Union Formation, Firm Selection and Aggregate Industry Productivity

Christian J. Bauer∗ Marco de Pinto†
University of Munich and CESifo IAAEU Trier and Trier University

Joerg Lingens‡
University of Muenster

July 25, 2015
Preliminary and incomplete
Please do not cite without permission of the authors.

Abstract
This paper develops a simple model with heterogeneous firms and labor unions to analyze the bidirectional relationship between unionization and industry productivity. The model shows how general wage setting reforms, measured by the costs of organizing workers, may have a non-monotonic impact on aggregate industry productivity. The paper also indicates how reallocation caused by international trade impacts unionization rates, thus highlighting an effect of trade on industry productivity not studied in the existing literature.

Keywords: endogenous unionization, firm heterogeneity, trade

JEL Classification: J51, F16, L11, L16

∗Contact details
†Institute for Labour Law and Industrial Relations in the European Union (IAAEU), Trier University, Behringstr. 21, D-54296, Germany; Phone: +49651/2014762; Email: depinto@iaaeu.de
‡University of Muenster, Stadtgraben 9, D-48143 Muenster; Phone: +49251/8322923; Email: joerg.lingens@wiwi.uni-muenster.de
1 Introduction

Union wage setting is an important characteristic of most western labor markets. As such, it does not come as a surprise that the allocative effects are heavily researched and fairly well understood: union wage setting results in higher wages (see Oswald, 1985 for a survey) while the employment effects are ambiguous and depend on the unions’ effect on productivity. There are arguments in favor of lowering productivity (basically caused by a hold-up effect, see Grout, 1984), but also empirical evidence that finds positive productivity effects as, for example, in Freeman and Medoff (1984). In addition, unions’ wage demands lead to changes in the industry structure by driving low productivity firms out of the market, which in turn causes an increase in average industry productivity (see Agell and Lommerud, 1993).

An important fact in the discussion about unionization is that union formation, i.e. the cooperation and organization of workers, is a costly process (see Voos, 1983 or more recently Economist, 2013). Unionization, hence, results from an endogenous choice of the workforce within a firm, which in turn is influenced by the industry structure. Since the pioneering papers by Hopenhayn (1992) and Melitz (2003), the consequences of different industry structures, e.g. in terms of productivity distribution as well as firms’ exit and entry, have been analyzed in many studies. However, its effect on how the labor market is organized has not been put under closer scrutiny. This is true although there are many different costs associated with organizing the workforce, e.g. through legal barriers to unionization or institutional arrangements, and a wide variety of policies that try to change these costs.¹

In addition, many western economies have experienced a substantial decline in unionization rates since the early 1980s (see Visser, 2013). The increase in globalization over this period of time aroused suspicion that this might have played an important role for the de-unionization process. Indeed, empirical studies do find a correlation between globalization and union density, but the evidence is mixed (see Wallerstein and Western, 2000 or Dreher and Gaston, 2007). This literature treats globalization (often measured as the amount of traded goods) as exogenous. Trade and unionization, however, are endogenous. As such, a model that accounts for this endogeneity would be helpful for a deeper understanding of the globalization/de-unionization nexus and could guide empirical work.

¹One of the more recent, prominent examples of attempts to change barriers to unionization was the discussion of the Employee Free Choice Act in the USA. But also the ban on closed shop unions in the UK in the late 1980s or more recently the ‘Tarifeinheitsgesetz’ in Germany (which only allows the largest union in a firm to set the wage) can be interpreted as changes in the costs organizing workers.
To deal with the bidirectional relationship between unionization and the industry structure, we present a model in which both are determined simultaneously. Furthermore, we extend this framework to an open economy model with two symmetric countries to analyze the bidirectional relationship between unionization and international integration. To this end, we draw heavily from the classical literature on worker unions and applied theory and empirical work in international trade. In the model, firms in the relevant sector are heterogeneous in productivity, operate under conditions of monopolistic competition, face fixed market entry costs and trade in goods is costly (Melitz, 2003); firm-productivity is Pareto distributed (Axtell, 2001, Helpman et al., 2004) and for simplicity we abstract from wealth effects (following Chor (2009) and Cole and Davies (2011), among others). Unions enter at the firm-level as monopoly unions (MacDonald and Robinson, 1992); they face a cost of organizing workers (Kuhn, 1988, Abowd and Farber, 1990); and only organize a firm if it pays. In equilibrium, only sufficiently productive firms choose to enter, and only the more productive active firms are unionized. Less productive active firms are not unionized, but are affected by union effects through the labor market (in line with the evidence, and akin to Kuhn, 1988).

Motivated by recent developments in labor market policies both in the U.S. and in Europe (see above), we study how general wage setting reforms, measured by changing costs of organizing workers, impact on aggregate industry productivity. We show that such wage setting reforms may have a non-monotonic impact on aggregate industry productivity. On the one hand, a rise in (marginal) unionization costs raises the prices charged by more productive, unionized firms such that competition becomes softer. More low-productive firms are able to produce and aggregate productivity c.p. decreases. On the other hand, a smaller share of firms is unionized such that wages and prices decline. Thus, competition becomes more fierce and fewer low-productive firms survive in the market. Aggregate productivity c.p. increases. The model also suggests that extreme wage setting institutions, which either prevent unions or give rise to full unionization, are most conducive towards high industry productivity, with a better relative performance of non-unionized industries. Intuitively, if no firm is unionized initially (because unionization costs are infinite), somewhat reducing unionization costs will induce unions to organize the most productive firms, reduce their out-

---

2We follow recent literature in assuming firm-level unions. In OECD countries, union wage bargaining occurs on several levels, including the national, sectoral, and the firm level. However, there is a tendency towards more decentralized wage determination also in countries where wages traditionally are determined at the sectoral level (see, e.g., Ochel, 2005 and Gürtzgen, 2009 for Germany).
put, thus the competitive real wage, and thereby the cutoff productivity. In sum, aggregate industry productivity goes down. If, however, all firms are unionized initially (because unionization costs are zero), raising unionization costs will prevent unionization in the least productive firms and raise wage demands in the most productive firms, again causing average productivity to deteriorate.

We then analyze how unionization varies with industry productivity as determined by the underlying productivity distribution. We show that increasing productivity dispersion (a mean preserving spread of the productivity distribution) has an ambiguous effect on the mass of unionized firms, but that the share of unionized firms increases. The intuitive argument for the ambiguity concerning the mass of unionized firms is that although the mass of higher productivity firms increases (and thereby those firms that are unionized), less firms may want to enter the market. Finally, we allow for intra-industry trade and analyze how this affects unionization.\textsuperscript{3} We show that the trade induced competition erodes rents such that firms are driven out of the market. This implies that the mass of unionized firms decline while its share remains constant.

In analyzing the bidirectional effects of unionization and industry structure (understood as productivity distribution), the paper contributes to two strands of research. First, a recent set of papers exploits firm selection as a quantitative relevant source of industry productivity (see e.g. Alfaro et al., 2008 and Bartelsman et al., 2008). In this context, union wage hikes may drive less productive firms out of the market, and thus reallocate resources towards surviving, more productive firms. Braun (2011) demonstrates that the selection effect differs between industry- and firm-level union bargaining. He shows that bargaining at the industry-level induces more demanding firm selection, generating, in equilibrium, higher aggregate industry productivity. This literature again treats the prevalence of unions as exogenous.

The second strand of related research recognizes unionization as endogenous, and thus highlights the allocative effects of unions that go beyond the direct effects of union wage hikes. Corneo (1993), Corneo (1995), and Corneo and Lucifora (1997) analyze the strategic employment choice and ensuing distortions when firms face the threat of becoming unionized. More recently, Taschereau-Dumouchel (2010) analyzes this idea in a search and matching framework, showing that the distortion of hiring decisions brought about by a union threat reduces wage dispersion. Corneo and Lucifora (1997) also

\textsuperscript{3}We focus on intra-industry trade because of its dramatic rise (Broda and Weinstein, 2006) and its pronounced effect on industry reallocations (Aw et al., 2000, Pavcnik, 2002, Bernard and Jensen, 1999).
provide indicative evidence of union threat effects by showing that wages in non-unionized firms are affected by union behavior. In our framework, we complement this notion but instead highlight the transmission of union effects through the common labor market. In this regard, our paper is most closely related to Kuhn (1988), who considers an occupational choice model with a common resource constraint on firms and workers. He argues that, by raising wages, unions reduce the incentive to become an entrepreneur and thus raise worker supply. Comparing economies with and without unions, he shows that large firms are more likely to be unionized, and that unions reduce efficiency by distorting the choice between becoming an entrepreneur or a worker.

2 The Model

2.1 Set-up

2.1.1 Demand

We study an economy in which a homogenous numeraire good \( Y \) and a differentiated good \( X \) are produced. There is an exogenously given mass of workers who are homogeneous and mobile across sectors. Without the loss of generality, we can consider a representative consumer to describe the demand side of the model. The representative consumer values the two private goods \( X \) and \( Y \) according to a quasi-linear utility function which is given by:

\[
U = \mu \ln X + Y^D, \quad X \equiv \left[ \int_{\omega \in \Omega} x(\omega)^{\frac{\epsilon-1}{\epsilon}} d\omega \right]^{\frac{\epsilon}{\epsilon-1}}.
\]

(1)

In Eq. (1), \( Y^D \) is the quantity consumed of the numeraire good and \( X \) is the Dixit-Stiglitz composite of all varieties in the differentiated good sector that are available. The set of these varieties is given by \( \Omega \). The relative market size of the differentiated good sector is captured by \( \mu \) with \( \mu > 0 \). Varieties are consumed in quantity \( x(\omega) \), where \( \omega \) is the index for the firm producing this variety. Varieties are substitutes and the elasticity of substitution between any two varieties is given by \( \epsilon \) with \( \epsilon > 1 \). Note that we choose good \( Y \) as numeraire, i.e. \( P_Y \equiv 1 \).

Utility maximization requires that the ratio of marginal utilities for the two private goods equals their relative price. Using (1), we obtain:

\[
X = \frac{\mu}{P}.
\]

(2)
Thus, the expenditures for the differentiated good $X$ are fixed at $\mu$. The remainder of income $I$ is spent on good $Y$

$$Y^D = I - \mu,$$

which implies that indirect utility is linear in income $I$.

Moreover, utility maximization yields isoelastic demand functions for each variety:

$$x(\omega) = \left[ \frac{p(\omega)}{P} \right]^{-\epsilon} \frac{\mu}{P}.$$  (4)

The CES price index $P$ for good $X$ is given by:

$$P = \left[ \int_{\omega \in \Omega} p(\omega)^{1-\epsilon} d\omega \right]^\frac{1}{1-\epsilon},$$  (5)

where $p$ denotes the price for variety $\omega$. The price index is an inverse measure of (horizontal) competition. If $P$ increases (decreases), demand increases (decreases) implying that competition between operating firms becomes softer (more fierce).

### 2.1.2 Production

In the numeraire sector $Y$, goods and labor markets are competitive by assumption. Each firm uses one unit of labor to produce one unit of output, and we have $P_Y \equiv 1$ (see above). Due to the linear production technology, wages in the $Y$, $\hat{w}$, are equal to one. Following Melitz (2003), firms in the differentiated good sector $X$ are heterogeneous with respect to their productivity $\phi$, and operate under conditions of monopolistic competition.

To model firm heterogeneity in the $X$ sector, we assume the following procedure (see Chaney, 2008 for a similar approach). There is a bounded pool of potential entrants (‘entrepreneurs’), $M^e$, who are capable of producing a variety of the differentiated good. Each of these entrepreneurs draws a productivity level from a given distribution at no cost. We assume that productivities are Pareto distributed with the distribution function:

$$G(\phi) = 1 - \left( \frac{\phi_{\text{min}}}{\phi} \right)^k, \quad \phi \geq \phi_{\text{min}} > 0, \quad k > \epsilon - 1,$$  (6)

where $\phi_{\text{min}}$ is the lower bound of the support and $k$ is the shape parameter. After each entrepreneur is assigned with a specific productivity level,
s/he decides whether to enter the market. In case of market entry, the entrepreneur becomes an active firm and starts production. Because the pool of entrants, $M^e$, is exogenously given, we can normalize it to unity without loss of generality.\footnote{In the Melitz framework, there is – in contrast to our approach – an unbounded pool of potential entrants. In addition, entrepreneurs must pay for drawing a productivity level. Both features imply that the mass of entrepreneurs which participate in the 'productivity draw' is determined endogenously.}

Consider an entrepreneur who is assigned with some productivity $\phi$, enters the $X$ sector (i.e. becomes an active firm), and produces the variety $\omega$.\footnote{Because firms are only heterogeneous with respect to $\phi$, we do not use any firm index for notational simplicity.} The production technology of this firm is given by:

$$x = \phi \cdot l,$$

where $l$ denotes the number of employed workers. Production also requires a (non-sunk) payment of uniform overhead fixed costs $F$ (in units of the numeraire). These fixed costs can be interpreted as ‘entrepreneurial services’ and capture the amount of ‘raw capital’ needed for start-up and production. Profits are

$$\pi = p\phi l - wl - F,$$

where $w$ represents the wage rate which the firm has to pay to its workers.

### 2.1.3 Unionization

We introduce unionization following Kuhn (1988) and MacDonald and Robinson (1992): unions act at the firm-level as separate players in the differentiated good sector. In contrast to standard approaches of modeling unionization, we assume that organizing workers of a firm is costly for the union. As argued by Voos (1983), organization costs consist of direct costs, e.g. salaries and expenses of the union’s representatives, and indirect costs, e.g. unpaid time which current members spend to organize new members. We summarize these costs in a union cost function denoted by $C(l)$. In general, various specifications for $C$ appear plausible. To keep our model analytically tractable, we assume that marginal unionization costs are constant, and that there are no fixed costs: $C(l) = cl$ with $c > 0$.

Due to organization costs, unionization is endogenous in our framework, i.e. the union that is allocated to a specific firm decides whether to enter
the market and become active.\footnote{The benefit of this stylized approach is that we do not need to make critical assumptions about how workers are allocated to firms initially, or how workers allocated to a given firm would make and respond to the unionization decision. Basically, we assume that the separate institutional body 'firm-level' union cares about workers in general, but not about specific workers.} When active, it organizes the firm’s workers (which determines union membership), sets wages, and leaves the firm with the right to manage employment. If the union decides to stay inactive, there is no unionization for the respective firm implying that this firm pays the competitive wage.

We assume that the union is utilitarian such that its utility\footnote{Usually, union’s utility is given by the aggregation of the union members’ indirect utilities. The formulation in Eq. (9) can be considered as an approximation for such an approach which we can use, because indirect utility is linear in income.} is given by:

\[ U^U = l(w - 1) - cl. \]  

As shown by Eq. (9), the outside option of the firm’s workers (and hence union members) is to be employed in the $Y$ sector and receive a wage payment of $\hat{w}$. Notably, the assumptions of perfect labor markets in the $Y$ sector and worker mobility imply that the economy wide unemployment rate is zero.

### 2.1.4 Timing

The timing structure of our model is as follows. At the first stage, each entrepreneur draws a productivity level and decides about market entry. At the second stage, the union associated with each active firm decides whether to organize the workers in the firm. Conditional on organizing the firm’s workers, the union sets the wage rate at stage three.

At the fourth stage, the firm hires the optimal amount of workers, and production and consumption occur. Note that the firm can also choose to employ no workers at all if the union’s wage demands are too high, i.e. they cause negative profits. We solve the model by backwards induction.

### 2.2 Solution

#### 2.2.1 Stage 4: Employment

At the fourth stage, the firm with productivity $\phi$ maximizes its profit (8) by choosing production $x$ (i.e. employment $l$) subject to the demand function (4) and the wage $w$ which has been determined at stage 3. The first-order
condition of that problem reads:

\[ p\phi \left( \frac{\partial p}{\partial x} + 1 \right) = w. \] (10)

With the inverse price elasticity of demand given by \(-1/\epsilon\), the profit-maximizing pricing rule is the standard mark-up \(1/\alpha\) (with \(\alpha \in (0, 1)\)) over (firm-specific) variable costs:

\[ p(\phi, w) = \frac{1}{\alpha} \frac{w}{\phi}, \quad \alpha \equiv 1 - \frac{1}{\epsilon}. \] (11)

Note that variable costs differ both because firms are heterogeneous with respect to the productivity level, and because they may face different wage rates.

Combining pricing rule and consumer demand, the optimal employment level satisfies:

\[ l(\phi, w) = \frac{\mu}{\phi} \frac{P^{\epsilon - 1}}{\left( \frac{1}{\alpha} \frac{w}{\phi} \right)^\epsilon}. \] (12)

These choices pin down maximized profits as:

\[ \pi(\phi, w) = (\alpha^{-1} - 1)l(\phi, w)w - F. \] (13)

However, the firm with productivity \(\phi\) sets employment equal to (12) if and only if maximized profits are non-negative: \(\pi(\phi, w) \geq 0\). Otherwise, the firm would be better off to employ no worker at all and leave the market. Consequently, we can define a feasible wage, \(\tilde{w}\), which is the maximum wage that a firm equipped with productivity \(\phi\) is willing to pay. This is implicitly defined by:

\[ \pi(\phi, \tilde{w}) = 0 \iff (\alpha^{-1} - 1)\tilde{w}l(\phi, \tilde{w}) = F. \] (14)

Using (12), we can explicitly calculate the feasible wage rate as:

\[ \tilde{w}(\phi) = \phi \alpha P \left( \frac{F \epsilon}{\mu} \right)^{-\frac{1}{\epsilon - 1}}. \] (15)

Note that the firm’s feasible wage is upward sloping in \((\tilde{w}, \phi)\)-space because, intuitively, more productive firms survive at higher wage costs.

### 2.2.2 Stage 3: Union wage setting

At the third stage, the union decides about the wage rate conditional on the fact that the firm’s workers are organized (which is determined at stage 2). There are two constraints of the union’s optimization problem. First,
the union recognizes the firm’s right-to-manage employment, and takes its optimal hiring decision into account. Second, given its objective, the union is not interested in forcing firms to set employment equal to zero. To ensure the firm’s participation in the wage setting (i.e. the firm stays in the market), it cannot demand a wage higher than the firm’s feasible wage $\tilde{w}$.

Formally, we can write the union’s optimization problem as:

$$\max_w U^U = l(w - 1) - cl$$

s.t. $l = l(\phi, w)$ and $w \leq \tilde{w}(\phi)$.

The Lagrange function is given by:

$$L = l(\phi, w)(w - 1) - cl(\phi, w) - \eta(\tilde{w}(\phi) - w),$$

where $\eta$ is the shadow value of the firm’s participation constraint. The first-order conditions read:

$$\frac{\partial l(\phi, w)}{\partial w} \frac{w}{l(\phi, w)}(w - 1) + w - c \frac{\partial l(\phi, w)}{\partial w} \frac{w}{l(\phi, w)} + \eta \frac{w}{l(\phi, w)} = 0,$$

$$\leftrightarrow 1 - \epsilon(w - 1) + w + \epsilon c + \eta \frac{w}{l(\phi, w)} = 0,$$

$$\tilde{w}(\phi) - w \geq 0 \quad \text{with} \quad \eta \geq 0,$$

where $^1$ follows from the fact that the labor demand elasticity is constant and given by $-\epsilon$ under CES preferences. $^9$

When the firm’s participation constraint is not binding, i.e. $\tilde{w}(\phi) - w > 0$ (which implies $\eta = 0$ by complementary slackness), we find that the wage rate is a constant mark-up over the sum of the competitive wage and the marginal organization costs (i.e. the opportunity costs of being unionized and working in the specific firm):

$$w = \frac{1}{\alpha}(1 + c) \equiv w^u,$$

We refer to this wage rate as the unconstrained wage rate denoted by $w^u$.

---

$^9$This implies that the (relative) employment response to wage increases is the same across firms with heterogeneous productivities. The benefit of this property is that it allows an insightful dissection of the impact of unions on industry productivity. In particular, when marginal unionization costs are constant, unions optimally choose identical wages as shown by (19). When marginal organization costs are not constant, wage differences in these firms are the result of changes in marginal organization costs. Away from constant labor demand elasticities, firm-size effects would reinforce the pattern of unionization and productivity, see Kuhn (1988), and strengthen our results.
If the firm’s participation constraint is binding, the union sets the wage rate equal to the feasible wage rate as shown by (18), \( w = \tilde{w}(\phi) \). Hence, the union follows a limiting-wage strategy which leaves the firm with zero profits.\(^{10}\) Then, by complementary slackness, \( \eta > 0 \) holds. By the second-order condition for a maximum and the fact that employment is c.p. increasing \( \phi \), we find that the shadow value is decreasing in \( \phi \). Thus, there exists some threshold productivity \( \phi^+ \) such that the participation constraint of the firm stops binding. This productivity level is implicitly given by:

\[
\eta(\phi^+) = 0 \iff l(\phi^+, \tilde{w}(\phi^+)) \left[ 1 - \frac{\epsilon}{\tilde{w}(\phi^+)} - 1 - c \right] = 0.
\]

Rearranging this expression as well as using (15) imply:

\[
\phi^+ = \left( \frac{F\epsilon}{\mu} \right)^{\frac{1}{\alpha - 2}} \alpha^{-2} P^{-1}(1 + c).
\]

Hence, a very simple wage pattern emerges. Firms with \( \phi < \phi^+ \) face the limit-wage \( \tilde{w}(\phi) \) while firms with \( \phi \geq \phi^+ \) face the unconstrained union wage \( w^u \).

**Lemma 1 (Union wage setting)** The productivity threshold \( \phi^+ \) (at which the firm’s participation constraint is binding at the margin) decreases in the price index \( P \) and increases in organization costs \( c \).

Because the price index \( P \) is an inverse measure for (horizontal) competition, firms’ profits and hence rents increases in \( P \). Therefore, firms survive at higher wage costs which implies that \( \tilde{w}(\phi) \) increases. The union is thus able to set the unconstrained wage for a higher share of firms without the risk of the shutdown; \( \phi^+ \) decreases. On the contrary, a rise of organization costs \( c \) implies that marginal costs of unionization increase. For the union, it is utility maximizing to raise the unconstrained wage. Because firms’ profits are unaffected by changes in \( c \), this implies, however, that fewer firms are able to pay \( w^u \) without making negative profits; \( \phi^+ \) increases.

### 2.2.3 Stage 2: Union formation

At the second stage, the union decides whether to organize the firm’s workers. A firm will only be unionized if doing so is profitable, implying that the associated utility stream must be non-negative: \( U^U \geq 0 \). By analogy with

\(^{10}\)This is in line with Freeman and Kleiner (1999) and Hirsch (2004), who find that unions generally reduce firms’ profits, but only to an extent that prevents firm failure.
the feasible wage for the firm, we can define a feasible wage for union’s formation, $w^c$, as:

$$U^U = 0 \iff w = 1 + c \equiv w^c$$  \hspace{1cm} (21)

This is the minimum wage such that average organization costs are at least covered.

Because the union anticipates the wage outcome at stage 3, there will be unionization if and only if $w \geq w^c$ holds. As pointed out in the previous section, the union can either set the unconstrained wage $w^u$ or the limit wage $\tilde{w}(\phi)$. Comparing (19) and (21), we find that $w^u > w^c$. If the union can set the unconstrained wage, unionization is always profitable. Consequently, high-productive firms with $\phi \geq \phi^+$ are unionized. This is, however, not necessarily true for firms that pay the limit wage. Because $\tilde{w}(\phi)$ increases in productivity while $w^c$ is constant, the limit wage falls short of the union’s minimum wage for firms that are equipped with relatively low productivity levels. The firm’s profit is so low that it cannot support the formation of a union.

Determining which firms’ profits cannot support unionization completes the description of stage 2. The firm’s productivity which leads to unionization at the margin – denoted by $\phi^-$ – is calculated by the intersection of $w^c$ and the limit wage $\tilde{w}(\phi^-)$. Using (15) and (20), we obtain:

$$\phi^- = \alpha \phi^+ = \left(\frac{F \epsilon}{\mu}\right)^{\frac{1}{1-\alpha}} \alpha^{-1} P^{-1}(1 + c).$$  \hspace{1cm} (22)

For $\phi < \phi^-$, there is no unionization, while firms with $\phi \geq \phi^-$ face a union. Note that we call the firm with productivity $\phi^-$ the marginal unionized firm.\footnote{The fact that the fraction $\frac{\phi^-}{\phi^+}$ is constant is due to the linear union costs function. This greatly facilitates the model and helps us to find closed form solutions.}

**Lemma 2 (Marginal unionized firm)** The productivity of the marginal unionized firm $\phi^-$ is decreasing in the price index $P$ and increasing in organization costs $c$.

An increase in the price index indicates that competition becomes softer. Firms’ profits rise such that they can survive at higher wage costs; $\tilde{w}(\phi)$ increases. Because the union’s minimum wage is not affected, a higher share of (low-productive) firms can cover average organization costs; $\phi^-$ decreases. On the contrary, an increase in organization costs leads to a rise of the union’s minimum wage while firms’ profits are unaffected. A smaller share of (low-productive) firms can cover average organization costs; $\phi^-$ increases.
2.2.4 Stage 1: Market entry

After the entrepreneur draws a productivity level $\phi$, s/he decides to enter the market (and hence becomes an active firm) at stage 1. Market entry requires that the respective profits are non-negative. Given this condition, we can calculate a cutoff productivity for entry denoted by $\phi^*$: $\pi(\phi^*, w) = 0$. We call a firm with productivity $\phi^*$ the marginal active firm in the market.

To calculate $\phi^*$ explicitly, we must know which wage the marginal active firm faces. Because organization costs are positive, unionization requires that the union can adopt (at least parts of) the firms’ profits to cover average organization costs. The marginal active firm has, however, zero profits by definition and there is, hence, no rent to redistribute. Therefore, the marginal active firms is not unionized, implying that (a) $\phi^* < \phi^-$ and (b) $w(\phi^*) = \hat{w} \equiv 1$ hold. The former inequality also shows that a fraction of low-productive firms earns positive profits such that they are active in the market but these profits are not high enough to support unionization. Formally, we thus obtain:

$$\pi(\phi^*, 1) = 0.$$  \hfill (23)

Using (23), (13) and (12), we can calculate the cutoff productivity as:

$$\phi^*(P) = (\alpha P)^{-1} \left( \frac{F \epsilon}{\mu} \right)^{\frac{1}{\alpha}}. \hfill (24)$$

**Lemma 3 (Cutoff productivity and price index)** The cutoff productivity $\phi^*$ is a negative function of the price index $P$.

Intuitively, a higher $P$ and thus softer competition imply that more profits can be earned by firms. This in turn allows less productive firms to cover their fix costs and enter the market; $\phi^*$ decreases.

Note that we can use the last expression as well as (20) and (22) to formulate the productivity thresholds $\phi^-$ and $\phi^+$ as functions of the cutoff productivity:

$$\phi^- = (1 + c)\phi^*, \hfill (25)$$

$$\phi^+ = \frac{1 + c}{\alpha} \phi^*.$$  

The derived pattern of productivity, wages, and unionization at the firm-level (i.e. given the price index $P$) results in a wage curve, i.e. the curve which depicts the wage that firms with different productivities face. The wage curve is illustrated in Figure 1 by the red line and is summarized in the following Lemma.
Lemma 4 (Wage curve) Firms with productivity

- \( \phi^* \leq \phi < \phi^- \) are not unionized and pay the competitive wage \( w = 1 \),
- \( \phi^- \leq \phi < \phi^+ \) are unionized and pay the limit wage \( w = \tilde{w}(\phi) \),
- \( \phi \geq \phi^+ \) are unionized and pay the unconstrained wage \( w = w^u \).

Figure 1: Wage curve

We complete the solution of our model by describing the pattern of unionization. To that end, we introduce two indicators: the mass of unionized firms \( M_U \), and the share of unionized firms \( s_M \). These are given by, respectively:

\[
M_U = 1 - G(\phi^-) = \left( \frac{\phi_{\min}}{\phi^-} \right)^k = \left( \frac{\phi_{\min}}{(1+c)\phi^*} \right)^k, \tag{26}
\]

\[
s_M \equiv \frac{M_U}{M} = \frac{1 - G(\phi^-)}{1 - G(\phi^*)} = \frac{1 - G((1+c)\phi^*)}{1 - G(\phi^*)} = (1+c)^{-k}. \tag{27}
\]
3 Equilibrium

3.1 Price index and cutoff productivity

As in Melitz (2003) and Chaney (2008), the determination of the equilibrium requires the calculation of the cutoff productivity level $\phi^*$. As pointed out by Lemma 3, the cutoff productivity is a decreasing function of $P$. At the industry level, however, this is only part of the story because $P$ is endogenous in itself and by definition a function of the mass of active firms $M$ and their individual prices $p(\phi)$. This follows by rewriting (5) as:

$$P = \left[ M \int_{\phi^*}^{\infty} p(\phi, w)^{-\xi} \frac{dG(\phi)}{1 - G(\phi^*)} \right]^{-\frac{1}{\xi}},$$

where $\xi \equiv \epsilon - 1$ (with $k > \xi$) is a measure for the price elasticity of demand. The probability that an entrepreneur draws a productivity $\phi \geq \phi^*$ is given by $1 - G(\phi^*)$. Hence, $dG(\phi)/(1 - G(\phi^*))$ represents the distribution of productivities conditional on a successful draw.

To derive the equilibrium at the industry level, we need an expression for the price index $P$ depending on $\phi^*$. For this purpose, we use the price structure of individual firms which can be calculated by using (11) and the wage structure described in Lemma 4.

**Lemma 5 (Price structure)**

Firms with productivity

- $\phi^* \leq \phi < \phi^-$ are not unionized and set the price $p(\phi) = (\alpha\phi)^{\alpha - 1}$,
- $\phi^- \leq \phi < \phi^+$ are unionized and set the price $p(\phi) = (\alpha\phi)^{\alpha - 1} \tilde{w}(\phi)$,\(^{12}\)
- $\phi \geq \phi^+$ are unionized and set the price $p(\phi) = (\alpha\phi)^{-1} w_u$.

Due to the normalization of the mass of potential entrepreneurs $M^e$ at unity, we obtain for the mass of active firms: $M = 1 - G(\phi^*)$. Using this as well as Lemma 4 and 5, we can rewrite the price index (28) as:

$$P = \left[ \int_{\phi^*}^{\phi^-} (\alpha\phi)^{\xi} dG(\phi) + \int_{\phi^-}^{\phi^+} \frac{P^{-\xi} F\epsilon}{\mu} dG(\phi) + \int_{\phi^+}^{\infty} (\alpha\phi)^{\xi} \left( \frac{1 + \epsilon}{\alpha} \right)^{-\xi} dG(\phi) \right]^{-1/\xi}.$$

\(^{12}\)Inserting (15) reveals that firms facing the limit wage charge a uniform, productivity independent price. Intuitively, a higher productivity decreases marginal costs such that price and profit can rise. This in turn raises the firm’s feasible wage and the union can demand a higher limit wage. Consequently, marginal costs increase and the price falls to its original level.
For firms paying limit wages, the respective price distribution is a function of the price index. The reason for this is that $P$ indicates the size of firms’ profits and hence the value of the limit wage. Thus, the price index is a function of itself.

As shown in Appendix A.1, we can rearrange this expression and obtain:

$$P(\phi^*) = (\alpha \phi^*)^{-1} \left( \frac{\vartheta \phi^k_{\min} \left(- (1 + c)^{-(k-\vartheta)} + 1 + \left(\frac{1+c}{\alpha}\right)^{-k}\right)}{(\phi^*)^k - \frac{F_{\xi}}{\mu} (\phi_{\min})^k (1 + c)^{-k}} \right)^{-1/\vartheta}, \quad (30)$$

with $\vartheta \equiv k/(k - \xi) > 1$. Eq. (30) defines an upward sloping curve in the $(\phi^*, P)$-space which represents the endogenous CES price index of all available varieties. Hence, an increase in the cutoff productivity $\phi^*$ is associated with an increase in $P$. Intuitively, if there are fewer low-productivity firms in the market, the remaining firms set higher prices on average and the aggregate price index increases.\(^{13}\) The formal proof for $\partial P(\phi^*)/\partial \phi^* > 0$ is given in Appendix A.2.

With (24) and (30) at hand, we can determine the equilibrium tuple $\{\bar{\phi}; \bar{P}\}$, i.e. the equilibrium cutoff productivity and the equilibrium price index. For $\bar{\phi}$, we obtain (see Appendix A.3):

$$\bar{\phi}^* = \left( \frac{\vartheta + (1 + c)^{-k}(1 + (\vartheta - 1)\alpha^k - \vartheta(1 + c)^\xi)}{\left(\frac{F_{\xi}}{\mu}\right)^{-1}} \right)^{1/k} \phi_{\min}. \quad (31)$$

Note that we allow only parameter constellations that ensure $\bar{\phi}^* > \phi_{\min}$. Otherwise, the economy was characterized by a corner solution implying that all potential entrepreneurs would enter the market.\(^{14}\)

Inserting (31) into (30), we could determine the equilibrium price index $\bar{P}$ and also the equilibrium price structure. This in turn pins down equilibrium employment and output.

### 3.2 Welfare

As usual, we measure welfare by the indirect utility $V$. Inserting (2) and (3) into (1), we obtain: $V = \mu \ln \mu - \mu - \mu \ln P + I$, which shows that indirect utility is linear in income $I$. We define income as the sum of aggregate net

\(^{13}\)The price index reflects the average price when using firms’ outputs as weights.

\(^{14}\)Essentially, the respective restriction is a function of the fixed costs, product market competition as well as organization costs.
profits and aggregate net labor income, i.e. without fixed costs for unions and firms, because only this share is disposable for consumption. Formally, we have:

\[ I = \int_{\phi^*}^{\infty} (p(\phi)x(\phi) - w(\phi)l(\phi)) \, dG(\phi) + \int_{\phi^*}^{\infty} w(\phi)l(\phi) \, dG(\phi). \]

Because \( \int_{\phi^*}^{\infty} p(\phi)x(\phi) \, dG(\phi) = PX \) holds by definition, we can use (2) to obtain: \( I = \mu \). Inserting this result into the indirect utility function leads to:

\[ V = \mu \ln \mu - \mu \ln P. \]

Solving (24) with respect to \( P \) and inserting the result into the last expression yields:

\[ V(\phi^*) = \mu \ln \mu + \mu \ln \phi^* - \mu \ln \left( \frac{1}{\alpha} \left( \frac{\epsilon F}{\mu} \right)^{\frac{1}{\xi}} \right). \]

As a result, welfare is an increasing function of the cutoff productivity \( \phi^* \). This is because a higher \( \phi^* \) is driven by a lower level of the price index (see Lemma 3) such that individuals can increase their consumption level. Thus, indirect utility and welfare rise.

4 Unionization and industry structure

In our framework, we embed endogenous unionization into an industry structure that is characterized by heterogeneous firms and monopolistic competition. This approach allows us to answer two questions: (i) How does endogenous unionization affect the pattern of an industry and (ii) how does a change of the industry structure influence the pattern of unionization?

4.1 Organization costs

Unionization varies with average organizations costs \( c \). To investigate the impact of endogenous unionization on the industry structure, we therefore analyze variations in average costs \( c \) (which also measures marginal organization costs in our model).

First, we look at the impact on the productivity distribution of active firms. This means that we investigate how variations in \( c \) influence the well-known firm-selection effect (see Melitz, 2003 who establishes this mechanism as a result of trade liberalization). Formally, we can calculate this effect by differentiating \( \phi^* \) with respect to \( c \) and find:
Proposition 1  The equilibrium cutoff productivity is u-shaped in organization costs, i.e. it is decreasing in $c$ at zero and low organization costs, reaches a minimum at $c = \hat{c}$ and is increasing from there on.

Proof 1  Differentiating $\bar{\phi}^*$ with respect to $c$ (see Appendix A.4) yields

$$\frac{\partial \bar{\phi}^*}{\partial c} \leq (>)0 \iff c \leq (>)\hat{c},$$

$$\hat{c} \equiv \left[ 1 + (\vartheta - 1)\alpha k \right]^\frac{1}{\xi} - 1.$$

Intuitively, an increase in organization costs has two countervailing effects. On the one hand, the union sets a higher unconstrained wage as it must cover higher marginal costs, i.e. $w^u$ increases (see Lemma 1). Firms that pay the unconstrained wage react to this increase of their marginal costs by raising their prices. At the industry level, the price index $P$ increases and competition becomes softer; $\bar{\phi}^*$ c.p. decreases (see Lemma 3). On the other hand, a smaller share of firms pays the unconstrained wage (see Lemma 1) and is unionized (see Lemma 2). These firms pay lower wages ($\tilde{w}$ instead of $w^u$ and $\tilde{w}$ instead of $\bar{w}$) and reduce their prices. At the industry level, the price index decreases and competition becomes more fierce; $\bar{\phi}^*$ c.p. increases (see Lemma 3). It is evidently from Proposition 1 that for relatively high (low) values of $c$, the second (first) effect dominates and $\bar{\phi}^*$ increases (decreases).

As a result, the firm-selection effect of a variation in organization costs is ambiguous, which implies that the effect on the productivity distribution of active firms and aggregate industry productivity is ambiguous as well. If $\bar{\phi}^*$ decreases (increases), low-productive firms benefit (are worse off) and aggregate industry productivity declines (rises). Moreover, the u-shape relationship between $c$ and $\bar{\phi}^*$ shows that both the productivity distribution and aggregate industry productivity could be similar for different values of organization costs. This may help to justify similar productivity patterns in industries with vastly different degrees of unionization, and may also be relevant for guiding empirical studies.

The u-shape relationship between the organization costs and the cutoff productivity has another implication. Looking at the two extreme scenarios ‘no unionization’ ($c \to \infty$) and ‘full unionization’ ($c = 0$), we find that both are conducive towards high cutoff productivity levels. In addition, we can prove that in non-unionized industries the cutoff productivity takes the highest value.

Proposition 2  The cutoff productivity in an industry with no unionization ($c \to \infty$) is higher than in an industry with full unionization ($c = 0$).
Proof 2 Using (31), we obtain:

\[
\frac{\tilde{\phi}^*_{c \to \infty}}{\tilde{\phi}^*_{c = 0}} = \left[ \frac{\vartheta}{1 + (\vartheta - 1) \alpha^k} \right]^{\frac{1}{k}} > 1,
\]

which is unambiguously greater than one because of \(0 < \alpha < 1\).

In an industry with no unionization, all firms pay the (relatively low) competitive wage such that firm-level prices are relatively moderate. This implies that the price index \(P\) is relatively low and competition is relatively fierce. Firms with low productivity levels cannot survive under this condition which is captured by a high level of \(\tilde{\phi}^*\). If all firms in the industry are unionized, they pay relatively high wages implying that the price index is relatively high respectively competition is relatively soft. More low-productive firms can survive under these conditions such that \(\tilde{\phi}^*_{c \to \infty} > \tilde{\phi}^*_{c = 0}\) holds. This result also shows that aggregate industry productivity is higher without unionization because the share of low-productive active firms is less pronounced in this case.

Next, we analyze the effects of variation in \(c\) on welfare. Because \(V\) depends positively on \(\tilde{\phi}^*\), we can use Propositions 1 and 2 to conclude:

**Corollary 1** Welfare is u-shaped in organization costs. In addition, an industry with no unionization has a higher level of welfare compared to an industry with full unionization.

As argued above, an increase in the cutoff productivity is caused by a decline in the price index \(P\) which in turn raises the consumption level and hence welfare. Therefore, unionization affects welfare in the same direction as it affects \(\tilde{\phi}^*\).

### 4.2 Elasticity of substitution

Besides the mentioned results, unionization also leads to a new mechanism through which the elasticity of substitution \(\epsilon\) affects the cutoff productivity and hence welfare.

**Proposition 3** With unions \((c < \infty)\), the degree of substitutability among varieties \(\epsilon\) has an ambiguous effect on \(\tilde{\phi}^*\). Without unions \((c \to \infty)\), the cutoff productivity unambiguously increases with \(\epsilon\).

**Proof 3** The first part of the proposition follows immediately from (31).

Moreover, note that \(\tilde{\phi}^*_{c \to \infty} = \left( \frac{\sigma_g F}{\mu} \right)^\frac{1}{k} \phi_{min}\). Differentiating this with respect to \(\epsilon\) proves the second part.
If there are no unions in the industry, an increase in $\epsilon$ leads to a lower price mark-up which reduces the firms’ profits and hence monopoly rents. Consequently, a higher fraction of low-productive firms cannot cover their fix costs and leave the market; $\tilde{\phi}^*$ increases. This mechanism is the standard effect in heterogeneous firm models. With unionization, however, there are two additional effects. On the one hand, the unconstrained union wage decreases because of the decline in monopoly rents. Firms can reduce their prices which works against the aforementioned standard effect. On the other hand, the pattern of unionization is also affected because the firm’s feasible wage changes. However, the sign of this change is a priori unclear (see (15)). If $\tilde{w}$ decreases, unionization is profitable for a smaller share of firms implying that more firms pay the competitive wage. The price index decreases which would also work against the described standard effect.

### 4.3 Firm heterogeneity

As a variation in the industry structure, we consider a change in the degree of firm heterogeneity. We measure this variation by a mean preserving spread (i.e. holding the expected value of the distribution $E$ constant while increasing its variance $\sigma^2$). As shown in Appendix A.5, the mean preserving spread leads to a decline of the shape parameter $k$:

$$\frac{\partial k}{\partial \sigma^2}_{E=0} = k_\sigma < 0. \tag{33}$$

With respect to the mass of unionized firms, we find:

**Proposition 4** Increasing heterogeneity (by a mean preserving spread of the productivity distribution) has ambiguous effects on the mass of unionized firms.

**Proof 4** Using (31), we can rewrite (26) as:

$$M_U = \frac{\mu}{cF} \left[ (1+c)^{k_\vartheta} + 1 + (\vartheta - 1)\alpha^k - \vartheta(1+c)^k \right]^{-1}.$$  

Because $\partial \vartheta / \partial k < 0$, the sign of $\partial M_U / \partial k$ is in general ambiguous. Together with (33), this proves the Proposition.

Looking at the share of unionized firms, we find instead a clear-cut effect.

**Proposition 5** Increasing heterogeneity (by a mean preserving spread of the productivity distribution) raises the share of unionized firms.
Proof 5 Differentiating (27) with respect to \( k \) yields:

\[
\frac{\partial s_M}{\partial k} = -(1 + c)^{-k} \ln(1 + c) < 0.
\]

Using this and (33), we find:

\[
\frac{\partial s_M}{\partial \sigma^2} = \frac{\partial s_M}{\partial k} \kappa > 0,
\]

which proves the Proposition.

Intuitively, the decline in the shape parameter \( k \) implies that the probability mass of low productivity levels decreases. Consequently, more entrepreneurs draw on average a productivity level which is at least as high as the threshold level \( \phi^- \) such that the share of unionized firms increases.

5 Open Economy

5.1 Modifications

So far, we have considered a closed industry without any access to foreign markets. In this section, we consider instead a trade open industry and analyze the influence of trade openness, i.e. the transition from autarky to trade, and trade liberalization on the pattern of unionization.

To that end, we have to extend our model to an open economy setting. Following Melitz (2003), we assume that the world economy consists of two fully symmetric countries and the economy within each country is structured as explained in section 2. We assume that the homogenous good \( Y \) is non-tradable and that the differentiated good \( X \) is tradable. Moreover, exporting is accompanied with fixed costs \( F_x \) (measured in terms of the numeraire) and with variable (iceberg) transport costs \( \tau > 1 \). An entrepreneur with productivity \( \phi \) decides whether to enter the domestic market as before. If the entrepreneur becomes an active firm, it additionally decides whether to export some varieties abroad.

Conditional on the fact that a firm with productivity \( \phi \) enters the export market, profit maximization implies that we can express all export variables as functions of their domestic counterparts (which is a consequence of the given CES preferences and also the standard result in the corresponding literature). Export prices, number of exported varieties and the number of workers who are employed to produce the export goods are hence given by,
respectively:

\[ p_x(\phi, w) = \tau_p(\phi, w), \]
\[ x_x(\phi, w) = \tau^{-\epsilon} x(\phi, w), \]
\[ l_x(\phi, w) = \tau^{1-\epsilon} l(\phi, w), \]

where the subscript \( x \) denotes export variables. Note that the balance of trade is equalized due to our assumption of fully symmetric countries.

Export profits are defined as \( \pi_x(\phi, w) = p_x x_x(\phi, w)\). Using the last results as well as (7) and (11) lead to:

\[ \pi_x(\phi, w) = \tau^{1-\epsilon}(\alpha^{-1} - 1) l(\phi, w) w - F_x. \]  (34)

There is a marginal exporting firm with productivity \( \phi_x^* \) at which export profits are equal to zero: \( \pi_x(\phi_x^*, w) = 0 \). To calculate \( \phi_x^* \), we use (34), (12) as well as the unchanged zero cutoff profit condition for firm entry into the domestic market (24) and arrive at:

\[ \phi_x^* = \tau \left( \frac{F_x}{F} \right)^{\frac{1}{\xi}} w^*, \]  (35)

where \( w \geq \hat{w} \equiv 1 \) holds. If we assume \( \tau^\xi F_x > F \) (as in Melitz (2003)), we find a partitioning of firms by export status: Firms with productivities \( \phi^* \leq \phi < \phi_x^* \) serve the domestic market only, while firms with productivities \( \phi \geq \phi_x^* \) additionally export.

For our further steps, it is important to note that we are not able to analytically clarify which wage rate the marginal exporting firm must pay. It could be that this wage rate is the unconstrained wage rate, i.e. \( \phi_x^* \geq \phi^+ \), but it is also possible that this wage rate is just the limit wage, i.e. \( \phi^- \leq \phi_x^* < \phi^+ \). As an extreme case, we also cannot eliminate that the marginal exporting firm is not unionized, i.e. \( \phi^* < \phi_x^* < \phi^- \).

5.2 Autarky versus trade

As in the closed economy setting, the equilibrium is fully described if the cutoff productivity \( \phi^* \) is known. Eq. (24) gives the cutoff productivity as a function of the price index. A second relationship between \( P \) and \( \phi^* \) can be obtained by using the definition of the price index as before. In the open economy setting, the latter is equal to:

\[ P = \left( \int_{\phi^*}^{\infty} p(\phi, w)^{-\xi} dG(\phi) + \int_{\phi_x^*}^{\phi^*} p_x(\phi, w)^{-\xi} dG(\phi) \right)^{-\frac{1}{\xi}}. \]  (36)
For the first summand in the bracket, we can use the same method as in section 3.1 and arrive at an analogous expression to (30) which suggest that the price index is a positive function of $\phi^*$. The problem here is the second summand in the bracket. Without knowledge about the wage rate for the marginal exporting firm (see above), we are not able to solve for the integral and hence cannot provide a closed form solution.

Nevertheless, we are able to show how the price index changes compared to the autarky case. Let us define $P_d \equiv \int_{\phi^*}^{\infty} p(\phi, w)^{-\xi}dG(\phi)$ and $P_x \equiv \int_{\phi^*}^{\infty} p_x(\phi, w)^{-\xi}dG(\phi)$, where the index $d$ stands for domestic. Then, we can rewrite (36) as: $P = (P_d + P_x)^{-1/\xi}$. If we compare this expression with the autarky level (for which we use the subscript $a$), we find:

$$\frac{P}{P_a} = \left( \frac{P_d + P_x}{P_d} \right)^{-\frac{1}{\xi}} < 1,$$

where we use $P_a = P_d^{-1/\xi}$ and $P_x > 0$. Hence, $P$ decreases if trade is allowed. Therefore, we can conclude:

**Lemma 6** If an industry switches from autarky to trade, the cutoff productivity increases.

Intuitively, there are more varieties available if trade is allowed such that consumers spread their income over a greater range of varieties. This implies, however, that the demand for each variety drops, the price index decreases and competition among domestic firms becomes more fierce. The least productive firms are not able to compensate for the implied reduction of their profits and leave the market; $\phi^*$ increases.

With this finding at hand, we are able to analyze the impact of trade openness on the unionization structure.

**Proposition 6** Trade openness reduces (i) the mass of unionized firms while (ii) the share of unionized firms remains constant.

**Proof 6** From (25) and (26), we find $\partial M_U/\partial \phi^* < 0$. Together with the increase in $\phi^*$, this proves the first part of the Proposition. The second part follows immediately from (27).

Trade openness leads to higher competition such that firms’ profits decline. Because organizations costs are unaffected, this implies that profits of some firms are no longer high enough to cover average organization costs. Therefore, fewer firms are unionized. However, higher competition also drives the least productive firms out of the market such that the mass of active firms decreases. With respect to the share of unionized firms, both effects offset each other and $s_M$ is constant.
5.3 Trade liberalization

Finally, we analyze the effects of trade liberalization. Due to variable and fixed trade costs, trade can be liberalized by either reducing $\tau$ or $F_x$. A decline in variable trade costs $\tau$ has two effects. On the one hand, exporting firms are able to lower their prices $p_x$, i.e. the effect on the intensive margin. On the other hand, the productivity of the marginal exporting firm $\phi^*_x$ decreases implying that more firms enter the foreign market, i.e. the effect on the extensive margin. Using (36), we immediately find that $\partial P/\partial p_x > 0$ and $\partial P/\partial \phi^*_x > 0$. Both effects hence imply a reduction of the price index which in turn leads to an increase in the cutoff productivity as in the previous discussed trade openness case.

If fixed trade costs $F_x$ decline, $\phi^*_x$ decreases and more firms are able to export while the export price $p_x$ and hence the intensive margin is unaffected. Nevertheless, the price index also decreases and the cutoff productivity increases. Therefore, we can conclude:

**Corollary 2** Trade liberalization either by a reduction of $\tau$ or $F_x$ yields to an increase in the cutoff productivity and has hence the same qualitative effects on the mass and share of unionized firms as in the trade openness case. Because a reduction of variable trade costs affects the extensive and intensive margin, the quantitative effect on the mass of unionized firms is higher compared to a reduction of fixed trade costs.

6 Conclusion

In this paper, we studied the bidirectional relationship between unionization and industry productivity. This relationship is relevant because unions may have a pronounced effect on the allocation of resources within an industry, and hence on aggregate industry productivity. In addition, industry characteristics may have an effect on the unionization rate. We introduce endogenous unionization into a model with firm heterogeneity and monopolistic competition.

Using the model we show how general wage setting reforms, measured by the costs of organizing workers, may have a non-monotonic impact on aggregate industry productivity. This highlights potentially unintended consequences of labor market reforms: raising barriers to unionization at high levels may induce unions to organize less firms, but resource-reallocation may imply that more productive unionized firms become larger, and that employment of unionized workers in the industry increases. Moreover, we show that the share of unionized firms decreases if the productivity dispersion among
firms declines. Reallocation caused by international trade implies a decline in the mass of unionized firms while its share remains constant. Our model thus offers an explanation for the substantial decline in unionization rates.

While the strong decline in unionization rates in some countries and the beneficial effect of trade on industry productivity are substantiated by thorough empirical research, the quantitative relevance of the mechanisms proposed in this paper remains unclear. For example, we have assumed that firms can only choose to produce or refrain from doing so. But firms in a global economy may gain a strategic edge through the options of outsourcing or FDI, which will impact the emergence and wage setting behavior of unions. While we leave these theoretical considerations for future work, we believe that the evidence for the bias in unionization rates towards more productive firms, and the plausible response of labor market institutions to increasing economic integration suggest that endogenous unionization may be a relevant lever for industry productivity and, thus, for welfare.

A Appendix

A.1 Derivation of the price index as given by Eq. (30)

First, we rearrange (29) to obtain:

\[ P^{-\xi} = \alpha^{\xi} \left( \frac{\int_{\phi^*}^{\phi^*} \phi^{\xi} dG(\phi) + \left( \frac{1+\epsilon}{\alpha} \right)^{-\xi} \int_{\phi^*}^{\phi^*} \phi^{\xi} dG(\phi)}{1 - \frac{E_{\xi}}{\mu} \int_{\phi^*}^{\phi^*} dG(\phi)} \right). \]  

(A-1)

Inserting (25) into (A-1), we can formulate the price index as a function of the cutoff productivity level \( \phi^* \) (and model’s parameters) only:

\[ P^{-\xi} = \alpha^{\xi} \left( \frac{\int_{\phi^*}^{(1+\epsilon)\phi^*} \phi^{\xi} dG(\phi) + \left( \frac{1+\epsilon}{\alpha} \right)^{-\xi} \int_{\phi^*}^{(1+\epsilon)\phi^*} \phi^{\xi} dG(\phi)}{1 - \frac{E_{\xi}}{\mu} \int_{\phi^*}^{(1+\epsilon)\phi^*} dG(\phi)} \right). \]  

(A-2)

In a next step, we use the fact that the price index is inter alia a function of the \( \xi \)th moment of subintervals of the productivity distribution Given the distribution function (6), we can derive an explicit expression of this moment for the subset of productivities between, say, \( a \) and \( b \), which reads:

\[ \tilde{\phi}(a, b) \equiv \int_{a}^{b} \phi^{\xi} dG(\phi) = \frac{\phi^{\xi}}{\phi_{\text{min}}^{\xi}} \left( b^{-(k-\xi)} - a^{-(k-\xi)} \right) > 0, \]  

(A-3)
with $\vartheta \equiv k/(k - \xi) > 1$. Inserting (A-3) into (A-2), we obtain:

$$P^{-\xi} = \alpha^\xi \left( \frac{-\vartheta \phi_{\text{min}}^k \phi^*^{-k} (k - \xi) \left( (1 + c)^{-k} (1 - \frac{1+c}{\alpha})^{-k} \right)}{1 - \frac{F_x}{\mu} (\phi^*)^{-k} (\phi_{\text{min}})^k \left( (1 + c)^{-k} - \frac{(1+c)}{\alpha}^{-k} \right)} \right).$$

Rearrangements yield:

$$P^{-\xi} = \alpha^\xi \left( \frac{-\vartheta \phi_{\text{min}}^k \phi^*^{-k} (k - \xi) \left( - (1 + c)^{-k} \left( (1 - \frac{1+c}{\alpha})^{-k} + 1 + \frac{(1+c)}{\alpha}^{-k} \right) \right)}{1 - \frac{F_x}{\mu} (\phi^*)^{-k} (\phi_{\text{min}})^k \left( (1 + c)^{-k} - \frac{(1+c)}{\alpha}^{-k} \right)} \right)^{-1/\xi},$$

$\Leftrightarrow P = (\alpha \phi^*)^{-1} \left( \frac{-\vartheta \phi_{\text{min}}^k \phi^*^{-k} (k - \xi) \left( - (1 + c)^{-k} \left( (1 - \frac{1+c}{\alpha})^{-k} + 1 + \frac{(1+c)}{\alpha}^{-k} \right) \right)}{1 - \frac{F_x}{\mu} (\phi^*)^{-k} (\phi_{\text{min}})^k \left( (1 + c)^{-k} - \frac{(1+c)}{\alpha}^{-k} \right)} \right)^{-1/\xi},$ where the last Eq. is equal to (30).

**A.2 Proof for $\partial P/\partial \phi^* > 0$**

Differentiating (30) with respect to $\phi^*$ yields after some rearrangements:

$$\frac{\partial P}{\partial \phi^*} = P \left( k \frac{\phi_{\text{min}}^k \phi^*^{-k} (1+c)^{-k} - \frac{\epsilon F}{\mu} \phi_{\text{min}}^k \phi^*^{-k} (1+c)^{-k} - \frac{(1+c)}{\alpha}^{-k}}{\left( (1 + c)^{-k} - \frac{(1+c)}{\alpha}^{-k} \right)} - 1 \right).$$

Hence, we get:

$$\frac{\partial P}{\partial \phi^*} > 0 \Leftrightarrow \frac{k}{\xi} > 1 - \frac{\epsilon F}{\mu} \left( \frac{\phi_{\text{min}}^k}{\phi^*^k} \right) \left( (1 + c)^{-k} - \left( \frac{1+c}{\alpha}^{-k} \right) \right). \quad (A-4)$$

Due to our assumption that $k > \xi$, this condition is fulfilled if the term in the squared bracket is non-negative. Since $0 < \alpha < 1$ holds by definition, we find $\partial P/\partial \phi^* > 0$ for any permitted parameter constellation.

**A.3 Derivation of $\bar{\phi}^*$ as given by Eq. (31)**

Inserting Eq. (30) into (24), we find after some rearrangements:

$$1 = \frac{\vartheta \phi_{\text{min}}^k \left( - (1 + c)^{-k} + 1 + \frac{(1+c)}{\alpha}^{-k} \right) \epsilon F}{\left( \phi^*]^k - \frac{F_x}{\mu} \phi_{\text{min}}^k \left( (1 + c)^{-k} - \frac{(1+c)}{\alpha}^{-k} \right) \right)^2} \mu.$$

(A-5)
This expression can be solved with respect to $\phi^*$:

$$
\phi^* = \left( \frac{\vartheta \left( - (1 + c)^{(k - \xi)} + 1 + \left( \frac{1 + c}{\alpha} \right)^k \right) + (1 + c)^{-k} }{\left( \frac{\mu F}{\epsilon} \right)^{-1}} \right)^{1/k} \phi_{\min}.
$$

Further rearrangements yield:

$$
\bar{\phi}^* \equiv \phi^* = \left( \frac{\vartheta + (1 + c)^{-k} \left( 1 + (\vartheta - 1)\alpha^k - \vartheta(1 + c)^{\xi} \right) }{\left( \frac{\mu F}{\epsilon} \right)^{-1}} \right)^{1/k} \phi_{\min},
$$

which is identical to Eq. (31).

**A.4 Proof of Proposition 1**

Differentiating (31) with respect to $c$ yields:

$$
\frac{\partial \bar{\phi}^*}{\partial c} = \frac{1}{k} \frac{\bar{\phi}^* \partial \Psi}{\Psi} \frac{\partial \Psi}{\partial c},
$$

(A-6)

with

$$
\Psi \equiv \vartheta + (1 + c)^{-k} \left[ 1 + (\vartheta - 1)\alpha^k - \vartheta(1 + c)^{\xi} \right].
$$

Using the definition of $\vartheta$, we obtain:

$$
\frac{\partial \Psi}{\partial c} = - \frac{\epsilon F}{\mu} k (1 + c)^{-k-1} f,
$$

(A-7)

with

$$
f \equiv 1 + (\vartheta - 1)\alpha^k - (1 + c)^{\xi}.
$$

Using (A-7), we can rewrite (A-6) as:

$$
\frac{\partial \bar{\phi}^*}{\partial c} = - \frac{\bar{\phi}^* \epsilon F}{\Psi} \frac{\partial \Psi}{\partial c} (1 + c)^{-k-1} f.
$$

(A-8)

Therefore, we find that $\partial \bar{\phi}^*/\partial c < (\geq)0$ if and only if $f > (\leq)0$. Rearrangements yield:

$$
\frac{\partial \bar{\phi}^*}{\partial c} < (\geq) \iff c < (\geq) \hat{c},
$$

(A-9)

where $\hat{c}$ is defined as in the main text. This proves Proposition 1.
A.5 Mean preserving spread

The density function of the Pareto distribution is given by

\[ g(\phi) = \frac{k}{\phi_{\text{min}}} \left( \frac{\phi_{\text{min}}}{\phi} \right)^{k+1}. \] (A-10)

With (A-10) at hands, we can compute the mean \( E \) and the variance \( \sigma^2 \) of the distribution and obtain, respectively:

\[ E = \int_{\phi_{\text{min}}}^{\infty} \phi g(\phi) d\phi = \frac{k}{k-1} \phi_{\text{min}}, \] (A-11)

\[ \sigma^2 = \int_{\phi_{\text{min}}}^{\infty} \phi^2 g(\phi) d\phi = (\phi_{\text{min}})^2 \frac{k}{(k-2)(k-1)^2}, \] (A-12)

where we assume that \( k > 2 \).

We can 'solve' these two equations for \( k \) and \( \phi_{\text{min}} \) which leads to:

\[ k = 1 + \sqrt{\frac{E^2}{\sigma^2} + 1}, \] (A-13)

\[ \phi_{\text{min}} = \frac{E}{\left( \sqrt{\frac{E^2}{\sigma^2} + 1} \right)^{-1} + 1}. \] (A-14)

From (A-13), we see that a mean preserving spread (holding \( E \) constant and increasing \( \sigma^2 \)) implies a decrease in the shape parameter \( k \):

\[ \frac{\partial k}{\partial \sigma^2}_{dE=0} < 0. \] (A-15)
References


Dreher, Axel and Noel Gaston, “Has globalisation really had no effect on unions?,” *Kyklos*, 2007, 60 (2), 165–186.


