The Union Threat*

Mathieu Taschereau-Dumouchel†
Princeton University

December 31, 2010

Job Market Paper

Abstract

This paper studies the impact of unions on wage inequality, output and unemployment. To do so, it proposes a search and matching model of union formation in which unions arise endogenously through a voting process within firms. In a union firm, workers bargain their wages collectively. In a nonunion firm, each worker bargains individually with the firm. Because of this wage setting asymmetry, a union lowers the profit of a firm and compresses the wage distribution of the workers. Furthermore, to prevent unionization, nonunion firms distort their hiring decisions in a way that also lowers the dispersion of wages. After being calibrated on the United States, the model shows that, even though a partial equilibrium estimate would predict a small impact of unions on inequality, removing the threat of unionization increases the variance of wages substantially. Completely outlawing unions increases wage inequality further. Moreover, outlawing unions increases welfare and output, and lowers unemployment. These results suggest that, even with a small membership, unions might have a significant impact on the economy through general equilibrium mechanisms and the way they distort firms' decisions.

*I am greatly indebted to Nobuhiro Kiyotaki, Esteban Rossi-Hansberg and Oleg Itskhoki for their invaluable advice and support throughout this project. Special thanks go to Esteban for suggesting the title. I also wish to thank Manuel Amador, Orley Ashenfelter, Francisco Buera, Jeff Campbell, Pablo Fajgelbaum, Henry Farber, Bariş Kaymak, David Lee, Thomas Lemieux, Benjamin Moll, Justinas Pelenis, Edouard Schaal, Andrew Shephard, Christopher Sims, Liz Tennant, as well as seminar participants at Princeton University and the Chicago Fed Rookie Seminar for helpful comments and suggestions. I gratefully acknowledge financial support from the Fonds québécois de recherche sur la société et la culture.

†Email: mtascher@princeton.edu; Address: 001 Fisher Hall, Department of Economics, Princeton University, Princeton, NJ 08544. For the most recent version go to: http://sites.google.com/site/mathtd/.

1 Introduction

What is the impact of unions on the economy? On the one hand, the unionization rate in the United States is at one of its lowest levels in decades. In 2005, only 9% of American private sector workers were unionized.[1] This low union membership limits the scope of collective bargaining agreements to a small fraction of the workforce. At face value, this suggests that unions have a restricted impact on the economy. On the other hand, there is a large empirical literature suggesting that unions are responsible for lowering wage inequality.[2] The bulk of this literature has, however, been limited to measuring partial equilibrium effects only. In particular, it is generally assumed that the union and nonunion wage schedules remain unchanged by a modification of union policies. This approach abstracts completely from the decision process of the firms. For instance, one could think that, if unions were outlawed, the previously unionized firms would demand workers with different characteristics. Also, nonunion firms might be modifying their behavior in response to a *threat* of unionization. Indeed, even if we observe that a firm is union free, its workers still have the *legal option* to unionize. If unionization lowers profit, a firm might distort its behavior to prevent the formation of a union. A change in union laws would change this threat and therefore affect the behavior of nonunion firms. Finally, partial equilibrium estimates obviously neglect general equilibrium mechanisms that influence unemployment and the way wages are set.

Because of the low variability of union policies across time[3] and the big differences in union laws across countries, it is hard to imagine a non-structural empirical exercise that could identify the global effects of unions on an economy. We therefore need a general equilibrium theory of firms' decision and union formation. This paper proposes such a theory.

The model features risk neutral heterogeneous agents who randomly meet with heterogeneous firms in a labor market characterized by search frictions. Once a firm has hired its new workers, its employees vote on the creation of a union to represent them. If this vote is successful, a union is established and wages are determined through a collective bargaining process between the firm and the union. On the other hand, if the majority of workers votes against unionization, the firm stays union free and each employee bargains his wage individually with the employer. The interaction between the two bargaining structures and the production technology implies that firms have, in general, a higher profit when a union does not represent the workers. The average wage among workers is, however, higher when a union is in charge of the negotiation for all employees. This leads some firms to distort the distribution of the workers they hire in order to prevent unionization. By doing so, they naturally influence the wages of the workers in such a way as to compress the wage distribution.

Several elements distinguish this paper from the existing literature. First, I consider the economy in general equilibrium. In such a setup, the presence of unions influences the way nonunion wages are set through the aggregate variables of the economy (for instance, the unemployment rate and the expectations that workers have about their future wages). These mechanisms cannot be captured by traditional empirical estimates and might have an important influence on the way unions influence the economy. Second, the threat of potential unionization is featured prominently in the model. This implies that, even in an economy in which no union actually exists, the possibility of unionization alone influences wages, unemployment and output through the distorted

[1] Source: Merged Outgoing Rotation Groups of the 2005 Current Population Survey. I define a worker as unionized if this worker is a union member or if he is covered by a union contract.

[2] Card et al. (2004) provides a nice historical survey of that literature together with its own estimates.

[3] The last major change in union regulations in the US is the Taft-Hartley Act of 1947.

behavior of the firms. Third, a wage compression effect of unions arises naturally from the model and its importance is influenced by the state of the economy. Fourth, the ultimate determinant of the union status of a firm is its production technology. In particular, firms with lower labor intensity tend to be more unionized.

After introducing the model and highlighting the various mechanisms that influence the behavior of the firms, I calibrate the model on the private sector of the United States and do an empirical exercise to see how a partial equilibrium estimator would measure the effects of unions on wage inequality. To do so, I give to each union worker the counterfactual wage that he would get if he were working in a nonunion job. This is a procedure that has been used in the empirical literature. This exercise suggests that, in the calibrated economy, the reallocation of workers in partial equilibrium lowers the variance of log wages by about 0.4%. I then perform two general equilibrium exercises to evaluate the full impact of unions on the economy. The first one consists in removing the threat of unionization. In other words, nonunion firms do not have to worry about the vote on the formation of a union anymore. Firms that are unionized remain unionized and vice versa. In the new equilibrium, the variance of log wages goes up by 5.7% when compared to the calibrated model. This shows that the threat of unionization alone might have an important impact on inequality. Welfare also goes up by 1.7%, suggesting that the firms' departure from their optimal hiring decision has a negative impact on the economy as a whole.[4] The second exercise is to eliminate unions completely. All wages are then negotiated on a one-on-one basis with the firms. In this scenario, the variance of log wages goes up by 6% with respect to the calibrated economy. Total production goes up by 2.5% and welfare also increases by 2.2%. The unemployment rate goes down by 2.5 percentage points.

These results suggest that, even with low membership, unions seem to have an impact on wage inequality, output and unemployment through the threat they exert and through general equilibrium mechanisms. Also, this paper shows that partial equilibrium estimates are likely to miss important channels through which unions influence the economy.

After a brief literature review, I introduce the model and explain how firms behave in an environment with unions. In particular, I highlight the distortion of the firm's behavior caused by the union threat. A discussion of the link between a firm's technology and its union status follows. I then calibrate the model on the US economy and do counterfactual policy experiments to see how unions affect the economy.

1.1 Related literature

There is a large empirical literature that evaluates the impact of unions on wage inequality. Freeman (1980) analyzes data from the first half of the 1970s about private sector male workers in the United States and finds that unions are responsible for an important equalizing effect of wages of union workers inside a given sector. This effect is particularly important in manufacturing. I compute the estimator of Freeman (1980) for the private sector of the US in 2005. It suggests that unions are responsible for lowering the variance of log wages by 0.4%.[5]

[4] I define welfare as the sum of the utilities of all the agents in the economy.

[5] Card et al. (2004) write this estimator 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, U is the unionization rate, Δ_v is the difference in the variance of log union and nonunion wages and Δ_w is the difference between the mean log of union and nonunion wages. For consistency with the calibrated economy, I clean the data by removing agricultural workers as well as workers earning an hourly wage of less than \$5 or more than \$150.

DiNardo et al. (1996) uses a semiparametric approach to estimate the impact of labor market institutions on the distribution of wages. They find that the decline in the unionization rate during the 1980s accounts for 10-15 percent of the rise in wage dispersion for men.

Some studies have explicitly accounted for unobserved productivity in their estimates. Among them, Lemieux (1993) and Card (1996) use longitudinal data on Canadian and American workers respectively to evaluate the impact of unions on job switchers. They find that low-skill union workers tend to have higher unobserved productivity than their nonunion counterparts. This implies that the smoothing effect of wages across skill groups that is observed in the raw data is exaggerated when compared to the causal effect of unions. Lemieux (1993) finds that in the late 1980s, unions were responsible for lowering the variance of male wages by 15% in Canada.

Card et al. (2004) provides more recent estimates of the impact of unions on wage inequality. They find that unions were responsible for lowering the variance of log wages of men by 4.5%.[6] Their sample includes private and public sector workers. The same estimate for women is about 2.4%.

All these studies consider what the variance of wages would be if each union worker was paid according to the nonunion wage structure. They do not take into account how unions could affect the structure of wages itself. Indeed, Card et al. (2004) clearly acknowledges this point in their summary of the literature. The model proposed in the current paper explicitly includes these general equilibrium effects.

Two papers propose a model of unions in general equilibrium. Açikgöz and Kaymak (2008) builds a tractable search and matching model of endogenous union formation to estimate the impact of a rising skill premium on the decline of union membership in the United States. They assume that the degree of wage compression is determined by an exogenous parameter that needs to be estimated. In the current paper the compression arises naturally and varies as a function of the aggregate conditions of the economy. Also, their model abstracts from studying the modification of nonunion firms' behavior in response to the unionization threat. Finally, my focus extends also to unemployment, output and welfare. Acemoglu et al. (2001) shows that deunionization had an amplifying effect on the rise in wage inequality during the last 25 years of the 20th century. They propose a model of union formation but abstract from the role of the firm in this process, therefore abstracting completely from the threat that unions exert. Other papers modeling unions include, among many, Farber (1978) and Ashenfelter and Johnson (1969).

Hirsch (2004) summarizes the state of current research on the impact of unions on productivity and profitability. He states that "empirical evidence on unions and productivity was rather sketchy in 1984; it remains less than clear-cut today". The research on profitability is more conclusive. According to Hirsch, "evidence points unambiguously to lower profitability among union companies," a feature that arises in the benchmark case of the model I propose.

Finally, Nickell and Layard (1999) finds a correlation between high union density and unemployment in OECD countries between 1983 and 1994 but explains that this correlation is offset by controlling for the level of coordination between unions and firms.

2 The Model

The model incorporates six main elements. First, workers are ex-ante heterogeneous in their skill. Second, firms have decreasing returns to scale and hire multiple workers. Third, the model is built

[6] This estimates comes from their model with skill groups.

along the lines of the search and matching literature (Mortensen and Pissarides, 1994; Pissarides, 2000). Fourth, the formation of a union is decided at the level of the firm by a vote cast by the workers.[7] Fifth, if a union is established, wages are bargained collectively. Otherwise, workers bargain individually with the firm.[8] Sixth, the whole model is in general equilibrium.

The skill heterogeneity interacts with the labor market friction to generate a union wage gap that varies with skill, a feature generally associated with unions (Farber and Saks, 1980). Also, the skill heterogeneity implies that different workers contribute differently to the firm's production. By bargaining individually, a worker's wage is a function of his own characteristics. By bargaining collectively, the union combines the characteristics of all the firm's employees and redistributes what it extracts from the firm among its members. This implies that workers with valuable attributes, for instance those who have a high marginal product, obtain a higher wage by bargaining individually and therefore vote against the unionization of the firm.

The firm is not indifferent between having its workers unionized or not. The decreasing returns to scale interacts with the union status of the firm to influence its profit. When bargaining individually, the firm treats each worker as if it is the marginal one. If negotiations break down, the firm can still produce with the rest of its employees. When bargaining collectively, on the other hand, the union prevents any production from taking place if an agreement on wages is not reached. Because of this threat, the union benefits from the high marginal surplus generated by the infra-marginal workers. This asymmetry of the threats gives a natural disadvantage to a firm that is bargaining collectively with its employees. This creates an incentive to influence the workers' decision in order to avoid unionization. To do so, the firm distorts the distribution of its employees by hiring more high-skill workers and less low-skill workers. This leads to a non-uniform modification of the workers' marginal products and leads to a compression of the range of wages paid by the firm.

2.1 Preferences and technology

There is a single good and time is discrete. There are no savings. I focus on steady state equilibria. The economy is populated by a continuum of ex-ante heterogeneous agents, each endowed with a specific type of labor $s \in [0, 1]$. I refer to s as the skill. Firms use the different skills for production. The exogenous density of skills in the economy is $N(s)$ with $N(s) > 0$ for all s. An agent's skill is constant over time. Agents live forever. They are risk neutral and maximize a linear utility function

$$\mathcal{U}(c) = \mathrm{E}_0 \sum_{t=0}^{\infty} \gamma^t c_t$$

where c_t denotes consumption in period t and $0 < \gamma < 1$ is the discount factor.

Firms combine the labor provided by workers of different skills to produce goods. To do so, they use heterogeneous production technologies, indexed by $j \in \{1, \ldots, j_{\max}\}$. There is a mass 1 of firms endowed with each technology. A firm of type j employing a (non-normalized) distribution

[7] Modeling unionization as a firm-level process is consistent with evidence presented by Traxler (1994) and Nickell and Layard (1999) that suggest that the coverage of union contracts is mostly at the enterprise level in the US, Canada and the UK.

[8] Cahuc and Wasmer (2001) builds a search and matching model with firms with decreasing returns to scale in which wages are set through individual bargaining.

of workers $g(s)$ produces goods according to the production function

$$F_j(L_j(g)) = A_j L_j^{\alpha}(g) = A_j \left\{ \exp\left(\int z_j(s) \log g(s) ds \right) \right\}^{\alpha_j}$$

where $0 < \alpha_j < 1$, $A_j > 0$ and where L_j is a Cobb-Douglas aggregator that describes how firm j combines the different types of labor for production. The function $z_j : [0,1] \to \mathbb{R}_+^*$ represents the relative intensity of skill utilization and is therefore normalized such that $\int z_j(s)\, ds = 1$. The parameter α_j describes the returns to scale of the production function. To avoid cluttering the notation, I omit the subscript j when referring to a single firm. Also, I sometimes write $F(g)$ directly instead of the more cumbersome $F(L(g))$. Notice that, since $z(s) > 0$, the marginal product of a worker of type s goes to infinity as $g(s) \to 0$. If the cost of hiring is finite, a firm therefore employs workers of every type.

2.2 Labor markets

There is a continuum of labor markets in which unemployed agents look for jobs and firms post vacancies. Each vacancy has a cost of κ. Each market is indexed by the skill s of agents searching in it. Agents can only search in the labor market corresponding to their skill.[9] Firms, on the other hand, are free to post a continuum of vacancies that covers all the markets. Figure 1 represents this structure. In each market, matches happen randomly at a rate determined by aggregate conditions. If, in a given period, u agents are searching and v vacancies have been posted, $m(u, v)$ matches are made. The matching function $m(\cdot, \cdot)$ is identical across labor markets and is homogenous of degree one. By defining the labor market tightness $\theta \equiv v/u$, the probability that a vacancy is filled in a given period is $q(\theta) \equiv m(u, v)/v$. Similarly, the probability that an unemployed agent finds a job is $\theta q(\theta)$. Notice that q is a strictly decreasing function of θ. Search is free and requires no effort. Every unemployed agent is therefore searching.

All types of firms are posting vacancies in each market. A searching worker can therefore be matched with firms using different technologies and with different union status.

This segmentation of the labor market has two main consequences. First, it allows the firm to control precisely the skill composition of its workforce and, through this channel, influence the unionization vote. Second, it allows me to study the effects of unionization on unemployment rates across skill groups.

I use the skill index as a way to characterize the heterogeneity of the workers. It is uniquely an index and has no meaning in itself. Later, in the empirical part of the paper, I calibrate the index such that wages are increasing in s and the unemployment rate is decreasing in s. This makes the interpretation of the results more intuitive and explains the name of this index.

2.3 Agents

Agents provide labor to firms in exchange for a wage. In each period, an agent is either employed or unemployed. An employed worker loses his job with exogenous probability $\delta > 0$, in which case he goes to the labor market corresponding to his type. With probability $1 - \delta$, the agent remains

[9] In the calibrated model, the unemployment rates across labor markets are decreasing with s while expected wages are increasing with s. Therefore, even if agents were allowed to search in markets with a lower s than their own, they would choose not to do so.

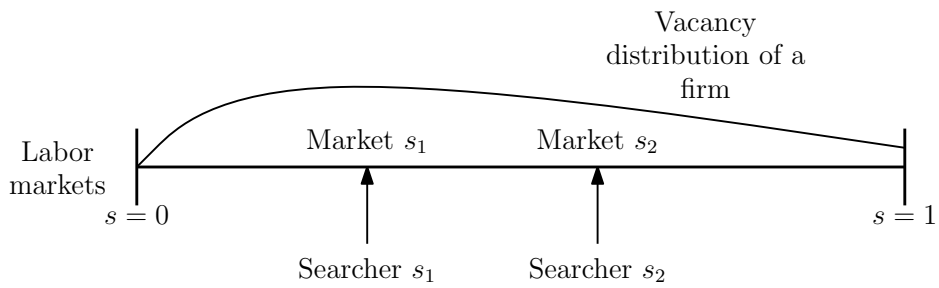

Figure 1: Continuum of labor markets. Agents can only search in the market corresponding to their type. Firms can post vacancies in all markets.

employed in his current job. Therefore, the lifetime discounted expected utility of a worker of type s who has been matched with a firm of type j and who is earning a wage w is

$$W_e(s, w) = w + \gamma \left[(1-\delta)W_e(s, w_j(s)) + \delta W_u(s)\right]$$

where $W_u(s)$ is the lifetime utility of being unemployed and $w_j(s)$ is the *equilibrium* wage of a worker of type s provided by a job in firm j. Wages are bargained every period. Therefore the negotiations with the firm are over w only. Both parties consider that $w_j(s)$ is fixed at its equilibrium value.

Every period, an unemployed agent s receives $b_0(s)$ from unemployment benefits and home production. He finds a job with probability $\theta(s)q(\theta(s))$. His lifetime discounted utility is therefore

$$W_u(s) = b_0(s) + \gamma \left\{\theta(s)q(\theta(s))\mathrm{E}(W_e(s, w)) + (1 - \theta(s)q(\theta(s)))W_u(s)\right\}$$

where the expectation $\mathrm{E}(W_e(s, w))$ is taken over all the possible wages offered to a searching agent of type s.

An agent will accept to work only if the utility provided by employment exceeds the utility of continuing the search for a job. This never happens in equilibrium. By combining the last two equations we can characterize the utility gain provided by employment:

$$W_e(s, w) - W_u(s) = w + \frac{\gamma(1-\delta)w_j(s) - (1-\gamma)W_u(s)}{1-\gamma(1-\delta)}. \tag{1}$$

In equilibrium, the utility provided by a job in firm j is

$$W_e(s, w_j(s)) - W_u(s) = \frac{w_j(s) - (1-\gamma)W_u(s)}{1-\gamma(1-\delta)}.$$

It is useful to define the flow utility of being unemployed by $b(s)$. Therefore,

$$b(s) \equiv (1-\gamma)W_u(s) = \frac{(1-\gamma(1-\delta))b_0(s) + \gamma\theta(s)q(\theta(s))\mathrm{E}(w(s))}{1-\gamma(1-\delta)+\gamma\theta(s)q(\theta(s))}. \tag{2}$$

The utility of an unemployed worker takes into account the fact that this worker will spend a part of his time employed in the future. It is therefore a weighted average of $b_0(s)$ and of the wage this agent expects to receive.

To simplify the notation, it is also convenient to define the equilibrium quantity

$$c_j(s) \equiv \frac{b(s) - \gamma(1-\delta)w_j(s)}{1-\gamma(1-\delta)} \tag{3}$$

which is the net outside option of the worker. By writing the gain from employment in a firm j at a wage w as

$$W_e(s,w) - W_u(s) = w - c_j(s)$$

we see that $c_j(s)$ is what the worker gets if the bargaining breaks down.

2.4 Firms

A firm that employed a distribution of workers g_{-1} during the previous period loses a fraction δ of all of its workers and therefore starts the current period with the distribution $(1-\delta)g_{-1}$. It then posts a schedule of vacancies 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.

Once the new hires have joined the firm, the workers vote on the formation of a union and the firm's optimal behavior will depend on the specifics of the unionization process as well as on how the union and nonunion wages are set. These will be described shortly. For now, it is sufficient to use an abstract function $\int w(s,g)$ to denote the wages that the firm pays as a function of the current workers distribution. Define the current period profit of a firm as $\pi(g) \equiv F(g) - \int w(s,g) \cdot g\, ds$. With this notation, the problem of a firm is

$$\tilde{J}(g_{-1}) = \max_v \pi(g) - \kappa \int v(s)\, ds + \gamma \tilde{J}(g)$$

subject to

$$\begin{cases} g(s) = g_{-1}(s)(1-\delta) + v(s)q(\theta(s)) \\ v(s) \geq 0 \end{cases}$$

where $\tilde{J}(g_{-1})$ is the value function of a firm that ended the previous period with workers g_{-1}. The first constraint is simply the law of motion of the stock of workers. The second constraint states that job separations are exogenous. Firms cannot post negative vacancies.

In a steady state equilibrium in which the aggregate variables remain constant, it is possible to simplify the firm's problem substantially. Suppose that in such an equilibrium, a firm's optimal distribution of workers is given by $g^*(s)$. Such a distribution exists because of the decreasing returns to scale. Two events might move the firm away from $g^*(s)$. First, every period, it loses a fraction of its workers. Second, if one of the wage bargaining sessions breaks down without an agreement, the firm loses additional workers.[10] 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 $v(s) = 0$ in the other markets. We can therefore substitute v from the law of motion of the workers directly into the objective function. The problem of the firm can be simplified as

$$J\left(\int \frac{g_{-1}}{q(\theta)} ds\right) = \max_g \pi(g) - \kappa \int \frac{g - g_{-1}(1-\delta)}{q(\theta)} ds + \gamma J\left(\int \frac{g}{q(\theta)} ds\right) \tag{4}$$

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

where $\int \frac{g_{-1}}{q(\theta)}\, ds$ is a new state variable that represents the value of the stock of workers with which the firm enters the period.[11]

This last value function has two additively separable pieces: one that depends on the distribution of previous period g_{-1} and a second one that depends on the firm's decision in the current period. This implies that, in a steady state, the firm's current period decision is independent of its state variable. The following lemma simplifies the firm's problem.

Lemma 1. *In a steady-state, the firm's dynamic problem can be written as the static optimization*

$$\max_g \pi(g) - \kappa(1-(1-\delta)\gamma) \int \frac{g}{q(\theta)}\, ds. \tag{5}$$

Proof. All proofs are relegated to the appendix. □

This result comes directly from the linearity of the hiring costs, the constant value of θ and the fact that, at the steady state, a firm never wants to downsize in response to a shock.

We now need to describe the wage schedule $w(s, g)$. Figure 2 details the sequence of events that occurs once a firm has recruited its new workers. First, the workers vote to decide whether to form a union or not. Then, if a union is established, wages are bargained collectively. The outcome of this bargaining is a wage schedule $w_u(s, g)$ and a profit function $\pi_u(g)$. If the union is rejected, wages are bargained individually. This generates the wage schedule $w_n(s, g)$ and the profit $\pi_n(g)$. Notice that when the vote takes place and when wages are bargained, the distribution of workers g is fixed. Also, when the workers cast their vote, they know exactly what wages they will get if the union is created or not. I first describe the two bargaining procedures and then come back to the voting process.

2.5 Wage setting

In both a union and a nonunion firm, wages are set using Nash bargaining to share the surplus generated by the match. The surplus that is bargained over is, however, different in both cases. If the firm is unionized and an agreement on wages cannot be reached, the whole workforce quits the firm and no production takes place. In a non unionized firm, if the bargaining with a single worker breaks down, this specific worker goes back to unemployment but the firm can still produce with the other workers. In a nonunion firm, the bargaining therefore takes place over the marginal surplus generated by each worker. In a union firm, the workers and the firm bargain over the total surplus generated by the whole workforce. This asymmetry between the two surpluses interacts with the decreasing returns of the production function and has important consequences for the firm's profits.

Union bargaining

If the workers vote in favor of unionization, the union is the only group authorized to bargain with the firm. Consider the firm's gain if it reaches an agreement with the union. In a steady-state, the difference in discounted profits for the firm, denoted by Δ^u, is

$$\Delta^u(w) = \left[\pi(g^*, w) + \gamma J\left(\int \frac{g^*}{q(\theta)}\, ds\right)\right] - [\pi(0) + \gamma J(0)] \tag{6}$$

[11] Notice that J and $\tilde{J}$ are two different objects but they give the same first order conditions in a steady state equilibrium.

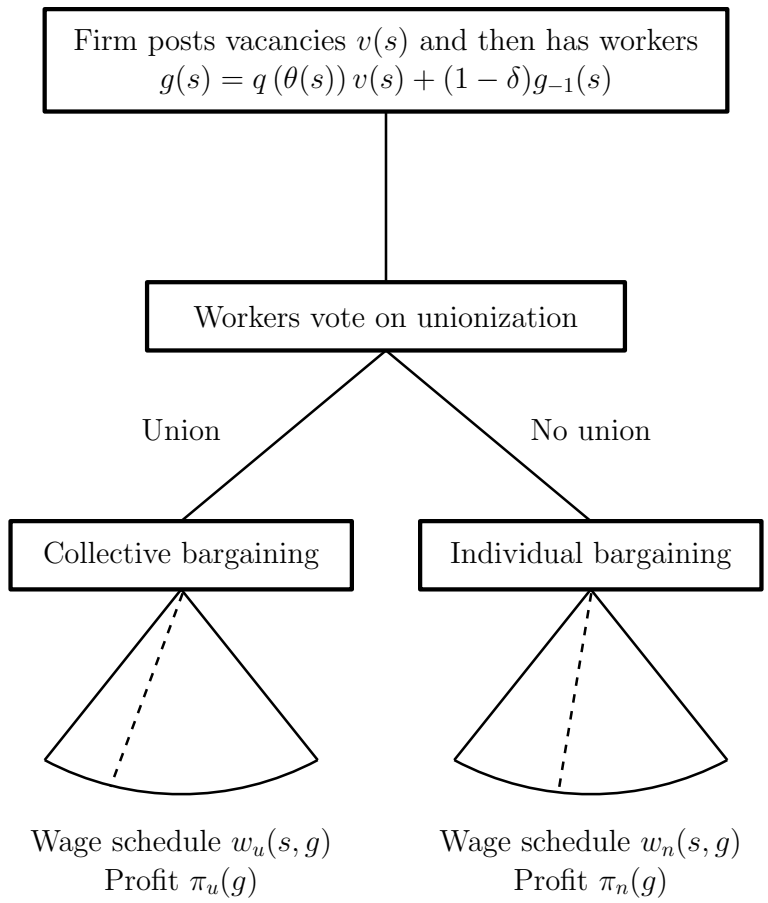

Figure 2: Sequence of events

where the first term in brackets is discounted profit if an agreement is reached and $\pi(0) + \gamma J(0)$ is the firm's discounted profit if negotiations break down. Notice that in such a scenario, the firm has no worker; it produces nothing and pays no wage. Therefore, the one-period profit $\pi(0)$ is equal to zero. $J(0)$ is the value function of a firm that starts the period with no workers. Because the firm's employment decision is independent of the distribution of its workers, the firm hires back to its steady-state optimal level g^* right away. Therefore,

$$J(0) = \pi(g^*, w^*) - \kappa \int \frac{g^*}{q(\theta)}\, ds + \gamma J\left(g^*\right)$$

where w^* is the equilibrium wage schedule for this firm. This last expression is identical to $J\left(g^*\right)$ except for the fact that the firm hires back all of its workforce in that period and therefore pays a higher vacancy cost. We can rewrite the difference in discounted profit as

$$\Delta^u(w) = \pi(g^*, w) + \gamma J\left(g^*\right) - \gamma \left(\pi(g^*, w^*) - \kappa \int \frac{g^*}{q(\theta)}\, ds + \gamma J\left(g^*\right) \right)$$

But, at the steady state, the firm's value function is

$$J(g^*) = \pi(g^*, w^*) - \kappa\delta \int \frac{g^*}{q(\theta)}\, ds + \gamma J(g^*) \tag{7}$$

and therefore the firm's surplus from reaching an agreement is

$$\Delta^u(w) = \pi(g^*, w) + (1-\delta)\gamma\kappa \int \frac{g^*}{q(\theta)}\, ds.$$

The intuition is straightforward. If negotiations breaks down, the firm loses the current period profit π 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 the firm next period.

We now need to specify the surplus of the union. To do so, an assumption needs to be made about how the workers divide among themselves the rent extracted from the firm. A natural assumption is to have them split that amount by solving a Nash bargaining problem in which each worker has the same bargaining power. In this case, the log of the Nash surplus of the union is given by

$$\int \frac{g(s)}{n} \log(W_e(s,w) - W_u(s))\, ds$$

with $n = \int g(s)\, ds$ and where $W_e(s,w) - W_u(s)$ is given by equation 1.[12,13] The union simply aggregates the individual surpluses of each worker.

With this way of sharing the surplus among the workers, the bargaining problem between the firm and the union is simply

$$\max_w \left[\exp\left(\int \frac{g}{n} \log(W_e(s,w) - W_u(s))\, ds\right)\right]^{\beta_u} \left[F(g) - \int w \cdot g\, ds + (1-\delta)\kappa\gamma \int \frac{g}{q(\theta)}\, ds\right]^{1-\beta_u}. \tag{8}$$

where $0 < \beta_u < 1$ denotes the bargaining power of the union. This coefficient is exogenous to the model and could possibly be influenced by labor market policies.

Lemma 2. *Assume that g is strictly positive on $[0,1]$. Then the following function solves the bargaining problem:*

$$w_u(s,g) - c(s) = \frac{\beta_u}{n}\left(F(g) - \int c \cdot g\, ds + \gamma(1-\delta)\kappa \int \frac{g}{q(\theta)}\, ds\right). \tag{9}$$

The solution is unique if the joint surplus of the match is strictly positive at the point w_u.

Also, in a union firm with equilibrium distribution of workers g^* and technology j, the equilibrium wage schedule $w_j(s) = w_u(s, g^*)$ is

$$w_u(s,g^*) - b(s) = \frac{1-\gamma(1-\delta)}{1-\beta_u\gamma(1-\delta)} \frac{\beta_u}{n^*}\left(F(g^*) - \int b \cdot g^*\, ds + \gamma(1-\delta)\kappa \int \frac{g^*}{q(\theta)}\, ds\right) \tag{10}$$

where $n^* = \int g^* ds$ is the optimal size of the firm.

[12] To see where this equation comes from, consider the discrete case in which there are k different skill groups, each with a weight $\epsilon > 0$, and that g_i workers are of type i. The surplus of a worker of type i, if an agreement is reached at a wage w, is $W_{ei}(w) - W_{ui}$. The log of the joint Nash surplus can be written as

$$\log\left\{(W_{e1} - W_{u1})^{\frac{g_1\epsilon}{n}} \times \cdots \times (W_{ei} - W_{ui})^{\frac{g_i\epsilon}{n}} \times \cdots \times (W_{ek} - W_{uk})^{\frac{g_k\epsilon}{n}}\right\} = \sum_{i=1}^{k} \frac{g_i\epsilon}{n} \log(W_{ei} - W_{ui})$$

where $n = \sum_{i=1}^{k} g_i\epsilon$ is the total number of workers in the firm and where $g_i\epsilon/n$ is the sum of the bargaining power of all the workers of type i. Taking the limit as $k \to \infty$ and $\epsilon \to 0$, we get the log of the union surplus.

[13] Nash's axiomatic theory of bilateral bargaining extends unchanged to a context with numerous players. Krishna and Serrano (1996) provides a strategic approach to multilateral bargaining.

Equation 10 implies that, in equilibrium, all the workers are getting the same transfer over their reservation wage $b(s)$. The union is basically mixing together the characteristics of all its members. Therefore, the variance of wages comes from the reservation wage schedule $b(s)$. The macroeconomic conditions, through $b(s)$, have a direct influence on the dispersion of wages in a unionized firm.

It is straightforward to show that the one-period profit of a union firm employing the distribution of workers g is given by

$$\pi_u(g) = (1-\beta_u)F(g) - (1-\beta_u)\int c \cdot g\, ds - \beta_u(1-\delta)\kappa\gamma \int \frac{g}{q(\theta)}\, ds. \tag{11}$$

Individual bargaining

If the workers vote against unionization, they each bargain individually with the firm. The workers cannot interact with each other. In particular, they cannot create a coalition. Once again, the worker and the firm use Nash bargaining to split the surplus created by the match. These surplus are, however, not identical across all workers. Because of the decreasing returns to scale, the surplus generated by hiring the first worker is higher than the one generated by the marginal worker. To solve this issue I follow the approach of Stole and Zwiebel (1996a,b). They introduce a game in which Nash bargaining is used to split the *marginal* surplus generated by hiring an extra worker. In this setup, the firm negotiates with each of its workers in turn. If any of the one-on-one negotiations breaks down, wages are renegotiated with all the workers remaining in the firm. When considering the marginal surplus generated by an additional worker, the firm is aware that if the negotiations break down, the other workers might want to rebargain their wages differently.

I show in the appendix (see proof of Lemma 3) that the marginal surplus of the firm from hiring a worker of type s is given by

$$\begin{aligned}\Delta_n(s,w) = \frac{\partial L}{\partial g(s)}\frac{dF}{dL} - \frac{\partial L}{\partial g(s)}\int \frac{\partial w(s,g(s),L)}{\partial L} g(s)\, ds - \frac{\partial w(s,g(s),L)}{\partial g(s)} g(s) \\ - w(s,g(s),L) + \gamma(1-\delta)\frac{\kappa}{q(\theta(s))}.\end{aligned}$$

This equation is fairly intuitive. The first term is the extra output produced by the additional worker. The two following terms represent the marginal effects of the worker on the wages of other members of the workforce. The fourth term is simply the wage paid to the worker and the fifth term is the vacancy costs saved from retaining a fraction $1-\delta$ of today's hire in the next period.

The Nash bargaining implies that the nonunion wage must solve the following equation:

$$\Delta_n(s,w) = \frac{1-\beta_n}{\beta_n}\left(W_e(s,w) - W_u(s)\right). \tag{12}$$

Lemma 3. *The wage schedule*

$$w_n(s,g) - c(s) = \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g(s)}F(g) - \beta_n\ c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))} \tag{13}$$

solves the bargaining problem (equation 12) of a firm employing the distribution of workers g.

Also, in a nonunion firm with equilibrium distribution of workers g^* and technology j, the equilibrium wage schedule $w_j(s) = w_n(s, g^*)$ is

$$w_n(s, g^*) - b(s) = \frac{1-\gamma(1-\delta)}{1-\beta_n\gamma(1-\delta)}\left(\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g^*(s)}F(g^*) - \beta_n\ b(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}\right) \tag{14}$$

It follows directly from the wage of nonunion workers that the one-period profit of the firm is

$$\pi_n(g) = \frac{1-\beta_n}{1-(1-\alpha)\beta_n}F(g) - (1-\beta_n)\int c\cdot g\,ds - \beta_n(1-\delta)\kappa\gamma\int\frac{g}{q(\theta)}\,ds. \tag{15}$$

Collective vs individual bargaining

The workers and the firm are not indifferent between the two types of bargaining. For a given distribution of workers g, the difference in wages is

$$\begin{aligned} w_n(s,g) - w_u(s,g) = F(g)&\left(\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g(s)} - \frac{\beta_u}{n}\right) - (\beta_n c_j(s) - \beta_u \mathrm{E}_g(c_j)) \\ &+ \kappa\gamma(1-\delta)\left(\beta_n\frac{1}{q(\theta(s))} - \beta_u\mathrm{E}_g\left(\frac{1}{q(\theta)}\right)\right) \end{aligned}$$

where $\mathrm{E}_g(x) = \int x\cdot g\,ds/\int g\,ds$ for any function x. It follows directly that

$$\begin{aligned} \mathrm{E}_g(w_n) - \mathrm{E}_g(w_u) = \frac{F(g)}{n}&\left(\frac{\alpha\beta_n}{1-(1-\alpha)\beta_n} - \beta_u\right) - (\beta_n - \beta_u)\,\mathrm{E}_g(c_j) \\ &+ \kappa\gamma(1-\delta)\,(\beta_n-\beta_u)\,\mathrm{E}_g\left(\frac{1}{q(\theta)}\right) \end{aligned}$$

and, in the case with equal bargaining power ($\beta_n = \beta_u \equiv \beta$):

$$\mathrm{E}_g(w_n) - \mathrm{E}_g(w_u) = -\frac{\beta(1-\beta)(1-\alpha)}{1-(1-\alpha)\beta}\frac{F(g)}{n} < 0.$$

For any distribution g, the workers prefer, on average, to have a unionized firm.

Similarly, with equal bargaining power, the difference in one-period profit is

$$\pi_n(g) - \pi_u(g) = \frac{(1-\alpha)\beta}{1-(1-\alpha)\beta}F(g) > 0.$$

Notice that, as $\alpha \to 1$, the differences in profits and in average wages go to zero.

In general, 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. When bargaining individually, the firm considers producing with or without the *marginal* worker, who has a relatively small impact on the total production. On the other hand, when the firm bargains with the union, the surplus is a function of *total* production, which includes the relatively high production generated by the infra-marginal workers. By forming a union, the workers can extract a part of these high marginal products, which lowers the firm's profit.

In the calibrated model, even though $\beta_n \neq \beta_u$, firms always prefer to be union free. This is consistent with evidence presented by Kleiner (2001) suggesting that firms generally oppose unions. Bronfenbrenner (1994) also details various tactics used by firms to prevent unionization. Hirsch (2004) summarize the literature on union and profitability and concludes that unions have a negative impact on firms' profits.

2.6 Voting procedure

Once a firm has welcomed its new workers, the vote on unionization takes place (see Figure 2). The distribution of workers is now fixed and workers are therefore fully aware of the wages they would get in both outcomes of the vote. Workers are rational and vote only to maximize their own individual utility. Each worker has random preferences on the union status of the firm. One can think that some workers have a negative or positive opinion of unions for reasons that are exogenous to the model. Specifically,

$$\text{Worker } s \text{ votes for a union} \Leftrightarrow w_u(s,g) - w_n(s,g) > \epsilon$$

where ϵ is a logistic random variable drawn independently across all workers. It has mean 0 and scale parameter $1/\rho$.[14]

A law of large numbers applies when aggregating the workers of a given skill. Therefore, a fraction

$$\frac{1}{1+\exp\{-\rho(w_u(s,g)-w_n(s,g))\}}$$

of workers of type s will vote in favor of unionization. By summing up all the voters, we can denote the excess number of workers in favor of unionization by

$$V(g) \equiv \int \frac{g}{1+\exp\{-\rho(w_u(s,g)-w_n(s,g))\}}\, ds - \frac{1}{2}n. \tag{16}$$

With that notation, we get the following condition for unionization:

$$\text{Firm is unionized} \Leftrightarrow V(g) > 0. \tag{17}$$

which simply states that a firm is unionized if a majority of its workers vote for it.

Notice that even though the preferences 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. In fact, the firm is deciding to be unionized or not. Notice also that, as the curvature parameter ρ goes to infinity, the outcome of the vote is decided by the median voter.[15]

The wage equations derived in the last section provide information on what types of workers vote in favor of unionization. Indeed, notice that the union wage (equation 9) is a function of the average characteristics of the workforce while the nonunion wage (equation 13) is a function of the individual characteristics of a worker. In particular, the union wage depends on the average production $F(g)/n$ while the nonunion wage is a function of the marginal product of each worker

$$\frac{\alpha z(s)}{g(s)} F(g).$$

This implies that a worker with valuable characteristics, for instance a high marginal product, would rather bargain individually with the firm than to share his advantage with the other employees.

[14] The CDF of ϵ is $P(\epsilon < x) = 1/(1+\exp(-\rho x))$.

[15] I use random preferences mainly for numerical purposes. The gradient methods used for the optimization perform much better this way.

2.7 Steady state equilibrium

In a steady-state equilibrium, the flows in and out of unemployment in all sub-market s need to be equal:

$$[N(s) - u(s)]\delta = u(s)\theta(s)q(\theta(s)). \tag{18}$$

Using the fact that, in the steady state, a firm j posts vacancies $v(s) = \delta g_j(s)/q(\theta(s))$, it is possible to rewrite the last condition as

$$\frac{N(s)\theta(s)q(\theta(s))}{\delta + \theta(s)q(\theta(s))} = \sum_{j=1}^{j_{\max}} g_j(s, \theta). \tag{19}$$

Definition 1. *A* stationary competitive equilibrium *in this economy is a reservation wage schedule $b(s)$, a labor market tightness schedule $\theta(s)$, a set of workers distributions $\{g_j\}_{j=1}^{j_{\max}}$ and a set of wage schedules $\{w_j\}_{j=1}^{j_{\max}}$ such that,*

1. *g_j solves the optimization problem of firm j,*
2. *w_j solves the collective bargaining problem (equation 9) if firm j is unionized or solves the individual bargaining problem (equation 13) if firm j is not unionized,*
3. *$b(s)$ satisfies equation 2,*
4. *unemployment is stationary in each labor market: equation 19 is satisfied,*
5. *the union status of each firm is consistent with equation 17.*

The full general equilibrium in which firms are constrained by the unionization vote cannot be solved analytically. The numerical algorithm that I use in this paper is detailed in appendix B.[16]

3 Firm's behavior

Now that we have derived the wage schedules $w_n(s, g)$ and $w_u(s, g)$, we can go back to the firm's problem. Fix a firm with a given technology. In what follows, I omit the index j and I use the subscript $i = \{u, n\}$ to denote this firm in a union or nonunion situation respectively. Remember that at the steady state, the problem of a firm is given by equation 5:

$$\max_g \pi_i(g) - \gamma\kappa(1 - (1 - \delta)) \int \frac{g}{q(\theta)}\, ds \tag{20}$$

where π_i is given by equation 15 if the firm is not unionized (condition 17 does not holds) or by equation 11 if the firm is unionized (condition 17 holds). Notice that when a firm hires its workers, it knows whether a given distribution g will lead to a unionized firm or not. In particular, the firm knows the profit it will get.[17]

[16] The question of the uniqueness of the equilibrium is unresolved analytically. Numerically, the equilibrium always seems unique. See Appendix B for more details.

[17] In both the union and the nonunion cases, the current period profit π_i is a strictly concave function of g. It is not however clear whether the unionization constraint defines a convex set or not.

Consider a firm that hires its workers at the beginning of a new period, i.e. a firm in the first stage of the sequence of events shown in Figure 2. Because of the decreasing returns to scale, the firm generally prefers to be union free. In such a case, profits are given by $\pi_n(g)$ and denote by

$$g_n^*(s) = \underset{g}{\text{argmax}} \;\; \pi_n(g) - \gamma\kappa(1-(1-\delta)) \int \frac{g}{q(\theta)}\, ds$$

the distribution of workers that blindly maximizes the firms discounted profit in that situation. Given this distribution, the workers can react in two ways. Either, $V(g_n^*) \leq 0$ and they reject the union. Then, $g_n^*(s)$ is actually the solution of the firm's problem. Or, $V(g_n^*) > 0$ and the workers form a union. Such a firm is *constrained* by the unionization vote and, in this case, $g_n^*(s)$ is not a solution to the firm's problem.

A constrained firm can try to fight the union. To do so, it distorts the distribution g_n^* in the least costly way possible to avoid unionization. Denote that distorted distribution by g_n. Therefore,

$$\begin{aligned} g_n = &\underset{g}{\text{argmax}} \;\; \pi_n(g) - \gamma\kappa(1-(1-\delta)) \int \frac{g}{q(\theta)}\, ds \\ &\text{subject to: } V(g) \leq 0. \end{aligned}$$

Obviously, by imposing the voting constraint on the firm's problem, the discounted profit of the firm goes down. The question is how far down? In particular, if the constraint is important enough, the profit of the firm is higher with the optimal union distribution of workers g_u^*. Such a firm will therefore be unionized. Notice that the firm is *rationally* choosing to be unionized. It is an optimal reaction to the threat imposed by the union.

In the next section, I first consider the firm's behavior when the unionization constraint is not binding and then move to the full constrained problem as a deviation from this case. Solving the full distorted problem of the firm must be done numerically.

3.1 Behavior of an unconstrained firm

I consider in this section the decision of a firm j that is not constrained by the unionization vote. In other words, suppose a firm can decide on the outcome of the unionization vote. How would it behave? First, the firm compares the profit it would make in the union and in the nonunion cases and then picks the case providing the highest profit.

The goal of this exercise is to understand the decision process of the firm. I therefore assume that the equilibrium is fixed: $b(s)$, $\theta(s)$ and $c_j(s)$ are fixed.

By combining equations 11 and 15, we can write the problem of a firm that is not constrained by the unionization vote as:

$$\max_g \Gamma_i F_j(g) - (1-\beta_i) \int c_j \cdot g\, ds - (1-\gamma(1-\delta)(1-\beta_i))\kappa \int \frac{g}{q(\theta)}\, ds$$

where

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

is the share of output retained by the firm. By defining,

$$\mathrm{MC}_i^j(s) \equiv (1-\beta_i)c_j(s) + (1-\gamma(1-\delta)(1-\beta_i)))\frac{\kappa}{q(\theta(s))} \tag{21}$$

as the marginal cost paid by a firm j to hire a worker s, the firm's optimal hiring decision, g_i^* is given by

$$\mathrm{MC}_i^j(s) = \Gamma_i \frac{\alpha F(g_i^*)z(s)}{g_i^*(s)}. \tag{22}$$

Notice that the firm has a different hiring strategy whether it expects its workers to unionized or not. The right-hand side of this last equation is simply the marginal cost of hiring an extra worker of type s, which includes the wage paid to the worker, while the left-hand side is the share of the marginal product of the worker that the firm retains. Notice that MC_i^j depends on the firm's equilibrium wage, on aggregate variables and on the union status of the firm. From equation 22, we see that workers who are rare ($\theta(s)$ high) or who have attractive outside options ($b(s)$ high) are expensive to hire ($\mathrm{MC}_i^j(s)$ high) and the firm therefore relies less on them for production ($g_i^*(s)$ small). The equilibrium wage schedule w_j also affects the marginal cost through c_j: a worker who knows he will get a high wage in the next period has more to lose if the bargaining breaks down.

It is straightforward to compare the discounted profit of the firm in both the union and nonunion scenarios:

Lemma 4. *An unconstrained firm j prefers to be union free if and only if*

$$\log\left(\frac{\Gamma_n}{\Gamma_u}\right) \geq \alpha \int z(s) \log\left(\frac{\mathrm{MC}_n^j(s)}{\mathrm{MC}_u^j(s)}\right) ds.$$

The term on the left hand side of this equation is a measure of the relative hiring costs in both the union and nonunion scenarios. The right hand side is a measure of the relative share of output that the firm retains. If $\beta_n = \beta_u$, $\mathrm{MC}_n = \mathrm{MC}_u$ and this condition is automatically satisfied. Also, the firm prefers to be union free when the union is very strong ($\beta_u \to 1$) and it would gladly welcome a union if individual workers have a strong bargaining power ($\beta_n \to 1$), as the intuition would predict. As we will see, this condition holds for all firms in the calibrated economy.

We now focus on wages and on how workers vote.

Lemma 5. *Assume that the labor market tightness schedule $\theta(s)$ and the outside option schedule $b(s)$ are increasing functions of the skill and that the bargaining powers β_n and β_u are equal. Then, in an unconstrained firm hiring according to g_i^* for $i = \{u, n\}$, the nonunion wage schedule $w_n(s, g_i^*)$ is an increasing function of s.*

The intuition for this lemma is straightforward. The firm hires until the marginal product of a worker is equal to his marginal cost. Under the lemma's assumption, the marginal cost is increasing in s and the result follows since nonunion wages depend directly on the marginal products. This lemma states that for all firms (those that are unionized and non unionized in equilibrium), the nonunion wage they would pay if they hire according to the optimal distribution g_u^* and g_n^* is an increasing function of s.

The following lemma characterizes the vote of the workers:

Lemma 6. *Assume that the labor market tightness schedule $\theta(s)$ and $c_j(s)$ are increasing functions of the skill.*[18] *Under the optimal hiring decision of unconstrained firms g_i^* for $i = \{u, n\}$, the union wage gap $w_n(s, g_i^*) - w_u(s, g_i^*)$ is increasing with s.*

Once again the intuition is straightforward. Since the marginal cost is increasing with skill, the nonunion wage of high-skill workers is higher then the one of low-skill workers. Instead, in a union firm, wages are determined by the average marginal product. This generates a union wage gap that is an increasing function of s.

Lemma 6 is consistent with the findings from Farber and Saks (1980) that the desire to be unionized goes down with the position in the intrafirm earnings distribution. It is also consistent with the large empirical literature suggesting that a union compresses the wage distribution of a firm.

Finally, it is possible to compare the variance of union and nonunion wages in unconstrained firms:

Lemma 7. *Assume that the labor market tightness schedule $\theta(s)$ and $c_j(s)$ are increasing functions of the skill and that the bargaining powers β_n and β_u are equal. Then, for a given firm j,*

$$\mathrm{Var}_{g_n^*}\left(w_n(s, g_n^*)\right) \geq \mathrm{Var}_{g_u^*}\left(w_u(s, g_u^*)\right)$$

where $\mathrm{Var}_g(x)$ is the traditional variance operator taken with the normalized distribution $g/\int g\, ds$.

This lemma characterizes the variance of wages in firms evolving in a policy environment in which unions are mandatory (unconstrained union) or illegal (unconstrained nonunion).

Figure 3 shows the decision process of a firm. The parameters that generate this example are picked to emphasize the various mechanisms. Their magnitudes are not realistic but the mechanisms are the ones present in the calibrated model. Panel A presents the two optimal distributions g_n^* and g_u^*. Panel B and C show the wages that voters are considering when they cast their votes. The vertical lines show the position of the worker who is indifferent between a union and a nonunion firm. There is only one single skill for which this is true. As the lemmas predicted, the wages and the union wage gap are all increasing function of s. Notice that because of this increasing wage gap, the unionization of a firm directly lowers the variance of wages.

In this example, if the firm hires according to g_n^* or g_u^*, the workers will vote in favor of a union. Therefore, g_n^* does not solve the problem of the firm. Notice also that g_n^* is simply a rescaled version of g_u^*. This comes directly from equation 22. Also, $g_n^*(s) > g_u^*(s)$ for all $s \in [0, 1]$. This is a direct consequence of the nonunion bargaining. Since the firm bargains individually with each worker over the marginal surplus, it increases the number of workers to lower this marginal product. This effect was described in Stole and Zwiebel (1996a,b).

3.2 Fighting the union

Because of the additional profit they generate by bargaining individually, firms generally prefer to be union free. The workers, however, have a strong incentive to form a union: by bargaining collectively, they extract a bigger share of the match surplus. In this section, we consider the case of a firm that is constrained by the unionization vote: $V(g_n^*) \geq 0$.

[18] The two schedules θ and c_j are increasing in the calibrated model.

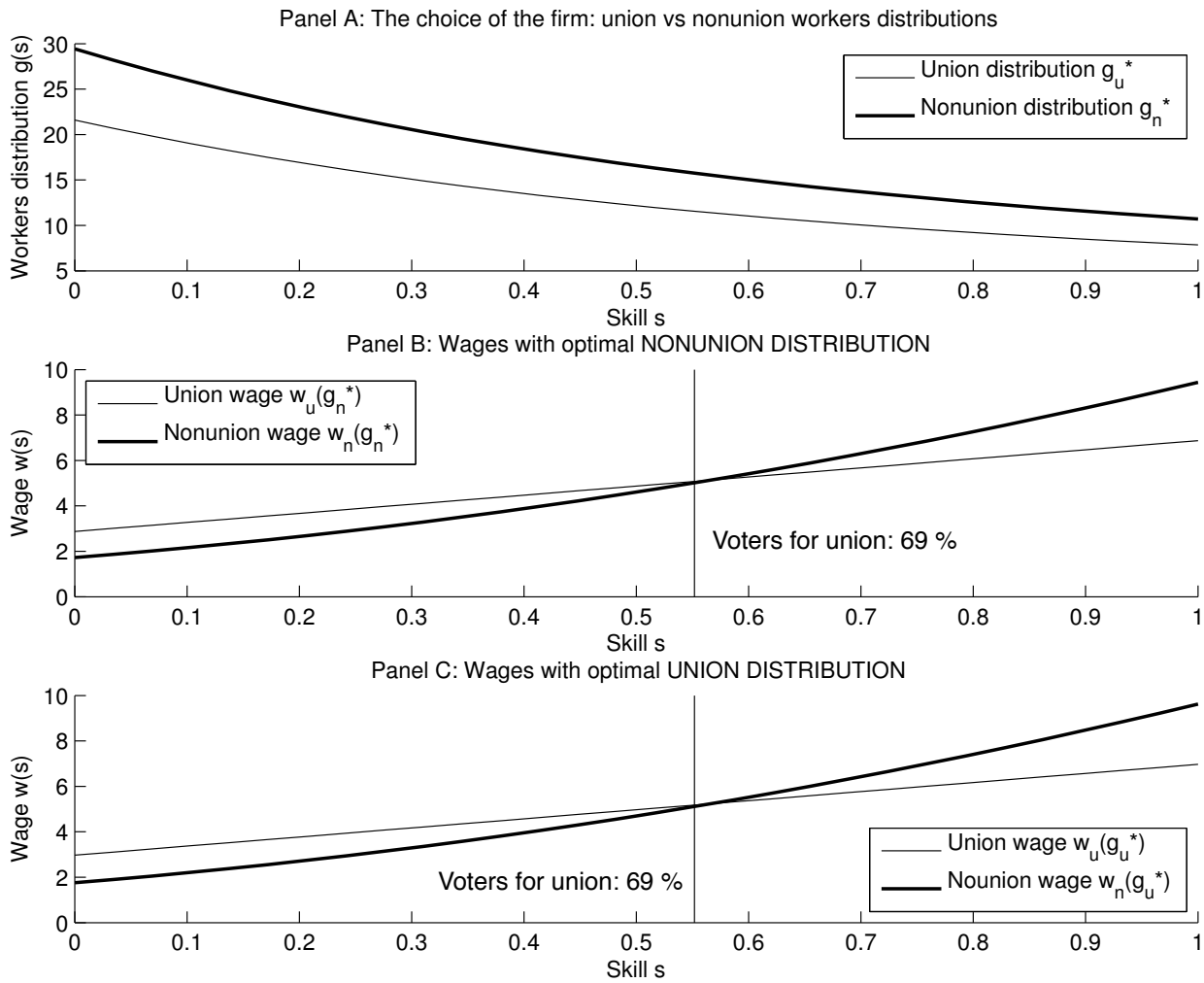

Figure 3: The hiring decision of an unconstrained firm. The vertical lines represent the position of the indifferent voter. Parameters of the economy: $\delta = 0.025$, $\gamma = 0.995$, $\beta_n = 3/10$, $\beta_u = 3/10$, $c_j(s) = 1 + 4s$, $\theta(s) = 1 + 9s^2$. Firm characteristics: $A = 10$, $\alpha = 0.8$, $z(s) = (s + 1)/2$.

When a firm is constrained by the vote on unionization, the first order condition given by equation 22 is modified to include the impact of the additional worker on the outcome of the vote:

$$\mathrm{MC}_n^j(s) = \Gamma_n \frac{\alpha F(g_n) z(s)}{g_n(s)} - \lambda_n \frac{\partial V(g_n)}{\partial g_n(s)} \tag{23}$$

where $\lambda_n > 0$ is the Lagrange multiplier of the voting constraint and where g_n is the optimal distribution for which workers actually reject the union.

When distorting the unconstrained distribution g_n^* the firm takes into consideration three mechanisms:

1. **Fraction of voters for union** By adding more workers at the top of distribution, or by removing workers at the bottom of the distribution, the firm directly lowers the fraction of workers in favor of union.

2. **Effect on nonunion wages** By increasing the number of workers of a given skill s the firm lowers the marginal product of these workers which, in turn, lowers their nonunion wage. If the firm increases the number of workers who vote against the union, it needs to make sure that their nonunion wage stays higher than their union wage. Otherwise, these workers will change their vote.

3. **Effect on union wages** While nonunion wages are determined by the marginal products, union wages are determined by the average product. By increasing the number of high-skill workers, for instance, the firm increases the number of high marginal product workers which shifts the union wage upwards. This could make some workers change their vote in favor of unionization. This effect, through the union wages, implies that when the firm wants to increase the numbers of workers against the union it will first do so with the workers of the *lowest skill possible*.

Figure 4 shows the decision process of the constrained firm that was represented on figure 3. The firm considers the profit it makes in two scenarios: optimal distribution under which the workers will unionize and optimal distribution under which the workers will reject the union. These are featured on Panel A by the thin and thick line respectively. Panel B shows the wages that voters are considering if the firm imposes g_n. The dashed line represents the nonunion wages when there was no unionization constraint (it is the same curve as the thick line of the Panel B of figure 3). Panel C shows the union and nonunion wages the workers are voting on when the distribution is g_u^*. Notice that the firm does not have to distort g_u^* here. The workers gladly form a union. In Panel B and C, the vertical lines represent the indifferent voter. On Panel B, the fraction of workers against unionization is 50%.

We can see on Panel A that the distorted distribution g_n has a lower number of low-skill workers and a higher number of high-skill workers then the distribution g_n^* from figure 3. The effect of this distortion on wages is clear by looking at Panel B. The constraint lowers the nonunion wage of high-skill workers and increases the wages of low skill workers. By reacting to the fact that the workers *can* unionize, the firm compresses the range of wages it is paying its workers. Notice that this threat effect could be present in an economy in which no firms are unionized but in which the legal system allows the workers to create unions.

We can see on Panel B that the union and nonunion wages of the workers with skill between 0.48 and 0.76 are almost identical. For this range of skills, the behavior of the firm is clearly constrained. The firm would like to hire more of these workers: they vote against the union and their relatively small marginal product has a smaller effect on the union wage schedule than the workers with higher skill. However, if the firm were to hire an additional worker s in this zone, all the workers of type s would change their vote in favor of unionization.

Table 1 compares different characteristics of the firm under the three following scenarios:

1. Unions are mandatory
2. Unions are legal
3. Unions are illegal

First, notice that the firm's profit is highest when unions are illegal and that, when unions are legal, the firm still manages to find a nonunion distribution with higher profit then in the union case. This firm is therefore union free in scenario 2. Second, the fraction of voters in favor of a

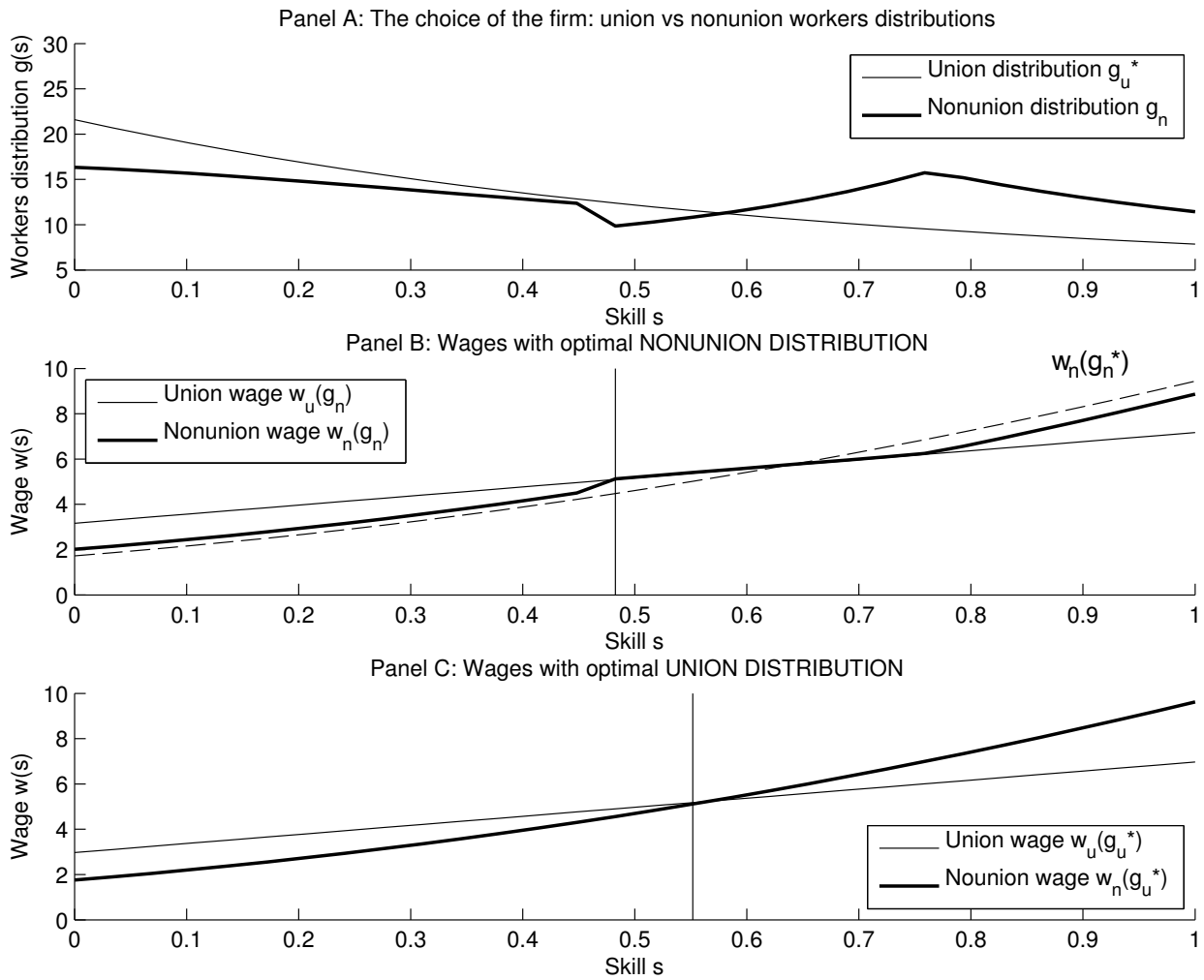

Figure 4: The hiring decision of the firm facing a unionization threat. The vertical lines represent the position of the indifferent voter. Parameters of the economy: $\delta = 0.025$, $\gamma = 0.995$, $\beta_n = 3/10$, $\beta_u = 3/10$, $c_j(s) = 1 + 4s$, $\theta(s) = 1 + 9s^2$. Firm characteristics: $A = 10$, $\alpha = 0.8$, $z(s) = (s + 1)/2$.

union is the same in scenario 1 and 3. In scenario 2, the firm pushes workers to vote against the union until it reaches 50%. Third, the mean of wages is the lowest in scenario 3. This comes from the differences in the bargaining structure. When bargaining individually, the firm is able to retain a higher fraction of the joint surplus. Making unions legal leads to an increase in the mean of wages. Fourth, the variance of wages is the highest when unions are illegal. Allowing the presence of unions brings down the variance. This is the wage compression effect of the unionization threat. Finally, the variance of wage is lowest in scenario 1. The differences in the bargaining structures are such that a union firm has a lower variance than nonunion firms.

Notice finally that, in this example, the unionization threat lowers the *range* of nonunion wages when compared to an unconstrained firm. However, the impact of the threat on the *variance* of wages needs to also take into account the changes in the distribution of workers. In general, the threat lowers the variance but this might be reversed in firms that are extremely constrained by

the vote. In these firms, the number of high-skill workers needs to be increased so much that w_n is almost equal to w_u for all the workers rejecting the union. The firm therefore tries to shift the w_u schedule downward by adding to its workforce a large number of workers with very low-skill. This leads to a U-shaped distribution of workers in these firms.

	1. Union mandatory	2. Union legal	3. Union illegal
Union status of the firm	Union	Nonunion	Nonunion
Firm discounted profit	2179	2397	2969
Fraction of voters for union	69%	50%	69%
Mean of wages	4.50	4.88	4.18
Variance of wages	1.37	3.69	4.76

Table 1: The behavior of a firm under the three scenarios.

3.3 Impact of technology on unionization

Remember that, with equal bargaining powers, a firm always prefers to be union free and that this preference is independent of its technology. Technology has, however, a strong influence on the vote of the workers and, through that channel, on the union status of the firm. The following lemma characterizes how the returns to scale α affect the workers vote.

Lemma 8. *For an unconstrained firm employing a distribution of workers g_n^* (given by equation 22):*

$$\frac{d(w_n(s, g_n^*) - w_u(s, g_n^*))}{d\alpha} = \frac{\beta_u}{\alpha^2 \int z(s)/\mathrm{MC}_n^j(s) ds} > 0. \tag{24}$$

Also, the fraction of voters in favor of a union, $V(g_n^*)$, is such that

$$\frac{dV(g_n^*)}{d\alpha} = -\frac{\beta_u}{\alpha^2 \left(\int z/\mathrm{MC}_n^j \, ds\right)^2} \int \frac{z}{\mathrm{MC}_n^j} \frac{\exp\{-\rho(w_u - w_n)\}}{(1 + \exp\{-\rho(w_u - w_n)\})^2} \, ds < 0$$

Increasing α increases the gap between w_n and w_u uniformly across skills. All else equal, workers in a firm with a low return to labor α tend to have a bigger advantage to be unionized. The second part of the lemma implies that, as α gets bigger, the share of workers in favor of forming a union goes down. Also, since the gap in wages is smaller, it is easier for the firm to convince the workers to vote against unionization by distorting the distribution of workers. This preference of the workers is consistent with the findings of Hirsch and Berger (1984) that industries that are more capital intensive have a higher share of union workers.

Figure 5 shows the firm's decision as a function of its technology. Panel A presents the contour curves of the ratio of nonunion (constrained) profit to union profits as a function of α and $z(s)$. The distributions $z(s)$ used to draw this picture are shown on Panel B, so that a small beta distribution parameter (on the horizontal axis of Panel A) indicates a distribution z that is skewed to the left (for instance, $d = -0.5$ on panel B). It is clear from this figure that the ratio of profits is an increasing function of the return to scale α. This result holds true in the calibrated economy.

The impact of the skill intensity $z(s)$ on the union status of the firm is less obvious. Two effects are competing. The first one relates to the number of voters, the second one to the average

marginal product of the workers and therefore to the union wage schedule. Consider a firm with a skill intensity z highly skewed towards low skill workers (for instance, $d = -0.5$ on panel B of figure 5). In such a firm, the median voter has a lower skill than the average voter, which tends to push the firm towards unionization. However, since most workers have low marginal product, the average marginal product is small and so is the union wage that workers would get in case of unionization. These two effects compete with each other. A firm that has a production technology skewed towards high-skill workers has to deal with the exact same two effects: many high-skill workers tend to vote against unionization but their high marginal products pushes union wages up. By looking at figure 5, it appears that the voting effect dominates for most technologies. In other words, by moving the median of $z(s)$ to the right, it gets easier and easier for the firm to fight unionization. However, when the median passes a certain point, the wage effect dominates. Union wages get so high that compensating for them becomes harder and harder. The ratio of profits therefore goes down.

Finally, the following lemma shows that the parameter A of the production function has no influence on the union status of a firm or on the wages it pays.

Lemma 9. *Consider two firms, identified by the subscripts 1 and 2, that have identical technologies except for $A_1 \neq A_2$. There exist an equilibrium such that: if g_1 solves the problem of firm 1, then*

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

solves the problem of firm 2. Also, both firms have the same union status and pay the same wages.

This lemma will also be useful in aggregating firms of the same type in the calibration.

4 Data and calibration

I calibrate the model on the private sector of the United States in 2005. One period is one month and the unit of all monetary amounts is one thousand dollars. I set the monthly discount rate to $\gamma = 0.996$ and the probability of job destruction to $\delta = 0.027$. For the matching function, I follow Hagedorn and Manovskii (2008) and use $q(\theta) = (1 + \theta^{\mu})^{-1/\mu}$. To estimate μ, I use data from the Job Opening and Labor Turnover Survey (JOLTS) for 2005 together with the probability of job finding from Shimer (2007).[19] The estimate for μ is 1.33. I set the scale parameter of the random preferences for unionization to $\rho = 20$.[20] This is a strong curvature that brings the firm close to the median voter case. Table 2 summarizes the parameters.

In what follows, all data about individuals is coming from the Merged Outgoing Rotation Groups of the Current Population Survey (CPS) as it is made available by the National Bureau of Economic Research (NBER). Industry data comes from the Bureau of Economic Analysis (BEA).

[19] I take the yearly average of the job finding probability for 2005, its value is 0.4136. Robert Shimer constructed this data. For additional details, please see Shimer (2007) and his webpage. I find the aggregate labor market tightness by using the vacancies and job searchers numbers provided by the JOLTS. I take the average over all months in 2005 of the seasonally adjusted non-farm job opening series together with the seasonally adjusted number of job searchers for part-time and full-time jobs. I find μ by solving $0.4136^{\mu} = \theta^{\mu}(1 + \theta^{\mu})^{-1}$.

[20] The bigger ρ is, the closer we get to the median voter scenario. However, with ρ too high, the numerical algorithm struggles.

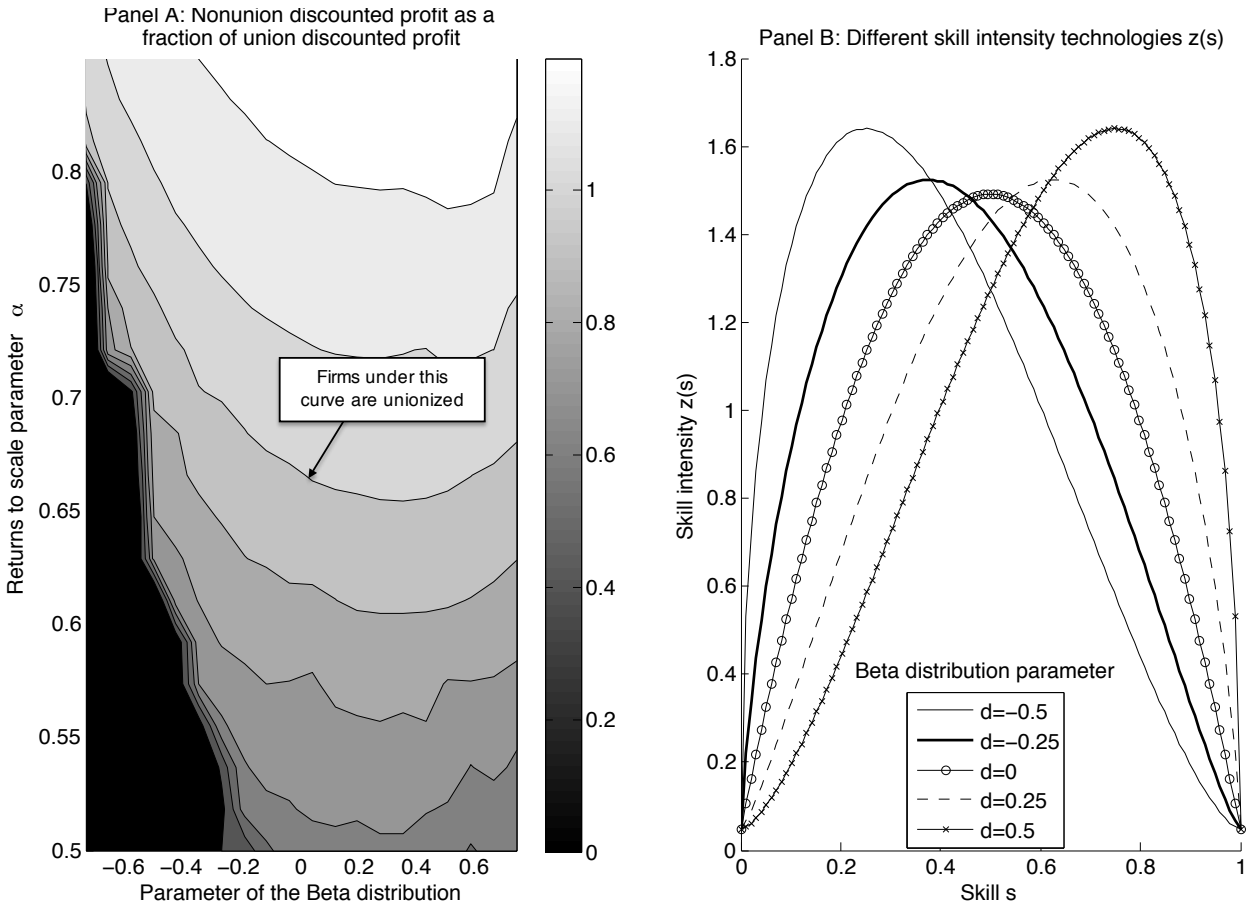

Figure 5: The profits of a firm as a function of its production technology. The left panel shows the ratio of constrained nonunion profit to union profit. The lines are contour lines showing where the ratio crosses specific thresholds. The mapping of technology to profits ratio is continuous. The vertical axis is labor share α. The horizontal axis represents different function $z(s)$ taken from the Beta distributions shown in Panel B. The distributions are Beta$(2-d, 2+d)+\epsilon$ where d is the parameter on the horizontal axis and $\epsilon > 0$ is a small number to prevent the distribution from reaching 0 at $s=0$ and $s=1$. The parameters of the economy: $\delta = 0.025$, $\gamma = 0.995$, $\beta_n = 3/10$, $\beta_u = 3/10$, $c_j(s) = 1+4s$, $\theta(s) = 1+9s^2$.

Calibration strategy

For simplicity, I assume that firms are endowed by two types of technology. In equilibrium, one type of firm is unionized while the other is not. Lemma 9 shows that it is equivalent to change the number of firms of a certain type or the parameter A_j of these firms' technology. We can therefore normalize the number of firms of each type to one and change A_j to adjust their size. I denote the technologies of union and nonunion firms by (A_u, α_u, z_u) and (A_n, α_n, z_n) respectively. The CPS provides information on the union status of the workers as well as the industry they are working in. By using this data together with the BEA data on industry, I set α_u and α_n to match the labor

Parameter	Definition	Value	Source/reason
γ	Discount factor	0.996	5% annual interest rate
δ	Probability of job destruction	0.027	Shimer (2007)
μ	Parameter of the matching function	1.33	JOLTS with Shimer (2007)
ρ	Parameter of preference for union	20	Strong curvature

Table 2: Parameters taken directly from the data or the literature

shares. The values of the returns to scale parameters are $\alpha_u = 0.5$ and $\alpha_n = 0.6$.[21]

In what follows, the adjective *empirical* (for instance, the empirical union wage distribution) designates a variable taken directly from the data. These variables are denoted by the superscript *emp*. On the other hand, a *calibrated* variable designates that variable as generated by the calibrated model. The superscript *cal* designates them.

The first step of the calibration is to define a skill index. Then, I use the CPS data to construct the labor market tightness $\theta(s)$ and an the outside option schedule $b(s)$. I then minimize a loss function to find the remaining parameters of the model.

Skill distribution

The skill index is only, well, an index. Throughout this model it is used to characterize the heterogeneity of the agents and to identify variables that are related to them ($\theta(s), b(s)$, N(s), etc.). Nowhere is s appearing alone; s does not mean anything by itself. I first define a skill index from the data and then, when minimizing the loss function, identify the firms' technology using equations from the model. This way, the skill index and the firms technology are consistently determined to make the model match the wage schedules and the distributions of workers.

I use data from the CPS to build the skill index. To do so, I run a regression of the log of normalized monthly *nonunion* wages on two types of variables. The first type includes variables related to each individual. The second type of variables depend on the industry in which the individual works. I then use the predicted variable given by the OLS estimator of the individual characteristics alone as the skill index. Explicitly, denote by w_i the log monthly wage of agent i, who is working in industry $j(i)$. The regression is

$$w_i = \Gamma X_{1,i} + \Psi X_{2,j(i)} + \epsilon_i.$$

and the skill index is therefore given by the predicted values $\hat{s}_i = \hat{\Gamma} X_{1,i}$. The individual characteristics X_1 are *sex*, *age*, *race*, *education* and *occupation* (set of dummy variables). The job related characteristics X_2 are *industry* (dummy variables) and the current *US state* in which the agent lives.[22] I drop from the sample individuals with skill index below the first percentile and above the 99th percentile. I then scale the index using a linear transformation such that s is between 0 and 1. Notice that even though the regression is run only on nonunion workers, the predicted values

[21] I pick the α's by hand to have less degrees of freedom in the loss function. Once the model is fully calibrated I compare the calibrated labor shares to the empirical ones. The biggest difference is of 3%.

[22] Including *US state* as an individual characteristic instead has minimal impact on the distribution. For *industry* and *occupation*, I use the variables generated by the NBER. Both are at the 3-digit level. I clean the sample by removing agricultural workers and individuals with hourly wage higher than 150$ or lower than 5$. I also remove individuals younger than 20 or older than 65 years old.

$\hat{s}_i$ are computed for *all* members of the labor force. Figure 6 shows the distribution of $\hat{s}_i$ for the whole sample.

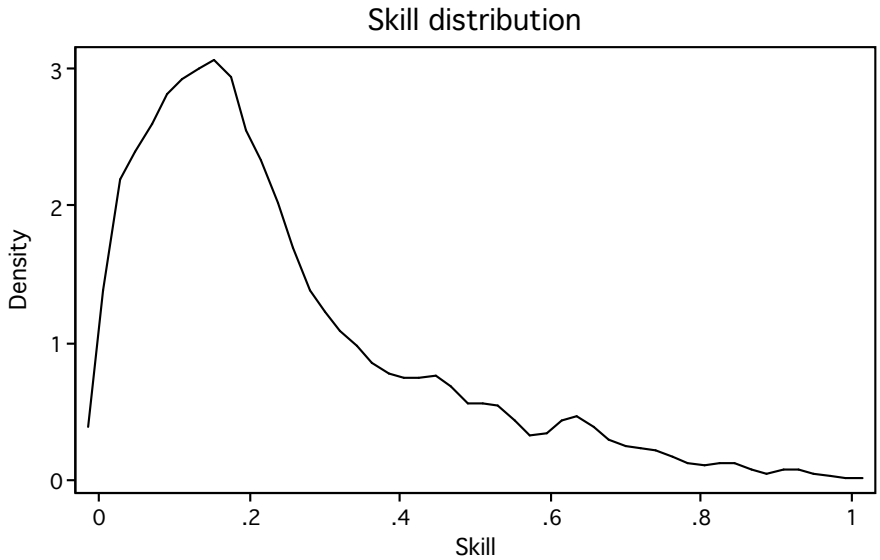

Figure 6: Skill distribution

This way of defining the skill distribution has the advantage of making the empirical wages and the empirical labor market tightness increasing with s. This makes the interpretation of the impact of unionization on different workers more intuitive.

Labor market tightness and value of outside option

In the United States, unemployment insurance programs are administered by the states. Krueger and Meyer (2002) provides the main characteristics of benefits for some US states in 2000. The replacement ratio is about 50% in every state but the maximum weekly benefits vary considerably. In the model, the variable b_0 also takes into account home production and the value of the extra leisure provided by unemployment, two elements that are harder to quantify. Hall and Milgrom (2008) uses an estimate of the flow value of non-work that is essentially equivalent to a replacement ratio of 71%. I therefore calibrate $b_0(s)$ to be 71% of the mean wage earned by workers of skill s.[23]

It is straightforward to identify the empirical value of some of the aggregate variables. I split the support of the skill distribution in 20 bins of equal sizes and use equation 18 together with the

[23]This estimate takes into consideration the value of the extra leisure associated with unemployment. One might however suspect that the replacement ratio changes with skills. The unemployment insurance programs tend to be more generous with agents earning low incomes. To evaluate the effect of this possible bias, I used data from the Uniform Extracts of the U.S. Census Survey of Income and Program Participation (SIPP) and looked at the unemployment benefits of individuals going from employment to unemployment. The SIPP also provides information on wages and it is therefore possible to build a measure of the replacement ratio as a function of the wage. The fitted replacement ratio of an average worker earning a monthly wage of 1000$ and moving to unemployment is 65%. The ratio then decreases quadratically to reach about 10% for monthly wages of 8000 $. Because high-skill workers stay unemployed for a very short time, their reservation wage is basically the same as when b_0 is taken to be 71% of the mean wage. In Hall and Milgrom (2008), 71% corresponds to the ratio of the value of unemployment on productivity.

observed unemployment rates by bin to compute the labor market tightness θ for workers in each of these bins.[24] Using the mean wages of union and nonunion workers together with the fact that firms hire a fraction δ of their workforce every period, I compute the expected wage of a worker who just found a job. I then use equation 2 to compute the outside option b for each of the skill bins. Similarly, by summing the number of agents in each bin, I compute the empirical skill distribution $N^{\text{emp}}(s)$. I also compute $b_0^{\text{emp}}(s)$ in each of the bins.

Loss function

I pick κ, β_n, β_u to minimize the following loss function:

$$\text{Loss} = \int \left(N^{\text{cal}} - N^{\text{emp}}\right)^2 ds + \lambda_b \int \left(b_0^{\text{cal}} - b_0^{\text{emp}}\right)^2 ds + \lambda_n \left(\text{LS}_n^{\text{cal}} - \text{LS}_n^{\text{emp}}\right)^2 + \lambda_u \left(\text{LS}_u^{\text{cal}} - \text{LS}_u^{\text{emp}}\right)^2$$

where LS_n and LS_u are the labor share[25] in the nonunion and union firms respectively and where the λ's are weighting constants picked such that the terms of the loss function have similar magnitudes.

For any κ, β_n, β_u, I use equations 22 and 13 to identify the firms technology. By using the empirical variables, these equations are

$$g_u^{\text{emp}}(s) = \frac{\alpha_u(1-\beta_u)F_u(g_u^{\text{emp}})z_u(s)}{(1-\beta_u)c_u(g_u^{\text{emp}}, s) + \frac{\kappa}{q(\theta(s))}(1-\gamma(1-\beta_u)(1-\delta))}$$

$$w_n^{\text{emp}}(s) - b(s) = \frac{1-\gamma(1-\delta)}{1-\beta_n\gamma(1-\delta)}\left(\frac{\beta_n}{1-(1-\alpha_n)\beta_n}\frac{\alpha_n z_n(s)}{g_n^{\text{emp}}(s)}F_n(g_n^{\text{emp}}) - \beta_n b(s) + (1-\delta)\frac{\beta_n\kappa\gamma}{q(\theta(s))}\right)$$

where

$$c_u(g_u^{\text{emp}}, s) = b(s) - \frac{\beta_u\gamma(1-\delta)}{1-\beta_u\gamma(1-\delta)}\frac{1}{n^{\text{emp}}}\left(F_u(g_u^{\text{emp}}) - \int b \cdot g_u^{\text{emp}}\, ds + \gamma(1-\delta)\kappa\int \frac{g_u^{\text{emp}}}{q(\theta)}\, ds\right).$$

I use a fixed point algorithm to find the technologies (A_u, z_u) and (A_n, z_n) from these equations (remember that the integral of z is normalized to one and that F depends on $z(s)$ and A). Notice that by defining (A_u, z_u) this way, the model replicates the distribution of workers in union firms perfectly. That is, $g_u^{\text{emp}} = g_u^{\text{cal}}$.

The idea behind the calibration is straightforward. For any vector of parameters $(\kappa, \beta_n, \beta_u)$, I identify the firms' technology using the model and I can compute the equilibrium. The schedules N^{cal} and b_0^{cal} are those that support this equilibrium. I pick $(\kappa, \beta_n, \beta_u)$ to make N^{cal} and b_0^{cal}, as well as the labor shares, as close as possible to their empirical counterparts.[26] This amounts to calibrating the wage schedules (through b_0) and the distributions of workers (through N).

Table 3 shows the parameter values that minimize the loss function.

[24] One point of the θ schedule departs strongly from the trend. I therefore use a moving average to smooth it.

[25] The empirical labor shares are computed by dividing total workers' compensation by value added using the data provided by the Bureau of Economic Analysis. The calibrated labor shares are computed by dividing the wage bill by production minus hiring costs. I also calibrated the model using a loss function without the labor share terms. The calibrated parameters have different values but the policy exercises give very similar results.

[26] I also calibrated the model using the inverse strategy: fix b_0 and $N(s)$ to their observed values and find the vector $(\kappa, \beta_n, \beta_u)$ that makes θ^{cal} and b^{cal}, as well as the labor shares, as close as possible to their empirical counterpart. With this other calibration, the fit is a bit worse and the effects of unions on the economy are very similar qualitatively and quantitatively. This other approach is however more computationally intensive, which limits the size of the skill grid.

β_n	β_u	κ
0.29	0.06	0.14

Table 3: The value of the calibrated parameters. The units of κ is thousand dollars per month.

Notice that $\beta_n > \beta_u$. This difference in bargaining powers is necessary to compensate the fact that the decreasing returns provide the unionized workers with more leverage in the negotiations. In the calibrated model, workers always prefer to form a union and the firms need to fight to prevent unionization. Figure 7 shows the calibrated technologies $z(\cdot)$ of the firms. We see that nonunion firms are more intensive in high-skill workers. This comes from the fact that, in the data, the distribution of nonunion workers has a fatter tail than the one of union workers.

Figure 8 shows how the model fits the distributions of workers and the wage schedules. We can see that the model fits the union workers distribution perfectly. This is a direct consequence of the way the technology of the union firms is identified. The model also fits the nonunion wage schedule and nonunion distribution of workers quite well. The fit of the union wage schedule is however less precise. This comes from the heavy structure imposed on union wages by equation 9. In particular, the shape of w_u is tightly linked to the shape of $b(s)$. Union wages in the calibrated economy are more unequal than in the data. Suggesting that the real equalizing effect of unions might be stronger than the one captured by the calibration. A better fit could be obtained by allowing different workers to have different bargaining powers when they divide what the union extracted from the firm.

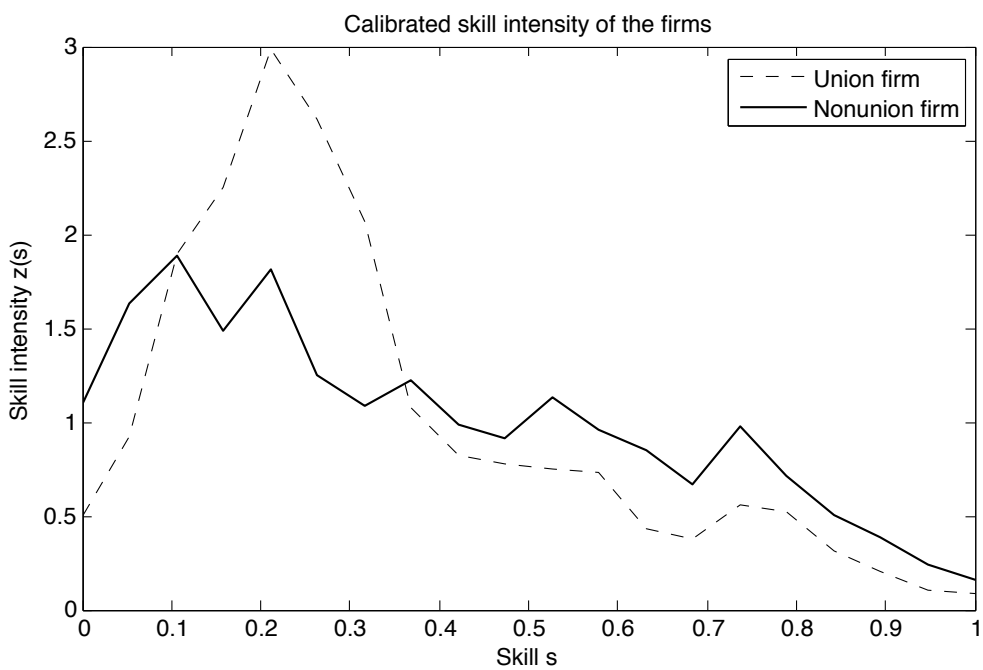

Figure 7: Calibrated skill intensities $z_n(s)$ and $z_u(s)$.

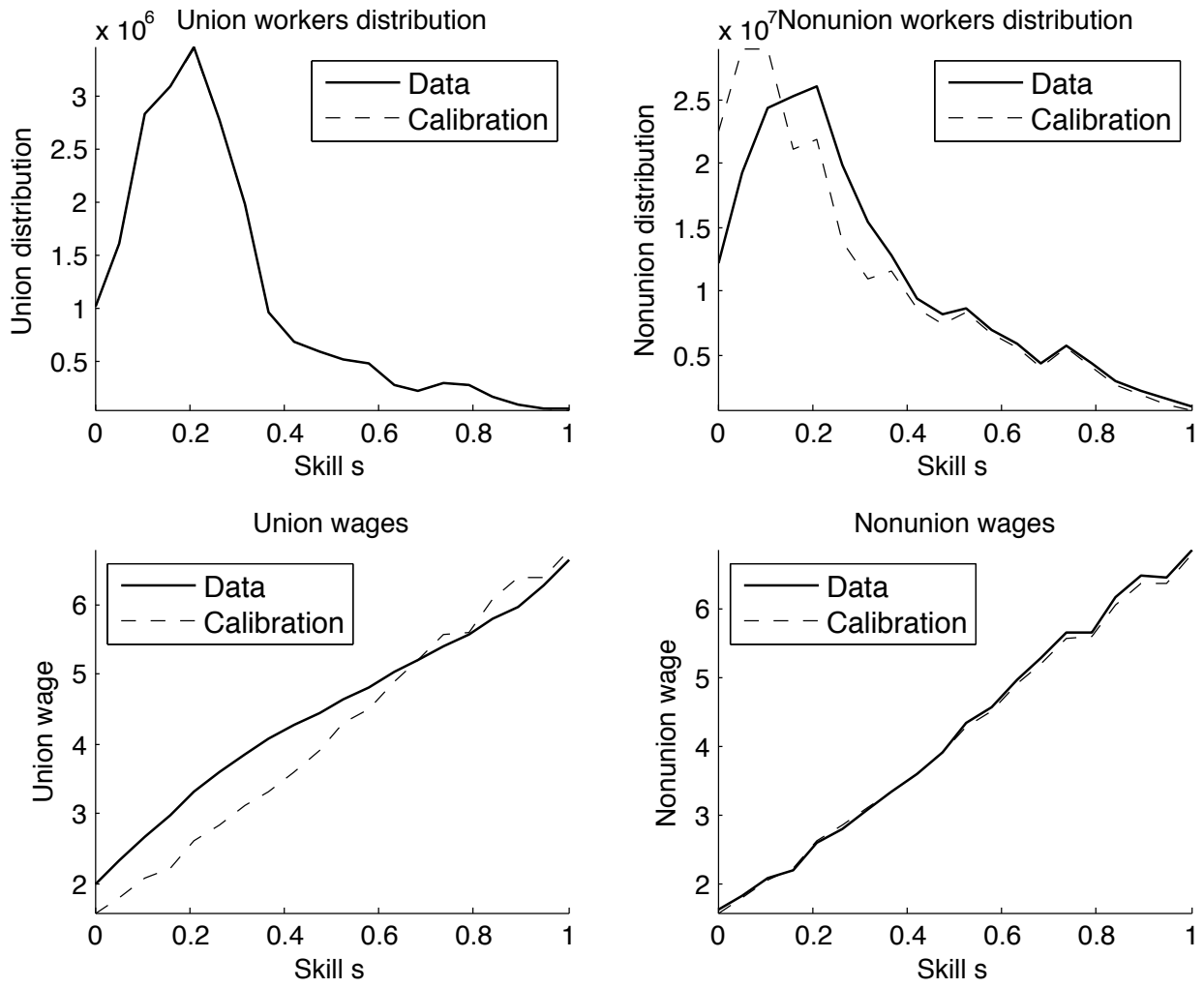

Figure 8: Fit of the calibrated model.

Partial equilibrium estimate

With the calibrated model, we can use the partial equilibrium estimator to compute the impact of unions on wage inequality according to a conventional econometric technique. To do so, I compute a new counterfactual wage distribution by giving to each union worker the wage paid to workers of his skill working in a nonunion job. Therefore each union worker s is given his nonunion wage $w_n(s)$. The new wage distribution has a slightly smaller variance. According to this estimate, unions would be responsible for a reduction in the variance of log wages of 0.4%. This implies that the inside-group effect of unions is larger than the between-group effect. In other words, the effect of the smaller variance of union wages is bigger than the effect coming from the difference in means between union and nonunion wages. This estimate is somewhat different from the ones found in the literature. These differences might come from the fact that the present model abstracts completely from the public sector, in which the unionization rate is much higher than in the private sector. In fact, the classical two-sector estimator of Freeman (1980), when applied to the cleaned data set, also finds that unions lower the variance of log wages by 0.4%.

5 Impact of unions

I do two comparative statics exercises using the calibrated economy:

1. Removing the union threat All union firms stay unionized and all nonunion firms stay union free but the latter do not have to worry about the unionization vote anymore.

2. Outlawing unions Unions are completely eliminated from the economy.

In both cases, I compute the new steady state general equilibrium. Figure 9 shows the two new equilibria. The top two graphs show the percentage change in union and nonunion wages from the calibrated wage schedules. We see that removing the threat of unionization increases wage inequality by increasing high wages more then low ones. Consider first the modification of the nonunion wage schedule when the union threat is gone. The change comes directly from the reaction of nonunion firms. They change their demand for workers, which leads to higher wages for high-skill workers and lower wages for low-skill workers. This has a direct effect on the value of unemployment $b(s)$, which, in turn, modifies the wages paid by union firms. The impact of the threat removal on wages paid by union firms is purely through a general equilibrium mechanism.

Outlawing unions amplifies the effect on wages further. In this policy exercise, the firms that were previously unionized now bargain wages individually with their workers. This leads to an increase in the slope of the schedule of wages paid by these firms. This, in turn, increases the slope of $b(s)$ which leads to further inequality in wages paid by firms that were previously union free, further amplifying the effect on inequality.

The unemployment rates of all skills go down in both policy exercises. These higher labor market tightnesses have in turn a positive impact on wages. This explains why the total changes in nonunion wages are positive for everyone except for the workers with very low skill. Without this general equilibrium feedback, the wages of low-skill workers would go down much more when unions are outlawed. If the elasticity of the matching function was different, such that the probability at which vacancies are filled reacted less to a change in labor market tightness, outlawing unions would have a more important negative effect on the wage of low-skill workers.

In the calibrated economy, removing the union threat and outlawing unions has a positive impact on the welfare of almost every agent.[27] The only exception is for workers with very low skills who are better off in an economy with unions.

Table 5 presents the variance of wages, the unemployment rate, total output as well as welfare after removing the union threat and outlawing unions. Removing the threat increases the variance of wages, lowers unemployment and increases output and welfare. These effects are further amplified when unions are outlawed. The variance of log nonunion wages decreases slightly in the last column. This suggests that the variance of nonunion wages is naturally higher in firms producing with technology (α_n, z_n) then in firms producing with technology (α_u, z_u).

Overall, outlawing unions increases total production by 2.5%, welfare by 2.2% and lowers the unemployment rate by 2.5 percentage points. It also increases the variance of log wages by 6%, 15 times more then the partial equilibrium estimate would suggest. This big difference comes directly from the impact of the union threat on nonunion firms and from general equilibrium mechanisms. Importantly, these effects are substantial even if the union membership is small (9% in the calibrated economy).

[27] The welfare schedule is computed by summing the welfare of all the agents, employed or unemployed, of a specific skill. In particular, the profits of the firms are not redistributed to the workers. The impact of unions on welfare is similar if profits are redistributed.

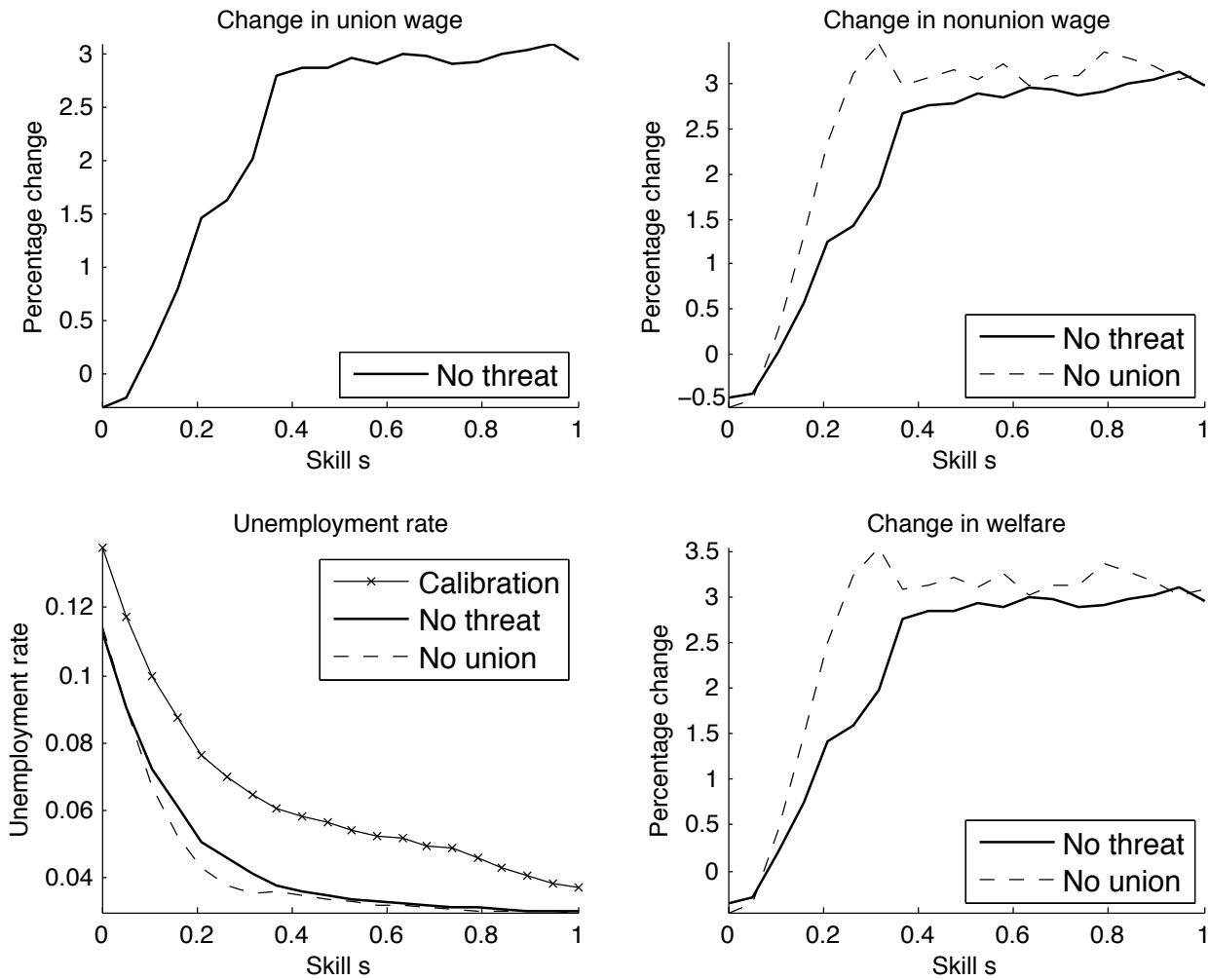

Figure 9: Policy exercises: removing the threat of unionization and outlawing unions.

6 Conclusion

Empirical estimators of the effects of unions on inequality generally abstract from the decision process of the firms and from general equilibrium mechanisms. In particular, they neglect the possible consequences that the unionization threat exerts on firms. This threat is created by the legal system and therefore may be present even in economies with low union membership.

This paper proposes a general equilibrium theory of firms' decisions and union formation to study the impact of unions on the economy. Workers and firms meet in a labor market characterized by frictions. Each period, the workers of a firm vote to create a union. If a union is created, wages are bargained collectively. Otherwise, each worker bargains his wage individually with the firm. This asymmetry of wage setting mechanisms causes unions to compress the wage distribution inside a firm. Furthermore, by fighting the threat of unionization, firms distort their hiring decisions in a way that also compresses wages.

I calibrate the model on the United States and show that outlawing unions increases the variance of wages substantially. This increase is much bigger than a partial equilibrium estimate would

	Calibration	No union threat	Outlawing unions
Variance log union wages	0.0966	0.1023	-
Variance log nonunion wages	0.1564	0.1648	0.1602
Variance log all wages	0.1510	0.1596	0.1602
Unemployment rate	8.2%	6%	5.7%
Union rate	8.9%	8.6%	0%
Total output ($\times 10^7$)	5.45	5.57	5.59
Welfare ($\times 10^9$)	9.36	9.53	9.58

Table 4: Effects of removing the union threat and outlawing unions on the variance of wages, unemployment, output and welfare.

suggest. Furthermore, outlawing unions increases welfare and output while lowering unemployment. The welfare gains are more important at the top of skill distribution while workers at the bottom of the distribution are worse off when unions are outlawed.

This paper only deals with the private sector of the economy. Since the public sector is heavily unionized in the United States, it is likely that the counterfactual policy exercises underestimate the full impact of unions.

One possible extension of the model would be to include a government in which the bargaining power of unions is different than in the rest of the economy. Another possible direction for future research would be to allow bargaining at the country level in order to compare the union systems in some European countries with the US system. Also, this theory could be used to study the interaction between the rise in inequality and the strong deunionization that has been observed in the United States during the last decades. In particular, it would be interesting to observe how a change in production technologies or in the skill distribution would impact the unionization rate.

References

Açikgöz, O. T. and B. Kaymak (2008): "The Rising Skill Premium and Deunionization in the United States," Working paper.

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

Ashenfelter, O. and G. E. Johnson (1969): "Bargaining Theory, Trade Unions, and Industrial Strike Activity," *The American Economic Review*, 59, 35–49.

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.

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.

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

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Farber, H. S. (1978): "Individual Preferences and Union Wage Determination: The Case of the United Mine Workers," *The Journal of Political Economy*, 86, 923–942.

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.

Freeman, R. B. (1980): "Unionism and the Dispersion of Wages," *Industrial and Labor Relations Review*, 34, 3–23.

Hagedorn, M. and I. Manovskii (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, 98, 1692–1706.

Hall, R. E. and P. R. Milgrom (2008): "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98, 1653–1674.

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.

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," *The 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.

Lemieux, T. (1993): "Unions and Wage Inequality in Canada and the United States," in *Small Differences That Matter: Labor Market and Income Maintenance in Canada and the United States*, ed. by D. Card and R. B. Freeman, University of Chicago Press, 69–107.

Luenberger, D. G. (1997): *Optimization by vector space methods*, Wiley-IEEE.

Mortensen, D. T. and C. A. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *The Review of Economic Studies*, 61, 397–415.

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.

Ok, E. A. (2007): *Real Analysis with Economic Applications*, Princeton University Press.

Pissarides, C. A. (2000): *Equilibrium Unemployment Theory - 2nd Edition*, The MIT Press, 2 ed.

Schmidt, K. D. (2003): "On the Covariance of Monotone Functions of a Random Variable." *Technische Universität Dresden*, manuscript.

Shimer, R. (2007): "Reassessing the Ins and Outs of Unemployment," *National Bureau of Economic Research Working Paper Series*, No. 13421.

Stole, L. A. and J. Zwiebel (1996a): "Intra-Firm Bargaining under Non-Binding Contracts," *The Review of Economic Studies*, 63, 375–410.

——— (1996b): "Organizational Design and Technology Choice under Intrafirm Bargaining," *The American Economic Review*, 86, 195–222.

Traxler, F. (1994): "Collective Bargaining: Levels and Coverage," in *Employment Outlook 1994*, OECD, 167–194.

Appendices

A Proofs

Here are the proofs from the previous sections.

A.1 Proof of lemma 1

Lemma 1. *In a steady-state, the firm's dynamic problem can be written as the static optimization:*

$$\max_g \pi(g) - \kappa(1-(1-\delta)\gamma)\int \frac{g}{q(\theta)}\,ds.$$

Proof. At a steady-state, the firm's problem is given by equation 4, which we can rewrite

$$J\left(\int \frac{g_{-1}}{q(\theta)}ds\right) = (1-\delta)\kappa\int \frac{g_{-1}}{q(\theta)}\,ds + \max_g\left\{\pi(g) - \kappa\int \frac{g}{q(\theta)}ds + \gamma J\left(\int \frac{g}{q(\theta)}ds\right)\right\}.$$

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

$$J\left(\int \frac{g}{q(\theta)}ds\right) = (1-\delta)\kappa\int \frac{g}{q(\theta)}\,ds + B.$$

The firm therefore solves

$$\max_g \pi(g) - \kappa\int \frac{g}{q(\theta)}ds + \gamma\left((1-\delta)\kappa\int \frac{g}{q(\theta)}\,ds + B\right)$$

and the result follows. □

A.2 Proof of lemma 2

Lemma 2. *Assume that g is strictly positive on $[0,1]$. Then the following function solves the bargaining problem:*

$$w_u(s,g) - c(s) = \frac{\beta_u}{n}\left(F(g) - \int c\cdot g\,ds + \gamma(1-\delta)\kappa\int \frac{g}{q(\theta)}\,ds\right).$$

The solution is unique if the joint surplus of the match is strictly positive at the point w_u.

Also, in a union firm with equilibrium distribution of workers g^* and technology j, the equilibrium wage schedule $w_j(s) = w_u(s,g^*)$ is

$$w_u(s,g^*) - b(s) = \frac{1-\gamma(1-\delta)}{1-\beta_u\gamma(1-\delta)}\frac{\beta_u}{n^*}\left(F(g^*) - \int b\cdot g^*\,ds + \gamma(1-\delta)\kappa\int \frac{g^*}{q(\theta)}\,ds\right).$$

where $n^* = \int g^* ds$ is the optimal size of the firm.

Proof. To keep a light notation, define

$$\Gamma \equiv F(g) + (1-\delta)\kappa\gamma \int \frac{g}{q(\theta)}\, ds \geq 0.$$

I work in the Lebesgue space $L^2[0,1]$. Two functions are identical if the measure of the set on which they differ is zero. By taking the log of the bargaining problem, we can define the objective function $P(w)$ as

$$P(w) = \beta_u \int \frac{g}{n} \log\left(w - c(s)\right)\, ds + (1-\beta_u) \log\left(\Gamma - \int w \cdot g\, ds\right)$$

and write the collective bargaining problem as

$$\max_w P(w) \tag{25}$$

Define the set of admissible functions

$$M = \left\{ w \in L^2[0,1] : w(s) - c(s) \geq 0 \,\forall\, s \in [0,1],\, \Gamma - \int w \cdot g\, ds \geq 0 \right\}.$$

M is the set of wage schedules w which might be agreed upon. For a wage schedule outside of M, some workers are better off unemployed or the firm will have negative surplus.

I first prove four preliminary results that characterize the set M and the function P.

Result 1. *The set of admissible functions M is convex.*

Proof. If M is a singleton then it is convex. If not, take any $w_1, w_2 \in M$ and consider the convex combination $w_a = aw_1 + (1-a)w_2$ with $0 \leq a \leq 1$. Then $w_a(s) \geq c(s)$ for all $s \in [0,1]$ and $\Gamma - \int w_a \cdot g\, ds \geq 0$. Since w_1 and w_2 are in $L^2[0,1]$, w_a is also in $L^2[0,1]$ and therefore M is convex. □

Result 2. *The function P is strictly concave on M.*

Proof. Take any $w_1, w_2 \in M$, $w_1 \neq w_2$ and consider the convex combination $w_a = aw_1 + (1-a)w_2$ with $0 < a < 1$. Since logarithm is a strictly concave function and $g > 0$, we can write

$$\begin{aligned}
P(w_a) &= \beta_u \int \frac{g}{n} \log(w_a - c)\, ds + (1-\beta_u) \log\left(\Gamma - \int w_a \cdot g\, ds\right) \\
&> \beta_u \int a\frac{g}{n} \log(w_1 - c)\, ds + (1-\beta_u) a \log\left(\Gamma - \int w_1 \cdot g\, ds\right) \\
&+ \beta_u \int (1-a)\frac{g}{n} \log(w_2 - c)\, ds + (1-\beta_u)(1-a) \log\left(\Gamma - \int w_2 \cdot g\, ds\right) \\
&= aP(w_1) + (1-a)P(w_2).
\end{aligned}$$

So P is strictly concave on M.[28] □

[28] It is important here that the measure of the set on which the two functions w_1 and w_2 are different is bigger than zero.

Result 3. *The wage function*

$$w_u(s,g) - c(s) = \frac{\beta_u}{n}\left(F(g) - \int c \cdot g\, ds + \gamma(1-\delta)\kappa \int \frac{g}{q(\theta)}\, ds\right)$$

is such that the functional derivative of P at point w_u, $\delta P[w_u]/\delta h$, is zero for every test function h and therefore w_u is a stationary point of P.

Proof. It is straightforward to show that

$$\begin{aligned}\int \delta P(w) \cdot h\, ds &= \left(\frac{d}{d\epsilon} P(w + \epsilon h)\right)\Big|_{\epsilon=0} \\ &= \int g \cdot h \left(\frac{\beta_u}{n(w-c)} - \frac{1-\beta_u}{\Gamma - \int w \cdot g\, ds}\right) ds\end{aligned}$$

Since $g > 0$ by assumption, the last equation is equal to zero for all h if and only if $w = w_u$. The idea is simply to see how P would vary around a point w if it is distorted in the direction h. If w is an optimum, the change in P should be 0 in all directions. □

Result 4. *If the joint surplus is strictly positive at w_u, then w_u is an interior point of M.*

Proof. The assumption on joint surplus is

$$F(g) - \int c \cdot g\, ds + (1-\delta)\kappa\gamma \int \frac{g}{q(\theta)}\, ds > 0.$$

This implies that $w_u(s) > c(s)$ for all $s \in [0,1]$. Furthermore, a simple calculation shows that the firm's surplus is equal to the fraction $(1-\beta_u)$ of the joint surplus. Therefore,

$$\Gamma - \int w_u \cdot g\, ds > 0$$

and w_u is in the interior of M. □

Putting the pieces together, P is a strictly concave function on the convex set M of the Hilbert space $L^2[0,1]$. It has the stationary point w_u and w_u is in the interior of M. Therefore, w_u is the unique global maximum of the function P.[29]

The equilibrium wage schedule follows directly by setting $w_u(s, g^*) = w_j$ and using the definition of c given by equation 3. □

A.3 Proof of lemma 3

Lemma 3. *The wage schedule*

$$w_n(s,g) - c(s) = \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g(s)}F(g) - \beta_n\ c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}$$

solves the bargaining problem (equation 12) of a firm employing the distribution of workers g.

Also, in a nonunion firm with equilibrium distribution of workers g^* and technology j, the equilibrium wage schedule $w_j(s) = w_n(s, g^*)$ is

$$w_n(s,g^*) - b(s) = \frac{1-\gamma(1-\delta)}{1-\beta_n\gamma(1-\delta)}\left(\frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g(s)}F(g) - \beta_n\ b(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}\right).$$

[29] See Ok (2007) and Luenberger (1997) for an exposition of calculus in a functional space.

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. I discretize the number of skills (each skill has a size ϵ) and the number of agents of a given skill (each agent has a size h). I start by writing the surplus of the firm and of the worker. After some manipulation and using the definition of the bargaining solution, taking the limits has $\epsilon, h \to 0$ will yield equation 12.

In the discretized framework, the production function is

$$F(L) = A\,L^{\alpha} = A \exp\left\{\alpha \sum_i \epsilon z_i \log(g_i)\right\}$$

where g_i is the number of workers of type i multiplied by the size of one worker, h.

The bargaining takes place when all vacancies have been posted. When bargaining with a worker, the firm compares two scenarios. Either an agreement is reached, in which case production takes place has planned, or the negotiations break down and the firm produces without this individual worker. In this last case, the 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 distribution of workers. In equilibrium, an agreement is always reached. The marginal discounted profit from hiring a worker of type j is

$$\begin{aligned}\Delta_j^n =& F(L) - \sum_i \epsilon w_i(\dots, g_j, \dots) g_i - \left(A \exp\left\{\alpha \sum_{i \neq j} \epsilon z_i \log(g_i) + \alpha \epsilon z_j \log(g_j - h)\right\}\right.\\ &\left. - \sum_{i \neq j} \epsilon w_i(\dots, g_j - h, \dots) g_i - w_j(\dots, g_j - h, \dots)(g_j - h)\epsilon - h\epsilon\gamma(1-\delta)\frac{\kappa}{q(\theta_j)}\right)\end{aligned}$$

where the notation $w_i(\dots, g_j, \dots)$ denotes the fact that w_i is a function of the whole distribution g. Δ_j^n is simply the difference between current period profits in the case of an agreement and in the case in which negotiations break down. Notice that in the latter case, the firm faces additional hiring costs in the next period. After using a Taylor's expansion on $\log(g_j - h)$ and rearranging, we get

$$\begin{aligned}\Delta_j^n =& A \exp\left(\alpha\epsilon \sum_i z_i \log(g_i)\right)\left(1 - \exp\left\{-\epsilon\alpha \frac{z_j}{g_j} h + \epsilon\alpha z_j O(h^2)\right\}\right)\\ &- \left(\sum_i \epsilon w_i(\dots, g_j, \dots) g_i - \sum_i \epsilon w_i(\dots, g_j - h, \dots) g_i\right)\\ &- h\epsilon w_j(\dots, g_j - h, \dots) + h\epsilon\gamma(1-\delta)\frac{\kappa}{q(\theta_j)}.\end{aligned}$$

The solution to the Stole and Zwiebel bargaining is the wage function that solves

$$\frac{\beta_n}{1-\beta_n}\Delta_j^n = (W_e(s,w) - W_u(s))\epsilon h$$

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

$$\lim_{h \to 0} \frac{\Delta_j^n}{h} = \alpha\epsilon \frac{z_j}{g_j} F(L) - \sum_i \epsilon g_i \frac{\partial w_i(\dots, g_j, \dots)}{\partial g_j} - \epsilon w_j(\dots, g_j, \dots) + \epsilon\gamma(1-\delta)\frac{\kappa}{q(\theta_j)}.$$

Because of the symmetry of the production function, the marginal product of a worker j depends only on g_j and on L. Its dependence on the whole distribution g is only through L. We can therefore impose more structure on the wage function and write $w_i(g_i, L)$ instead of $w_i(\dots, g_j, \dots)$. Therefore,

$$\lim_{h\to 0} \frac{\Delta_j^n}{h} = \alpha\epsilon \frac{z_j}{g_j} F(L) - \sum_i \epsilon g_i \frac{\partial w_i(g_i, L)}{\partial L} \frac{\partial L}{\partial g_j} - \epsilon g_j \frac{\partial w_j(g_j, L)}{\partial g_j} - \epsilon w_j(g_j, L) + \epsilon\gamma(1-\delta)\frac{\kappa}{q(\theta_j)}.$$

By dividing the previous expression by ϵ and taking the limit $\epsilon \to 0$, we get

$$\begin{aligned}\lim_{h,\epsilon\to 0} \frac{\Delta_j^n}{h\epsilon} = \frac{\partial L}{\partial g(s)} \left(\frac{dF}{dL} - \int \frac{\partial w(s, g(s), L)}{\partial L} g(s)\, ds \right) - \frac{\partial w(s, g(s), L)}{\partial g(s)} g(s) \\ - w(s, g(s), L) + \gamma(1-\delta)\frac{\kappa}{q(\theta(s))}.\end{aligned}$$

where $\frac{\partial L}{\partial g(s)}$ is a short notation for $z(s)/g(s) \cdot L(g)$. Therefore, the wage function is the solution to the following partial differential equation:

$$\begin{aligned}\frac{\partial L}{\partial g(s)} \left(\frac{dF}{dL} - \int \frac{\partial w(s, g(s), L)}{\partial L} g(s)\, ds \right) - \frac{\partial w(s, g(s), L)}{\partial g(s)} g(s) - w(s, g(s), L) + \gamma(1-\delta)\frac{\kappa}{q(\theta(s))} \\ = \frac{1-\beta_n}{\beta_n} \left(w(s, g(s), L) - c_j(s) \right)\end{aligned}$$

It is straightforward to show that w_n solves this equation. □

A.4 Proof of lemma 4

Lemma 4. *An unconstrained firm prefers to be union free if and only if*

$$\log\left(\frac{\Gamma_n}{\Gamma_u}\right) \geq \alpha \int z(s) \log\left(\frac{\mathrm{MC}_n^j(s)}{\mathrm{MC}_u^j(s)}\right) ds.$$

Proof. It is straightforward to show that, using the optimal hiring decision g_i^*, the production of a firm is given by

$$F(g_i^*) = A^{\frac{1}{1-\alpha}} (\alpha\Gamma_i)^{\frac{\alpha}{1-\alpha}} \exp\left\{ \frac{\alpha}{1-\alpha} \int z(s) \log\left(\frac{z(s)}{\mathrm{MC}_i(s)}\right) ds \right\}.$$

The proof follows directly from writing equation 5 with the optimal union and nonunion distributions (given by equation 22) and by simplifying the inequality. □

A.5 Proof of lemma 5

Lemma 5. *Assume that the labor market tightness schedule $\theta(s)$ and the outside option schedule $b(s)$ are increasing functions of the skill and that the bargaining powers β_n and β_u are equal. Then, in an unconstrained firm hiring according to g_i^* for $i = \{u, n\}$, the nonunion wage schedule $w_n(s, g_i^*)$ is an increasing function of s.*

Proof. Consider a firm j that is not unionized in equilibrium. Then by using equation 14 and the definition of MC_n^j, we find that the equilibrium wage paid by that firm is

$$w_n(s, g_n^*) = b(s) + \frac{\beta}{1-\beta} \frac{\kappa}{q(\theta(s))} (1 - \gamma(1-\delta)).$$

□

It is then straightforward to show that

$$c_j(s) = \frac{b(s) - \gamma(1-\delta)w_n(s, g_n^*)}{1-\gamma(1-\delta)} = b - \gamma(1-\delta)\frac{\beta}{1-\beta}\frac{\kappa}{q(\theta(s))}.$$

Then

$$\mathrm{MC}_u^j(s) = \mathrm{MC}_n^j(s) = (1-\beta)b(s) + \frac{\kappa}{q(\theta(s))}(1-(1-\delta)\gamma)$$

is an increasing function of s. Given the shape of c_j, the wage function can be written as

$$w_n(s, g_i^*) = \frac{\beta}{1-(1-\alpha)\beta}\frac{\mathrm{MC}_i^j(s)}{\Gamma_i} + (1-\beta)b(s)$$

and since $\mathrm{MC}_i^j(s)$ and $b(s)$ are increasing, so is $w_n(s, g_i^*)$.

We now need to show the result for a firm j that is unionized in equilibrium. Equation 10 shows that $w_u(s, g_u^*) - b(s)$ is equal to a constant that is independent of s. Denote that constant D; then $w_u(s, g_u^*) - b(s) = D$. Therefore,

$$c_j(s) = \frac{b(s) - \gamma(1-\delta)w_u(s, g_u^*)}{1-\gamma(1-\delta)} = b(s) - \frac{\gamma(1-\delta)D}{1-\gamma(1-\delta)}$$

and we see that c_j is increasing in s. This directly implies that MC_i^j is increasing. We can write the nonunion wage as

$$w_n(s, g_i^*) = \frac{\beta}{1-(1-\alpha)\beta}\frac{\mathrm{MC}_i^j(s)}{\Gamma_i} + (1-\beta)c_j(s) + \beta\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}$$

and since c_j, MC_i^j and θ are increasing, so is $w_n(s, g_i^*)$. We have shown that all firms (those that are unionized and non unionized in equilibrium) pay increasing nonunion wages when they hire according to their optimal distribution g_u^* and g_n^*.

A.6 Proof of lemma 6

Lemma 6. *Assume that the labor market tightness schedule $\theta(s)$ and $c_j(s)$ are increasing functions of the skill. Under the optimal hiring decision of unconstrained firms g_i^*, the union wage gap $w_n(s, g_i^*) - w_u(s, g_i^*)$ is increasing with s.*

Proof. With simple algebra, we find that

$$\begin{aligned} w_n(s, g_n^*) - w_u(s, g_n^*) = &\frac{\kappa\beta_n}{q(\theta(s))(1-\beta_n)} \\ &- \frac{\beta_u}{\int z(s)/\mathrm{MC}_n^j(s)ds}\int\left(\frac{\mathrm{MC}_n^j(s)}{\alpha\Gamma_n} - c_j(s) + (1-\delta)\kappa\gamma\frac{1}{q(\theta(s))}\right)\frac{z(s)}{\mathrm{MC}_n^j(s)}ds \end{aligned}$$

such that the variation in union wage premium across skill is coming exclusively from the labor market tightness. If $\theta(s)$ is increasing, the workers with the lowest skill are the ones who are the most likely to vote against unionization in non unionized firms.

Similarly, in a union firm:

$$w_n(s,g_u^*) - w_u(s,g_u^*) = \frac{c_j(s)\beta_n^2(1-\alpha)}{1-(1-\alpha)\beta_n} + \frac{\beta_n}{1-(1-\alpha)\beta_n}\left(\frac{1-\gamma(1-\delta)(1-\beta_u)}{1-\beta_u}\frac{\kappa}{q(\theta(s))}\right) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}$$
$$- \frac{\beta_u}{\int z(s)/\mathrm{MC}_u^j(s)ds}\int\left(\frac{\mathrm{MC}_u^j(s)}{\alpha(1-\beta_u)} - c_j(s) + (1-\delta)\kappa\gamma\frac{1}{q(\theta(s))}\right)\frac{z(s)}{\mathrm{MC}_u^j(s)}ds.$$

and since c_j is increasing, the result follows. □

A.7 Proof of lemma 7

Lemma 7. *Assume that the labor market tightness schedule $\theta(s)$ and $c_j(s)$ are increasing functions of the skill and that the bargaining powers β_n and β_u are equal. Then, for a given firm j,*

$$\mathrm{Var}_{g_n^*}\left(w_n(s,g_n^*)\right) \geq \mathrm{Var}_{g_u^*}\left(w_u(s,g_u^*)\right)$$

where $\mathrm{Var}_g(x)$ is the traditional variance operator taken with the normalized distribution $g/\int g\,ds$.

Proof. With equal bargaining powers and under the optimal hiring decisions given by equation 22, the union and nonunion wages are given by

$$w_n(s,g_n^*) = c_j(s) + \frac{\beta}{1-\beta}\frac{\kappa}{q(\theta(s))}$$
$$w_u(s,g_u^*) = c_j(s) + \text{Constant}.$$

Therefore the only source of variability in union wages comes from $c_j(s)$ while the nonunion wage schedule has an additional term coming from the labor market tightness. Under equal bargaining powers, $\mathrm{MC}_u^j(s) = \mathrm{MC}_n^j(s)$ for all s and therefore g_u^* is equal to g_n^* multiplied by a constant. The normalized distribution are therefore identical. We find,

$$\mathrm{Var}_{g_u^*}\left(w_u(s,g_u^*)\right) = \mathrm{Var}_{g_u^*}\left(c_j(s)\right) = \mathrm{Var}_{g_n^*}\left(c_j(s)\right)$$
$$\mathrm{Var}_{g_n^*}\left(w_n(s,g_n^*)\right) = \mathrm{Var}_{g_n^*}\left(c_j(s) + \frac{\beta}{1-\beta}\frac{\kappa}{q(\theta(s))}\right)$$
$$= \mathrm{Var}_{g_n^*}\left(c_j(s)\right) + \mathrm{Var}_{g_n^*}\left(\frac{\beta}{1-\beta}\frac{\kappa}{q(\theta(s))}\right) + 2\times\mathrm{Cov}_{g_n^*}\left(c_j(s), \frac{\beta}{1-\beta}\frac{\kappa}{q(\theta(s))}\right)$$

Since c_j and θ are increasing, the result follows.[30] □

A.8 Proof of lemma 8

Lemma 8. *For an unconstrained firm employing a distribution of workers g_n^* (given by equation 22):*

$$\frac{d(w_n(s,g_n^*) - w_u(s,g_n^*))}{d\alpha} = \frac{\beta_u}{\alpha^2\int z(s)/\mathrm{MC}_n^j(s)ds} > 0.$$

Also, the fraction of voters in favor of a union, $V(g_n^*)$, is such that

$$\frac{dV(g_n^*)}{d\alpha} = -\frac{\beta_u}{\alpha^2\left(\int z/\mathrm{MC}_n^j\,ds\right)^2}\int\frac{z}{\mathrm{MC}_n^j}\frac{\exp\{-\rho(w_u-w_n)\}}{\left(1+\exp\{-\rho(w_u-w_n)\}\right)^2}ds < 0$$

[30] See Schmidt (2003) for a proof that the covariance of two increasing functions of a random variable is positive.

Proof. With simple algebra, we find that

$$w_n(s, g_n^*) - w_u(s, g_n^*) = \frac{\kappa\beta_n}{q(\theta(s))(1-\beta_n)} - \frac{\beta_u}{\int z(s)/\mathrm{MC}_n^j(s)ds} \int \left(\frac{\mathrm{MC}_n^j(s)}{\alpha\Gamma_n} - c_j(s) + (1-\delta)\kappa\gamma \frac{1}{q(\theta(s))} \right) \frac{z(s)}{\mathrm{MC}_n^j(s)} ds$$

The first result of the lemma follows directly by taking the derivative of this last equation with respect to α. The second result comes as a consequence of the first result and the definition of $V(g)$ given by equation 16. □

A.9 Proof of lemma 9

Lemma 9. *Consider two firms, identified by the subscripts 1 and 2, that have identical technologies except for $A_1 \neq A_2$. There exist an equilibrium such that: if g_1 solves the problem of firm 1, then*

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

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 c_1 and c_2 are identical and denote that schedule by c. I will show this result later in the lemma.

The optimal distribution of workers for firm 2, g_2, is

$$g_2 = \arg\max_g F_2(g) - \int w_2(g) \cdot g\, ds - \gamma\kappa(1-(1-\delta)) \int \frac{g}{q(\theta)}\, ds$$

where w_2 is given by

$$w_{n2}(s, g) = \frac{\beta_n}{1-(1-\alpha)\beta_n} \frac{\alpha z(s)}{g(s)} F_2(g) + (1-\beta_n)c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}$$

if the firm is union free and by

$$w_{u2}(s, g) = c(s) + \frac{\beta_u}{n}\left(F_2(g) - \int c \cdot g\, ds + \gamma(1-\delta)\kappa \int \frac{g}{q(\theta)}\, ds\right)$$

otherwise.

The strategy for the proof is to rewrite firm 2's problem as a transformation of firm 1's problem. To do so, rewrite the problem of firm 2 by using the following transformation:

$$g = \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \tilde{g}.$$

Using this notation, the objective function of firm 2 is

$$F_2(g) - \int w_2(s, g) \cdot g\, ds - \gamma\kappa(1-(1-\delta)) \int \frac{g}{q(\theta)}\, ds \tag{26}$$

$$= \frac{A_2}{A_1} F_1\left(\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \tilde{g}\right) - \int w_2(s, g) \cdot \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \tilde{g}\, ds - \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \gamma\kappa(1-(1-\delta)) \int \frac{\tilde{g}}{q(\theta)}\, ds$$

$$= \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \left(F_1(\tilde{g}) - \int w_2\left(s, \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}} \tilde{g}\right) \cdot \tilde{g}\, ds - \gamma\kappa(1-(1-\delta)) \int \frac{\tilde{g}}{q(\theta)}\, ds\right) \tag{27}$$

where F_1 and A_1 identify the technology of firm 1. Notice that it is equivalent for firm 2 to pick g to maximize the expression on line 26 or to pick $\tilde{g}$ to maximize the expression on line 27.

Now, for nonunion wages in firm 2

$$\begin{aligned}
w_{n2}(s,g) &= \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{g(s)}F_2(g) + (1-\beta_n)c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))} \\
&= \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{(A_2/A_1)^{\frac{1}{1-\alpha}}\,\tilde{g}}\frac{A_2}{A_1}F_1\left((A_2/A_1)^{\frac{1}{1-\alpha}}\,\tilde{g}\right) + (1-\beta_n)c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))} \\
&= \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{\tilde{g}(s)}F_1(g_1) + (1-\beta_n)c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))} \\
&= w_{n1}(s,\tilde{g})
\end{aligned}$$

where the last equality denotes the fact that firm 2 using g and firm 1 using $\tilde{g}$ would be paying the same wages if they were union free. Similarly, for union wages in firm 2:

$$\begin{aligned}
w_{u2}(s,g) &= c(s) + \frac{\beta_u}{n}\left(F_2(g) - \int c\cdot g\,ds + \gamma(1-\delta)\kappa\int\frac{g}{q(\theta)}\,ds\right) \\
&= c(s) + \frac{\beta_u}{n}\left(\frac{A_2}{A_1}F_1\left(\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}}\tilde{g}\right) - \int c\cdot g\,ds + \gamma(1-\delta)\kappa\int\frac{g}{q(\theta)}\,ds\right) \\
&= c(s) + \frac{\beta_u}{\tilde{n}}\left(F_1(\tilde{g}) - \int c\cdot\tilde{g}\,ds + \gamma(1-\delta)\kappa\int\frac{\tilde{g}}{q(\theta)}\,ds\right) \\
&= w_{u1}(s,\tilde{g})
\end{aligned}$$

where $\tilde{n} = \int \tilde{g}\,ds$. Again, the last equality denotes the fact that firm 2 using g and firm 1 using $\tilde{g}$ would be paying the same wages if they are unionized.

We can now rewrite the problem of firm 2:

$$\left(\frac{A_2}{A_1}\right)^{\left(-\frac{1}{1-\alpha}\right)} g_2 = \arg\max_{\tilde{g}}\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}}\left(F_1(\tilde{g}) - \int w_2\left(s,\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}}\tilde{g}\right)\cdot\tilde{g}\,ds - \gamma\kappa(1-(1-\delta))\int\frac{\tilde{g}}{q(\theta)}\,ds\right)$$

where w_2 is given by

$$w_{n2}\left(s,\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}}\tilde{g}\right) = \frac{\beta_n}{1-(1-\alpha)\beta_n}\frac{\alpha z(s)}{\tilde{g}(s)}F_1(\tilde{g}) + (1-\beta_n)c(s) + \beta_n\gamma(1-\delta)\frac{\kappa}{q(\theta(s))}$$

if the firm is union free and by

$$w_{u2}\left(s,\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\alpha}}\tilde{g}\right) = c(s) + \frac{\beta_u}{\tilde{n}}\left(F_1(\tilde{g}) - \int c\cdot\tilde{g}\,ds + \gamma(1-\delta)\kappa\int\frac{\tilde{g}}{q(\theta)}\,ds\right)$$

otherwise.

Because of the link between the wage schedules we can transform the unionization condition for firm 2 such that:

$$\begin{aligned}
&\int\frac{g}{1+\exp(-\rho(w_{u2}(s,g)-w_{n2}(s,g)))}\,ds - \frac{1}{2}n \\
&= \int\frac{g}{1+\exp(-\rho(w_{u1}(s,\tilde{g})-w_{n1}(s,\tilde{g})))}\,ds - \frac{1}{2}n \\
&= \int\frac{\tilde{g}}{1+\exp(-\rho(w_{u1}(s,\tilde{g})-w_{n1}(s,\tilde{g})))}\,ds - \frac{1}{2}\tilde{n}.
\end{aligned}$$

We have rewritten the problem of firm 2 as a transformation of the problem of firm 1. Since $\tilde{g} = g_1$ solves the problem of firm 1 by assumption, we get that

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

which is the desired result. Notice that since the two firms have the same union status and are paying the same wages, we find $c_1(s) = c_2(s)$ for every s. □

B Solving the general equilibrium

Here is the algorithm I use to find a general equilibrium of the economy.

Given the parameters of the economy $(\beta_n, \beta_u, \delta, \gamma, \kappa, \rho, \mu, b_0(s), N(s))$ as well as the firms technology $(z_j(s), A_j, \alpha_j)$ for $j \in \{1, \dots j_{max}\}$, we need to find the aggregate variables $\theta(s)$ and $b(s)$ that sustain this equilibrium.

The algorithm to solve for an equilibrium uses the following strategy:

1. Fix the union status of each type of firm (either union or nonunion).
   (a) Make an initial guess on the aggregate variables: $\theta^0(s)$, $b^0(s)$.
   (b) Given the guess, compute the decision of each firm according to its union/nonunion status.
   (c) From wages and the distribution of hired workers, compute the new $\theta^{i+1}(s)$, then use it to compute the new $b^{i+1}(s)$.
   (d) Measure the distance between (θ^{i+1}, b^{i+1}) and (θ^i, b^i).[31] If the distance is smaller than some criterion $\epsilon > 0$ we found an equilibrium candidate. If not, go back to step 1b using $\theta^{i+1}(s)$ and $b^{i+1}(s)$ as the current guess.
2. Once we have an equilibrium candidate, we need to make sure that firms do not want to deviate from the union status they were assigned at step 1. If no firm wants to deviate, we have an equilibrium. Otherwise, fix the firm union status differently and repeat the steps.

[31] The distance I use is the integral of the square of the differences between the two functions.