowest possible value of $q(\theta_l)$, $q(\theta_l) = z_h q(\theta_h) z_l^{-1}$, this inequality is satisfied. As $q(\theta_l)$ increases from there, so does the left-hand side of the inequality, so it is satisfied for any $q(\theta_l)$ that satisfies our assumptions. This completes the proof that total employment is smaller when the voting constraint binds. □

Proposition 6. *The union threat increases the average wage and decreases wage inequality, defined as the ratio of high-skill to low-skill wages, in nonunion firms.*

Proof. We use (30), (31) and (35) to find that average wage in the unconstrained firm is

$$\begin{aligned}\frac{g_l^{n*} w_l^n + g_h^{n*} w_h^n}{g_l^{n*} + g_h^{n*}} &= \frac{\frac{\beta_n}{1-(1-\alpha)\beta_n}\alpha F\left(\boldsymbol{g}^{n*}\right)}{\frac{\alpha B_n}{\kappa} F\left(\boldsymbol{g}^{n*}\right) z_l q\left(\theta_l\right) + \frac{\alpha B_n}{\kappa} F\left(\boldsymbol{g}^{n*}\right) z_h q\left(\theta_h\right)} \\ &= \frac{\beta_n}{1-\beta_n}\frac{\kappa}{z_l q\left(\theta_l\right) + z_h q\left(\theta_h\right)}\end{aligned}$$

Similarly, for the constrained firm, we find

$$\begin{aligned}\frac{g_l^{n} w_l^n + g_h^{n} w_h^n}{g_l^{n} + g_h^{n}} &= \frac{g^n \frac{\beta_n}{1-(1-\alpha)\beta_n}\alpha F\left(\boldsymbol{g}\right)\frac{z_l}{g^n} + g^n \frac{\beta_n}{1-(1-\alpha)\beta_n}\alpha F\left(\boldsymbol{g^n}\right)\frac{z_h}{g^n}}{2g^n} \\ &= \frac{\beta_n}{1-\beta_n}\frac{1}{2}\left(\frac{\kappa}{q\left(\theta_l\right)} + \frac{\kappa}{q\left(\theta_h\right)}\right).\end{aligned}$$

We therefore need to show that

$$\frac{\beta_n}{1-\beta_n}\frac{1}{2}\left(\frac{\kappa}{q\left(\theta_l\right)} + \frac{\kappa}{q\left(\theta_h\right)}\right) \geq \frac{\beta_n}{1-\beta_n}\frac{\kappa}{z_l q\left(\theta_l\right) + z_h q\left(\theta_h\right)}$$

which, after simplifying, becomes

$$z_l \frac{q\left(\theta_l\right)}{q\left(\theta_h\right)} + z_h \frac{q\left(\theta_h\right)}{q\left(\theta_l\right)} \geq 1.$$

As explained in the previous proof, this inequality is true under our assumptions so that the average wage is higher in the constrained firm.

For the ratio of wages, simplify using the equations for the wages and the optimal employment we find that

$$\frac{w_h\left(\boldsymbol{g^{n*}}\right)}{w_l\left(\boldsymbol{g^{n*}}\right)} = \frac{q(\theta_l)}{q(\theta_h)} \text{ and } \frac{w_h\left(\boldsymbol{g^{n}}\right)}{w_l\left(\boldsymbol{g^{n}}\right)} = \frac{z_h}{z_l}$$

so that $w_h\left(\boldsymbol{g^{n*}}\right)/w_l\left(\boldsymbol{g^{n*}}\right) > w_h\left(\boldsymbol{g^{n}}\right)/w_l\left(\boldsymbol{g^{n}}\right)$ is true under our assumptions. □

Proposition 7. $\Delta\Pi$ *is maximized when worker heterogeneity is minimal, in the sense that* $z_l q\left(\theta_l\right) = z_h q\left(\theta_h\right)$.

Proof. We know from (34) that

$$\Pi\left(\boldsymbol{g^{u*}}\right) = \left(1-\alpha\right) B_u \left(\frac{\alpha B_u}{\kappa}\right)^{\frac{\alpha}{1-\alpha}} \left(\left(z_l q\left(\theta_l\right)\right)^{z_l}\left(z_h q\left(\theta_h\right)\right)^{z_h}\right)^{\frac{\alpha}{1-\alpha}}.$$

Similarly, for the constrained firms we use (38) and (39) to find that

$$\begin{aligned}\Pi\left(\boldsymbol{g^n}\right) &= F\left(\boldsymbol{g^n}\right) - \sum_{s\in\mathcal{S}} g_s w_s^n\left(\boldsymbol{g^n}\right) - \kappa \sum_{s\in\mathcal{S}} \frac{g_s^n}{q\left(\theta_s\right)} \\ &= \left(1-\alpha\right)\alpha^{\frac{\alpha}{1-\alpha}} B_n^{\frac{1}{1-\alpha}} \left(\frac{\kappa}{q\left(\theta_l\right)} + \frac{\kappa}{q\left(\theta_h\right)}\right)^{-\frac{\alpha}{1-\alpha}}.\end{aligned}$$

After simplification, we can write $\Delta\Pi$ as the product of two terms:

$$\Delta\Pi = \frac{\Pi\left(\boldsymbol{g^n}\right)}{\Pi\left(\boldsymbol{g^{u*}}\right)} = \left(\frac{B_n}{B_u}\right)^{\frac{1}{1-\alpha}} \left(\underbrace{\frac{\left(\left(z_l q\left(\theta_l\right)\right)^{-1}\right)^{z_l}\left(\left(z_h q\left(\theta_h\right)\right)^{-1}\right)^{z_h}}{q\left(\theta_l\right)^{-1} + q\left(\theta_h\right)^{-1}}}_{\Gamma}\right)^{\frac{\alpha}{1-\alpha}}. \tag{41}$$

The term Γ in the equation is the ratio of the geometric and the arithmetic means of $\left(z_l q\left(\theta_l\right)\right)^{-1}$ and $\left(z_h q\left(\theta_h\right)\right)^{-1}$, with weights z_l and z_h. By the inequality of arithmetic and geometric means, that ratio is weakly smaller than 1 and it reaches it maximum when $z_l q\left(\theta_l\right) = z_h q\left(\theta_h\right)$. □

Proposition 8. *There is a threshold $\bar{\alpha} \in [0,1]$ such that $\Delta\Pi > 1$ for $\alpha < \bar{\alpha}$ and $\Delta\Pi < 1$ for $\alpha > \bar{\alpha}$. In addition, there is a threshold $\hat{\alpha} \in [0,1]$ such that the firm cannot prevent unionization if $\alpha < \hat{\alpha}$.*

Proof. The expression for $\Delta\Pi$ is given by (41). As shown in the proof of Proposition 7 we have that $\Gamma \leq 1$. We will first show that $\Delta\Pi$ is either decreasing in α or that it lies above 1. Taking the derivative of the log we find,

$$\frac{d}{d\alpha}\log\Delta\Pi = \frac{1}{\left(1-\alpha\right)^2}\log\left(\frac{B_n}{B_u}\right) - \frac{1}{1-\alpha}\frac{\beta_n}{1-\left(1-\alpha\right)\beta_n} + \frac{1}{\left(1-\alpha\right)^2}\log\Gamma.$$

Note that this derivative is strictly negative as long as B_u is large enough (recall that $\Gamma < 1$). The smallest B_u such that the derivative is negative, defined as $\hat{B}_u$, is such that

$$\log\left(\frac{B_n}{\hat{B}_u}\right) = \frac{\left(1-\alpha\right)\beta_n}{1-\left(1-\alpha\right)\beta_n} - \log\Gamma$$

Notice that $\log\Delta\Pi$ is decreasing in B_u so if $\log\Delta\Pi > 0$ for $B_u = \hat{B}_u$ then $\log\Delta\Pi > 0$ for $B_u < \hat{B}_u$. Compute $\log\Delta\Pi\left(\hat{B}_u\right)$

$$\begin{aligned}\log\Delta\Pi &= \frac{1}{1-\alpha}\log\left(\frac{B_n}{\hat{B}_u}\right) + \frac{\alpha}{1-\alpha}\log\Gamma \\ &= \frac{\beta_n}{1-\left(1-\alpha\right)\beta_n} - \log\Gamma\end{aligned}$$

which is strictly positive. Since $\log\Delta\Pi > 0$ for $B_u < \hat{B}_u$ we have shown that, depending on B_u,

$\Delta\Pi$ is strictly decreasing in α or that it lies above 1. As a result, there is a threshold $\bar{\alpha} \in [0,1]$ such that $\Delta\Pi > 1$ for $\alpha < \bar{\alpha}$ and $\Delta\Pi < 1$ for $\alpha > \bar{\alpha}$. If $B_u < \hat{B}_u$ that threshold is $\bar{\alpha} = 1$.

We now turn to establishing the threshold $\hat{\alpha}$. Comparing the union and nonunion wages of workers of skill s we find

$$w_s^u \left(\boldsymbol{g^n}\right) - w_s^n \left(\boldsymbol{g^n}\right) = \left(\frac{1}{2}\beta_u - \frac{\alpha\beta_n}{1-(1-\alpha)\beta_n} z_s\right) (g^n)^{\alpha-1}. \tag{42}$$

For high-skill workers to vote against the union, the inside of the parenthesis must be nonpositive for $s = h$. This term as the same sign as

$$B_n - B_u + \frac{\alpha\beta_n (1-2z_s)}{1-(1-\alpha)\beta_n}$$

Since $z_l < 0.5$ and $B_n > B_u$, low-skill workers always vote in favor of the union. High-skill workers, on the other hand, can vote for or against the union, depending on the parameters. Under our assumptions, the parenthesis in (42) is strictly decreasing in α so that there exits a threshold $\hat{\alpha} \in [0,1]$ such that the parenthesis is negative for $\alpha > \hat{\alpha}$, which is the result. □

Proposition 9. *There are thresholds $\bar{\beta}_u \in [0,1]$ and $\bar{\beta}_n \in [0,1]$ such that the firm cannot prevent unionization if $\beta_u > \bar{\beta}_u$ or if $\beta_n < \bar{\beta}_n$.*

Proof. From equation (42), we see that the the union wage of high-skill worker exceeds their nonunion wage if β_u is large enough or if β_n is small enough (recall that $z_h > 0.5$). As a result, we have the thresholds $\bar{\beta}_u \in [0,1]$ and $\bar{\beta}_n \in [0,1]$ such that the firm cannot prevent its unionization if $\beta_u > \bar{\beta}_u$ or if $\beta_n < \bar{\beta}_n$. □

F.2 Results related to the quantitative exercises

Lemma A1. *Consider two firms, identified by the subscripts 1 and 2, that have identical technologies except for $A_1 \neq A_2$. In equilibrium, if g_1 solves the problem of firm 1, then $g_2 = (A_2/A_1)^{\frac{1}{1-\alpha}} g_1$ solves the problem of firm 2. Also, both firms have the same union status and pay the same wages.*

Proof. Assume first that the equilibrium schedules $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ are identical and denote that schedule by $\boldsymbol{c}$. This result will be shown later in the lemma. We can write the problem of firm $j \in \{1,2\}$ as

$$\max_g \Pi\left(A_j, \boldsymbol{w}\left(A_j, \boldsymbol{g}\right), \boldsymbol{g}\right)$$

such that

$$w\left(A_j, \boldsymbol{g}\right) = \begin{cases} w^u\left(A_j, \boldsymbol{g}\right) & \text{if } \Lambda\left(A_j, \boldsymbol{g}\right) > 0 \\ w^n\left(A_j, \boldsymbol{g}\right) & \text{if } \Lambda\left(A_j, \boldsymbol{g}\right) \leq 0 \end{cases}$$

where $\boldsymbol{w^u}$ is the union wage function, $\boldsymbol{w^n}$ is the nonunion wage function and V is the excess number of workers for unionization.

The proof proceeds by showing that if g_1 solves the FOCs of firm 1 then

$$\boldsymbol{g_2} = \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \boldsymbol{g_1}$$

solves the FOCs of firm 2.

We therefore start with the FOC of firm 1 given by equations (20) and (21). First, notice that

$$\begin{aligned}\frac{A_1}{(g_{1,s})^{1/\sigma}}\left(\sum_{k\in\mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}} &= \frac{A_1\left(\frac{A_1}{A_2}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma(1-\alpha)}}}{\left(\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\alpha}} g_2\right)^{1/\sigma}}\left(\sum_{k\in\mathcal{S}} z_k g_{2,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}} \\ &= \frac{A_2}{(g_{2,s})^{1/\sigma}}\left(\sum_{k\in\mathcal{S}} z_k g_{2,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}} .\end{aligned}$$

This also implies that $\boldsymbol{w^n}(A_1, \boldsymbol{g_1}) = \boldsymbol{w^n}(A_2, \boldsymbol{g_2})$. It is also straightforward to show that $F(A_1, \boldsymbol{g_1})/n_1 = F(A_2, \boldsymbol{g_2})/n_2$ such that $w^u(A_1, g_1) = w^u(A_2, g_2)$. We have so far shown that the terms not multiplied by the Lagrange multiplier in equation (20) are the same, which completes the proof if firm 1 is unconstrained ($\lambda^1 = 0$). In which case, firm 2 is also unconstrained.

We now consider the derivatives in equation (21). Notice that, for any $s' \neq s$, we have

$$\begin{aligned} g_{1,s'}\frac{\partial w^n_{s'}(A_1, \boldsymbol{g_1})}{\partial g_{1,s}} &= \frac{\alpha\beta_n(1-\sigma(1-\alpha))}{(1-\beta_n(1-\alpha))\sigma} A_1 \left(\sum_{k\in\mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\alpha\sigma-2\sigma+2}{\sigma-1}} z_s g_{1,s}^{-1/\sigma} z_{s'} g_{1,s'}^{\frac{\sigma-1}{\sigma}} \\ = \frac{\alpha\beta_n(1-\sigma(1-\alpha))}{(1-\beta_n(1-\alpha))\sigma} A_2 &\left(\sum_{k\in\mathcal{S}} z_k g_{2,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\alpha\sigma-2\sigma+2}{\sigma-1}} z_s g_{2,s}^{-1/\sigma} z_{s'} g_{2,s'}^{\frac{\sigma-1}{\sigma}} = g_{2,s'}\frac{\partial w^n_{s'}(A_2, \boldsymbol{g_2})}{\partial g_{2,s}}. \end{aligned}$$

Similarly,

$$\begin{aligned} g_{1,s}\frac{\partial w^n_s(A_1, \boldsymbol{g_1})}{\partial g_{1,s}} &= \frac{\alpha\beta_n z_s A_1}{1-\beta_n(1-\alpha)}\left(-\frac{1}{\sigma}\left(\sum_{k\in\mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\alpha\sigma-\sigma+1}{\sigma-1}} g_{1,s}^{-\frac{1}{\sigma}}\right. \\ &\left. +g_{1,s}^{\frac{\sigma-2}{\sigma}}\frac{1-\sigma(1-\alpha)}{\sigma}\left(\sum_{k\in\mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\alpha\sigma-2\sigma+2}{\sigma-1}} z_s\right) \\ &= g_{2,s}\frac{\partial w^n_s(A_2, \boldsymbol{g_2})}{\partial g_{2,s}}. \end{aligned}$$

Similar computations yield that for any $s' \in \{1, \ldots, S\}$

$$g_{1,s'} \frac{\partial w^u_{s'}(A_1, \boldsymbol{g_1})}{\partial g_{1,s}} = g_{2,s'} \frac{\partial w^u_{s'}(A_2, \boldsymbol{g_2})}{\partial g_{2,s}}.$$

Combining these results, it follows that $V(A_1, g_1) = V(A_2, g_2)$ and that

$$\frac{\partial \Lambda(A_1, \boldsymbol{g_1})}{\partial g_{1,s}} = \frac{\partial \Lambda(A_2, \boldsymbol{g_2})}{\partial g_{2,s}}$$

for all s. This completes the proof since, if firm 1 is constrained, there exists a $\lambda^2 = \lambda^1 \geq 0$ such that g_2 solves the problem of firm 2 and $\Lambda(A_2, \boldsymbol{g_2}) = 0$. Notice that firm 2 is also constrained. Notice also that since the two firms have the same union status and are paying the same wages, we find $c_1 = c_2$, which justifies our initial assumption. □

Lemma A2. *If ϕ can be written as $\phi(x) = ax + 0.5$ with $a > 0$ then the optimal decision of a firm is independent of a.*

Proof. The problem of the firm is

$$F(\boldsymbol{g}) - \sum_{s \in \mathcal{S}} g_s w_s(\boldsymbol{g}) - \kappa(1 - (1 - \delta)\gamma) \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)}$$

subject to

$$\sum_{s \in \mathcal{S}} g_s \phi(w^u_s(\boldsymbol{g}) - w^n_s(\boldsymbol{g})) - \frac{1}{2} n \leq 0.$$

The constraint can be written as

$$a \sum_{s \in \mathcal{S}} g_s (w^u_s(\boldsymbol{g}) - w^n_s(\boldsymbol{g})) \leq 0.$$

Dividing the constraint by a, notice that a does not show up in the optimization problem and has therefore no impact on the firm's decision. □