

The Union Threat*

Mathieu Taschereau-Dumouchel[†]

Cornell University

October 17, 2017

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. As unionized workers bargain collectively with the firm, unionization compresses the wage distribution and lowers profits. To prevent unionization, nonunion firms distort the skill composition of their workforce by over-hiring high-skill workers, who vote against the union, and under-hiring low-skill workers, who vote in its favor. Because of decreasing returns to labor, this change in hiring lowers output while reducing the range of wages paid. In the calibrated economy, removing the threat of unionization, by freezing the union status of firms, reduces unemployment and increases output and the variance of wages. Removing, in addition, all unions from the economy leads to a larger increase in wage inequality but does not further affect output and unemployment. These results suggest that the threat that unionization exerts on nonunion firms, more than the fact that some firms are actually unionized, is the main channel through which unions affect output and unemployment in the U.S. economy.

JEL Classifications: J51, E24

Abstract

*I thank the editor Michèle Tertilt and three anonymous referees for their valuable suggestions. I am grateful to Nobuhiro Kiyotaki, Esteban Rossi-Hansberg and Oleg Itskhoki for their advice and support. I also thank Andy Abel, Zvi Eckstein, Lukasz Drozd, Pablo Fajgelbaum, Henry Farber, João Gomes, Lawrence Kahn, Morris Kleiner, Thomas Lemieux, 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 comments.

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

1 Introduction

As unions are now covering only 7% of private sector jobs in the United States, many observers have argued that their impact on the aggregate economy must be small.¹ In opposition to this view, this paper investigates how unions can nonetheless have a sizable impact on the macroeconomy through the influence they have on *nonunion* firms. Indeed, if unionization lowers profits, vulnerable nonunion firms would distort their behavior to prevent their own unionization. Through this channel, unions may influence employment, wages and output in nonunion firms and, as most firms are union free, in the aggregate economy.

To analyze the impact of this threat of unionization, 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 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 all of 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 the scope of the wage bargaining, unionization generates two conflicts within the firm. First, as collective bargaining compresses the distribution of wages, high-productivity workers vote against the creation of the union while low-productivity workers vote in its favor. Unionization therefore creates a conflict between workers. Second, as collective bargaining allows the workers to extract a higher share of the production surplus, unionization increases the average wage and lowers profits, thereby creating a second conflict, this time between the firm and its workforce. The union threat affects the decisions of the firm through the interaction of these two conflicts: to prevent profits-reducing unionization, the firm hires more high-skill workers and fewer low-skill workers, thereby adjusting the outcome of the vote in its favor.²

In the theory, the distortion created by the union threat interacts with the decreasing returns in production to push the firm towards lower output and employment. As a result, the average marginal product of the workers increases which leads to a higher average wage. The union threat also affects wage inequality. As the firm over-hires high-skill workers, their marginal product declines which reduces their wage. Since the opposite happens to low-skill workers, nonunion firms pay a narrower range of wages in response to the threat of unionization.

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

¹See [Hirsch and Macpherson \(2003\)](http://www.unionstats.com/) and their database at <http://www.unionstats.com/>.

²Empirical studies have found that, consistent with the model, low-skill workers favor unions while high-skill workers do not ([Farber and Saks, 1980](#)). It is therefore natural that a firm that wants to avoid unionization modifies its employment decisions to increase its odds of remaining union free.

By providing a microfounded bargaining theory of unionization, the model 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 use the parametrized model to conduct three policy experiments in general equilibrium. The first experiment consists of removing the threat of unionization: the union status of all firms is fixed and cannot be changed anymore. As a result, nonunion firms stop distorting their behavior to prevent unionization. This first policy experiment captures the impact of the threat of unionization alone, as the union status of all firms remain unchanged. In the new equilibrium, the variance of log wages goes up by 0.6% and output increases by 1.0%, while the unemployment rate decreases by 1.2 percentage points. If, in addition to removing the threat of unionization, all union firms are forced to become union free, the variance of log wages goes up by an additional 4.9%, but output and unemployment are not further affected. This second policy experiment therefore suggests that the threat of unionization alone, more than the fact that some firms are actually unionized, is the main channel through which unions affect output and unemployment in the U.S. economy. Finally, in the third policy experiment, all firms are forced to be unionized. Comparing this new equilibrium to the calibrated economy, the variance of log wages goes down by 27% while output and employment increase as much as in the no-union experiment. These policy experiments confirm that the union threat has a substantial effect on output and unemployment while the impact of the union status of firms is mostly on wage inequality.

The paper also shows that often-used empirical estimators underestimate the impact of labor unions on wage inequality. For instance, the classical estimator of Freeman (1980) finds that, in the calibrated economy, unions reduce the variance of log wages by 3.47% while their true impact, as measured using the policy experiments, is of 5.5%. More sophisticated estimators that take into account the heterogeneity between workers do worse by suggesting that unions lower the variance of wages by only 14% of their true impact. These large differences between the empirical 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 empirical estimators do not capture this channel.

The theory also provides an explicit mechanism to explain why regression discontinuity studies, such as DiNardo and Lee (2004), find little impact of unionization on firms. These studies compare firms before and after unionization. According to the theory, before unionization, these firms are actively distorting their behavior in response to the threat of unionization. As a result, regression discontinuity estimators only capture part of the impact of unions on firms.³

³DiNardo and Lee (2004) discuss how the union threat may contribute to their results. Frandsen (2011) uses a regression discontinuity approach to estimate the impact of unionization on the full wage distribution.

1.1 Using changes in right-to-work laws to evaluate the impact of the union threat

The passage of right-to-work (RTW) legislations by U.S. states can be used to estimate the impact of the union threat on firms. As [Farber \(2005\)](#) mentions, these laws are thought to weaken the union threat:

A credible source of variation in the threat of union organization is the adoption of RTW laws. Union-shop and agency-shop agreements reached in collective bargaining between unions and employers make it a requirement of continued employment that workers either become dues-paying members of the union or pay a continuing fee in lieu of membership dues. These contract provisions are important facilitators of a stable union presence. RTW laws make it illegal for labor unions and employers to negotiate such agreements while requiring unions to represent, negotiate on behalf of, and provide services to even those who choose not to join the union or provide financial support. Effectively, unions are prevented from taxing workers to pay for benefits, yet they are required to provide workplace public goods. Not surprisingly, there is a substantially larger free-rider problem in states with RTW laws. For this reason, it is likely that the threat of unionization is lower after the passage of a RTW law.

I propose new estimates of the impact of these laws on earnings and employment.⁴ The data comes from the Merged Outgoing Rotation Groups of the Current Population Survey. The sample covers the period from January 1989 to July 2017 and includes the passage of right-to-work laws in Indiana (2012), Kentucky (2017), Michigan (2012), Missouri (2017), Oklahoma (2001), Texas (1993), West Virginia (2016) and Wisconsin (2015).⁵

Table 1 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 that has enacted a right-to-work law in the current or in a previous year, and zero otherwise. We see that the passage of these laws is associated with a significant decline in the earnings of union and nonunion workers. The impact on union workers is straightforward to understand. Since workers in a bargaining unit do not have to be member of the union anymore — and therefore no longer have to pay dues — the bargaining position of the union is weakened which leads to lower wages. Similarly, since unions are now weaker after the passage of the law, nonunion firms are less worried about unionization and they do not have to keep wages high to influence a union vote. Controlling for individual fixed effects (column 5), the passage of a right-to-work law is associated with a 1% decline in nonunion

⁴The exercise complements the previous literature in two ways (see next section for a summary of the literature). First, since several right-to-work laws have been passed in recent years, this 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.

⁵I use the version of the CPS provided by IPUMS ([Flood et al., 2015](#)). January 1989 is the first month in which there are consistent weekly earnings data.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Earnings	Earnings	Earnings	Earnings	Earnings	Earnings
Workers in the sample	All	Nonunion	Union	All	Nonunion	Union
Right-to-work law	-0.012*** (0.003)	-0.006** (0.003)	-0.026*** (0.008)	-0.009*** (0.002)	-0.010*** (0.002)	-0.010* (0.006)
Number of observations	3,921,320	3,547,728	373,592	3,921,320	3,547,728	373,592
State & year fixed effects	yes	yes	yes	yes	yes	yes
Individual fixed effects				yes	yes	yes

Notes: The dependent variable is the log of weekly earnings for all workers (columns 1 and 4), nonunion workers (columns 2 and 5) or union workers (columns 3 and 6). Standard errors are in parenthesis. Individual fixed effects are industry (3-digit), occupation (3-digit), age, sex, education (32 categories), full/part-time status and union coverage. The data covers the adult civilian population in the Current Population Survey Merged Outgoing Rotation Groups between January 1989 and July 2017. I remove from the sample government employees and individuals older than 65. Union workers includes union members as well as workers covered by a union. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1: Impact of right-to-work laws on weekly earnings

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Earnings	Earnings	Earnings	Earnings	Earnings	Earnings
Workers in the sample	All	Nonunion	Union	All	Nonunion	Union
Right-to-work law	-0.025*** (0.004)	-0.019*** (0.004)	-0.014 (0.011)	-0.02*** (0.003)	-0.0108*** (0.003)	-0.016** (0.008)
Number of observations	2,421,264	2,224,105	197,159	2,421,264	2,224,105	197,159
Years in sample	≥ 2000	≥ 2000	≥ 2000	≥ 2000	≥ 2000	≥ 2000
State & year fixed effects	yes	yes	yes	yes	yes	yes
Individual fixed effects				yes	yes	yes

Notes: The dependent variable is the log of weekly earnings for all workers (columns 1 and 4), nonunion workers (columns 2 and 5) or union workers (columns 3 and 6). Standard errors are in parenthesis. Individual fixed effects are industry (3-digit), occupation (3-digit), age, sex, education (32 categories), full/part-time status and union coverage. The data covers the adult civilian population in the Current Population Survey Merged Outgoing Rotation Groups between January 1989 and July 2017. I remove from the sample government employees and individuals older than 65. Union workers includes union members as well as workers covered by a union. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Impact of right-to-work laws on weekly earnings in recent years

earnings, suggesting that the union threat has a substantial impact on compensation in nonunion firms.

Perhaps surprisingly, the impact of the threat on nonunion earnings has been stronger in recent years. Table 2 provides the same ordinary least-square estimates as Table 1 but limits the sample to the years after 2000. All the estimations coefficients with individual fixed effects (columns 4 to 6) find a larger impact of right-to-work laws on earnings since 2000, suggesting that the threat remains an important driver of compensation in recent years.

One worry about the regressions of Tables 1 and 2 is that the passage of right-to-work legislations is not exogenous. In particular, law-makers might pass these laws in bad economic times — during which wages are already lower — in an attempt to sustain the economy. This narrative is however undermined by the regressions presented in Table 3 which show that the passage of a right-to-work law is associated with a significant increase in the probability that an individual is employed.

Dependent variable	(1) Prob. of Empl.	(2) Prob. of Empl.	(3) Prob. of Empl.	(4) Prob. of Empl.
Right-to-work law	0.002* (0.0012)	0.006*** (0.0007)	0.006*** (0.0019)	0.004*** (0.0009)
Number of observations	6,658,874	6,658,874	4,104,993	4,104,993
Years in sample	all	all	≥ 2000	≥ 2000
State and year fixed effects	yes	yes	yes	yes
Individual fixed effects		yes		yes

Notes: The dependent variable equals 1 if the individual is employed and 0 otherwise. Standard errors are in parenthesis. Individual fixed effects are industry (3-digit), occupation (3-digit), age, sex, education (32 categories) and union coverage. The data covers the adult civilian population in the Current Population Survey Merged Outgoing Rotation Groups between January 1989 and July 2017. I remove from the sample government employees and individuals older than 65. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Impact of right-to-work laws on the probability of being employed

The estimated coefficient in the regression with individual fixed effects (column 2) amounts to an increase in the probability of being employed of about 1%. Columns 3 and 4 show that the impact of right-to-work laws on employment has remained strong in recent years.

Since right-to-work laws only weaken the threat of unionization, instead of completely removing it, the empirical findings of Tables 1 to 3, as well as those already in the literature, only provide lower bounds on the full impact of the threat on the economy. Nonetheless, they suggest that the threat has a significant impact by increasing wages and lowering employment, and that this impact has remained strong even under the lower unionization rates of the recent years. The model introduced in this paper is consistent with these findings.⁶

1.2 Literature review

Several papers use variations in right-to-work laws across U.S. states and over time to evaluate the importance of the union threat. Holmes (1998) shows that firms prefer to locate their establishments in right-to-work states. Farber (2005) finds that nonunion wages fell by 4.2% after the passage of a right-to-work law in Idaho in 1981.⁷ More recent work has shown that the threat remains a substantial force 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

⁶The theoretical results of Section 3 and the quantitative results of Section 4 both find that the removal of the threat increases employment in partial and in general equilibrium. Wages are predicted to go down in partial equilibrium but can go up or down in general equilibrium depending on how elastic the labor market is. Since products and labor markets are highly integrated across states in the United States (Esipova et al., 2013) the exercises of Tables 1 to 3 are mostly indicative of the forces at work in partial equilibrium.

⁷Farber (2005) finds an insignificant impact of the Oklahoma right-to-work law in 2001 but mentions that this negative result might be attributable to data availability issues. His sample only includes one year of data after the passage of the law while Lee and Mas (2012) find that the impact of unionization on a firm can take between 15 and 18 months to fully materialize.

passed.^{8,9}

A literature also documents the negative impact of unionization on firm profitability. In a recent paper, [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. Interestingly, [Lee and Mas \(2012\)](#) find that this cumulative 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.¹⁰

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](#); [Bronfenbrenner, 1994](#); [Freeman and Kleiner, 1990](#)). 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 unionization ([Vieira, 2014](#)).¹¹ Recently, several private universities have improved graduate students salary and benefits substantially in anticipation of a decision by the National Labor Relations Board to allow them to unionized ([Elejalde-Ruiz, 2016](#); [Flaherty, 2016](#)).

Finally, survey evidence finds that, in 2004, workers wanted union representation more then ever before ([Freeman, 2007](#)) and that firms are increasingly attempting to prevent unionization, for instance there is a resurgence of pro-union workers firing during election campaigns ([Schmitt and Zipperer, 2007](#)).

The literature reviewed so far leads us to three conclusions. First, measures of the union threat

⁸Private sector union coverage for these states was 8.6% in 2015. 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. In his doctoral dissertation, [Rowe \(2016\)](#) summarizes the literature on the impact of right-to-work laws along with his own estimations and also concludes that these laws are associated with a decline in wages.

⁹Other studies have used union densities as measures of the importance of the union threat. [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 the dispersion of wages, [Kahn and Curme \(1987\)](#) find a lower nonunion wage dispersion in more heavily unionized industries. [Foulkes \(1980\)](#) documents from survey that, like in the model, large nonunion firms increase wages and working conditions preemptively to incentivize workers to reject unionization.

¹⁰Note that [Lee and Mas \(2012\)](#) probably underestimate the full impact of unions since firms are likely to increase wages, and therefore to sacrifice profits, before a union vote to try to prevent unionization. See Section 4.2 a discussion of regression discontinuity estimators in the context of the model.

¹¹Part of a handbook that Wal-Mart distributes to managers contains the following passage [Featherstone \(2004\)](#):

Staying union free is a full-time commitment. [...] The commitment to stay union free must exist at all levels of management—from the Chairperson of the “Board” down to the front-line manager. Therefore, no one in management is immune to carrying his or her “own weight” in the union prevention effort. The entire management staff should fully comprehend and appreciate exactly what is expected of their individual efforts to meet the union free objective... Unless each member of management is willing to spend the necessary time, effort, energy, and money, it will not be accomplished. The time involved is... 365 days per year...

are positively correlated with higher nonunion wages, suggesting that these firms pay their workers more to incentivize them to reject a potential union. Second, unionization is associated with a significant decline in the value of the firm, providing strong incentives for firms to take actions to prevent unionization. Third, the impact of the threat on wages, and of unionization on firm value, has remained strong even in recent years which suggest that the union threat still has a sizable impact on the economy.

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, this paper considers a dynamic, general equilibrium framework with heterogeneous workers to evaluate the impact of the union threat on wage inequality, output and unemployment. While the union threat often leads to wage increases in Dickens (1986), the opposite is generally true in the current paper. While the threat also leads to higher wages at the firm level, general equilibrium forces can undo this increase. Indeed, since threatened firms hire less, unemployment increases which lowers the bargaining position of the workers and leads to a decline in overall wages. 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. They 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. Açıkgöz and Kaymak (2014) estimate the impact of a rising skill premium on the decline of union membership in the United States. Boeri and Burda (2009) look into the impact of an endogenous bargaining regime on economic activity. Recently, Krusell and Rudanko (2012) have studied the dynamic problem of a monopoly union that sets wages with or without commitment. In contrast to the literature, this paper investigates the impact of the threat of unionization on decision makers and the macroeconomy.

The next section introduces the model. An explanation of how firms respond to the distortion created by the union threat follows. The model is then calibrated to the U.S. economy and policy experiments are conducted to evaluate the impact of unions. The last section concludes.

2 Model

2.1 Preferences and technology

The economy is populated by heterogeneous workers, each endowed with a skill $s \in \mathcal{S} = \{1, \dots, S\}$, constant over time. The exogenous measure of skills in the economy is given by a vector $\mathbf{N} \in \mathbb{R}^S$, with $N_s > 0$ for all s . Workers live forever, are risk-neutral and discount future consumption at the rate $0 < \gamma < 1$. Variables in bold are vectors.

Firms combine the labor provided by workers of different skills into consumption goods. To do so, each firm is endowed with one of J production functions, indexed by $j \in \mathcal{J} = \{1, \dots, J\}$. A firm of type j that employs a measure of workers \mathbf{g} produces goods according to the production function

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

where $A_j > 0$ is total factor productivity and $\sigma > 0$ is the elasticity of substitution between the different skills. The vector $z_{j,s} > 0$ represents 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 avoid cluttering the notation, the subscript j is often omitted when referring to a single firm.¹²

Decreasing returns to scale imply the presence of a fixed factor of production — for instance, managerial ability in the spirit of Lucas (1978) — whose returns go to the firm owner. Since these returns are only realized when a firm operates and that, as will become clear later, a union can prevent production from happening, the parameter α is an important determinant of the strength of unions in this economy.¹³ Overall, the technology of the firm determines its union status in equilibrium.

2.2 Labor markets

There are S labor markets 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, at a unit cost κ , in multiple markets. 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, influence the unionization vote.¹⁴

In a labor market where U unemployed workers are searching and V vacancies are posted, $m(U, V)$ random 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 $\theta = V/U$, the probability that a vacancy is filled is $q(\theta) = m(U, V)/V = m(1/\theta, 1)$ and the probability that an unemployed worker finds a job is $p(\theta) = m(U, V)/U = m(1, \theta)$. Since search requires no effort, all unemployed workers are searching. At the end of each period, a fraction δ of jobs are exogenously destroyed.

¹²Acemoglu et al. (2001), Açıkgöz 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.

¹³In the quantitative exercises of Section 4, firms also use capital to produce. This additional factor of production is straightforward to include in the current setup, as Appendix A.1 shows.

¹⁴The presence of search frictions generates a gap between union and nonunion wage. This union wage gap, which is well documented in the empirical literature. The assumption of segmented markets is not necessary for the main mechanism. As long as a firm has some control over the skill of the workers it hires the threat of unionization might influence its decision.

2.3 Firms

A firm that employed a measure of workers \mathbf{g}_{-1} during the previous period starts the current period with $(1 - \delta)\mathbf{g}_{-1}$ workers. It then posts a schedule of vacancies \mathbf{v} 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 number of successful matches is deterministic.

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

$$J(\mathbf{g}_{-1}) = \max_{\mathbf{v} \geq 0} \pi(\mathbf{g}) - \kappa \sum_{s \in \mathcal{S}} v_s + \gamma J(\mathbf{g}) \quad (2)$$

where \mathbf{v} denotes the vector of vacancies posted in each submarket, and subject to the law of motion for employment,

$$\mathbf{g} = \mathbf{g}_{-1} (1 - \delta) + \mathbf{v} q(\boldsymbol{\theta}),$$

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

At a steady state, we can simplify the problem of the firm substantially. At the beginning of a period, the firm has a fraction $1 - \delta$ of its optimal measure of workers and, because of the linear hiring costs, it immediately hires back to its optimal level. The constraint $\mathbf{v} \geq 0$ 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*

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

Proof. All the proofs are in the appendix. □

This equation states that a firm sets its employment \mathbf{g} to maximize its present-period profit (first term) net of some vacancy posting costs (second term), and taking into account that the $(1 - \delta)\mathbf{g}$ 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 earning a wage w is therefore

$$W_{j,s}^E(w) = w + \gamma [\delta W_s^U + (1 - \delta) W_{j,s}^E(w_{j,s})] \quad (4)$$

where W_s^U is the lifetime utility of being unemployed and $w_{j,s}$ is the 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 w only. Both parties take the 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 $\mathbb{E}(W_{j,s}^E)$, 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 increasing in s . The lifetime discounted utility of an unemployed worker is therefore

$$W_s^U = p(\theta_s) \mathbb{E}(W_{j,s}^E) + (1 - p(\theta_s)) [b_s + \gamma W_s^U]. \quad (5)$$

By combining the last two equations we can characterize the utility gain provided by employment at wage w :

$$W_{j,s}^E(w) - b_s - \gamma W_s^U = w - c_{j,s} \quad (6)$$

where

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

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 usually at the enterprise level (Traxler, 1994; Nickell and Layard, 1999).¹⁵ 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 integrates these features of the institutional environment. The sequence of events that occurs once a firm has hired its new workers is represented 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 $\mathbf{w}^u(\mathbf{g})$ and a profit function $\pi^u(\mathbf{g})$. Instead, if the union vote fails, wages are bargained *individually*. This generates the wage schedule $\mathbf{w}^n(\mathbf{g})$ and the profit function $\pi^n(\mathbf{g})$. 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

¹⁵A large fraction 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).

that is bargained over is different in both cases. 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.

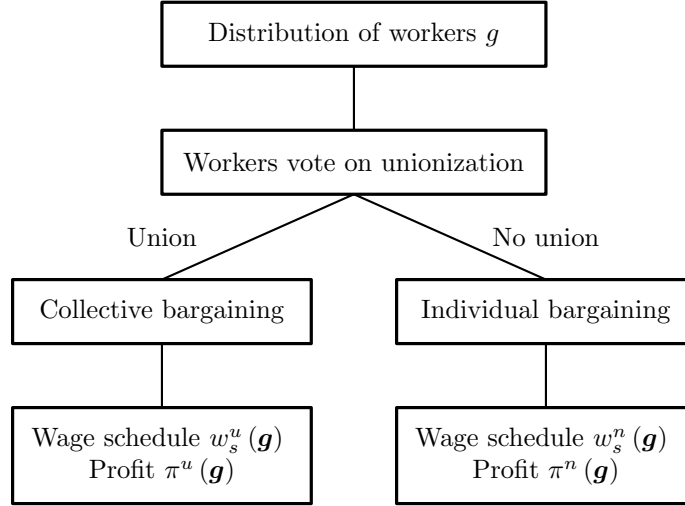


Figure 1: Timing of events in a firm after hiring

Collective bargaining

Collective bargaining is modeled as an n -player Nash bargaining between the firm and all its workers.¹⁶ If an agreement on a wage schedule \mathbf{w} is reached, a worker s receives $W_s^E(\mathbf{w})$, otherwise he or she receives home production b_s today and starts the next period as unemployed, which has value γW_s^U . The net benefit of an agreement to a worker is therefore $W_s^E(\mathbf{w}) - b_s - \gamma W_s^U$. On the firm side, if an agreement is reached production takes place and wages are paid. Otherwise, the firm loses 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*

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

¹⁶Nash bargaining with more than two players is microfounded in axiomatic bargaining theory (Roth, 1979) and in game theory (Krishna and Serrano, 1996).

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

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

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 with it next period if negotiations had been successful.

From (9), it is straightforward to compute the one-period profit of a union firm employing a measure \mathbf{g} of workers

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

Other work in the literature also uses Nash bargaining to model wage setting in a unionized firm (Açikgöz and Kaymak, 2014; Bauer and Lingens, 2010). In general, these papers assume that the union seeks to maximize the sum of the workers' surplus. The current approach has two advantages over this alternative protocol. First, with risk-neutral heterogeneous workers, the alternative approach only pins down the *total* share of the surplus going to the workers, not how it is shared among them. Second, the alternative protocol requires specifying preferences for the union itself, usually that it seeks to maximize the sum of the workers' surpluses. In contrast, in the current approach a union is simply the collective of the agents entering in an n -person Nash bargaining with the firm. Each agent uses his or her own preferences in this bargaining and there is no need to model a union as a middleman between the workers and the firm. As a result, here, the "preferences of the union" are microfounded from the preferences of the individual workers. Another advantage of the current approach is that it is robust in the sense that other bargaining environments yield the same outcome. For instance, Appendix A.3 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.

Individual bargaining

If the workers vote against the union, they each bargain individually with the firm. The surplus to split is, however, dependent on the outcome of the bargaining between the firm and all the *other* workers. Indeed, if one of the bargaining sessions breaks down and a worker leaves the firm, the marginal product of the remaining workers will change. As a result, these workers or the firm might

want to reopen the bargaining. Brügemann et al. (2015) and Stole and Zwiebel (1996a,b) provide the theoretical foundation for this type of bargaining.¹⁷

In this context, the firm's marginal gain from employing an extra worker of type s is¹⁸

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

The first term is the extra output produced by the worker. The next one is simply the wage paid to the worker. The third term is the marginal effect of this worker on the wages of 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 must solve the system of partial differential equations

$$\Delta_s^n(\mathbf{w}) = \frac{1 - \beta_n}{\beta_n} (W_s^E(\mathbf{w}) - b_s - \gamma W_s^U), \quad (11)$$

for all s with the standard boundary conditions $\{\lim_{g_s \rightarrow 0} w_s^n(\mathbf{g}) g_s = 0\}_{s=1}^S$. The solution to this system is characterized in the following lemma.

Lemma 3. *The wage schedule*

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

solves the individual bargaining problem.

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

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

Comparing collective and individual bargaining

The wage equations (9) and (12), from the collective and the individual bargaining problems, have remarkably similar structures. They both consist of three terms: one related to production, one related to the outside option of the workers and one related to the hiring costs. They, however, differ in how these quantities influence wages. Indeed, the union wage is mostly a function of the *average* characteristics of the firm's workforce, while the nonunion wage is a function of the *individual* characteristics of each worker. In particular, the union wage depends on the average product

¹⁷Bertola and Garibaldi (2001) show that a standard search model with wages set through 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 Cahuc and Wasmer (2001), Elsby and Michaels (2013) and Acemoglu and Hawkins (2014) for other examples using Stole and Zwiebel bargaining in search models. Appendix A.2 shows that the key mechanisms are preserved in an alternative model in which firms can pick nonunion wages unilaterally instead of through individual bargaining.

¹⁸See the proof of Lemma 3 in the appendix for a derivation of this equation.

$F(g)/n$ while the nonunion wage is a function of the marginal product $\partial F(\mathbf{g})/\partial g_s$ of the worker. 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. Indeed, a worker with valuable characteristics, for instance a high marginal product, would rather bargain individually with the firm than share his or her advantage with the other employees.¹⁹ As a result, high-skill workers are more likely to be against unionization than low-skill ones.

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

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

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

where \mathbb{E}_g is the expectation across skills. Equivalently, the difference between nonunion and union profits is

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

This proposition shows that, for any hiring decision \mathbf{g} , the firm prefers to bargain individually, while the workers, on average, would rather be represented by a union. This conflict of preferences is a direct consequence of the decreasing returns to scale of the production function. Indeed, as $\alpha \rightarrow 1$, 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 the 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 infra-marginal workers. By forming a union, the workers can thus extract a bigger share of these high marginal products, which lowers the firm's profit.

To intuitively understand the importance of α , one can think of decreasing returns as arising because of a fixed factor of production (fixed capital or land, span of control, etc). Since the union can stop production, thereby preventing the firm from receiving the returns to the fixed factor, the firm is willing to increase the wage of the union workers. The more important the fixed factor is to production (lower α) the bigger is the bargaining advantage of the union. If there is no fixed factor ($\alpha = 1$) the union loses its bargaining advantage and, following Proposition 1, both workers and the firm are indifferent with regards to unionization.

¹⁹Verna (2005) discusses the literature on the relationship between the productivity of the workers and pay in union firms. Consistent with the theory, pay is more correlated with ability and performance in nonunion firms than in union firms.

Proposition 1 is consistent with evidence from Kleiner (2001) showing that firms generally oppose unions. Bronfenbrenner (1994) and Freeman and Kleiner (1990) 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. Hirsch and Berger (1984) finds that there is a positive correlation between the capital share of an industry and its union membership.²⁰

2.6 Voting procedure

When the union vote takes place, the measure of workers is fixed and the workers know the wages they will get after either outcome of the vote. Each worker has random preferences over the union status of the firm. Specifically, a worker of skill s votes for a union if and only if $w_s^u(\mathbf{g}) - w_s^n(\mathbf{g}) > \epsilon$, where ϵ is an idiosyncratic disutility cost of being part of a union. It has mean 0, is drawn independently across workers in each period and has CDF ϕ with $\phi(0) = 1/2$.²¹

A law of large numbers applies when aggregating the workers of a given skill. Therefore, a fraction $\phi(w_s^u(\mathbf{g}) - w_s^n(\mathbf{g}))$ of workers of type s votes for unionization. Denoting by

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

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

Notice that even though the preference shocks ϵ are random, the outcome of the vote is fully deterministic. Therefore, at the moment of posting vacancies, the firm knows whether the workers will form a union or not. 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 sub-market must be equal. This implies that the rate of job searchers in submarket s is

$$\frac{U_s}{N_s} = \frac{\delta}{\delta + p(\theta_s)(1 - \delta)}. \quad (15)$$

We can now define a steady-state equilibrium.

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

1. the value functions $\left\{W_{j,s}^E, W_s^U\right\}_{s \in \mathcal{S}}^{j \in \mathcal{J}}$ solve (4) and (5);

²⁰The findings of Hirsch and Berger (1984) are consistent with the theory as long as capital is not perfectly mobile in the short-run, such that some capital goes unused if the bargaining fails.

²¹Although not necessary for the results, this random disutility of being in a union smooths the firm's optimization problem and therefore allows numerical maximization algorithms to converge rapidly.

2. the measure $\{g_s^j\}_{s \in \mathcal{S}}$ solves the optimization problem of firm j given by 3 where $\mathbf{w}(\mathbf{g}) = \mathbf{w}^n(\mathbf{g})$ if $V(\mathbf{g}) \leq 0$ and $\mathbf{w}(\mathbf{g}) = \mathbf{w}^u(\mathbf{g})$ otherwise;
3. the wage schedule $\{w_s^j\}_{s \in \mathcal{S}}$ satisfies the collective bargaining wage (9), if firm j is unionized or the individual bargaining wage (12) if firm j is not unionized;
4. unemployment is stationary in each labor market such that (15) is satisfied.

3 Economic forces at work

We now consider how the union threat influences the decisions of a firm. As shown in Lemma 1, at a steady state, a firm solves

$$\max_{\mathbf{g}} \mathcal{J}(\mathbf{g}, \mathbf{w}(\mathbf{g})) \quad (16)$$

with

$$\mathbf{w}(\mathbf{g}) = \begin{cases} \mathbf{w}^u(\mathbf{g}) & \text{if } V(\mathbf{g}) > 0 \\ \mathbf{w}^n(\mathbf{g}) & \text{if } V(\mathbf{g}) \leq 0 \end{cases}$$

and where the objective function \mathcal{J} is

$$\mathcal{J}(\mathbf{g}, \mathbf{w}(\mathbf{g})) = F(\mathbf{g}) - \sum_{s \in \mathcal{S}} g_s w_s(\mathbf{g}) - \kappa(1 - (1 - \delta)\gamma) \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)}.$$

In an economy in which unions are weak, perhaps because of a low bargaining power β_u , firms do not have to worry about unions and simply hire to maximize discounted profits under the nonunion wages $\mathbf{w}^n(\mathbf{g})$. Let us denote this hiring decision by $\mathbf{g}^{n*} = \operatorname{argmax}_{\mathbf{g}} \mathcal{J}(\mathbf{g}, \mathbf{w}^n(\mathbf{g}))$.

As the strength of unions increases, workers might decide to form a union. If this happens, the firm is *constrained* by the unionization vote and hiring according to \mathbf{g}^{n*} is not optimal anymore. A constrained firm will consider distorting \mathbf{g}^{n*} to influence the workers into rejecting the union. This distorted measure, denoted by \mathbf{g}^n , maximizes $\mathcal{J}(\mathbf{g}, \mathbf{w}^n(\mathbf{g}))$ subject to the workers rejecting the union: $V(\mathbf{g}) \leq 0$. Because of this voting constraint, the threat of unionization modifies the hiring decisions and the wages paid by firms that are union free in equilibrium.

As the strength of unions increases even more, the firm begins to contemplate letting its workers unionize. In this case, its profits would be $\mathcal{J}(\mathbf{g}^{u*}, \mathbf{w}^u(\mathbf{g}^{u*}))$ where \mathbf{g}^{u*} is the optimal employment vector under collective bargaining. If unions are strong enough, such that $\mathcal{J}(\mathbf{g}^{u*}, \mathbf{w}^u(\mathbf{g}^{u*})) > \mathcal{J}(\mathbf{g}^n, \mathbf{w}^n(\mathbf{g}^n))$, the firm chooses to be unionized as an optimal reaction to the unionization threat.²²

To understand the mechanisms at work, it is useful to first consider an equilibrium in which the union status of each firm is given exogenously, such that no union vote takes place. In this case, we can characterize how the firm hires, the wages it pays as well as the workers' preference

²²It is possible to build examples in which $V(\mathbf{g}^{u*}) < 0$ but this requires extreme parameters.

for unionization. The full problem of a firm, in which its union status depends endogenously on the vote, can then be thought of as a deviation from the exogenous case. To focus the analysis on the empirically relevant cases, we assume from now on that the value of unemployment W_s^u and the labor market tightness θ_s are increasing in s .

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. Such a firm maximizes $\mathcal{J}(\mathbf{g}, w^i(\mathbf{g}))$ where $i = u$ if the firm is unionized and $i = n$ otherwise. By defining

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

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

$$\text{MC}_s^i = B_i \frac{\partial F(\mathbf{g})}{\partial g_s} \quad (18)$$

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 output retained by the firm. Equation (18) simply states that at the optimum the marginal revenue from hiring an extra worker of type s is equal to its marginal cost. Solving this equation, we find that the optimal hiring decision g^{i*} is

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

which 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 the firm therefore relies less on them for production (g_s^{i*} small). All else equal, nonunion firms are also larger than union firms 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 firms.

Proposition 2. *The equilibrium wage schedules $w_s^u(\mathbf{g}^{u*})$ and $w_s^n(\mathbf{g}^{n*})$ are increasing in s and the union wage gap $w_s^u(\mathbf{g}^{u*}) - w_s^n(\mathbf{g}^{n*})$ 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 they were to vote in

favor or against unionization in a given firm.

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

Proposition 3 is consistent with Farber and Saks (1980), who show that the desire to be unionized goes down with the position of the worker in the intra-firm earnings distribution.

Propositions 2 and 3 are direct consequences of the individual and collective bargaining protocol. As individually bargained wages depend more on individual characteristics, they 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. The firm therefore compares its profits under two employment vectors: the optimal one under which the workers unionize \mathbf{g}^{u*} and the optimal one under which the workers reject the union \mathbf{g}^n .

The optimal nonunion hiring decision \mathbf{g}^n solves a modified version of the first-order conditions (18) that takes into account the impact of the workers on the vote. The new conditions are

$$MC_s^n + \lambda \frac{\partial V(\mathbf{g}^n)}{\partial g_s^n} = B_n \frac{\partial F(\mathbf{g})}{\partial g_s} \quad (20)$$

where $\lambda \geq 0$ is the Lagrange multiplier associated with the voting constraint. We see that the voting constraint effectively increases the marginal cost of hiring workers who help the union vote. To better understand who these workers are, we can expand this additional term in the following way:

$$\frac{\partial V(\mathbf{g})}{\partial g_s} = \underbrace{\phi(\Delta_s(\mathbf{g})) - \frac{1}{2}}_{(a)} + \underbrace{g_s \frac{\partial \Delta_s(\mathbf{g})}{\partial g_s} \frac{\partial \phi(\Delta_s(\mathbf{g}))}{\partial \Delta_s(\mathbf{g})}}_{(b)} + \underbrace{\sum_{s' \neq s} g_{s'} \frac{\partial \Delta_{s'}(\mathbf{g})}{\partial g_s} \frac{\partial \phi(\Delta_{s'}(\mathbf{g}))}{\partial \Delta_{s'}(\mathbf{g})}}_{(c)} \quad (21)$$

where $\Delta_s(\mathbf{g})$ is short notation for the counterfactual union wage gap $w_s^u(\mathbf{g}) - w_s^n(\mathbf{g})$ and where $\phi(\Delta_s)$ is, as before, the fraction of workers of skill s who vote for the union when the wage gap is Δ_s . Each term in (21) represents one mechanism through which workers of type s influence the union vote.

(a) Fraction of voters for union As a fraction $\phi(\Delta_s)$ of workers of type 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) Wages of workers in the same skill group Adding an extra worker of type s changes the union wage gap for all workers of type s . In particular, it lowers their marginal product

which then lowers their nonunion wage and makes a larger fraction of them vote in favor of unionization.

(c) Wages of workers in other skill groups Adding an extra worker of type s also changes the union wage gap for all workers of type $s' \neq s$. For instance, since union wages are determined by the average product, increasing the number of high-skill workers shifts the union wage schedule upward, leading some workers to change their vote in favor of unionization. Similarly, if the firm hires a lot of low-skill workers, their relatively low marginal product pushes the union wage schedule downward, which increases the number of workers against unionization.

Mechanism (a) is the main driver of unionization in the model and it pushes the firm to hire more workers who vote against the union and fewer workers who vote in its favor.²³

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 while keeping the main mechanisms of the model active in a simplified, 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 ($\epsilon = 0$). For tractability, we further assume that the firm combines labor inputs using a Cobb-Douglas technology ($\sigma = 1$) and that there is no home production ($b_s = 0$).

The following proposition establishes conditions under which the union threat influences the behavior of nonunion firms in partial equilibrium.

Proposition 4. *If $B_n > B_u$ and $z_h q(\theta_h) < z_l q(\theta_l)$ then the union threat is binding in nonunion firms.*

The first assumption of the proposition ($B_n > B_u$) imposes that workers, as a group, prefer to be unionized. The second assumption ($z_h q(\theta_h) < z_l q(\theta_l)$) guarantees that the firm would hire more low-skill than high-skill workers in an environment without the voting constraint, thereby giving low-skill workers the majority of the votes in an election. As these workers vote in favor of the union, the firm must distort its hiring decision, from g^{n*} to g^n , to prevent unionization. The firm does so by over-hiring high-skill workers, who vote against the union, and under-hiring low-skill workers, who vote in its favor. The following propositions describe how this distortion in the skill composition of its workforce influences the firm and the wages it pays.

Proposition 5. *Under the same assumptions as Proposition 4, the union threat lowers the profits, employment and output of nonunion firms.*

This proposition highlights that the threat essentially lowers the size of the firm and its profitability. To understand why, it is useful to think of the threat as a constraint that forces the firm

²³Mechanism (c) can give rise to a “poison-pill” policy in which a nonunion firm hires many low skill workers to prevent unionization. While these workers vote in favor of unionization, their hiring can lower the union wages schedule so much that it pushes workers with higher skill to vote against unionization.

to use a set of workers with a less efficient skill composition. As a result, the marginal cost of producing one unit of goods increases which pushes the firm to produce less and therefore to hire fewer workers.²⁴

The threat also has an impact on wages, as the following proposition shows.

Proposition 6. *Under the same assumptions as Proposition 4, the union threat increases the average wage and decreases wage inequality, as defined as the ratio of the high-skill wage to the low-skill wage, in nonunion firms.*

As the firm reduces its size in response to the threat, the average marginal product of the workers increases, which leads to a higher average wage — further incentivizing workers to vote against the union. As the firm hires a higher ratio of high-skill to low-skill workers, the marginal product of high-skill workers falls relative to that of the low-skill workers. As a result, wage inequality decreases when a firm is subject to the union threat.

Finally, the next proposition characterizes which firms decide to prevent unionization and which ones let the workers unionize to maximize profit.

Proposition 7. *Under the same assumptions as Proposition 4, if*

$$\frac{B_n}{B_u} \geq \left(\frac{q(\theta_h)^{-1} + q(\theta_l)^{-1}}{\left(\frac{1}{z_l}q(\theta_l)^{-1}\right)^{z_l} \left(\frac{1}{z_h}q(\theta_h)^{-1}\right)^{z_h}} \right)^\alpha \quad (22)$$

and

$$\frac{\alpha\beta_n}{1 - (1 - \alpha)\beta_n} z_h \geq \frac{1}{2}\beta_u \quad (23)$$

it is optimal for the the firm to prevent unionization, otherwise it chooses to be unionized.

Equation (23) is a *feasibility* condition. If it is not satisfied, the firm cannot hire such that high-skill workers have the majority of the votes and also prefer to be in a union free firm.²⁵ Equation (22) is a *profitability* condition. Its left-hand side compares the share of the production surplus that the firm receives if it bargains individually with its workers to the share it receives under collective bargaining. It is a measure of the gain in profits associated with preventing unionization. The right-hand side of the equation measures the cost associated with preventing unionization. It depends on the amount of heterogeneity between workers, defined as $z_l q(\theta_l) - z_h q(\theta_h)$. When the

²⁴The voting constraint effectively increases the marginal cost of hiring a low-skill worker by a Lagrange multiplier and decreases the marginal cost of hiring a high-skill worker by the same multiplier. As a result of the concavity of the production function, the firm decreases the number of low-skill workers it employs by more than it increases the number of high-skill workers. Because of the complementarity between the two skill groups, the marginal product of high-skill workers goes down which leads to even less employment and less production.

²⁵The firm prevents unionization by hiring more high-skill workers and fewer low-skill workers. If inequality (23) does not hold, 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.

heterogeneity is minimal, $z_l q(\theta_l) = z_h q(\theta_h)$, the right-hand side equals 1 and it is always profitable to prevent unionization, under our assumption that $B_n > B_u$. As heterogeneity grows larger, the right-hand side increases and preventing unionization becomes more costly.

To better understand the link between worker heterogeneity and the cost of preventing unionization, consider the problem of a firm in which $z_l q(\theta_l)$ and $z_h q(\theta_h)$ are close to each other. In this case, without the voting constraint, the firm would hire a similar number of high-skill and low-skill workers. With the voting constraint, the firm therefore needs to hire only a few additional high-skill workers and a few less low-skill workers to prevent unionization. The distortion associated with this change is small and so is the cost of preventing unionization. As the gap between $z_l q(\theta_l)$ and $z_h q(\theta_h)$ becomes larger, so is the size of the distortion needed for the firm to win the union vote.

Panel (a) of Figure 2 shows the implications of Proposition 7 for the union status of the firm as a function of the bargaining powers β_u and β_n . For all β_u 's under the thick black curve, assumption $B_n > B_u$ is not satisfied and the firm prefers to be unionized. For the β_u 's above this curve, the firm prefers to be union free but low-skill workers, who have the majority of the vote, prefer to be in a union firm. The firm must therefore modify its hiring policy in order for the union vote to fail. In the dark grey zone, the firm finds it optimal to do so (inequalities 22 and 23 are satisfied) and it hires according to g^n . Outside the dark grey zone, the firm prefers to let the union win; preventing unionization would be too costly. In this case, the firm hires according to g^{u*} .²⁶

Panel (a) also shows how the set of (β_u, β_n) for which the firm prevents unionization evolves as a function of the firm's return to scale parameter α . As α moves from a low value to a high value, the dark grey zone becomes the light grey zone, and the black curve becomes the grey curve. Two mechanisms are at work. First, the asymmetry between individual and collective bargaining is more important when α is low. As explained after Proposition 1, a small α makes individual bargaining more attractive to the firm. This effect pushes down the thick black curve and the lower curve of the grey zone (equation (22)) as α declines. Second, a smaller α lowers the share of the surplus that the workers receive when bargaining individually. As a result, the nonunion wage of high-skill workers can easily fall under their counterfactual union wage, which makes it hard for the firm to prevent unionization. This effect pushes down the upper curve of the grey zone through its impact on inequality (23).

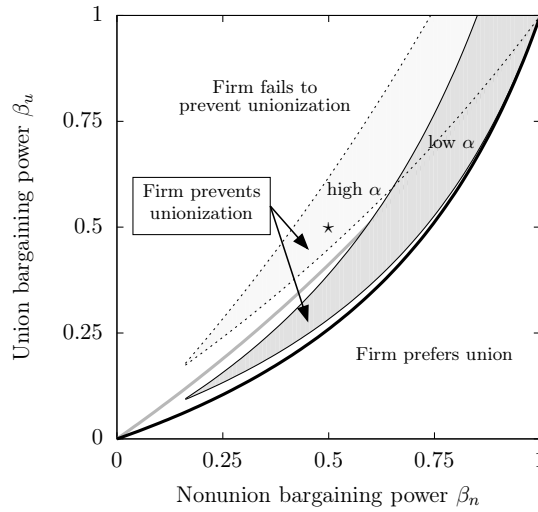
We can infer from Panel (a) which type of firms are more likely to be unionized in equilibrium. For instance, if the bargaining powers are both in the middle of the parameter space, indicated by the \star on the figure, a firm with a low α would be unionized while a firm with a high α would be union free. This prediction of the model is consistent with Hirsch and Berger (1984) who find that industries with lower labor shares are more unionized. Notice also that Proposition 7 implies that there is, in general, a range $[\underline{\alpha}, \bar{\alpha}]$ such that firms with these returns to scale parameters are union free but have their decisions influenced by the union threat. In particular, if the economy features a full distribution of firm types with various α 's then a positive measure of firm types would be

²⁶In Figure 2, the lower curves of the grey zones corresponds to inequality (22), while the upper curves corresponds to inequality (23).

influenced by the threat.

The bottom panel of Figure 2 shows the impact of an increase in the intensity of the high-skill workers z_h on the union status of a firm. This increase lowers the heterogeneity $z_l q(\theta_l) - z_h q(\theta_h)$ between workers and makes it easier for the firm to prevent unionization. As a result, the firm finds it optimal to be union free for a larger range of β_u 's and β_n 's.

(a) Impact of returns to scale α on unionization



(b) Impact of skill intensity z_h on unionization

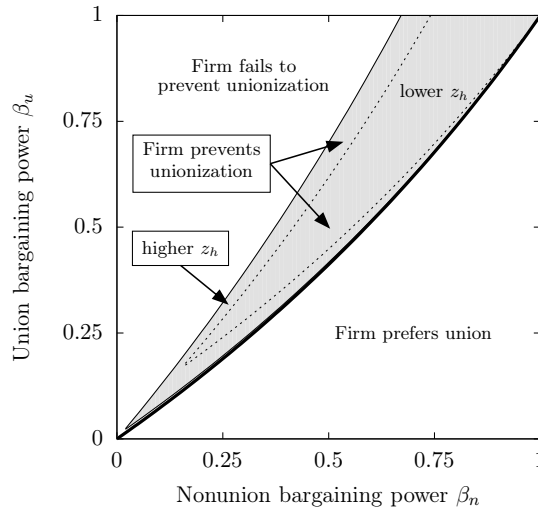


Figure 2: Technology and unionization

Numerical example

The results obtained in this simple economy hold more generally. To show this point, and to illustrate how the union threat affects firms that employ more than two types of workers, Figure 3

provides a numerical example of a firm's behavior with and without the threat of unionization.²⁷

Consider first the dashed line in Panel (a), which shows the optimal employment decision g^{n*} when there is no threat of unionization. In this case, the firm hires according to equation (19) and pays wages $w^n(g^{n*})$, represented by the dashed line in Panel (c). While this behavior is optimal when the workers cannot form a union, it might not be in an environment in which unionization is possible. In such an environment, workers would compare their current wage $w^n(g^{n*})$ to what they would receive if they were to unionize, represented by the dotted line in Panel (c), and would vote to maximize their income. Panel (b), which shows the cumulative distribution function of g^{n*} , illustrates how such a vote would play out. The thin dotted line on this figure indicates that the median voter has skill $s = 6$ and that he or she would receive a higher wage in a unionized firm than in a nonunion one. As a result, under g^{n*} , the workers would vote to unionize the firm.

Since profits are lower under collective bargaining, the firm would attempt to prevent unionization. To so do, it distorts employment (and, implicitly, wages) by over-hiring high-skill workers (who vote against the union) and under-hiring low-skill workers (who vote in favor of the union). This change in employment can be seen in Panel (a) and Panel (b). Notice that the marginal voter is now of skill $s = 9$ and that this worker just favors being in a nonunion firm, as can be seen by comparing his or her union and nonunion wages in Panel (d).

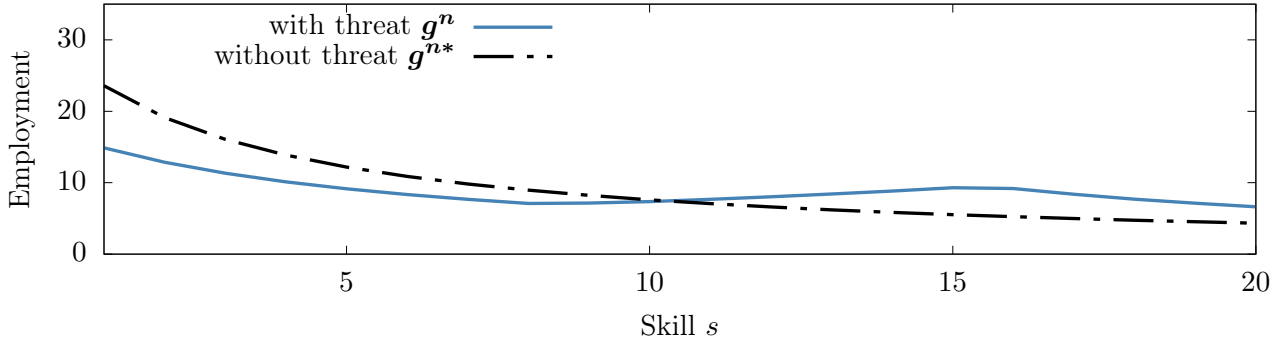
While the firm is changing employment to prevent unionization, it understands that it is implicitly changing the wages that the workers receive. In particular, hiring more high-skill workers lowers their marginal product which pushes their wages $w^n(g^n)$ to fall compared to the unconstrained wage schedule $w^n(g^{n*})$. In contrast low-skill workers see their wages increase because of the opposite mechanism. Notice that the range of nonunion wages paid by the firm is smaller when the workers have the possibility to unionize, contributing to a lower wage inequality within the firm. The relative movements in the employment vector (more high-skill and fewer low-skill workers) push, however, for more wage inequality.²⁸

The union threat also affects the firm in other ways. As in Proposition 5, the firm's output, employment and profits are lower under g^{n*} than under g^n . Also, the threat lowers the range of wages and increases average wages, as in Proposition 6.

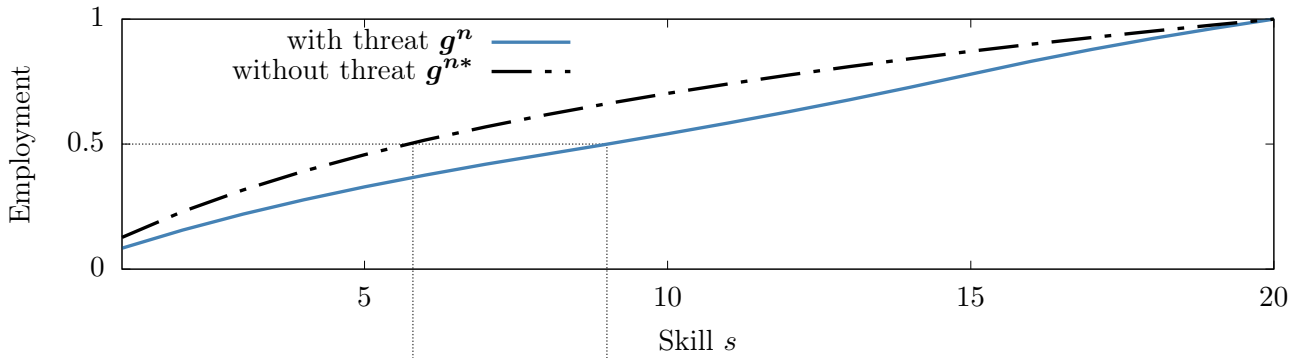
²⁷Appendix B.1 provides the details of this exercise.

²⁸We also see in Panel (d) that $w^u(g^n)$ and $w^n(g^n)$ are very close to each other for workers with skills in the middle of the distribution. The firm would like to hire more of these workers: they vote against unionization and their relatively small marginal product has a small impact on the union wage schedule. However, hiring additional workers in this zone would lower their marginal product and push their nonunion wage under their union wage. They would therefore change their vote to support the union, which the firm wants to avoid.

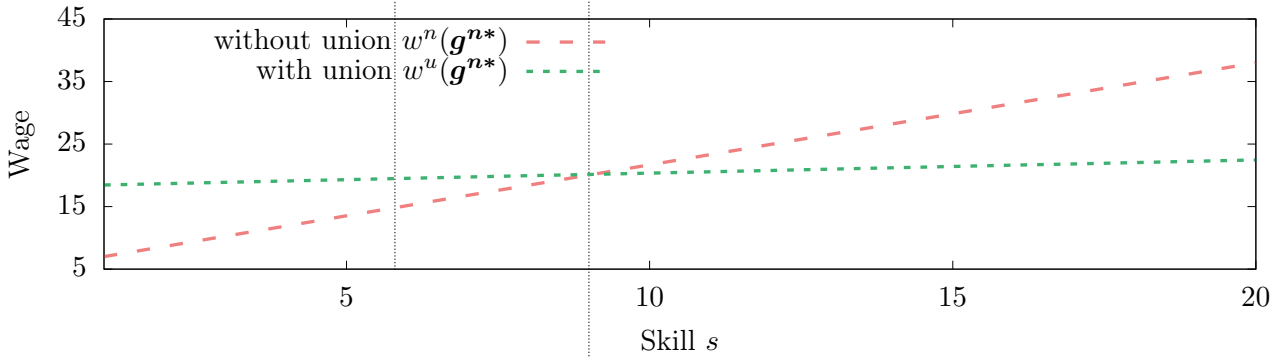
(a) Employment with (g^n) and without (g^{n*}) threat (probability density function)



(b) Employment with (g^n) and without (g^{n*}) threat (cumulative distribution function)



(c) Wages without threat



(d) Wages with threat

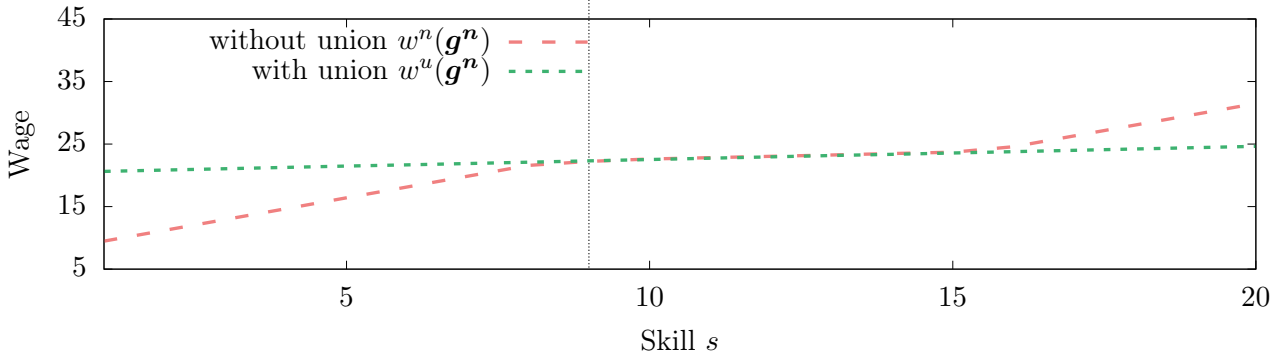


Figure 3: Employment and wages with the union threat

4 Quantitative exploration

I now estimate the model on the private sector of the United States in 2005. To do so, I assume that firms use one of two technologies u and n . Firms with the technology u are unionized in equilibrium while firms of type n are not.

Several parameters, mostly about preferences and the workings of the labor market, are taken directly from the literature. The remaining parameters (bargaining powers and technology parameters), which are the key determinants of the strength of the threat, are estimated directly from the data using a method of simulated moments. With the estimated model in hand, I conduct several policy experiments to evaluate the impact of unions on the economy and I also use the calibrated model to evaluate the predictions of several reduced-form estimators used in the empirical literature.

The last part of this section discusses policy implications that follow from the quantitative exercises. It also provides a version of the model estimated on the United States in 1983 to evaluate the importance of the union threat during a period of the U.S. history when unionization rates were much higher.

Appendix B provides additional details about the quantitative exercises along with several robustness tests.

4.1 Data and estimation

Data

Data on wages and the union status of workers come from the Merged Outgoing Rotation Groups of the Current Population Survey (CPS).²⁹ 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.³⁰

To build the skill index of 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 $j(i)$, the regression is

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

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

²⁹I clean the sample by removing agricultural, public sector and out-of-the-labor-force workers. I also remove individuals with an hourly wage higher than \$100 or lower than \$5, and individuals younger than 16 or older than 65.

³⁰For each industry in the BEA dataset, I divide total workers' compensation by value added to get an estimate of the labor share in that industry. 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. The estimation targets these moments.

$\hat{s}_i = \exp(\hat{\Lambda}X_i^1)$. 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 only on nonunion workers, the predicted values \hat{s}_i are computed for *all* members of the labor force.³¹ The support of the distribution is 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. Using this skill index, I compute the average wage and the mass of workers of each skill in the union and nonunion sectors of the economy.

Parameters from the literature

To reflect the typical duration of labor contracts the time period is set to one year. All monetary amounts are measured in thousands of dollars. The discount rate is set to $\gamma = 0.95$ and the job destruction rate to $\delta = 0.113$, as in Pries and Rogerson (2005). For the matching function, I use the functional form of den Haan et al. (2000) along with the parametrization of Krusell and Rudanko (2012) so that $m(U, V) = UV/(U + V)$. The elasticity of substitution between skills is set to $\sigma = 1.5$ as in Krusell et al. (2000). For the cost of posting a vacancy, I follow the analysis of Silva and Toledo (2009) who find that training and vacancy posting costs amount to 69% of quarterly wages, which translates to $\kappa = 1.8$.³² The value of these parameters is listed in Table 4. Finally, for the workers' preference for unionization, I use a linear function to approximate the CDF ϕ . This approach 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, so that one less parameter has to be estimated.³³ Finally, I have verified that, in the calibrated economy, using a logistic CDF instead only has a minimal impact on the results.³⁴

Definition	Parameter	Value	Source/reason
Discount factor	γ	0.95	5% annual interest rate
Job destruction probability	δ	0.113	Pries and Rogerson (2005)
Skill elasticity of substitution	σ	1.5	Krusell et al. (2000)
Cost of posting a vacancy	κ	1.8	Silva and Toledo (2009)
Number of skills	S	6	See text

Table 4: Parameters from the literature

³¹This 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.

³²Alternative 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.

³³See Lemma A2 in the Appendix. Note that this slope can be as flat as necessary to ensure that all probabilities remain between zero and one.

³⁴An earlier calibration used micro data from the 1970s about workers' preference for unionization to parametrize the logistic CDF. The current approach has the advantage of not requiring outdated data for the estimation. That earlier calibration found a similar impact of unions on the economy.

In the model, the value of non-work activities \mathbf{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, averaged across states, of \$19,280 in 2005 dollars. I include these benefits in \mathbf{b} . To capture the components associated with home production and leisure, I also include a second term in \mathbf{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).³⁵

Technologies

Firms are endowed with one of two technologies $\{\alpha_u, \mathbf{z}_u\}$ and $\{\alpha_n, \mathbf{z}_n\}$. Firms with technology u will be unionized in equilibrium while firms with technology n will be union free.

The skill intensities \mathbf{z}_u and \mathbf{z}_n are modeled as probability density functions of truncated log-normal distributions with mean parameters $\{\mu_u, \mu_n\}$ and variance parameters $\{\xi_u^2, \xi_n^2\}$ to be estimated. In the model, the parameters α_u and α_n capture the curvature of the production functions with respect to labor inputs. Appendix A.1 shows that if capital is also a factor of production, and that it is combined to labor in a Cobb-Douglas way, then α_j also depends on total returns to scale in addition to the labor intensity of the firm. To estimate α_j , I therefore assume that firms have total returns to scale of 0.80, in the range of values estimated by Burnside et al. (1995), and target the labor shares in the union and nonunion sectors.³⁶

A natural way to estimate the model would be to allow all firms of type n to be subject to the threat of unionization and to let the estimation determine whether that is the case or not. It is possible however that several of these firms evolve in industries or states in which unions are weak, perhaps for policy reasons, and where the union threat has therefore little to no impact on their decisions. I therefore add an additional degree of flexibility to the model and allow a fraction of the firms of type n to be completely immune to the union threat.^{37,38} To evaluate the size of this fraction, I use micro data from the National Labor Relation Board (NLRB) about the outcome of union elections. In 2005, about 50% of union elections, in workers-weighted terms, were successful. Abstracting from repeat elections, this indicates that a number of nonunion firms equal to the number of current union firms would have had an unsuccessful election in the past. I assume that these nonunion firms *might* be subject to the threat of unionization while the rest of the firms of type n are immune.³⁹

³⁵An alternative calibration of the model in which b_s is 85% of the average wage of workers of type s finds an overall stronger impact of the union threat on the economy.

³⁶Appendix A.1 shows that capital can be included in production by reinterpreting the parameters of the model.

³⁷An earlier version of the paper assumed that all nonunion firms were subject to the unionization threat. It found stronger effects of the threat on the economy.

³⁸Since the NLRB data does not contain information about the industry in which a firm operates or about the skill composition of its workforce, I cannot separately identify the technologies used by the threatened and non-threatened non-union firms.

³⁹Note that this assumption is likely to underestimate the impact of the union threat on the economy. Indeed, a firm that is, in reality, subject to the threat but that manages to avoid a union election would be counted as

Calibrated economy

I use a method of simulated moments to estimate the remaining parameters of the model: the bargaining powers β_u and β_n , and the technology parameters $\{\alpha_u, \mu_u, \xi_u^2\}$ and $\{\alpha_n, \mu_n, \xi_n^2\}$.⁴⁰ The estimation seeks to bring a set of model quantities as close as possible to their data counterparts. The targeted quantities are the wage schedules $\{w_s^u\}_{s=1}^S$ and $\{w_s^n\}_{s=1}^S$, employment in the union and nonunion sectors $\{g_s^u\}_{s=1}^S$ and $\{g_s^n\}_{s=1}^S$, and the labor shares in the union and nonunion sectors.^{41,42}

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 curvature parameters α_n and α_u are identified by the labor shares. The average wages in the union and nonunion sectors identify the bargaining powers β_n and β_u , and the number of workers of each skill in the union and nonunion sectors identify the remaining technology parameters.

Table 5 shows the parameter values that best fit the data.^{43,44} Importantly, the estimation finds that the voting constraint is actually binding and that the union threat therefore distorts decisions. The key features of the data that push the estimation to reach that conclusion are the union and nonunion wage schedules. Since union wages are relatively high in the data, the estimation finds a high value for β_u . 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, which leads to distortions in the way they hire and in the wages they pay.

Figure 4 shows how the estimated model fits wages and employment in the union and nonunion sectors. The labor shares are fitted perfectly.

immune in the current exercise and the threat would have no impact on its decisions. Appendix B provides several additional exercises that consider different assumptions related to immune firms. First, it contains an alternative parametrization of the economy in which, instead of being immune to unionization by assumption, the non-threatened nonunion firms are unaffected by the union threat because of a higher curvature α in their production function. The results are overall similar to those presented in this section. Second, Appendix B also provides a full estimation of an alternative model in which the firms that are potentially affected by the union threat are those that, in the NLRB union election data, have won their election by a margin of fewer than 10% of the votes. The impact of the threat is somewhat weaker in this case.

⁴⁰As Lemma A1 in the appendix shows, it is equivalent to change the mass of firms of type j or their productivity A_j , so that the mass of firms of each type is normalized to one and their productivity A_j is simply set to replicate employment in the data.

⁴¹Since there is no data to differentiate between the wages and employments vectors of the immune and non-immune nonunion firms, the targeted nonunion employment and wage vectors \mathbf{g}^n and \mathbf{w}^n are computed by grouping together all nonunion firms.

⁴²I have also estimated the model with the outcome of the Freeman (1980) estimator as a targeted moment. The policy exercises of Table 6 are almost identical in this case.

⁴³In the calibrated economy, 8.5% of the population works in a union firms, 8.5% works in a nonunion firm that is subject to the threat of unionization and 77% works in a nonunion firm that is unaffected by the threat. The unemployment rate is 6%.

⁴⁴The estimation sets 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 .

Definition	Parameter	Calibrated value
Nonunion bargaining power	β_n	0.50
Union bargaining power	β_u	0.39
Curvature of u technology	α_u	0.72
Mean of skill intensity z_u	μ_u	1.18
Variance of skill intensity z_u	ξ_u^2	0.57
Curvature of n technology	α_n	0.73
Mean of skill intensity z_n	μ_n	1.49
Variance of skill intensity z_n	ξ_n^2	0.79

Table 5: Estimated parameters

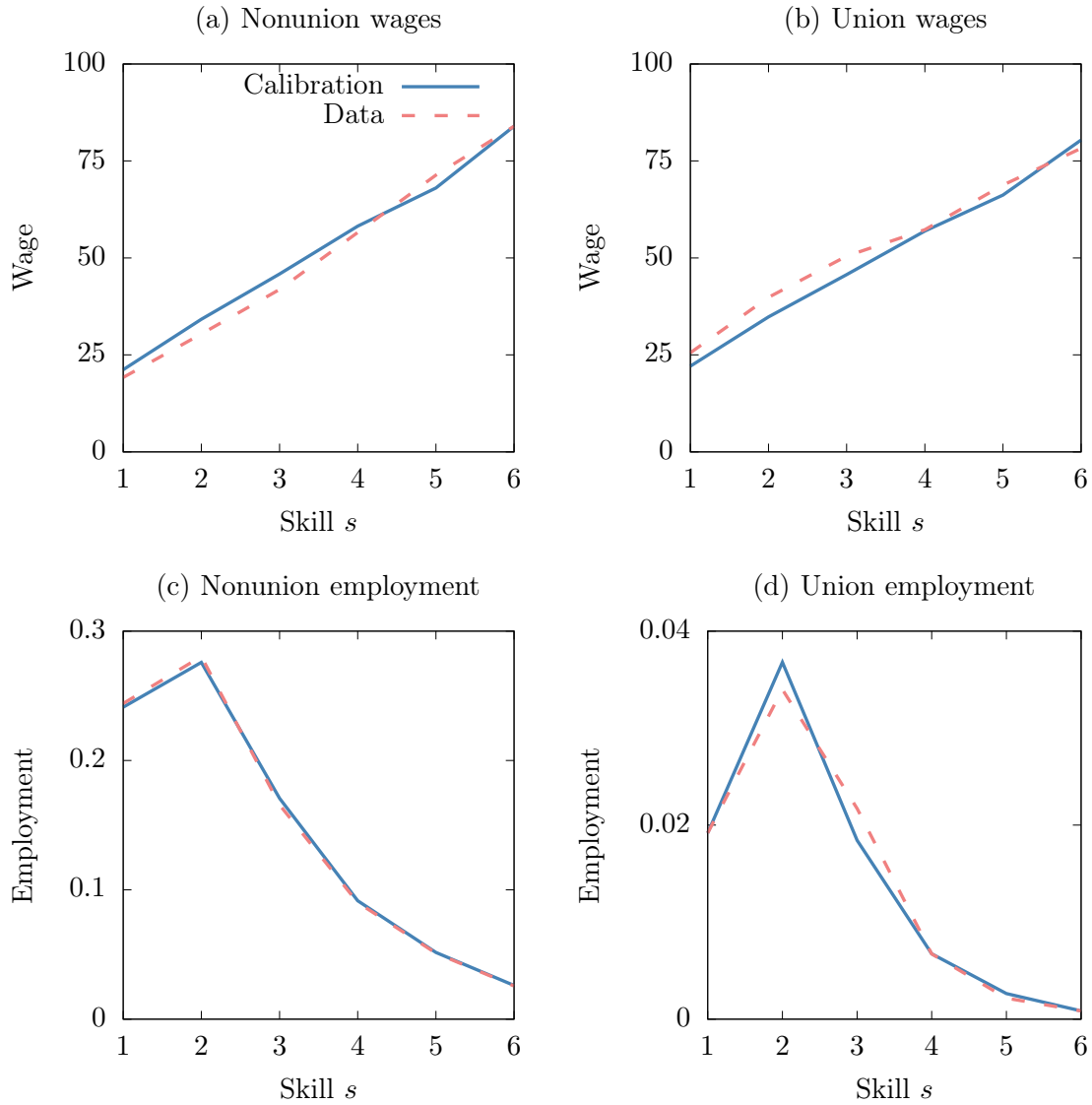


Figure 4: Fit of the calibrated model

4.2 Impact of unions on the economy

I now evaluate the impact of labor unions on the calibrated economy by considering three policy experiments in general equilibrium. First, I investigate the role played by the union threat alone (“no threat” policy). 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 policy experiment, I assume that unions are forbidden (“no union” policy). In this case, not only does the union threat disappears, but all firms that were previously unionized are now union free. This experiment captures the overall impact of labor unions on the economy. Finally, in the third policy experiment, I assume that all firms are now unionized (“all union” policy). In this case, all previously union-free firms become unionized. Notice that the union threat is inactive in all of these policy experiments.

The results of these experiments are summarize in Figure 5, which shows how they impact wages and unemployment across skills and in Table 6, which reports how they affect aggregate output, unemployment, welfare and wage inequality. The rest of this section describes how the economic forces at work in the model generate these results.

	No threat	No union	All unions
Output	1.0%	1.0%	1.0%
Unemployment rate	-1.2 pp	-1.2 pp	-1.1 pp
Welfare	0.3%	0.3%	0.6%
Var(log wages)	0.6%	5.5%	-49.0%
90-10 ratio log wages	1.2%	1.5%	-8.3%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 6: Impact of policy experiments

Impact of the union threat

We begin by considering the impact of the first experiment: the removal of the union threat. Figure 6 shows how wages and employment in nonunion firms react to the policy experiment. To better highlight the various mechanisms at work, the solid lines represent the changes that occur in partial equilibrium (when all aggregate quantities are kept unchanged) while the dashed lines represent the overall impact of the threat removal in general equilibrium.⁴⁵

Let us consider the partial equilibrium reaction of the firms first. From Panel (b), we see that, once the threat is gone, firms hire substantially more. This result is consistent with the theory and, in particular, with Proposition 5. Indeed, as the threat disappears, firms are no longer subject to a distortion and the effective marginal cost of production goes down. As a result firms increase their size to reach the flatter part of the production function. While this increase in hiring affects

⁴⁵To be precise, the partial equilibrium exercises keep the labor market tightness vector θ and the value of non-work activities vector \mathbf{b} fixed at their calibrated values.

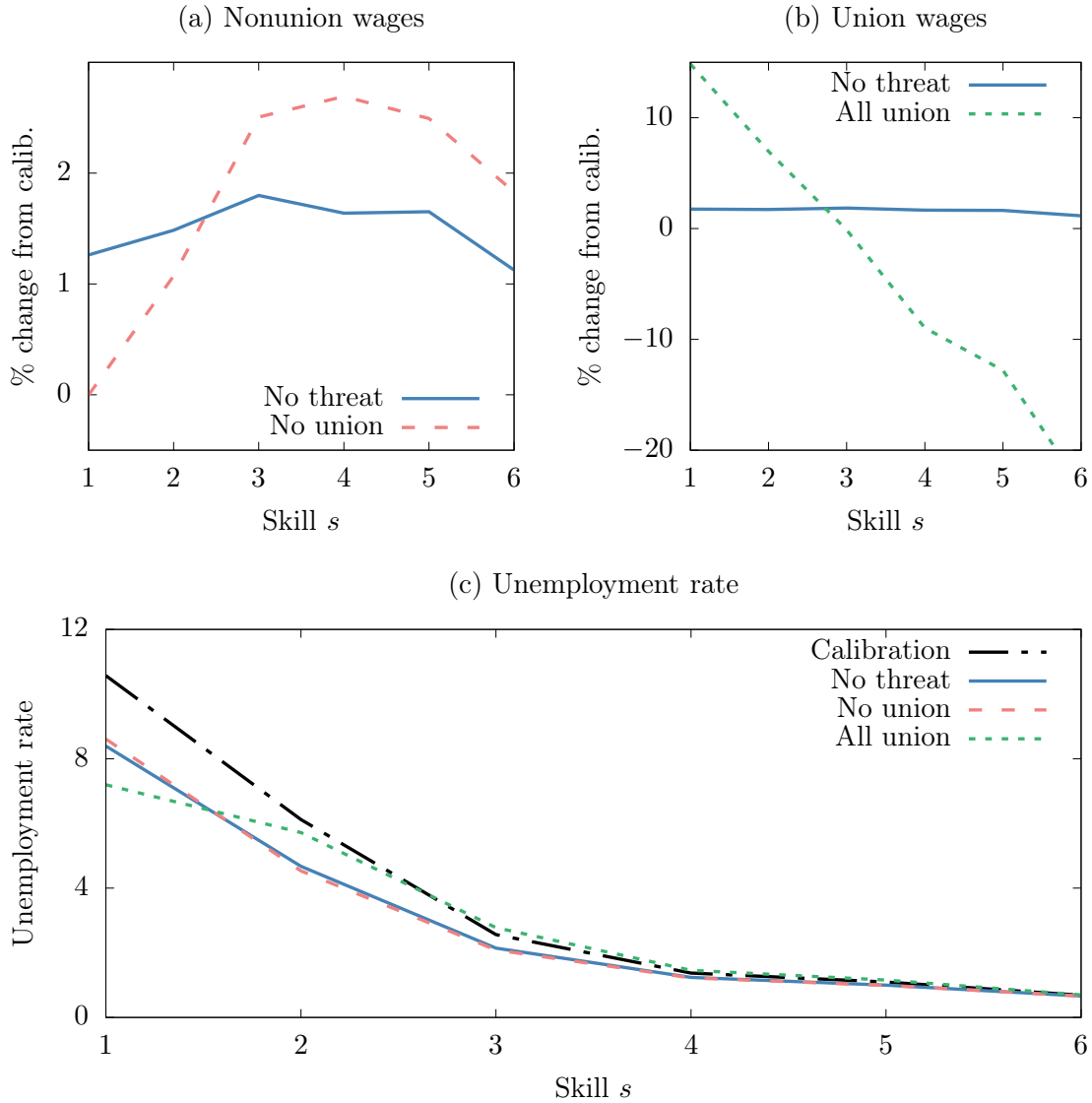


Figure 5: General equilibrium impact of policy experiments

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 (a) of Figure 6. Since firms now hire more workers, the marginal product of these workers declines which, through individual bargaining, adversely affects their wage. Notice that, in partial equilibrium, the disappearance of the union threat leads the firm to pay a broader range of wages, which pushes for an increase in wage inequality, as was predicted by Proposition 6.

In general equilibrium, the increase in hiring pushes unemployment down for all skill groups (bottom of Figure 5) for an overall decline in the unemployment rate of 1.2 percentage points. Since they can now find another job quickly if negotiations break down, these lower unemployment levels benefit the bargaining position of the workers and lead 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 all workers. Finally, these high wages temper 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 (b) of Figure 6).⁴⁶

As in partial equilibrium, the increase in nonunion wages in general equilibrium benefits workers above the median skill level more than those at its bottom. As a result, the threat removal increases the variance of log wages by 0.6% and the ratio of the 90th percentile to the 10th percentile of the wage distribution by 1.2%.⁴⁷ The removal also benefits output, with the increased in hiring that follows the disappearance of the threat pushing output up by 1%.

Welfare also benefits from the threat removal.⁴⁸ There are three inefficiencies that interact to yield this result.⁴⁹ First, there is the union threat which distorts the type of workers hired by the firms. Second, there is over- or under-hiring by firms since they do not internalize the impact of their vacancy posting decisions on labor market conditions (Hosios, 1990). Third, firms that are bargaining individually with their workers tend to over-hire to lower the workers' marginal products (Stole and Zwiebel, 1996a). In this first policy experiment, the first mechanism dominates and the removal of the threat increases welfare by 30 basis points. This number is smaller than the increase in output since, as many unemployed workers find employment once the threat is gone, the extra value of leisure b that unemployment brought is lost.

Mandating or prohibiting unions

Figure 5 and Table 6 also show the impact of the two other policy experiments: prohibiting all unions or, to the opposite, forcing all firms to be unionized. We see from Panel (a) in Figure 5 that removing all unions leads to a substantial increase in wage inequality (all workers are nonunion workers in this experiment). Indeed, since all bargaining is now done individually, the wage of low skill workers does not benefit from the high productivity of the high-skill workers anymore. As a result, the variance of log wages increases by 5.5% from its calibrated value while the 90-10 ratio of the wage distribution increases by 1.5%.

Perhaps surprisingly, output, welfare and the unemployment rate do not react much more to the

⁴⁶The unionization rate drops from 9.0% to 8.3% when the union threat disappears since, while the union status of firms do not change, the nonunion firms hire more workers.

⁴⁷Workers in the highest skill group benefits less than the workers in the lowest skill group but since there are very few workers with skill $s = 6$ their impact on overall wage inequality is minimal.

⁴⁸To avoid introducing Pareto weights, the reported welfare levels correspond to the discounted value of total output plus the value of non-work activities of the unemployed workers minus the hiring costs at the steady state. Appendix A.4 provides a discussion of the inefficiencies present in the environment.

⁴⁹Appendix A.4 provides a more technical discussion of these inefficiencies.

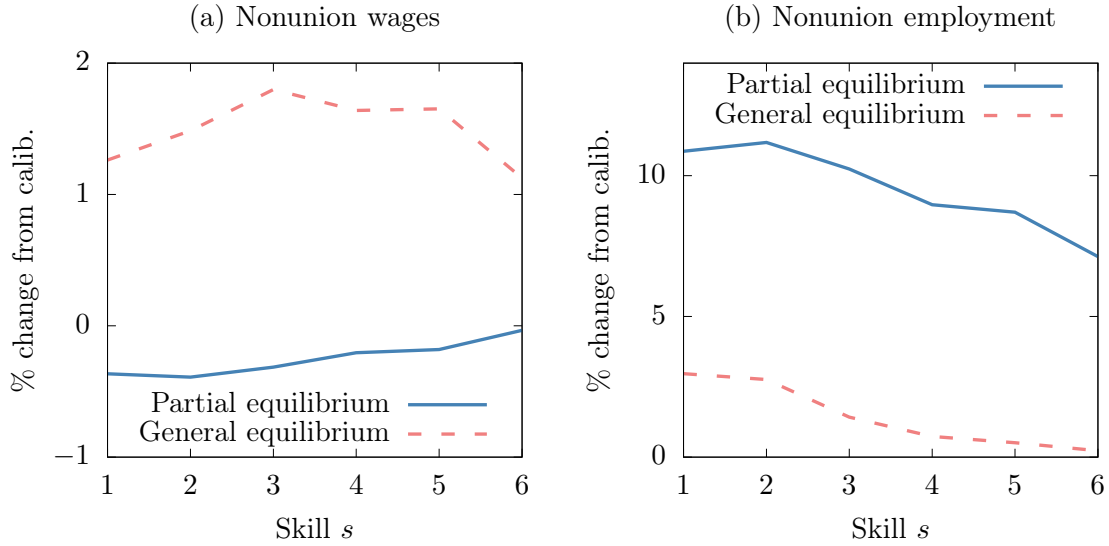


Figure 6: Removing the union threat: partial and general equilibrium impact

removal of all unions from the economy than under the “no threat” policy. We can conclude from this fact that the threat of unionization, more than the actual union status of the firms, is the the main channel through which unions affect output, welfare and unemployment in this economy.

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 5 (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 wage falls 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 49% and the 90-10 ratio declines by 8.3% under this policy. Similarly to the “no union” experiment, output and the unemployment rate are essentially the same as in the “no threat” experiment which reinforces the conclusion that the union threat is the key driver of these two variables.⁵⁰ The “all union” policy leads, however, to a larger increase in welfare than the other two policies. The different efficient properties of the bargaining protocols is responsible for this effect (Stole and Zwiebel, 1996a).

Comparison with empirical estimators

In the calibrated economy, the true impact of unions on wage inequality differs from what typical empirical estimators suggest. For instance, by computing the classical Freeman (1980) estimator on the calibrated economy, we find that unions are responsible for lowering the variance of log wages by 3.47%.⁵¹ In contrast, the “no union” policy experiment finds that unions are responsible for

⁵⁰Figures 8 and 9 in Appendix B.2 show how union and nonunion employment change in response to the policy exercises.

⁵¹This estimator can be written as $V - V^n = U\Delta_v + U(1 - U)\Delta_w^2$ 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

lowering wage inequality by 5.5%, or about 160% of the Freeman estimate. 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. \(2004\)](#)). In the model, the union and nonunion skill distributions are different. The idea is to attribute to every union worker a draw from the non-union wage distribution of workers of his or her skill. Taking this heterogeneity into account, these estimators predict that unions are responsible for lowering the variance of log wages by 0.80% or only about 14% of their true impact on wage inequality.⁵²

The key reason for these large discrepancies between empirical and model-based estimates is that the empirical estimators assume that the union and nonunion wage schedules are unaffected by the disappearance of unions. In the model, however, these schedules react to the change in policy for multiple reasons. First, 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 differs precisely because they use different technologies. Therefore, when the 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 generated by a mix of two technologies. Finally, the empirical 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 upwards. As this channel affects workers with different skills differently it leads to asymmetrical changes in the wages schedules. Putting all these mechanisms together explains the difference between the empirical and the model-based estimates of the impact of labor unions on wage inequality.

The model can also shed some light on regression discontinuity estimators. [DiNardo and Lee \(2004\)](#) use such a framework to compare firms who barely win a union election to firms who barely lose a union election. They find essentially no significant impact of unionization, although their estimates are often imprecise. To explain this small impact of unionization, they mention that a union threat effect, which would affect nonunion firms before a union election, would tend to bias the estimates against finding an impact of unionization. Since nonunion firms change their behavior to prevent unionization, union policies already have an effect even before a union election. As a result, comparing a firm before and after the election misses the impact of the union threat. While the current model does not fit directly in the [DiNardo and Lee \(2004\)](#) framework, we can still use it to evaluate the magnitude of the bias introduced by the threat of unionization.⁵³ To

the variance of log union and nonunion wages and Δ_w is the difference between the mean log of union and nonunion wages.

⁵²The Freeman estimator predicts a larger impact of unions on wages inequality than the more sophisticated estimators since the union skill distribution is more concentrated than the nonunion skill distribution (see Figure 4). 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 union on wage inequality.

⁵³The reason why [DiNardo and Lee \(2004\)](#) does not fit directly in the current framework is that the model does not feature any exogenous shock to the union vote. Here, I however assume that such an exogenous shock happens in the model, with measure zero probability as to not influence the behavior of firms, to evaluate the bias in the

do so, I consider, in the calibrated economy, a threatened nonunion firm that faces a nonunion bargaining power β_n such that, if it were to unionize, its employment would not change. Such a firm fits into the DiNardo and Lee (2004) framework: 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, like in DiNardo and Lee (2004). To evaluate the role played by the threat in this framework, I consider the impact of unionization on this threatened firm and compare it to the impact of unionization *on the same firm*, but assuming that it is not affected by the threat. The difference between these two outcomes is indicative of the impact of union policies that is not captured by the regression discontinuity estimator. Table 7 shows the results of the exercise. We see that for employment, output and wages the threatened firm reacts less to unionization than its non-threatened counterpart. These results suggests that union policies have a stronger impact on firms than what the regression discontinuity approach suggests.

Outcome	Threatened firm	Non-threatened firm	Difference
Employment	0%	-4.16%	4.16%
Output	-6.82%	-9.48%	2.66%
Wages	-2.52%	-6.56%	4.04%

Table 7: Impact of unionization on threatened vs non-threatened firm

4.3 Additional Exercises

Policy and the non-monotone link between unionization and welfare

The policy exercises of the last section highlight the 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 6). Figure 7 emphasizes this point by showing how welfare reacts when the union bargaining power β_u and, through it, the strength of the threat varies. Two forces are at work to create this non-monotonicity. First, when β_u increases from its calibrated value, all workers in nonunion firms understand that their wage, if they were to unionize, would be higher. As a result, unionization is harder to prevent and firms must distort the skill composition of their workforce more heavily, which is inefficient. As β_u keeps increasing, there comes a point at which nonunion firms do not wish to prevent unionization anymore. Doing so would reduce profits so much that they prefer to let the workers unionized. This point is reached around $\beta_u \approx 0.4$ in Figure 7. As we can see from the figure, at this point, the unionization rate and the welfare level jump. Indeed, since the threat no longer distorts the

regression discontinuity estimator.

behavior of the newly unionized firms, efficiency is improved.⁵⁴

The mechanisms at work in Figure 7 have important consequences for policy design. In particular, any policy that slightly weakens the bargaining position of unions from its calibrated value — for instance the passage of a right-to-work law — is welfare improving as it lessens the distortion created by the threat. In contrast, increasing β_u by a sizable amount, say to 0.41, would also be welfare improving as the threat would then affect fewer firms. In practice, the optimal policy design should weigh the positive impact of increasing the threat (fewer firms are subject to it) against its negative impact (stronger distortion for remaining nonunion firms).

Overall, the bargaining power β_u that maximizes welfare in this economy is $\beta_u = 0$, at which point the unionization rate is 0%. This is in sharp contrast to the policy experiments summarized in Table 6 and which suggest that a fully unionized economy yields a higher welfare than a union free one. The key distinction between these welfare-maximizing unionization rate is the set of tools available to the policymaker. In the policy exercises of the last section, it is assumed that the policymaker can impose broad policy changes such as mandating or prohibiting unions without changing the parameters of the model.⁵⁵ In contrast, the exercise of Figure 7 assumes that the policymaker only has access to the bargaining power β_u to influence the economy. In this environment, pushing all firms to unionization requires a large β_u , which significantly lowers the share of the production surplus that firms retains. As a result, vacancy posting declines to a worst level in terms of efficiency (Hosios, 1990). Instead, setting the unionization to 0% can be accomplished by lowering β_u to 0. In this case, the union threat does not affect any firm (no worker wants to be part of a union) and the bargaining power that actually splits the joint surplus is β_n , which is set at a better level in terms of efficiency.

While the best policy depends on the instruments available to the policy maker, the optimal unionization rate is in general 0% or 100%. These levels are achieved when unions are very strong (only union firms) or very weak (only unthreatened nonunion firms). Any intermediate situation makes the union threat, and the inefficient distortion that it creates, active.

The decline of unionization in the United States

One question that arises naturally from the non-monotonicity between the unionization rate and efficiency is whether the large decline in unionization that happened in the U.S. over the last decades led to efficiency gains. In the 1970s and the 1980s, since the unionization rate was much higher, fewer firms might have been subject to the union threat. It is therefore possible that the distortion created by the threat was less important then. At the same time, it is unclear whether the effect of the threat on any given firm was stronger or weaker than in 2005. To answer these

⁵⁴In a model with a full distribution of firm types, welfare in Figure 7 would still be non-monotonic but the transition would be smoother as different types of firms would progressively let their workforce unionize.

⁵⁵The policy experiments of Section 4.2 can be generalized by letting the policymaker pick the union status of each initially nonunion firm exogenously, thereby implicitly setting the unionization rate. The optimal policy implies a 100% unionization rate. Note that, since the policymaker picks the union status of each firm, the union threat does not affect any decisions in this exercise.

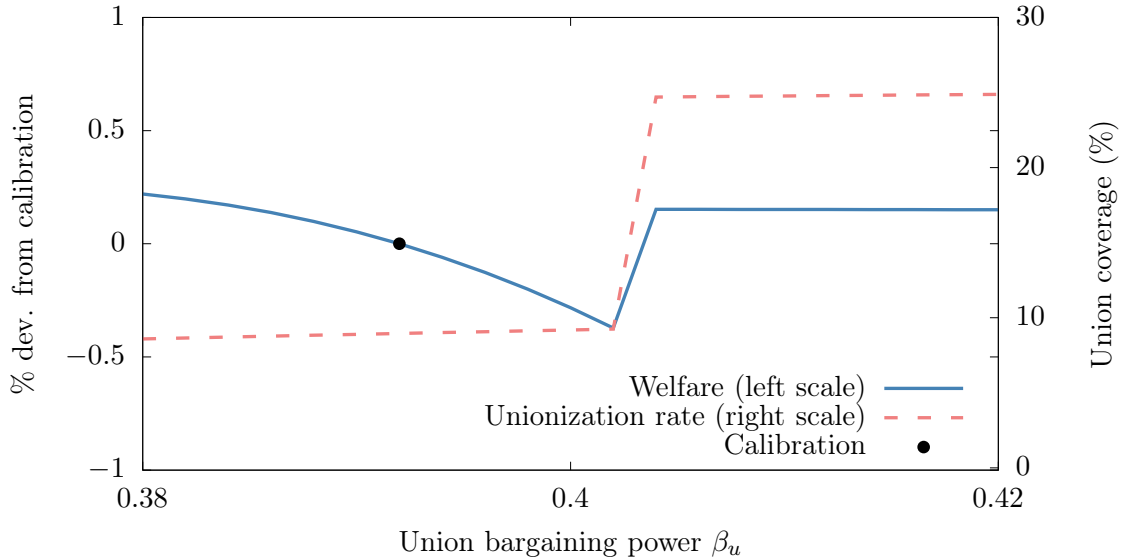


Figure 7: Nonlinear relationship between union coverage and welfare

questions, I estimate the model on the U.S. economy in 1983 and look at the impact of the threat on output, unemployment and welfare. The details of the exercise are in Appendix B.4.⁵⁶ Overall, the estimation finds that the impact of the union threat on output and welfare was about four times larger in 1983 than in 2005, while its impact on the 90-10 wage ratio was 175% larger. The moment in the data that pushes the estimation to this conclusion is the union wage gap, which was larger in 1983. The estimation therefore infers that the incentives for unionization were more important then and that the union threat was therefore stronger.

Robustness exercises

To evaluate the robustness of the benchmark policy experiments of Section 4.2, I include four additional exercises in Appendix B.3. First, I consider a model without capital such that the production function curvature parameters α are smaller. Second, I evaluate the impact of the experiments on an economy in which, instead of having a large fraction of nonunion firms exogenously immune to unionization, I assume that these firms are unaffected by the threat because of their higher curvature parameter α . Third, I reestimate the model by assuming that fewer firms are subject to the threat of unionization. Precisely, I go back to the NLRB micro data about the outcome of union elections and assume that a nonunion firm is subject to the threat if it won a union election by a margin of less than 10% of the eligible voters. Finally, I consider an environment in which the mass of firms in the economy adjusts after each policy experiment through a free-entry condition. In all

⁵⁶The key technology parameter that leads to a higher unionization rate in 1983 is the mass of firms in (relatively) capital intensive sectors. The share of employment in manufacturing, mining and transportation was higher in 1983. In contrast, the service industry (which is labor intensive) was much larger in 2005. Since, in the model (and in the data), labor intensive firms tend to be less unionized, this change in technology contributes to the change in the unionization rate between 1983 and 2005.

of these robustness exercises, the union threat as a substantial impact on the economy.

5 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 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.

Policy 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. Outlawing unions completely amplifies these effects. Forcing all firms to be unionized, on the other hand, reduces wage inequality substantially while still improving output, welfare and unemployment.

The theory could be used to evaluate the importance of the union threat in other countries and during different historical periods. In addition, the model could be estimated on each U.S. state and over time to better evaluate the impact of right-to-work laws on the economy. These projects are left to future research.

References

- AÇIKGÖZ, O. T. AND B. KAYMAK (2014): “The Rising Skill Premium and Deunionization,” *Journal of Monetary Economics*, 63, 37 – 50.
- ACEMOGLU, D., P. AGHION, AND G. L. VIOLANTE (2001): “Deunionization, technical change and inequality,” *Carnegie-Rochester Conference Series on Public Policy*, 55, 229–264.
- ACEMOGLU, D. AND W. B. HAWKINS (2014): “Search with multi-worker firms,” *Theoretical Economics*, 9, 583–628.
- ALVAREZ, F. AND M. VERACIERTO (2000): “Labor-Market Policies in an Equilibrium Search Model,” in *NBER Macroeconomics Annual 1999, Volume 14*, National Bureau of Economic Research, Inc, NBER Chapters, 265–316.
- 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, 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 *Restoring the promise of American labor law*, ed. by S. Friedman, R. Hurd, R. Oswald, and R. Seeber, IRL Press, 75–89.
- BRÜGEMANN, B., P. GAUTIER, AND G. MENZIO (2015): “Intra Firm Bargaining and Shapley Values,” Working Paper 21508, National Bureau of Economic Research.
- BURNSIDE, C., M. EICHENBAUM, AND S. REBELO (1995): *Capital Utilization and Returns to Scale*, MIT Press, 67–124.
- 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, 943–972.
- CAHUC, P. AND E. WASMER (2001): “Labor Market Efficiency, Wages and Employment when Search Frictions Interact with Intrafirm Bargaining,” *IZA Discussion Papers*.
- CARD, D. (1996): “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis,” *Econometrica*, 64, 957–979.
- (1998): “Falling Union Membership and Rising Wage Inequality: What’s the connection?” *National Bureau of Economic Research Working Paper Series*, No. 6520.
- (2001): “The Effect of Unions on Wage Inequality in the U.S. Labor Market,” *Industrial and Labor Relations Review*, 54, pp. 296–315.
- CARD, D., T. LEMIEUX, AND W. RIDDELL (2004): “Unions and wage inequality,” *Journal of Labor Research*, 25, 519–559.
- CORNEO, G. AND C. LUCIFORA (1997): “Wage formation under union threat effects: Theory and empirical evidence,” *Labour Economics*, 4, 265 – 292.
- DELACROIX, A. (2006): “A multisectorial matching model of unions,” *Journal of Monetary Economics*, 53, 573–596.
- DEN HAAN, W. J., G. RAMEY, AND J. WATSON (2000): “Job Destruction and Propagation of Shocks,” *American Economic Review*, 90, 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, pp. 560–575.
- (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*, Blackwell, 48–89.
- DINARDO, J. AND D. S. LEE (2004): “Economic Impacts of New Unionization on Private Sector Employers: 1984–2001,” *Quarterly Journal of Economics*, 119, 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, 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, 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, 1–48.
- ESIPOVA, N., A. PUGLIESE, AND J. RAY (2013): “381 Million Adults Worldwide Migrate Within Countries,” *Gallup*.
- FARBER, H. S. (1987): “The analysis of union behavior,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter and R. Layard, Elsevier, vol. 2 of *Handbook of Labor Economics*, chap. 18, 1039–1089.
- (2005): “Nonunion Wage Rates and the Threat of Unionization,” *Industrial and Labor Relations Review*, 58, 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, 349–369.
- FEATHERSTONE, L. (2004): “Under the gun, unions are realizing they have to think outside the big box.” *The Nation*.
- 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.
- FRANSEN, B. R. (2011): “Unionization and the Distribution of Employee Earnings: Evidence from Certification Elections,” Working paper.
- FREEMAN, R. (2007): “Do Workers Still Want Unions? Yes More than Ever!” in *Economics Policy Institute Policy Forum on Broadly Shared Prosperity, Work that Works.*, ePI BP182.
- FREEMAN, R. B. (1980): “Unionism and the Dispersion of Wages,” *Industrial and Labor Relations Review*, 34, 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, pp. 351–365.
- HALL, R. E. (2009): “Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor,” *Journal of Political Economy*, 117, 281–323.
- HIRSCH, B. T. (2004): “What Do Unions Do for Economic Performance?” *Journal of Labor Research*, 25, 415–456.
- HIRSCH, B. T. AND M. C. BERGER (1984): “Union Membership Determination and Industry Characteristics,” *Southern Economic Journal*, 50, 665–679.
- HIRSCH, B. T. AND D. A. MACPHERSON (2003): “Union Membership and Coverage Database from the Current Population Survey: Note,” *Industrial and Labor Relations Review*, 56, 349–354.
- 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, 138–148.
- HOLMES, T. J. (1998): “The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders,” *Journal of Political Economy*, 106, 667–705.
- HOSIOS, A. J. (1990): “On the Efficiency of Matching and Related Models of Search and Unemployment,” *The Review of Economic Studies*, 57, 279–298.
- KAHN, L. M. AND M. CURME (1987): “Unions and Nonunion Wage Dispersion,” *The Review of Economics and Statistics*, 69, 600–607.
- KLEINER, M. M. (2001): “Intensity of management resistance: understanding the decline of unionization in the private sector,” *Journal of Labor Research*, 22, 519–540.
- KRISHNA, V. AND R. SERRANO (1996): “Multilateral Bargaining,” *Review of Economic Studies*, 63, 61–80.
- KRUEGER, A. B. AND B. D. MEYER (2002): “Labor supply effects of social insurance,” in *Handbook of Public Economics*, ed. by A. J. Auerbach and M. Feldstein, Elsevier, vol. Volume 4, 2327–2392.

- KRUSELL, P., L. E. OHANIAN, J.-V. RÍOS-RULL, AND G. L. VIOLANTE (2000): “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, 68, pp. 1029–1053.
- KRUSELL, P. AND L. RUDANKO (2012): “Unions in a Frictional Labor Market,” Working Paper 18218, National Bureau of Economic Research.
- 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, 333–378.
- LUCAS, R. E. (1978): “On the size distribution of business firms,” *The Bell Journal of Economics*, 508–523.
- 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*.
- 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, pp. 20–38.
- NICKELL, S. AND R. LAYARD (1999): “Chapter 46 Labor market institutions and economic performance,” in *Handbook of Labor Economic*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. Volume 3, Part 3, 3029–3084.
- PISSARIDES, C. A. (1986): “Trade Unions and the Efficiency of the Natural Rate of Unemployment,” *Journal of Labor Economics*, 4, pp. 582–595.
- PRIES, M. AND R. ROGERSON (2005): “Hiring Policies, Labor Market Institutions, and Labor Market Flows,” *Journal of Political Economy*, 113, 811–839.
- ROSEN, S. (1969): “Trade Union Power, Threat Effects and the Extent of Organization,” *The Review of Economic Studies*, 36, 185–196.
- ROTH, A. E. (1979): *Axiomatic Models of Bargaining*, Springer-Verlag.
- ROWE, S. (2016): “The Impact of Right-to-Work Laws on Wages and Employment,” Ph.D. thesis, University of Maryland, Ann Arbor.
- SCHMITT, J. AND B. ZIPPERER (2007): “Dropping the Ax: Illegal Firings During Union Election Campaigns,” CEPR Reports and Issue Briefs 2007-01, Center for Economic and Policy Research (CEPR).
- SILVA, J. I. AND M. TOLEDO (2009): “Labor Turnover Costs And The Cyclical Behavior Of Vacancies And Unemployment,” *Macroeconomic Dynamics*, 13, 76–96.
- STOLE, L. A. AND J. ZWIEBEL (1996a): “Intra-Firm Bargaining under Non-Binding Contracts,” *Review of Economic Studies*, 63, 375–410.

- (1996b): “Organizational Design and Technology Choice under Intrafirm Bargaining,” *American Economic Review*, 86, 195–222.
- TRAXLER, F. (1994): “Collective Bargaining: Levels and Coverage,” in *Employment Outlook 1994*, OECD, 167–194.
- VERNA, A. (2005): “What Do Unions Do to the Workplace? Union Impact on Management and HRM Policies,” *Journal of Labor Research*, 26, 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, pp. 576–591.

Online Appendices

A Additional Analytical Exercises

A.1 Including capital in the model

This section shows how including capital in the benchmark model is equivalent to relabelling the parameters. Consider a firm j with production function

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

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 variable as in (1). Assuming that the firm has access to capital at a constant rental rate of $r > 0$ and taking the first-order condition on capital we can write steady-state profits as

$$(1 - (1 - \gamma_j) \omega_j) \tilde{A}_j^{\frac{1}{1-(1-\gamma_j)\omega_j}} \left(\frac{(1 - \gamma_j) \omega_j}{r} \right)^{\frac{\omega_j(1-\gamma_j)}{1-(1-\gamma_j)\omega_j}} L_j^{\frac{\gamma_j \omega_j}{1-(1-\gamma_j)\omega_j}} - \sum_{s \in \mathcal{S}} g_{j,s} w_{j,s} - \sum_{s \in \mathcal{S}} \kappa \frac{\delta g_{j,s}}{q(\theta_j)} \quad (24)$$

so that this formulation is equivalent to the one in the body of the text if we define A_j and α_j as

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

A.2 Alternative wage setting protocol

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

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

Consider first an environment without any possibility of unionization. The firm optimally sets nonunion wages to the worker's reservation value of 0 in this simplified environment. In which case, the firm's problem becomes

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

and the first-order conditions yield

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

Since all workers would earn more under collective bargaining g_s^{n*} is not optimal in this case and the firm must react to prevent unionization. To do so, the optimal strategy is to pay high-skill workers 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

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

Notice that, under our assumptions, the firm now hires relatively more high-skill workers than without the threat: $g_h^{n*}/g_l^{n*} < g_h^n/g_l^n$. We can also compare output in both cases and find that $F(\mathbf{g}^{n*}) > F(\mathbf{g}^n)$ if

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

which is true since $\left(1 - \frac{\beta_u}{2}\right)^{\frac{\alpha}{1-\alpha}} < 1$ and because of the inequality of arithmetic and geometric means.

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

$$\frac{MP_h^n}{MP_l^n} = \frac{z_h}{z_l} < \frac{q(\theta_l)}{q(\theta_h)} = \frac{MP_h^{n*}}{MP_l^{n*}}.$$

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

Overall, these predictions are consistent with the bargaining wage setup of the model as shown by Propositions 5 and 6. The predictions of the current model are therefore robust to this alternative wage setting mechanism.

A.3 Alternative procedures for the collective bargaining

This appendix solves two alternative bargaining procedures for the union wages. I introduce a 'union' 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 -person 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.

Suppose that, in this second stage, the firm bargains with a risk-neutral union. The bargaining problem is

$$\max_T T^\beta (F(\mathbf{g}) - T)^{1-\beta}$$

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 $T(g) = \beta F(g)$.

In stage 1, the workers have to split this surplus among them. There are two ways to model this. 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.

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. 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

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

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 \(1996a\)](#) and [Brügemann et al. \(2015\)](#) bargaining. This problem is very similar to the one encountered with the nonunion individual bargaining. We need to solve the following system of equations:

$$\frac{\partial T(\mathbf{g})}{\partial g_s} - w_s(\mathbf{g}) - \sum_k g_k \frac{\partial w_k(\mathbf{g})}{\partial g_s} = \frac{1-\epsilon}{\epsilon} w_s$$

where ϵ is the bargaining power of the workers. The solution to this system is very similar to the one of the individual bargaining (I'm using the same boundary conditions):

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

We see that, under these assumptions, this new union wage schedule is structurally identical to nonunion wages. In particular, inequality in log-wages is the same whether the firm is unionized or not, for a given distribution of workers. Therefore, this model does not seem appropriate to discuss the union-generated wage compression observed in the data (Frandsen, 2011). Furthermore, if we assume that the union leader gets a negligible share of the surplus ($\epsilon \rightarrow 1$), all workers would always vote for in favor of unionization. There would be no union free firms in the economy.

Collective bargaining between all workers and the union

We now consider the other way of splitting the surplus extracted from the firm. Here all workers and the union negotiate in a single bargaining session. Under our assumptions, this problem is

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

where again $0 < \epsilon < 1$ denotes the bargaining power of the workers. This problem is very similar to the actual union bargaining solved in the core of the paper except that we are using the union surplus $T(\mathbf{g})$ instead of the production function $F(\mathbf{g})$. But, since the second stage bargaining yielded $T(\mathbf{g}) = \beta F(\mathbf{g})$, this difference is minimal. In fact, we find that the wage with this procedure is

$$w_s(\mathbf{g}) = \frac{\epsilon}{n} \beta F(\mathbf{g}).$$

If we send the bargaining of the workers with the union to $\epsilon \rightarrow 1$, we find exactly the union wage equation from the core of the text (if we impose the same simplifying assumptions there too). This goes to show that adding an actual union as an intermediary between the workers and the firms does not affect the structure of wages.

A.4 Welfare analysis

To better understand the welfare results of Section 4.2 let us consider the problem of a social planner that seeks to maximize the steady-state welfare level in this economy. To keep the analysis tractable while keeping active the key mechanisms that affect welfare, I set $\delta = 1$.

The planner solves

$$\max_{\{g_{j,s}\}} \sum_{j \in \mathcal{J}} F_j(\mathbf{g}_j) + \sum_{j \in \mathcal{J}} \left(\sum_{s=1}^S (N_s - g_{j,s}) b_s \right) - \sum_{j \in \mathcal{J}} \left(\sum_{s=1}^S g_{j,s} \kappa \frac{1}{q(\theta_s)} \right)$$

where $g_{j,s}$ is the measure of employed workers of skill s in firm j . The first term is the total amount of output produced, the second term corresponds to home production and the last term refers to the hiring costs paid at the steady state. This maximization problem is subject to a steady-state

link between \mathbf{g} and $\boldsymbol{\theta}$. By adapting (15), the steady-state level of θ_s is

$$\theta_s \equiv \frac{V_s}{U_s} = \frac{1}{N_s} \sum_{j \in \mathcal{J}} \frac{g_{j,s}}{q(\theta_s)}.$$

Taking the first-order condition and simplifying we find the following optimally condition for the planner's problem:

$$\frac{\partial F_j}{\partial g_{j,s}} - b_s - \frac{\kappa}{p'(\theta_s)} = 0. \quad (25)$$

Let us now compare this equation with its equilibrium equivalent. Under our assumptions, $c_{j,s} = b_s$ and thus

$$\text{MC}_s^j = (1 - \beta_j) b_s + \frac{\kappa}{q(\theta_s)}.$$

The first-order condition of the firm (20) is then

$$B_j \frac{\partial F_j}{\partial g_{j,s}} - (1 - \beta_j) b_s - \frac{\kappa}{q(\theta_s)} - \lambda \frac{\partial V(g_j)}{\partial g_{j,s}} = 0. \quad (26)$$

where, as before, $\lambda > 0$ is the Lagrange multiplier associated with the union threat.

Comparing the planner's first-order condition (25) with firm j 's first-order condition in the equilibrium (26) we can identify the three sources of inefficiencies in the environment. First, without the union threat ($\lambda = 0$) and under collective bargaining ($B_i = 1 - \beta_u$) we see that the equilibrium is efficient if and only if $\beta_u = -\frac{\theta q'(\theta)}{q(\theta_s)}$ which is the usual [Hosios \(1990\)](#) condition. When, instead, wages are bargained individually between the workers and the firm, we have that $B_j \neq (1 - \beta_n)$ and an additional wedge is introduced. Under this form of bargaining, the firm over-hires workers to lower their marginal product, leading to an inefficiency ([Stole and Zwiebel, 1996a](#)). Finally, the threat ($\lambda > 0$) inefficiently distorts the type of workers used into production.

B Additional Quantitative Exercises

B.1 Parameters for the simulations in partial equilibrium

Table 8 contains the parameters used for the partial equilibrium simulations of Figure 3 as well as for Table 9 below, while Table 9 shows some characteristics of the firm under the three following scenarios:

1. Exogenously unionized: the firm hires according to \mathbf{g}^{u*} .
2. Endogenous union status: the firm compares profits under \mathbf{g}^n and \mathbf{g}^{u*} . In this example, profits are maximized under \mathbf{g}^n .
3. Exogenously union free: the firm hires according to \mathbf{g}^{n*} .

Definition	Parameter	Value
Number of skills	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	1/2
Union bargaining power	β_u	1/2
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

These scenarios can be thought of as policy environments in which unions are mandatory, allowed or forbidden. Table 9 shows how these policies affects the firm size, its profit and the wages it pays.

	1. Exogenously unionized	2. Endogenous union status	3. Exogenously union free
Union status of the firm	Union	Nonunion	Nonunion
Union threat affects decisions	No	Yes	No
Firm discounted profit ($\times 10^4$)	0.9	1.4	1.5
Mass of workers	108	151	185
Fraction of voters for union	66%	50%	66%
Mean of wages	22	20	18
Standard deviation of wages	1.2	7.1	9.4

Table 9: A firm under three union policies

Let us first compare the firm under scenarios 1 and 3. Consistent with Propositions 1 and 2, profits are higher and wages are lower and more dispersed when the firm is union free. Also, the firm is larger under policy 3 as it hires more workers to lower their marginal products, an important determinant of wages under individual bargaining.

Under policy 2, the firm is subject to the union threat. In this example, the best nonunion measure of workers is more profitable than its best union counterpart, and the firm is therefore union free. While the union status of the firm is the same under policies 2 and 3, the voting constraint has important implications for wages and employment. First, it increases the cost of producing an extra unit of goods. Since the firm wants to equalize the marginal cost and the marginal revenue of production, the constraint pushes the firm to hire fewer workers to increase their marginal products. In turn, this increase in marginal products raises the average wage paid to the workers. Second, the

constraint pushes the firm to hire more high-skill workers and fewer low-skill ones, which leads to a decline in wage dispersion. These results are consistent with Propositions 5 and 6.

B.2 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 6, 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 z .

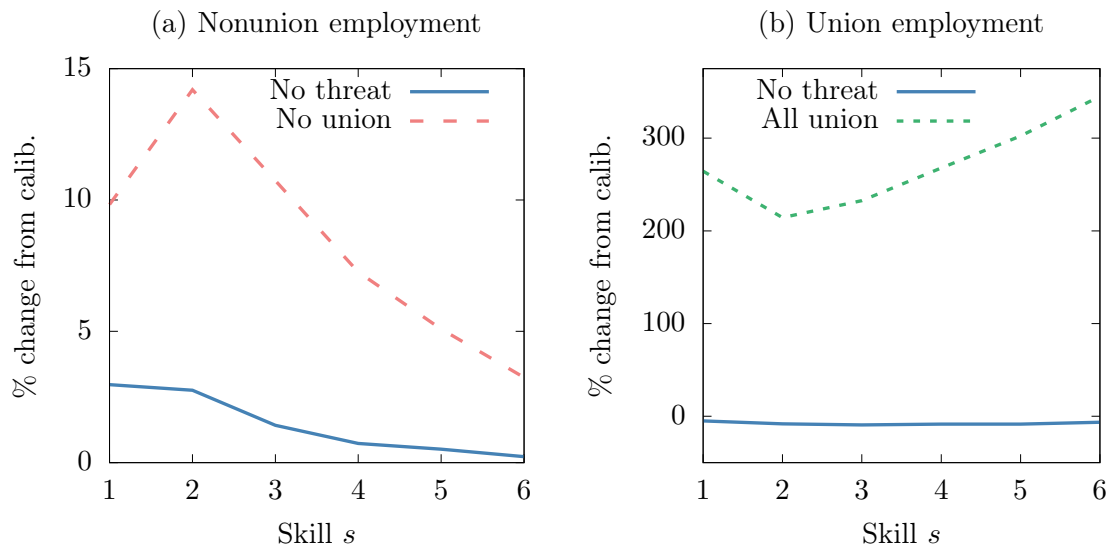


Figure 8: Impact of the policy changes on union and nonunion workers

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.

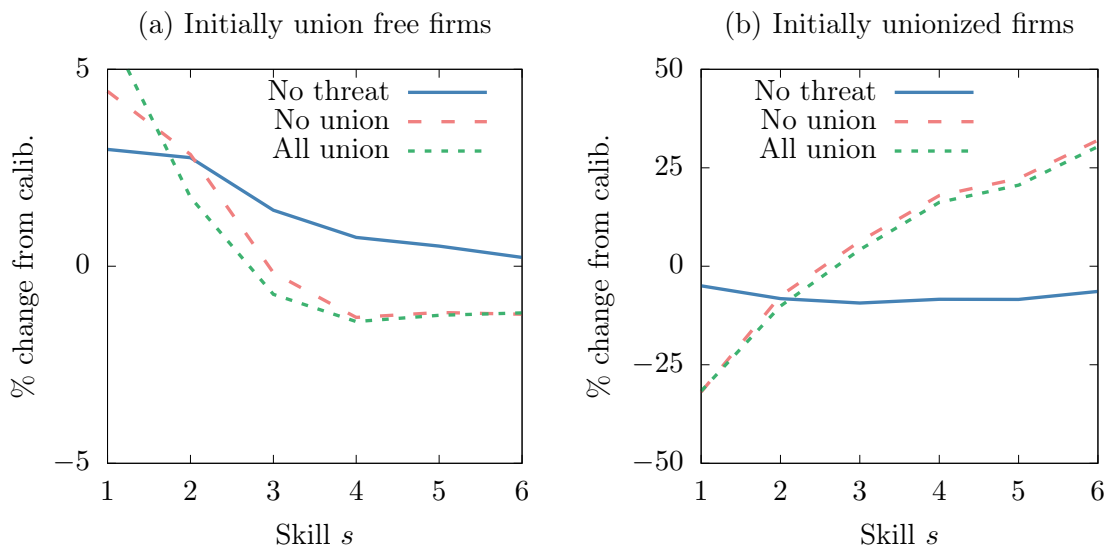


Figure 9: Impact of the policy changes on employment as a function of the technology of each firm

B.3 Robustness exercises

This section contains various additional numerical exercises to show the robustness of the quantitative results of Section 4. In all cases, the model is completely re-estimated using the new specification and the policy exercises are conducted again.

Model without capital

The results of Section 4 assume that firms use capital into production. Here, I consider an alternative specification of the model in which firms only use labor to produce. The estimated parameters are shown in Table 10. They are similar to the benchmark estimation except that α_n and α_u are set to lower values to match the labor shares in the data without the presence of capital. The bargaining power also take smaller values to compensate the change creates by the new α 's on the wages. The outcome of the policy experiments are reported in Table 11. We see that the results are broadly similar to the benchmark calibration. The union threat has however a larger impact on the unemployment rate and a somewhat smaller impact on output, welfare and wage inequality.

Definition	Parameter	Calibrated value
Nonunion bargaining power	β_n	0.43
Union bargaining power	β_u	0.25
Curvature of u technology	α_u	0.48
Mean of skill intensity z_u	μ_u	1.19
Variance of skill intensity z_u	ξ_u^2	0.56
Curvature of n technology	α_n	0.52
Mean of skill intensity z_n	μ_n	1.49
Variance of skill intensity z_n	ξ_n^2	0.79

Table 10: Estimated parameters in the model without capital

	No threat	No union	All unions
Output	0.7%	0.8%	0.7%
Unemployment rate	-1.6 pp	-1.6 pp	-0.4 pp
Welfare	0.1%	0.1%	0.7%
Var(log wages)	0.5%	6.4%	-50.0%
90-10 ratio log wages	1.1%	1.2%	-8.5%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 11: Impact of policy experiments in the model without capital

Endogenously unaffected by the union threat

The quantitative results of Section 4 assume that only a fraction of the firms using technology n are subject to the threat of unionization. Instead of assuming that the rest of the n -technology firms are immune to unionization, I now assume that their curvature parameter α is high enough that the threat is endogenously not binding. This is achieved by setting $\alpha = 0.81$. The rest of the parameters are unchanged. This economy still fits the data well with only a small difference in the loss function compared to the benchmark economy.

Table 12 shows how the policy experiments affect the economy in this case. As we can see, the threat accounts for most of the action in terms of output, welfare and the unemployment rate while the actual union status of firms is a key determinant of wage inequality.

	No threat	No union	All unions
Output	0.8%	0.8%	1.0%
Unemployment rate	-1.1 pp	-1.1 pp	-1.4 pp
Welfare	0.1%	0.2%	0.2%
Var(log wages)	0.3%	5.1%	-49.1%
90-10 ratio log wages	1.2%	1.6%	-8.1%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 12: Impact of policy experiments when some firms are endogenously immune to threat

Fewer nonunion firms affected by the union threat

This section proposes an alternative calibration of the model with fewer nonunion firms that are subjected to the union threat. Going back to the union election data from the NLRB, the assumption is now that only firms that won the union election with less than a margin of 10% of the vote are subject to the union threat.⁵⁷ Table 13 shows the estimated parameters and Table 14 shows the impact of the policy experiments. The estimated parameters are close to those of the benchmark model. The policy experiments have a somewhat smaller impact on output, welfare and unemployment, while their impact on wage inequality is similar.

Definition	Parameter	Calibrated value
Nonunion bargaining power	β_n	0.50
Union bargaining power	β_u	0.39
Curvature of u technology	α_u	0.72
Mean of skill intensity z_u	μ_u	1.18
Variance of skill intensity z_u	ξ_u^2	0.57
Curvature of n technology	α_n	0.73
Mean of skill intensity z_n	μ_n	1.50
Variance of skill intensity z_n	ξ_n^2	0.80

Table 13: Estimated parameters when fewer firms are affected by the threat

Model with free-entry of firms

This section replicates the policy 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.

⁵⁷From the NLRB data, the number of employees in 2005 in firms in which the union lost the election by less than 10% of eligible voters was 50% of the number of workers who unionized.

	No threat	No union	All unions
Output	0.5%	0.5%	0.5%
Unemployment rate	-0.6 pp	-0.6 pp	-0.5 pp
Welfare	0.1%	0.2%	0.4%
Var(log wages)	0.3%	5.2%	-48.4%
90-10 ratio log wages	1.4%	0.7%	-8.2%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 14: Impact of policy experiments when fewer firms are affected by the threat

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 $j \in \{u, n\}$. The probability of being assigned to a given type is proportional to the mass of firms of this type in the calibrated economy. 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.⁵⁸

The estimated parameters are identical to those of the benchmark exercise of Section 4. Table 15 shows the outcome of the policy 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, 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.

	No threat	No union	All unions
Output	+3.2%	3.3%	+2.4%
Unemployment rate	-0.5 pp	-0.5 pp	-0.7 pp
Welfare	+2.6%	+2.7%	+1.9%
Var(log wages)	+0.7%	+3.7%	-48.9%
90-10 ratio log wages	+1.3%	+1.4%	-8.1%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 15: Impact of policy experiments under free-entry of firms

B.4 Calibration on the 1983 United States

I estimate the model on the United States economy in 1983, the earliest year for which the CPS data about unionization is consistent with that used in the 2005 exercise.⁵⁹ Most of the parameters

⁵⁸An alternative specification in which each firm type enters separately gives rise to multiple equilibria.

⁵⁹Union coverage data was available from 1973 to 1981 from the May Population Survey instead of the Merged Outgoing Rotation Groups of the Current Population Survey. No union questions were asked in the 1982 CPS.

taken from the literature or directly from the data are the same as in the benchmark quantitative exercise with a few exceptions. Since the aggregate labor share was higher in 1983 by about 5%, I increase the targeted labor shares by that amount. The cost of posting a vacancy is adjusted to $\kappa = 1.33$ so that the cost of hiring a worker equals to 69% of quarterly wage, as in the benchmark calibration. The Annual Report of the NLRB for 1983 mentions that 43% of union elections were successful during that year. As in the benchmark calibration, I use this number to identify the mass of nonunion firms that are immune to the threat. The estimated parameters are listed in Table 16.

Definition	Parameter	Calibrated value
Nonunion bargaining power	β_n	0.48
Union bargaining power	β_u	0.38
Curvature of u technology	α_u	0.74
Mean of skill intensity z_u	μ_u	1.02
Variance of skill intensity z_u	ξ_u^2	0.57
Curvature of n technology	α_n	0.76
Mean of skill intensity z_n	μ_n	1.83
Variance of skill intensity z_n	ξ_n^2	0.97

Table 16: Estimated parameters for the 1983 U.S. economy

I repeat the three policy exercises of Section 4 on this economy. The results are shown in Table 17. The impact of the union threat is overall much more important in 1983 than in 2005.

	No threat	No union	All unions
Output	4.2%	4.2%	4.3%
Unemployment rate	-4.2 pp	-4.3 pp	-4.1 pp
Welfare	1.3%	1.3%	1.5%
Var(log wages)	15.5%	24.4%	-32.6%
90-10 ratio log wages	2.1%	3.2%	-3.2%

Notes: Differences from the calibrated economy. The numbers for output, welfare, var(log wages) and the 90-10 ratio represent percentage change. The unemployment rate number is the difference in percentage point. Output is measured as value added.

Table 17: Impact of policy experiments for the 1983 U.S. economy

C Proofs

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

$$\max_{\mathbf{g}} \pi(\mathbf{g}) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)} + \kappa(1 - \delta)\gamma \sum_{s \in \mathcal{S}} \frac{g_s}{q(\theta_s)}. \quad (27)$$

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.⁶⁰ 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

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

The term that is maximized is constant with respect to \mathbf{g}_{-1} . Denote that constant by B . Then, in particular

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

and the firm solves

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

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

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

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

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

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 \mathbf{g} is fixed and the hiring cost is sunk. In a steady state, the difference in discounted profits for the firm,

⁶⁰This does not happen in equilibrium but the value function needs to be defined along these paths to correctly specify the bargaining problems.

denoted by $\Delta^u(w)$, is

$$\Delta^u(\mathbf{w}) = [\pi(\mathbf{g}, \mathbf{w}) + \gamma J(\mathbf{g})] - [\pi(\mathbf{0}) + \gamma J(\mathbf{0})] \quad (28)$$

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,

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

where w^* is the equilibrium wage schedule for this firm. We can therefore rewrite equation (28) as

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

But the firm's value function is

$$J(\mathbf{g}) = \pi(\mathbf{g}^*, \mathbf{w}^*) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s^* - (1 - \delta)g_s}{q(\theta_s)} + \gamma J(\mathbf{g}^*) \quad (29)$$

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

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

On the workers' side, the net benefit of an agreement is $W_s^e(w) - b_s - \gamma W_s^u$. 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 $\epsilon > 0$ such that $h_s \times \epsilon \rightarrow g_s$ as we move to the continuum. The bargaining problem with *equal* bargaining power is

$$(W_1^E - b_1 - \gamma W_1^U)^{h_1} \times \dots \times (W_i^E - b_i - \gamma W_i^U)^{h_i} \times \dots \times (W_S^E - b_S - \gamma W_S^U)^{h_S} \times \Delta^u.$$

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

$$(W_1^E - b_1 - \gamma W_1^U)^{\frac{\beta_u \epsilon h_1}{\epsilon H}} \times \dots \times (W_S^E - b_S - \gamma W_S^U)^{\frac{\beta_u \epsilon h_S}{\epsilon H}} \times (\Delta^u)^{1 - \beta_u}$$

where $H = \sum_s h_s$. Taking the limit to the continuum, $\frac{\epsilon h_i}{\epsilon H} \rightarrow \frac{g_i}{n}$ 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

$$w_s^n(\mathbf{g}) - c_s = \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \frac{\alpha z_s}{g_s^{1/\sigma}} A \left(\sum_{k \in \mathcal{S}} z_k g_k^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1-\sigma(1-\alpha)}{\sigma-1}} - \beta_n c_s + \beta_n \gamma (1 - \delta) \frac{\kappa}{q(\theta_s)} \quad (30)$$

solves the individual bargaining problem.

Proof. The [Stole and Zwiebel \(1996a,b\)](#) solution to the bargaining problem is the wage function that gives the worker a share β_n of the joint surplus. The bargaining takes place when all vacancies have been posted and the vacancy costs are therefore sunk. When bargaining with a single worker, the firm compares two scenarios. Either an agreement is reached, in which case production takes place as planned, or the negotiations break down and the firm produces without this individual worker. In this last case, that worker departs from the firm and additional vacancies will have to be posted in the next period for the firm to go back to its optimal measure of workers. In equilibrium, an agreement is always reached.

To solve the problem, assume that each worker has size h . We will take the limit as $h \rightarrow 0$. The marginal discounted profit from hiring a worker of type s is proportional to

$$\begin{aligned} \Delta_s^n(\mathbf{w}) = & F(\mathbf{g}) - \sum_{k \in \mathcal{S}} w_k(\mathbf{g}) g_k - \left(F(\dots, g_s - h, \dots) - \sum_{k \neq s} w_k(\dots, g_s - h, \dots) g_k \right. \\ & \left. - w_s(\dots, g_s - h, \dots)(g_s - h) - h\gamma(1 - \delta) \frac{\kappa}{q(\theta_s)} \right) \end{aligned}$$

where the notation $(\dots, g_s - h, \dots)$ makes explicit the fact that we are considering the measure \mathbf{g} without a mass h of its s th member. Δ_s^n is simply the difference between the value of the firm with and without an agreement. Notice that in the latter case, the firm loses value since it faces an additional hiring cost in the next period to get back to its equilibrium size.

A solution to the individual bargaining is a wage vector w that solves

$$\frac{\beta_n}{1 - \beta_n} \Delta_s^n(\mathbf{w}) = (W_s^E(w) - W_s^U)h$$

where the right-hand side is the worker's surplus. By dividing Δ_s^n by h and taking the limit $h \rightarrow 0$, we get

$$\lim_{h \rightarrow 0} \frac{\Delta_s^n(\mathbf{w})}{h} = \frac{\partial F(\mathbf{g})}{\partial g_s} - \sum_{k \in \mathcal{S}} g_k \frac{\partial w_k(\mathbf{g})}{\partial g_s} - w_s(\mathbf{g}) + \gamma(1 - \delta) \frac{\kappa}{q(\theta_s)}.$$

Therefore, a solution must solve the following system of partial differential equations:

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

for all $s \in \mathcal{S}$. General solutions to this system are of the form

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

where C_s is a constant term that could depend on $\{g_j\}_{j \neq s}$. To fix the constants, I use the convenient boundary conditions

$$\left\{ \lim_{g_s \rightarrow 0} w_s^n(\mathbf{g}) g_s = 0 \right\}_{s=1}^S$$

which guarantees that $C_s = 0$ for all s .⁶¹ □

Proposition 1. If the bargaining powers are equal ($\beta \equiv \beta_n = \beta_u$) than the difference between the average of nonunion and union wages is

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

where \mathbb{E}_g is the expectation across skills. Equivalently, the difference between nonunion and union profits is

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

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

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

and

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

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

Proposition 2. The equilibrium wage schedules $w_s^u(\mathbf{g}^{u*})$ and $w_s^n(\mathbf{g}^{n*})$ are increasing in s and the

⁶¹Cahuc et al. (2008) study a similar bargaining problem with general production functions.

union wage gap $w_s^u(\mathbf{g}^{u*}) - w_s^n(\mathbf{g}^{n*})$ is decreasing in s .

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

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

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

Using equation (7) once again yields

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

Since W_s^U and θ_s are increasing in s so is $w_s^n(\mathbf{g}^{n*})$.

For the union wage gap, notice that

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

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

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

Proof. We first start with the unionized firm. This firm hires according to g_s^{u*} . From lemma 2, we know that $w_s^u(\mathbf{g}^{u*}) = c_s^u + D$ 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

$$w_s^n(\mathbf{g}^{u*}) = (1 - \beta_n)c_s^u + \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \frac{MC_s^u}{1 - \beta_u} + \beta_n\gamma(1 - \delta) \frac{\kappa}{q(\theta_s)}.$$

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

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

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

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

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

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

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

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

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

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

Proposition ??. An unconstrained firm prefers to be union free if and only if

$$\frac{B_n}{B_u} \left(\frac{\sum_{s \in \mathcal{S}} z_s^\sigma (\text{MC}_s^u)^{1-\sigma}}{\sum_{s \in \mathcal{S}} z_s^\sigma (\text{MC}_s^n)^{1-\sigma}} \right)^{\frac{\alpha}{\sigma-1}}.$$

Proof. We need to compare the value of a firm under g^{u*} and g^{n*} , for a given vector c_s . From Lemma 1, we know that we can compare

$$\pi(\mathbf{g}^{i*}) - \kappa(1 - (1 - \delta)\gamma) \sum_{s=1}^S \frac{g_s^{i*}}{q(\theta_s)}.$$

Using equations (10) and (13) together with equation (19) and after simplification we can now compare

$$(AB_i)^{\frac{1}{1-\alpha}} \alpha^{\frac{\alpha}{1-\alpha}} (1 - \alpha) \left(\sum_{s \in \mathcal{S}} z_s \left(\frac{z_s}{\text{MC}_s^i} \right)^{\sigma-1} \right)^{\frac{\alpha}{(1-\alpha)(\sigma-1)}}$$

for $i \in \{u, n\}$. A few simplifications yield the result. \square

Proposition 4. If $B_n > B_u$ and $z_h q(\theta_h) < z_l q(\theta_l)$ then the union threat is binding in nonunion firms.

Proof. Under the assumptions made in this section $c_s = 0$ for all s , $\text{MC}_s^n = \kappa/q(\theta_s)$ and $F(\mathbf{g}) = (g_l^{z_l} g_h^{z_h})^\alpha$. In addition, the nonunion wage is

$$\begin{aligned} w_s^n(\mathbf{g}) &= \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \frac{\partial F(\mathbf{g})}{\partial g_s} \\ &= \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \alpha F(\mathbf{g}) \frac{z_s}{g_s} \end{aligned}$$

where the $-s$ notation refers to the skill that is not s . We can rewrite this equation as

$$g_s w_s^n(\mathbf{g}) = \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \alpha F(\mathbf{g}) z_s$$

and the union wage is

$$w_s^u(\mathbf{g}) = \frac{\beta_u}{g_l + g_h} (g_l^{z_l} g_h^{z_h})^\alpha. \quad (31)$$

Consider the problem of a nonunion unconstrained firm:

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

Simplifying, we can write the Lagrangian as

$$\mathcal{L} = \frac{1 - \beta_n}{1 - (1 - \alpha)\beta_n} (g_l^{z_l} g_h^{z_h})^\alpha - \kappa \left(\frac{g_h}{q(\theta_h)} + \frac{g_l}{q(\theta_l)} \right)$$

and the first-order condition are

$$\begin{aligned} \frac{1 - \beta_n}{1 - (1 - \alpha)\beta_n} \alpha F(\mathbf{g}) \frac{z_l}{g_l} &= \frac{\kappa}{q(\theta_l)} \\ \frac{1 - \beta_n}{1 - (1 - \alpha)\beta_n} \alpha F(\mathbf{g}) \frac{z_h}{g_h} &= \frac{\kappa}{q(\theta_h)} \end{aligned} \quad (32)$$

yield $\frac{z_l}{g_l} q(\theta_l) = \frac{z_h}{g_h} q(\theta_h)$ which implies that $g_l > g_h$ given our assumption that $z_l q(\theta_l) > z_h q(\theta_h)$.

Now consider the union wage gap for low-skill workers

$$w_l^u - w_l^n = \frac{\beta_u}{g_l + g_h} F(\mathbf{g}) - \frac{\beta_n}{1 - (1 - \alpha)\beta_n} \alpha F(\mathbf{g}) \frac{z_l}{g_l} = \frac{\kappa}{\frac{1 - \beta_n}{1 - (1 - \alpha)\beta_n} \alpha q(\theta_l)} \left(\frac{\beta_u}{z_l + z_h \frac{q(\theta_h)}{q(\theta_l)}} - \frac{\alpha \beta_n}{1 - (1 - \alpha)\beta_n} \right)$$

such that the union wage gap is positive if

$$\beta_u > \frac{\alpha \beta_n}{1 - (1 - \alpha)\beta_n} \left(z_l + z_h \frac{q(\theta_h)}{q(\theta_l)} \right)$$

which is true under our assumptions that $B_n > B_u$, $z_l q(\theta_l) > z_h q(\theta_h)$ and $\theta_h > \theta_l$. Since the majority of the voters have a higher wage if the firm unionizes, this nonunion firm is threatened and the union vote is a binding constraint in its optimization problem \square

Proposition 5. Under the same assumptions as Proposition 4, the union threat lowers the profits, employment and output of nonunion firms.

Proof. We first characterize the hiring decisions \mathbf{g}^n and \mathbf{g}^{n*} . From (32) we have that

$$\frac{1 - \beta_n}{1 - (1 - \alpha) \beta_n} \alpha F(\mathbf{g}^{n*}) z_l \frac{q(\theta_l)}{\kappa} = g_l^{n*}$$

$$\frac{1 - \beta_n}{1 - (1 - \alpha) \beta_n} \alpha F(\mathbf{g}^{n*}) z_h \frac{q(\theta_h)}{\kappa} = g_h^{n*}$$

and

$$F(\mathbf{g}^{n*}) = \left(\frac{1 - \beta_n}{1 - (1 - \alpha) \beta_n} \alpha \right)^{\frac{\alpha}{1 - \alpha}} \left(\left(z_l \frac{q(\theta_l)}{\kappa} \right)^{z_l} \left(z_h \frac{q(\theta_h)}{\kappa} \right)^{z_h} \right)^{\frac{\alpha}{1 - \alpha}}.$$

Turning to \mathbf{g}^n , the problem of the firm is

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

subject to $g_h \geq g_l$ and $w_h^n(\mathbf{g}) \geq w_h^u(\mathbf{g})$.⁶² If the threat is binding then at least of these two constraint binds. Suppose that the first constraint binds ($g_h = g_l = g$), then the optimal hiring decision is

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

and the union wage gap is

$$w_s^u(\mathbf{g}) - w_s^n(\mathbf{g}) = \left(\frac{1}{2} \beta_u - \frac{\alpha \beta_n}{1 - (1 - \alpha) \beta_n} z_s \right) g^{\alpha - 1} \quad (33)$$

so that the inside of the parenthesis must be negative for the firm to successfully prevent unionization.⁶³

We can now compare the firm outcomes when the threat does and does not bind. It is obvious that the voting constraint lowers profits. For output, simply plugging into the production function, we are interested in showing

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

By the means inequality, this inequality is always satisfied. For employment,

$$\sum_{s \in \mathcal{S}} g_s^{n*} \geq \sum_{s \in \mathcal{S}} g_s^n$$

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

⁶²One can show that the two other scenarios under which the firm is union free ($w_l^n \geq w_l^u$ and $g_l \geq g_h$, or $w_h^n \geq w_h^u$ and $w_l^n \geq w_l^u$) are either impossible or not optimally chosen by the firm under our assumptions.

⁶³Assuming that only the second constraint $w_h^n(\mathbf{g}) \geq w_h^u(\mathbf{g})$ binds leads to less profits for the firm.

Again, by the means inequality, a sufficient condition for this inequality to hold is

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

For the lowest possible $q(\theta_l)$, $q(\theta_l) = z_h q(\theta_h) z_l^{-1}$, this inequality is satisfied. As $q(\theta_l)$ increases, so does the left-hand side of the inequality, so it is satisfied for any $q(\theta_l)$ that satisfies our assumptions. \square

Proposition 6. Under the same assumptions as Proposition 4, the union threat increases the average wage and decreases wage inequality, as defined as the ratio of the high-skill wage to the low-skill wage, in nonunion firms.

Proof. For the average wage, we need to show that

$$\mathbb{E}(\mathbf{w}(\mathbf{g}^n)) \geq \mathbb{E}(\mathbf{w}(\mathbf{g}^{n*}))$$

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

which is equivalent to

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

which, as explained in the previous proof, is true under our assumptions. For the ratio of wages, we know that

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

so that

$$\frac{w_h(\mathbf{g}^{n*})}{w_l(\mathbf{g}^{n*})} > \frac{w_h(\mathbf{g}^n)}{w_l(\mathbf{g}^n)}$$

is true under our assumptions. \square

Proposition 7. Under the same assumptions as Proposition 4, if

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

and

$$\frac{\alpha \beta_n}{1 - (1 - \alpha) \beta_n} 2z_h \geq \beta_u$$

it is optimal for the the firm to prevent unionization, otherwise it chooses to be unionized.

Proof. We need to compare the firm's profit under \mathbf{g}^n and \mathbf{g}^{u*} . Let us first derive \mathbf{g}^{u*} . From the results of the previous propositions, the firm's problem is

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

where the union wage is given by

$$\frac{\beta_u}{g_l + g_h} (g_l^{z_l} g_h^{z_h})^\alpha.$$

The Lagrangian is

$$\mathcal{L} = (1 - \beta_u) (g_l^{z_l} g_h^{z_h})^\alpha - \kappa \left(\frac{g_h}{q(\theta_h)} + \frac{g_l}{q(\theta_l)} \right)$$

the first-order conditions are

$$(1 - \beta_u) \alpha F(\mathbf{g}) \frac{z_s}{g_s} = \frac{\kappa}{q(\theta_s)}$$

so that

$$F(\mathbf{g}^{u*}) = ((1 - \beta_u) \alpha)^{\frac{\alpha}{1-\alpha}} \left(\left(\frac{z_l}{\kappa} q(\theta_l) \right)^{z_l} \left(\frac{z_h}{\kappa} q(\theta_h) \right)^{z_h} \right)^{\frac{\alpha}{1-\alpha}}$$

and the value of the objective function at \mathbf{g}^{u*} is thus

$$\begin{aligned} (1 - \beta_u) F(\mathbf{g}^{u*}) - \kappa \left(\frac{g_h^{u*}}{q(\theta_h)} + \frac{g_l^{u*}}{q(\theta_l)} \right) &= (1 - \beta_u) (1 - \alpha) F(\mathbf{g}^{u*}) \\ &= (1 - \beta_u) (1 - \alpha) ((1 - \beta_u) \alpha)^{\frac{\alpha}{1-\alpha}} \left(\left(\frac{z_l}{\kappa} q(\theta_l) \right)^{z_l} \left(\frac{z_h}{\kappa} q(\theta_h) \right)^{z_h} \right)^{\frac{\alpha}{1-\alpha}}. \end{aligned}$$

For the constrained firms, we find that

$$F(\mathbf{g}^n) - \sum_{s \in \mathcal{S}} g_s w_s^n(\mathbf{g}^n) - \kappa \sum_{s \in \mathcal{S}} \frac{g_s^n}{q(\theta_s)} = \left(1 - \frac{\beta_n}{1 - (1 - \alpha) \beta_n} \alpha \right) F(\mathbf{g}^n) - \left(\frac{1 - \beta_n}{1 - (1 - \alpha) \beta_n} \left(\frac{\kappa}{q(\theta_l)} + \frac{\kappa}{q(\theta_h)} \right)^{-1} \alpha \right)^{\frac{1}{1-\alpha}}$$

where

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

Then the firm prefers to be fight unionization if

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

otherwise it lets the workers unionize. Finally, we must make sure that the high-skill workers prefer to vote against the union. From 33

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

So that the inside of the parenthesis (inequality 23) must be negative for $s = h$. \square

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 \mathbf{c}_1 and \mathbf{c}_2 are identical and denote that schedule by \mathbf{c} . This result will be shown later in the lemma. We can write the problem of firm $j \in \{1, 2\}$ as

$$\max_g \mathcal{J}(A_j, \mathbf{w}(A_j, \mathbf{g}), \mathbf{g})$$

such that

$$\mathbf{w}(A_j, \mathbf{g}) = \begin{cases} \mathbf{w}^u(A_j, \mathbf{g}) & \text{if } V(A_j, \mathbf{g}) > 0 \\ \mathbf{w}^n(A_j, \mathbf{g}) & \text{if } V(A_j, \mathbf{g}) \leq 0 \end{cases}$$

where \mathbf{w}^u is the union wage function, \mathbf{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

$$\mathbf{g}_2 = \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \mathbf{g}_1$$

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 $\mathbf{w}^n(A_1, \mathbf{g}_1) = \mathbf{w}^n(A_2, \mathbf{g}_2)$. It is also straightforward to show that $F(A_1, \mathbf{g}_1)/n_1 = F(A_2, \mathbf{g}_2)/n_2$ 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

$$\begin{aligned} g_{1,s'} \frac{\partial w_{s'}^n(A_1, \mathbf{g}_1)}{\partial g_{1,s}} &= \frac{\alpha \beta_n (1 - \sigma(1 - \alpha))}{(1 - \beta_n (1 - \alpha)) \sigma} A_1 \left(\sum_{k \in \mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\alpha \sigma - 2\sigma + 2}{\sigma-1}} z_s g_{1,s}^{-1/\sigma} z_{s'} g_{1,s'}^{\frac{\sigma-1}{\sigma}} \\ &= \frac{\alpha \beta_n (1 - \sigma(1 - \alpha))}{(1 - \beta_n (1 - \alpha)) \sigma} A_2 \left(\sum_{k \in \mathcal{S}} z_k g_{2,k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\alpha \sigma - 2\sigma + 2}{\sigma-1}} z_s g_{2,s}^{-1/\sigma} z_{s'} g_{2,s'}^{\frac{\sigma-1}{\sigma}} = g_{2,s'} \frac{\partial w_{s'}^n(A_2, \mathbf{g}_2)}{\partial g_{2,s}}. \end{aligned}$$

Similarly,

$$\begin{aligned}
g_{1,s} \frac{\partial w_s^n(A_1, \mathbf{g}_1)}{\partial g_{1,s}} &= \frac{\alpha \beta_n z_s A_1}{1 - \beta_n (1 - \alpha)} \left(-\frac{1}{\sigma} \left(\sum_{k \in \mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\alpha \sigma - \sigma + 1}{\sigma - 1}} g_{1,s}^{-\frac{1}{\sigma}} \right. \\
&\quad \left. + g_{1,s}^{\frac{\sigma-2}{\sigma}} \frac{1 - \sigma(1 - \alpha)}{\sigma} \left(\sum_{k \in \mathcal{S}} z_k g_{1,k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\alpha \sigma - 2\sigma + 2}{\sigma - 1}} z_s \right) \\
&= g_{2,s} \frac{\partial w_s^n(A_2, \mathbf{g}_2)}{\partial g_{2,s}}.
\end{aligned}$$

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

$$g_{1,s'} \frac{\partial w_{s'}^u(A_1, \mathbf{g}_1)}{\partial g_{1,s}} = g_{2,s'} \frac{\partial w_{s'}^u(A_2, \mathbf{g}_2)}{\partial g_{2,s}}.$$

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

$$\frac{\partial V(A_1, \mathbf{g}_1)}{\partial g_{1,s}} = \frac{\partial V(A_2, \mathbf{g}_2)}{\partial g_{2,s}}$$

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 $V(A_2, \mathbf{g}_2) = 0$. 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. \square

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

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

subject to

$$\sum_{s \in \mathcal{S}} g_s \phi(w_s^u(\mathbf{g}) - w_s^n(\mathbf{g})) - \frac{1}{2}n \leq 0.$$

The constraint can be written as

$$a \sum_{s \in \mathcal{S}} g_s (w_s^u(\mathbf{g}) - w_s^n(\mathbf{g})) \leq 0.$$

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. \square