# Oligopsony Power and Factor-Biased Technology Adoption

Michael Rubens*

February 16, 2022

**Abstract**

I show that oligopsony power of firms over their input suppliers could either increase or decrease their technology adoption, depending on the direction of technical change and the technology's Hicks-neutral effects. I illustrate this in an empirical application that features oligopsonistic labor markets and a large technology shock: the introduction of mechanical coal cutters in the 19th century Illinois coal mining industry. By estimating a model of coal production and labor supply using rich mine-level data, I find that the returns to cutting machine adoption would have increased by 11% when moving from one to four firms per labor market.

**Keywords:** Oligopsony, Market Power, Innovation, Technological change, Productivity

**JEL codes:** L11, L13, O33, J42, N51

*KU Leuven, Department of Economics. Contact: michael.rubens@outlook.com

This paper previously circulated under the title 'Monopsony Power and Factor-Biased Technology Adoption'. I thank Jo Van Biesebroeck, Jan De Loecker, Frank Verboven, Chad Syverson, Otto Toivanen, Penny Goldberg, Steve Berry, Suresh Naidu, and Daniel Haanwinckel for feedback and discussion, and participants at the KU Leuven and Yale IO student workshops and various conferences. This is the second chapter of my Ph.D. dissertation at KU Leuven. Financial support from the Belgian American Educational Foundation, the Fulbright program, the Research Foundation Flanders (FWO grant 11A4618N), and the European Research Council (ERC grant 816638) is gratefully acknowledged.

# 1 Introduction

There is increasing empirical evidence for the existence of oligopsony power across various industries, countries, and types of factor markets.[1] When studying the welfare consequences of such oligopsony power, prior research has typically assumed that oligopsony power does not affect firms' technology choices. In contrast to this stands a large literature that studies the effects of imperfect *product* market competition on innovation incentives.[2] This paper fills this gap by examining how oligopsony power affects innovation. The focus of the paper lies on the *adoption* of new technologies, rather than on their invention, and on process innovations, which affect the cost side of production, rather than on product innovations, which shift the product demand curve.

I start the analysis with a theoretical model of a firm that produces a homogeneous good using two homogeneous inputs, and faces log-linear upward-sloping input supply curves. The firm is both a monopolist downstream and a monopsonist upstream, and hence sets the price of each input at a markdown below its marginal revenue product. I consider the introduction of a new technology that could have factor-biased effects, by changing the marginal rate of technical substitution, and/or could change Hicks-neutral productivity. A first key result is that the relative change in profits following technology adoption weakly increases with the markdown over the input towards which the technology is biased, but falls with the markdown over the other input. The intuition behind this result is that adopting a technology is more profitable if it increases demand for the factor of which the input price markdown is the highest, because that is the factor from which the firm extracts the highest surplus. A second key result is that technology adoption increases with the price markdown of the input towards which the technology is biased. The net effect of the price markdown of the other input on technology adoption depends on the relative size of the rotation and shift of the production isoquant.

Given that the effect of oligopsony power on technology adoption has an ambiguous sign, and to quantify its size, I turn to an empirical application. I study how the mechanization of the Illinois coal mining industry between 1884 and 1894 was affected by oligopsony power on the market for coal miners. There are three reasons why this provides an interesting setting to study the relationship between oligopsony power and innovation. Firstly, 19th century Illinois coal mining towns are a textbook example of classical oligopsony power, as local labor markets were isolated and highly concentrated due to prohibitive commuting distances between towns. Up to 1898, wages were set unilaterally by firms, without collective bargaining with labor unions. Secondly, the introduction of coal cutting machines in the U.S. in 1882, which started the transition from manual to mechanical mining, provides a large technological shock. The data set tracks the usage of these cutting machines over time, together with input and output quantities, wages and coal prices, all at the mine level. Thirdly, bituminous coal firms are single-product firms producing a nearly homogeneous product,

[1] See literature reviews by Ashenfelter et al. (2010) and Manning (2011), and recent papers by, among others, Naidu et al. (2016); Berger et al. (2019); Rubens (2020); Morlacco (2017); Lamadon et al. (2019); Kroft et al. (2020).

[2] Examples include, among many others, Schumpeter (1942), Aghion, Bloom, Blundell, Griffith, and Howitt (2005) and Igami and Uetake (2017).

which facilitates the empirical analysis.

The central counterfactual question of the paper is how changes in labor market competition would have affected cutting machine adoption.[3] In order to answer this question, I construct an empirical model of input supply and demand in the coal mining industry, which has three components. First, I specify a production function for coal with three factors: skilled miners who cut coal, low-skilled other workers who did a variety of tasks such as driving mules and sorting coal, and capital, in the form of cutting machines. I rely on a Cobb-Douglas production function in both labor types, but with output elasticities that are a function of cutting machine usage, and that vary flexibly across firms and over time.[4] This is crucial because anecdotal historical evidence strongly suggests that cutting machines were not Hicks-neutral, but biased towards unskilled workers, similarly to many other technologies throughout the 19th century (James & Skinner, 1985; Mokyr, 1990; Goldin & Katz, 2009). Second, I specify a coal demand model in which coal firms compete along the same railroad in a static Cournot game, assuming their output is undifferentiated given their location. The production function and coal demand model jointly determine the demand for all inputs. Thirdly, the supply for each labor type is modelled as a log-linear supply curve of which both the supply elasticity and vary flexibly across firms and over time. I assume that firms are homogeneous from the employees' point of view, and model labor market competition as a static Cournot employment-setting game. The capital market is assumed to be perfectly competitive, with perfectly elastic supply of cutting machines. In contrast to the theoretical model, which has monopsonistic labor markets, the empirical model features oligopsonistic labor markets, as most labor markets contained multiple coal firms.[5]

I estimate the production model with firm-level data on output and input quantities, and rely both on the profit maximization assumption and on input timing assumptions for identification. I find that cutting machines were unskill-biased, which confirms contemporaneous anecdotal evidence, and increased Hicks-neutral productivity.[6] The coal demand model is used using market-level price and quantity data, and is identified by exploiting geological variation in the thickness of coal seams as cost shifters that are excluded from consumer utility. Finally, the labor supply model is estimated using labor-market level data on wages and employment, and is identified using seasonal weather variation, which shifts labor demand but is assumed not to shift labor supply.[7] The labor supply estimates reveal a moderate degree of oligopsony power over skilled workers, but no oligopsony power over unskilled workers.[8]

[3] Such changes in competition could be due to competition policy, or to technological change that change labor supply elasticity, such as improved mobility infrastructure that expands the employment choice set of workers.

[4] In other words, I allow for the technology to change both $\beta$ and $A$ in $Y = AH^{\beta}L^{\nu-\beta}$, and I allow for unobserved variation across firms and time in both $A$ and $\beta$.

[5] The empirical model collapses to the theory model when assuming that labor market shares become one, meaning that every firm is its own market.

[6] With the aforementioned production function $Y = AH^{\beta}L^{\nu-\beta}$, cutting machines lowered $\beta$ and increased $A$.

[7] Evidence motivating this assumption is discussed in the paper.

[8] Miner skills, such as building mine roofs or knowing how thick pillars should be in order to avoid collapse, were not easily transferable to other industries. This explains why coal mines enjoyed some wage-setting power over their skilled

I combine the estimated labor supply and demand model to find the equilibrium, which is a function of labor market structure. Using the estimated model, I conduct the counterfactual exercise of how changes in labor market structure would affect the returns to, and adoption of, new technologies. I carry out this exercise both for the actual production technology, cutting machines, and for two counterfactual technologies: one that is skill-biased, such as mining locomotives, and another one that is unskill-biased without Hicks-neutral productivity effect, meaning that it only rotates but does not shift the production isoquant. I find that increasing labor market competition would increase the returns to cutting machine adoption: moving from one to ten employers per labor market would increase the average return to cutting machine adoption by 16%. The usage rate of cutting machines would, however, barely change with changing labor market competition. If the technology would have been skill-biased rather than unskilled-biased, changes in labor market competition would have very different effects. Moving from one to ten firms per labor market would now *decrease* the average return to machine adoption by 5%, and technology usage would drop by 50%. Finally, if cutting machines would have been purely unskill-biased, without any Hicks-neutral effect, the same increase in labor market competition would increase the returns to cutting machine adoption by 7% and increase machine usage by 23%. Hence, both the direction and Hicks-neutral productivity effects of the technology are crucial determinants of the size and direction of how labor market competition affects technology adoption.

Although the empirical setting of the paper is historical, the model has important current-day implications. For instance, it sheds new light on how oligopsony power on labor markets affects automation incentives. Although technologies were mostly unskill-biased throughout the 19th century, they have been skill-biased throughout the last part of the 20th century.[9] Hence, the effects of oligopsony power over low- and high-skilled workers on automation incentives may have inverted between the 19th and 20th century. Moreover, oligopsony power over high- and low-skill workers affects automation incentives differently. Knowing both the direction of technical change and the relative wage markdowns for different types of workers is therefore crucial to determine how oligopsonistic labor markets affect technological change today. Especially the latter is a mostly open empirical question: the labor literature has mainly focused on oligopsony power over low-skilled workers, such as Card and Krueger (1994), for instance due to a lack of outside options of workers (Schubert, Stansbury, & Taska, 2020). Non-compete clauses are, however, most frequent among high-skilled jobs in the U.S. (Starr et al., 2020). The model also has implications beyond the study of labor markets. Energy-saving production technologies are another example of directed technological change. If energy-intensive manufacturing firms have some local market power on energy markets, the model can be used to understand how such market power affects the incentives to adopt technologies that are more energy-efficient.

By examining the effect of oligopsony power on directed technological change, this paper con-

laborers, but not over their unskilled laborers, who could switch to other jobs at a lower financial loss.

[9] Or, or at least, hollowing out the center of the skill and income distribution Autor et al. (2006); Goos and Manning (2007); Goos et al. (2014). (Katz & Margo, 2014) argue this also held for technical change during the second industrial revolution.

tributes to four different sets of literature. First, it fits within a large literature that studies the relationship between competition and innovation (Schumpeter, 1942; Aghion et al., 2005; Collard-Wexler & De Loecker, 2015; Hashmi & Van Biesebroeck, 2016; Igami & Uetake, 2017). Whereas this literature studies the effects of product market power on innovation, the focus of this paper lies on the innovation effects of input market power. In their study of tomato harvesters, Just and Chern (1980) examine how oligopsony power of buyers affects technology adoption of their suppliers, and the same holds for Huang and Sexton (1996), Köhler and Rammer (2012), and Parra and Marshall (2021). In contrast, I focus on technology adoption by the buyers who exert oligopsony power. Inderst and Wey (2003) and Loertscher and Marx (2020) equally study investment and buyer power, but in bilateral oligopoly, rather than oligopsony. They also do not consider directed technological change. Finally, Goolsbee and Syverson (2019) find that monopsony power over tenure-track faculty induces universities to substitute these workers for adjunct faculty members. In contrast to their paper, I endogenize the choice of the production technology.[10]

Secondly, this paper builds on a large literature on directed technological change and factor bias. I contrast with the seminal models of directed technical change such as Autor et al. (2003); Acemoglu (2002, 2003) and Antras (2004), I do not focus on the invention of new technologies, but take the arrival of a new technology as exogenous, and examine differences in the adoption of new technologies between firms. Another important difference is that I allow input prices to be endogenous from the point of view of individual firms, meaning that labor supply functions are upward-sloping to the firms. By relaxing the assumption that input prices are exogenous to individual firms, I also contribute to the literature on production function identification with non-Hicks-neutral technical change, such as Doraszelski and Jaumandreu (2017), and Demirer (2020). However, I impose stronger assumptions on the substitution elasticity between production inputs and on the transition process of output elasticities and Hicks-neutral productivity. This paper is also related to contemporaneous work by Haanwinckel (2018) and Lindner et al. (2019), who examine the effects of skill-biased technologies on skill demand and wage inequality with imperfectly competitive labor markets. This paper is distinct from these papers by allowing for factor-biased technology choices that are endogenous to the degree of oligopsony power on factor markets.

Thirdly, this paper relates to the literature on the welfare effects of market power in general (De Loecker et al., 2020; Edmond et al., 2018), and of oligopsony power in particular (Manning, 2013; Berger et al., 2019; Morlacco, 2017). I contribute to this literature by showing that the adoption of new technologies is endogenous to the degree of oligopsony power, and hence also productivity growth. This is an additional channel through which (input) market power shapes aggregate outcomes and, ultimately, welfare. A subset of this literature focuses on the productivity consequences of market power through its effects on allocative efficiency (Harberger, 1954; Asker et al., 2019). This paper is complementary to these approaches: oligopsony affects aggregate productivity not only through

[10]Whereas Goolsbee and Syverson (2019) only allow for changes *along* the production isoquant, keeping the isoquant fixed, I allow for both changes along the isoquant and changes of the isoquant itself, due to technological change.

reallocation, but also by affecting technology choices, which in term determine productive efficiency.

Finally, this paper contributes to a literature on labor market power during the industrial revolution. In his 'induced innovation' hypothesis, Hicks (1932) posited that labor-saving technological change is more likely if wages are high, because cost savings are then higher as well. This theory has been forwarded as a reason why Britain was the first country to experience an industrial revolution (Allen, 2009)[11], and has been studied empirically by Hanlon (2015) and Dechezleprêtre, Hémous, Olsen, and Zanella (2019). However, the induced innovation hypothesis has been critized by Salter (1966) and Acemoglu (2002) because the notion 'expensive input' does make sense if factor prices are equal to marginal products. I solve this logical inconsistency by allowing for a wedge between factor prices and marginal products, and find that the relevant metric to understand innovation incentives is not which factor has the highest price, but which factor has the highest wedge between its price and marginal product. I also contribute to a body of work on oligopsony power during the late 19th century. Naidu and Yuchtman (2017) provides evidence of oligopsony power in 19th century U.S. labor markets and its relationship with labor market institutions. Boal (1995) estimates inverse miner supply functions in the context of West Virginian coal mines from 1897-1932, but uses aggregate data, a different labor supply identification strategy, and assumes exogenous production technologies.

The remainder of this paper is structured as follows: Section 2 contains the theoretical model, Section 3 the industry background, Section 4 the empirical model, its estimation, and the counterfactuals, and Section 5 concludes.

# 2 Theory

## 2.1 Primitives

### A Production

Consider a firm $f$ that produces $Q_f$ units of a homogeneous product using two variable inputs, of which the quantities are denoted $H_f$ and $L_f$. Production is given by a Cobb-Douglas function, in Equation (1a). The output elasticity of input $V \in \{H, L\}$ at firm $f$ is denoted $\beta_f^v$. Scale returns are parametrized as $\nu = \beta_f^h + \beta_f^v$, which is below, above or equal to one if returns to scale are decreasing, increasing, and constant. Total factor productivity is denoted $\Omega_f$. Firms can use a technology $K_f \in \{0, 1\}$, which has a common fixed cost $\Phi$.

$$Q_f = H_f^{\beta_f^h(K_f)} L_f^{\beta_f^l(K_f)} \Omega_f(K_f) \tag{1a}$$

I let both the output elasticities and the Hicks-neutral productivity residual be a function of technology usage. I call the technology $K$ 'H-biased' if $\frac{\partial \beta_f^h}{\partial K_f} > 0$, because $K$ then increases the marginal rate

[11] This hypothesis that has in turn been critized by, among others, Humphries (2013).

of technical substitution of $H$ for $L$, keeping factor proportions constant.[12] Conversely, $K$ is an 'L-biased' technology if $\frac{\partial \beta_f^h}{\partial K_f} < 0$. The technology is 'neutral' if $\frac{\partial \beta_f^h}{\partial K_f} = 0$, and 'directed' otherwise. It is possible that the technology changes only Hicks-neutral productivity $\Omega_f(K_f)$, only the output elasticities $\beta_f^v(K_f)$, or both. I assume that the technology does not change the returns to scale.

Using a Cobb-Douglas production function with technology-specific output elasticities departs from the canonical models on technical change, which usually rely on a constant elasticity of substitution (CES) production function. Although imposing a Cobb-Douglas elasticity between different types of workers is clearly a strong assumption, I allow for directed technical change by making the output elasticities a function of technology usage, and also allow for flexible variation in output elasticities across both firms and time in the empirical application. The main benefit of the Cobb-Douglas function is that it allows analytically expressing the market equilibrium even when labor supply functions are upward-sloping to the individual firms.

## B Input markets

Assume firm $f$ is a monopsonist that belongs to exactly one market for each input. Input prices are $W_f^h$ and $W_f^l$. Each firm faces its own inverse supply functions, one for each input, as in Equation (1b). The firm-level inverse supply elasticity, which is identical to the market-level supply elasticity due to the monopsony assumption, is given by $(\psi_f^h - 1)$ for input $H$, and by $(\psi_f^l - 1)$ for input $L$. Defining $\psi^v$ this way gives it the interpretation of being the ratio of the marginal product of an input $v$ over its price. A value of one implies that the input price is equal to its marginal product. I assume that the supply function for each input is weakly increasing in the input price, meaning that $\psi_f^h \geq 1$ and $\psi_f^l \geq 1$.

$$\begin{cases} W_f^h = {H_f}^{\psi_f^h - 1} \\ W_f^l = {L_f}^{\psi_f^l - 1} \end{cases} \tag{1b}$$

## C Output market

Output is sold at a price $P_f$. The firm is a monopolist on the output market, and faces a log-linear demand curve with inverse elasticity $\eta$, in Equation (1c). I assume that the demand curve is either horizontal or downward-sloping, which implies that $\eta \leq 0$.

$$P_f = Q_f^\eta \tag{1c}$$

[12] $MRTS_{hl} \equiv \frac{\frac{\partial Q}{\partial H}}{\frac{\partial Q}{\partial L}} = \frac{\beta^h}{\nu - \beta^h} \frac{H}{L}$

## 2.2 Behavior and equilibrium

### A Behavior

Variable profits are defined as $\Pi_f \equiv P_f Q_f - W_f^h H_f - W_n^l L_f$, while total profits are $\Pi_f^{tot} \equiv \Pi_f - \Phi K_f$. I assume that firms choose the variable input quantities $H$ and $L$ that maximize the objective function (1d), taking the technology $K$ as given. The parameters $\mu_f^h \in [1, \psi_f^h]$ and $\mu_f^l \in [1, \psi_f^l]$ measure the *actual* markdowns of each input price charged by the firm. Given that the firm is a monopsonist on the market for $H$, the profit-maximizing markdown is equal to the inverse supply elasticity, $\mu_f^h = \psi_f^h$, and similarly for the other input. The labor market equilibrium then lies in the point $M$ in Figure 1b. The lower bound of the markdown, $\mu_f^h = 1$, implies that the price of $H$ is equal to the marginal product of $H$, which corresponds to the competitive equilibrium on the market for $H$. In Figure 1a, this equilibrium is depicted by the point $C$. The markdowns $\mu_f^h$ and $\mu_f^l$ are 'reduced-form' parameters for the degree of market power on each factor market. With a monopsonistic labor market, it might seem strange that firms would ever set markdown different from the profit-maximizing markdown $\mu_f^h = \psi_f^h$.[13] Different types of labor market structure and competition imply, however, lower profit-maximizing markdowns. For instance, in the empirical model in Section 4, the difference between the actual markdown and the inverse input supply elasticity will be micro-founded by replacing the monopsony assumption by an oligopsonistic Cournot model of the labor market, in which markdowns are a function of labor market shares.

$$\max_{H_f, L_f} P_f Q_f - \frac{W_f^h H_f \mu_f^h}{\psi_f^h} - \frac{W_f^l L_f \mu_f^l}{\psi_f^l} \tag{1d}$$

Solving the first order conditions to the optimization problem 1d results in the input demand functions in Equation (1e):

$$\begin{cases} H_f^* &= \frac{P_f Q_f \beta_f^h (1+\eta)}{W_f^h \mu_f^h} \\ L_f^* &= \frac{P_f Q_f \beta_f^l (1+\eta)}{W_f^l \mu_f^l} \end{cases} \tag{1e}$$

By rewriting equation (1e), the parameters $\mu^h$ and 'markdown ratio' of the marginal product of an input over its price. Denote the marginal product of input suppliers $H$ as $MR_f^h \equiv \frac{\partial (P_f Q_f)}{\partial H_f} = \beta_f^h P_f Q_f (1+\eta)$. The 'markdown ratio' is the extent to which the marginal product of $H$ exceeds its price, which is equal to the parameter $\mu^h$:

$$\mu_f^h = \frac{MR_f}{W_f^h} \in [1, \psi_f^h]$$

[13](Goolsbee & Syverson, 2019) also distinguish the actual markdown from the maximum markdown, which is the inverse input supply elasticity.

## B Equilibrium

The supply and demand for goods is given by Equations (1a) and (1c), supply and demand for inputs by Equations (1b) and (1e). Solving this system of equations yields the equilibrium expression for output $Q_f^*$ in Equation (2a), at which both the goods and input markets are in equilibrium.

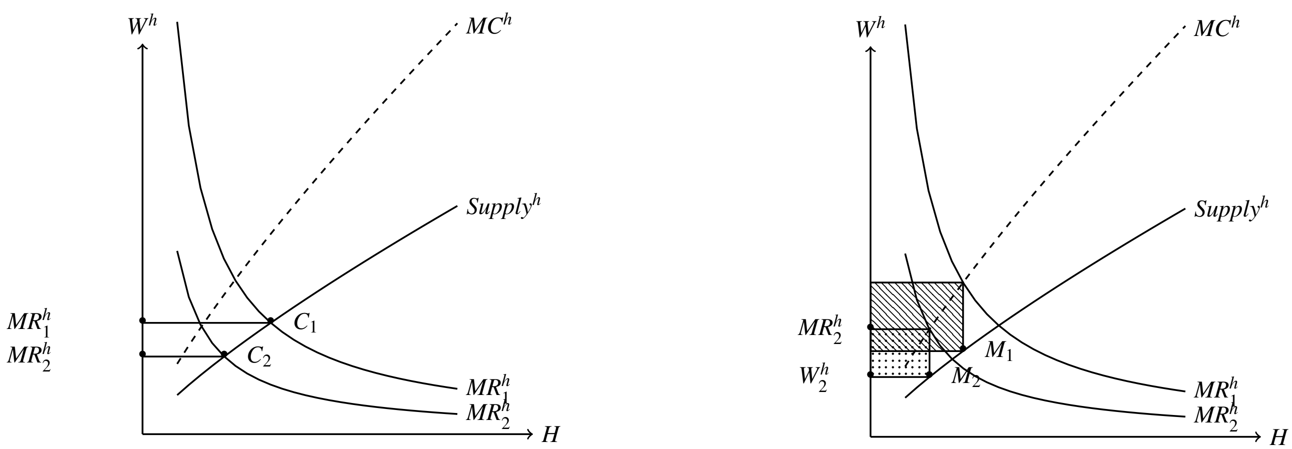
$$Q_f^* = \left[ \left( \frac{\beta_f^h (1+\eta)}{\mu_f^h} \right)^{\frac{\beta_f^h}{\psi_f^h}} \left( \frac{\beta_f^l (1+\eta)}{\mu_f^l} \right)^{\frac{\beta_f^l}{\psi_f^l}} \Omega_f \right]^{\frac{1}{1 - \frac{\beta_f^h (1+\eta)}{\psi_f^h} - \frac{\beta_f^l (1+\eta)}{\psi_f^l}}} \tag{2a}$$

The equilibrium goods price, input prices, and input quantities are functions of this equilibrium quantity. Equilibrium revenue is equal to $Q_f^{*(1+\eta)}$. Equilibrium variable profits $\Pi_f^*$ are equal to the product of equilibrium revenues $Q_f^{*(1+\eta)}$ and the variable profit margin $\left(1 - \frac{\beta_f^h(1+\eta)}{\mu_f^h} - \frac{(\nu - \beta_f^h)(1+\eta)}{\mu_f^l}\right)$:

$$\Pi_f^* = Q_f^{*(1+\eta)} \underbrace{\left( 1 - \frac{\beta_f^h (1+\eta)}{\mu_f^h} - \frac{(\nu - \beta_f^h)(1+\eta)}{\mu_f^l} \right)}_{\text{variable profit margin}} \tag{2b}$$

**Figure 1: Monopsony power and technology choice**

**(a) Competitive markdown:** $\mu^h = 1$ **(b) Monopsonistic markdown:** $\mu^h = \psi^h$



## 2.3 The returns to technology adoption

With the equilibrium expressions at hand, I now considering how the effects of technology usage $K$ on variable profits $\Pi$.

### A Relative profit return

I start by examining the *relative* profit returns to technology adoption, $\frac{\Pi(K=1)-\Pi(K=0)}{\Pi(K=0)}$. For small changes in profits, this ratio is approximated by the log of variable profits $\log(\Pi)$. Theorem 1 says that the markdown of an input price increases the relative returns to a technology that is biased towards that input, but decreases the relative returns to a technology that is biased towards the other input.

**Theorem 1** *Consider a firm $f$ that faces log-linear input supply and product demand curves (5)-(6), a production function (1a) with weakly decreasing returns to scale, $\nu \leq 1$, and is a monopolist on its output market. Then, the price markdown of an input weakly increases the relative variable profit return to a technology that is biased towards that input, but weakly decreases the returns to a technology that is biased towards the other input.*

$$\frac{\partial \beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\; \Rightarrow \frac{\partial^2(\ln(\Pi_f))}{\partial \mu_f^h \partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$$

**Proof:** see Appendix B.1.

The intuition behind this result becomes clear from Figure 1. If the markdown ratio $\mu$ is one, meaning that wages are equal to marginal products, in Figure 1(a), no profit is made from $H$ and $L$. Hence, a technology that is biased towards $L$ has no effect on variable profits, because these are zero to begin with.[14] If the profit-maximizing markdown is charged, $\mu^h = \psi^h$, which corresponds to Figure 1(b), the firm derives profits from the wedge between the marginal product and price of input $H$. Adopting the $L$-biased technology leads to lower usage of $H$, and hence lower profits extracted from $H$. In other words, the higher the input price markdown, the larger the loss in profits if a technology reduces usage of this input. This may seem counter-intuitive: in general, monopsony power leads to firms pushing down the usage of an input in order to decrease its equilibrium price, a point also made by Goolsbee and Syverson (2019). However, that is a movement *along* the input demand curve. I consider shifts of the input demand curve due to technological change.

### B Absolute profit return

Theorem 1 explained how markdowns affect the *relative* change in variable profits in response to technology adoption. However, in order to understand technology adoption, we need to know the

[14] With imperfect competition downstream, the technology can still increase profits.

effects of markdowns on the *absolute* change in *total* profits after machine adoption, $\Pi(K=1) - \Pi(K=0) - \Phi$. Before doing so, I need an intermediate result: Lemma 1 says that variable profits increase with markdowns.

**Lemma 1** *Variable profits increase with the markdown over any input:* $\frac{\Pi_f}{\mu_f^v} \geq 0 \forall v \in \{h, l\}$

**Proof:** see Appendix B.2

I start with examining the absolute change in variable profits, in expression (3). To understand the effect of markdowns on technology adoption, we need to know the sign of the left hand side of Equation (3), $\frac{\partial}{\partial \mu_f^h}\left(\frac{\partial \Pi_f}{\partial K_f}\right)$.

$$\frac{\partial}{\partial \mu_f^h}\left(\frac{\partial \Pi_f}{\partial K_f}\right) = \underbrace{\frac{\partial}{\partial \mu_f^h}\left(\frac{\partial \ln(\Pi_f)}{\partial K_f}\right)}_{A} \Pi_f + \underbrace{\frac{\partial \Pi_f}{\partial \mu_f^h}}_{B} \underbrace{\frac{\partial \Pi_f}{\partial K_f}}_{C} \Pi_f \tag{3}$$

Theorem 2 says that effect of a markdown on absolute technology returns is positive if the effect on relative technology returns is positive, and ambiguous if the effect on relative technology returns is negative.

**Theorem 2** *The markdown of an input price increases the absolute return to a technology that is biased towards that input. It can increase or decrease the absolute return to a technology that is biased towards the other input.*

The proof is as follows. Consider the effect of an input price markdown for input $H$, $\mu_f^h$. Variable profits $\Pi_f$ are assumed to be positive, $\Pi_f > 0$ , otherwise the firm would not operate. From Lemma 1, the term $B$ is positive, $B > 0$. The term $C$ is assumed to be positive, $C > 0$: if adopting the technology would decrease variable profits, firms would never adopt it. From Theorem 1, we know that $A > 0$ if the technology $K$ is $H$-biased, and that $A < 0$ if it is $L$-biased. Hence, if technology $K$ is $H$-biased, then the input price markdown of $H$ always increases its absolute effect on variable profits, because the entire right-hand side is positive. If, on the other hand, technology $K$ is $L$-biased, the effect of markdowns on the absolute profit return from $K$ is ambiguous, as term $A$ is negative. Whether the markdown increases or decreases the absolute return to technology adoption then depends on the relative size of the term $A$, which is negative and the product of terms $B$ and $C$, which is positive.

The intuition behind Theorem 2 is as follows. There are two reasons why the result in Theorem 1 will not necessarily translate to the absolute profit difference. First, the higher markdowns are, the lower variable profits. Even if the relative profit change $\frac{\pi(K=1)-\pi(K=0)}{\pi(K=0)}$ increases with the markdown $\mu^l$, still considering an $L$-biased technology $K$, the absolute profit change $\pi(K=1) - \pi(K=0)$ might be lower with a lower markdown as the baseline profit level $\pi(K=0)$ is lower to begin with. Secondly, if the technology lowers $\beta^h$, this reduces the relative demand for $H$ compared to $L$, as can be seen from the input demand function (1e). However, if the technology also increases Hicks-neutral productivity $\Omega$, it could be that the firm ends up using more of input $H$ after adopting the technology. In that case, a higher markdown over $H$ could increase the absolute profit return from the technology.

## C Hicks-neutral vs. factor-biased technology effects

Suppose the markdown decreases the relative return to technology adoption. Which factors then determine whether the effect of the markdown on the absolute return to adoption will be negative or positive? Lemma 3 says that the Hicks-neutral productivity effect is crucial, as was already explained in the intuition above.

**Lemma 2** *The higher the effect of a technology on Hicks-neutral productivity, the more likely that markdowns increase the absolute return to technology adoption.*

**Proof**: see Appendix B.3

Now consider the limiting case of a neutral technology that only increases Hicks-neutral productivity $\Omega$ but not the output elasticities $\beta$. Lemma 2 says that the markdown on any input market increases the absolute return from such a technology.

**Lemma 3** *The absolute profit effect of a technology that weakly increases Hicks-neutral productivity but does not change the output elasticity of any input then weakly increases with monopsony power on either input market.*

$$\frac{\partial \beta_f^h}{\partial K_f} = 0 \;;\; \frac{\partial \Omega}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \Rightarrow \frac{\partial^2 (\Pi_f)}{\partial \mu_f^h \partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$$

**Proof:** see Appendix B.4.

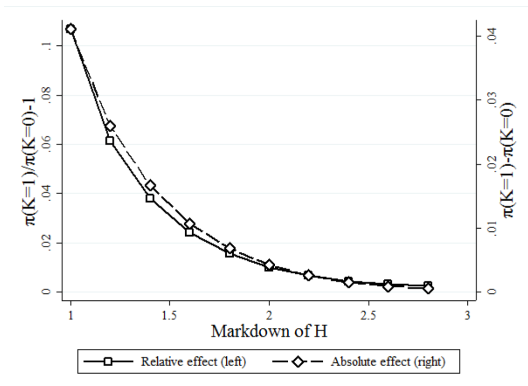
The intuition here is as follows. An increase in Hicks-neutral productivity results in higher equilibrium output produced by the firm. The higher the degree of monopsony power a firm has, the higher profits are, and hence the higher the change in the profit level due to an increase in productivity.

## D Illustration using the calibrated model

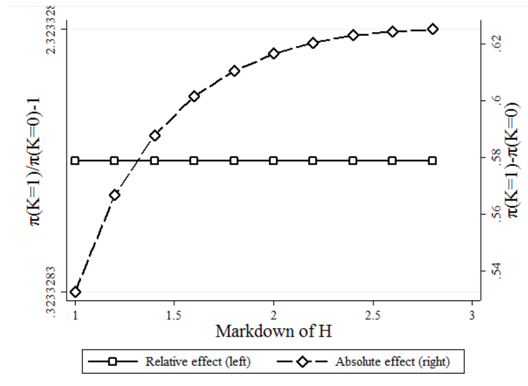
I carry out a calibration exercise to illustrate the results above. I let the degree of scale returns be $\nu = 0.9$, draw the output elasticity $\beta^h$ from a uniform distribution on the interval [0,$\nu$], with 1000 draws. I let the market-level inverse input supply elasticities be $\psi_m^h = \psi_n^h = 3$, which means that a monopsonist pays each of its input suppliers a third of their marginal product. I consider the relative and absolute variable profit effect of a technology at ten different markdowns on the interval $\mu^h \in [1, \psi^h]$. I set the inverse demand elasticity to $\eta = -\frac{1}{3}$. I consider four technologies. In Figure 2a, the technology reduces the output elasticity of $H$ by half, and is hence $L$-biased, but it does not change Hicks-neutral productivity $\Omega_f$. A second technology, in Figure 2b, doubles Hicks-neutral productivity, but does not change the output elasticity $\beta$: $\Omega(K=1) = 2\Omega(K=0)$. A third technology, in Figure 2c, both halves $\beta$ and doubles $\Omega$. A final technology, in Figure 2d, has the same effect on $\beta$ as technology (c), but triples Hicks-neutral productivity: $\Omega(K=1) = 3\Omega(K=0)$.

**Figure 2: Returns to an L-biased technology: calibration**

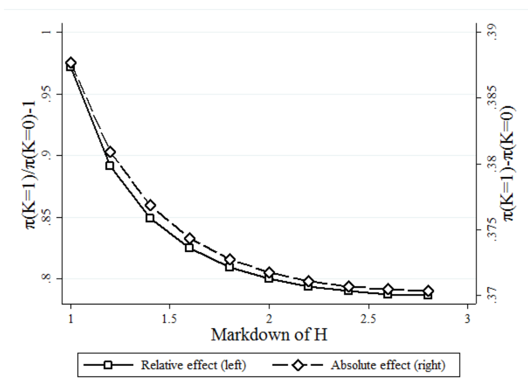
**(a) Only $\beta$ changes**



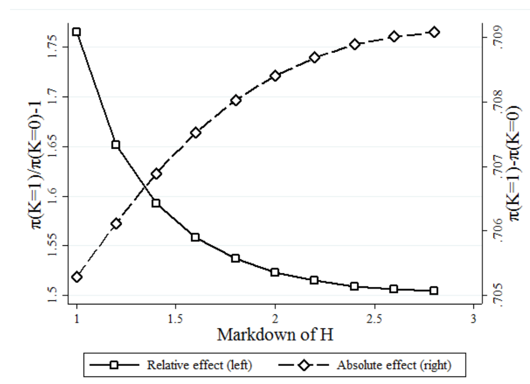

**(b) Only $\Omega$ changes**




**(c) Both $\beta$ and $\Omega$ change**




**(d) Both $\beta$ and $\Omega$ change**




**Notes:** Each panel plots the relative and absolute returns to mechanization, $\frac{\Pi(K=1)-\Pi(K=0)}{\Pi(K=0)}$ and $\Pi(K=1) - \Pi(K=0)$, for an $L$-biased technology $K$ against the markdown of input $H$ by step of 5 percentiles. Panel (a) considers a technology that only changes the output elasticity $\beta$ but not Hicks-neutral productivity $\Omega$. Panel (b) lets $\Omega$ change but not $\beta$. Panel (c) lets both $\Omega$ and $\beta$ change, with $\Omega(K=1) = 2\Omega(K=0)$. Panel (d) increases the Hicks-neutral effect to $\Omega(K=1) = 3\Omega(K=0)$.

For technology (a), moving from a competitive to a monopsonistic markdown lowers the relative return to technology adoption from 11% to 0%, in line with Theorem 1. The absolute return drops from 0.04 units to 0 units, but as explained by theorem 2, this effect could go in the other direction depending on the parametrization of the model. For technology (b), which only changes Hicks-neutral productivity, the relative returns to technology adoption do not change with the markdown, in line with Theorem 1, and the absolute returns to adoption increase with the markdown. Lemma 3 says that is is true more in general. Technologies (c) and (d) both decrease the output elasticity of $H$, but also increase Hicks-neutral productivity, by respectively $100\%$ and $200\%$. For technology (c), the absolute returns to technology adoption still fall with the markdown, but for technology (d), higher markdowns lead to higher returns to innovation. This is a result of Lemma 2: for technology (c), the Hicks-neutral productivity effect is too small to dominate the factor-biased effect of the technology, and markdowns decrease returns to innovation, whereas the Hicks-neutral effect dominates for technology (d), making markdowns increase the absolute return to innovation.

# 3 Coal mining in Illinois (1884-1902)

There are three reasons to complement the theoretical results in the previous section with empirical analysis. First, following Theorems 1 and 2, the effect of markdowns on the returns to innovation depend on the directed and Hicks-neutral productivity effects of the technology, and on which input market the firm has monopsony power. Knowing these primitives requires some empirical analysis. Secondly, in case the technology is biased towards the input over which the firm has monopsony power, we know from Theorem 2 that markdowns increase the incentives to adopt that technology. However, in order to know not just the direction but also the size of this effect, estimating a model of input supply and demand is needed. Finally, if the technology is biased away from the input over which the firm has monopsony power, we do not even know the sign of the effect of markdowns on technology adoption incentives. In that case, empirical analysis is needed to both know the direction and size of this effect.

As an empirical application, I study the adoption of coal cutting machines in the Illinois coal mining industry between 1884 and 1902. This is an interesting setting because it features isolated labor markets in the form of mining towns, which are likely to feature some oligopsony power, with a large factor-biased technological innovation due to the invention of coal cutting machines. Before jumping into the empirical model, I discuss the most important industry characteristics and the data sources.

## 3.1 Industry background

### A Extraction process

The coal extraction process consisted of three consecutive steps. First, the coal seam had to be accessed, which usually required either a vertical 'shaft', a diagonal 'slope' or a horizontal 'drift', depending on the geography of the mine. As large parts of Illinois are flat, $60\%$ of the mines were 'shaft' mines.[15] Second, upon reaching the seam, the coal wall was 'undercut', traditionally manually using picks, but from 1882 onward also with coal cutting machines. The mechanization of the cutting process is considered to be the most significant technological change during this time period

[15]Less than $2\%$ of the mines were surface mines that did not require any digging.

(Fishback, 1992).[16] Third, coal had to be transported back to the surface and sorted from impurities. The hauling was done using mules or underground locomotives. Over 90% of output was hauled using locomotives. Mines used two types of intermediate inputs. First, black powder was used to blast the coal wall. This was purchased by the miners, not by the firm. Secondly, coal itself was used to power steam engines, electricity generators, and air compressors.[17]

## B Technological change: the coal cutting machine

The first mechanical coal cutter in the U.S.A. was invented by J.W. Harrisson in 1877, but it was merely a prototype.[18] The Harrisson patent was acquired and adapted by Chicago industrialist George Whitcomb, whose 'Improved Harrison Cutting Machine' was released in 1882, of which the patent is pictured in Figure A3a. An illustration of how the coal cutting machine was used is in Figure A3b. Ninety percent of the cutting machines in the dataset are of this type. The spatial diffusion of cutting machines is shown in Figure A1. As shown in Figure 3, the share of mines using a coal cutting machine increased from below 2% to 9% between 1884 and 1902. Mechanized mines were larger: their share of output increased from 7 to 30% over this same time period. The mechanization of the hauling process, which replaced mules by underground locomotives, was another source of technical change. This was largely accomplished in Illinois: the share of output mined in locomotive mines was above 90%.

## C Occupations

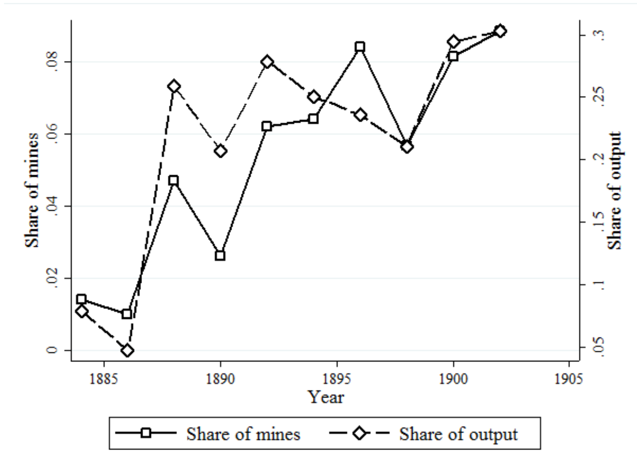
Coal mining involved a wide variety of different tasks. The inspector report from 1890 reports wages at the occupation-level, and this subdivision is reported in Appendix Table A1 for the 20 occupations with the highest employment shares, together covering 97% of employment. Three out of five workers were miners, who did the actual coal cutting. This required a significant amount of skill: in order to determine the thickness of the pillars, miners had to trade off lower output with the risk of collapse. The other 40% of workers did a variety of tasks such as clearing the mine of debris ('laborers'), hauling coal to the surface using locomotives or mules ('drivers' and 'mule tenders'), loading coal onto the mine carts ('loaders'), opening doors and elevators ('trappers'), etc. The skills required to carry out these tasks were usually less complex than those of the miners, and were moreover not specific to coal mining: tending mules or loading carts are general-purpose tasks, in contrast to

[16] Two techniques existed to cut the coal: nine out of ten mines used a 'rooms and pillars' technique in which miners excavated everything except pillars, which were left to sustain the roof. The other mines used so-called 'longwall' techniques in which miners temporarily constructed an artificial roof and allowed the room to collapse in a controlled way.

[17] A fraction of the mine's coal output was re-used as an energy input. I only observe reused coal inputs in 1902, and the fraction of output that was re-used as an input was on average 5%, and 0% for the median mine. As I do not observe this variable in all years, I do not take it into account in the model.

[18] Simultaneously, prototypes of mechanical coal cutting machines were invented in Northern England in the late 1870s (Reid, 1876; Ackermann, 1902).

**Figure 3: Cutting machine adoption**




**Notes:** This graph plots the share of Illinois coal mines using at least one cutting machine (solid line, left axis) and the share of output produced by mines with at least one cutting machine (dashed line, right axis) over time.

undercutting coal walls.[19]

Difference in industry-specific skills are reflected in daily wages: miners earned an average daily wage of $2.3, which was higher than any other employees except for 'pit bosses' (middle managers), and 'roadmen', who maintained and repaired mine tracks, but these two categories of workers represent barely 2% of the workforce. The higher wages of miners cannot be explained as a risk premium, because nearly all other occupations worked below the surface as well, and were hence subject to the same risks of mine collapse or flooding.

The biennial mine-level data set I will rely on for the model classifies workers into two types: miners and all other employees. Henceforth, I will call miners 'skilled labor', and other workers 'unskilled labor'.

## D Labor markets

Skilled workers received a piece rate per ton of coal mined, whereas unskilled workers were paid a daily wage.[20] Converting the piece rates to daily wages, the net salary of skilled labor was on average 23% higher compared to unskilled labor. 'Net salary' means net of material costs and other work-related expenses. At some of the mines, 'wage screens' were used, which means that skilled workers were paid only based on their output of large coal pieces, rather than on their total output. This introduces some measurement error in labor costs. However, the data set reports the usage of wage screens in 1898, and shows that they were used in merely 4 out of 52 counties, at mines that

[19] Some unskilled workers eventually became skilled, such as boys who started out as trappers but became miners at an older age. I abstract from such dynamic considerations on the labor supply side in the model.

[20] Piece rates were an incentive scheme in a setting with moral hazard, as permanent miner supervision would be very costly.

jointly represented merely 2.3% of employment.[21]

Rural Illinois was, and still is, sparsely populated: the median and average population sizes of the towns in the dataset was 1067 and 3090 inhabitants, and on average a third of the population were coal miners. Considering that women and children under the age of 12 did not work in the mines, almost the entire working population was employed in coal mining in most villages. Of all the villages, 50% had just one coal mine, and another 30% had two or three. Two thirds of all employees worked in a village with three or less coal mines. Although most of the villages in the data set were connected by railroad, these were exclusively used for freight: passenger lines only operated between major cities (Fishback, 1992). Given that the average village was 7.4 miles apart from the closest other village, and that skilled workers had to bring their own supplies to the mine, commuting between villages was not an option, and the mining towns can be considered as isolated local labor markets. Most roads were unpaved and automobiles not yet introduced. In order to switch employers, miners had to migrate to another town.[22]

First attempts to unionize the Illinois coal miners started around 1860, without much success (Boal, 2017). Unionism was countered by employers in various ways, for instance by including non-membership of a labor union as a requirement in labor contracts. These so-called 'yellow-dog' contracts were criminalized in Illinois in 1893, with fines of $100 USD, which was equivalent to on average six monthly miner wages. (Fishback, Holmes, & Allen, 2009). In 1886, 15% of mine workers in Illinois were member of a trade union. The first succesful labor union in Illinois was the *United Mine Workers of America*, founded in 1890. A major strike in 1897-1898 had important consequences: wages were raised and working hours reduced to a maximum of eight hours per day. Even more importantly, wages were determined during annual wage negotiations between the unions and a state-wide representation of employers after 1898, which took place in January (Bloch, 1922).Wage-setting was therefore done by each mine independently until 1898, and through collective bargaining afterwards. There was no minimum wage law. In contrast to other states, the mines in the data set did not pay for company housing of the miners (Lord, 1883, 75), which would otherwise be a labor cost in addition to miner wages.

## E Coal markets

Coal was sold at the mine gate, and there was no vertical integration with post-sales coal treatment, such as coking. On average 93% of the mines' coal output was either sold to railroad firms or transported by train to final markets. The remaining 7% was sold to local consumers. The main coal destination markets for Illinois mines were St. Louis and, to a lesser extent, Chicago, which was supplied with cheaper coal from fields in Ohio, Pennsylvania, and West Virginia using lake steamers (Graebner, 1974). Railway firms were also major coal consumers. Historical evidence points to in-

[21]Nevertheless, I test the robustness of the results for the non-inclusion of counties in which wage screens were used in Appendix C.4.

[22]Some more evidence supporting the isolated mining towns assumption is in Appendix C.3.

tense competition on coal markets during the last two decades of the 19th century, before the large consolidation wave in the early 1900s (Graebner, 1974). Nevertheless, there was still a considerable transportation cost of coal, which makes that coal markets were likely not entirely integrated. There are large differences in the coal price across Illinois: in 1886, for instance, it varied between 90 cents/short ton at the 10th percentile of the price distribution to 2 dollars/short ton at the 90th percentile, and this price dispersion slightly increased over time.

## 3.2 Data

I observe every bituminous coal mine in Illinois between 1884 and 1902 at two-year intervals, which results in 8356 observations. The data are obtained from the *Biennial Report of the Inspector of Mines of Illinois*. The dataset covers all mines, of which the yearly number fluctuated between 683 and 919. I observe the name of the mine, the mine owner, yearly coal extraction, average employee counts for both skilled and unskilled workers, days worked, and a dummy for cutting machine usage in every two-year period. Materials are measured as the total number of powder kegs used in a given year. Other technical characteristics are observed for a subset of years, such as dummies for the usage of various other technologies (locomotives, ventilators, longwall machines), and technical characteristics such as mine depth and the mine entrance type (shaft, drift, slope, surface). Not all of these variables are used in the analysis, given that some of these are observed in a small subset of years.

I observe the average piece rate for skilled labor throughout the year and the daily wage for unskilled labor from 1888-1896. Skilled wages and employment are separately reported for the summer and winter months between 1884 and 1894. For some years I observe additional variables such as mine capacities, the value of the total capital stock and a break-up of coal sales by destination. Wages and employee skill types are not observed in 1896. I deflate all monetary variables using historical CPI estimates from Hoover (1960). The reported monetary values are all in 1884 U.S. dollars.

In addition to the main biennial dataset, I use different other datasets. First, the inspection report from 1890 contains monthly data on wages and employment for both types of workers, and of production quantities are given for a sample of 11 mines that covers 15% of skilled and 9% of unskilled workers. Second, monthly free-of-board bituminous coal prices in the harbor of New York are collected for the years 1890-1900 from the NBER Macrohistory Database (National Bureau of Economic Research, n.d.). Third, town- and county-level information from the 1880 and 1900 population census and the censuses of agriculture and manufacturing are collected as well. Fourth, I collect information on coal cutting machine costs from Brown (1889). I refer to appendix A for more details regarding the data sources and cleaning procedures.

## 3.3 Key facts

**Fact 1** *Output and labor productivity increased, but skilled wages stagnated until 1898.*

The Illinois coal mining industry grew rapidly during the last two decades of the 19th century. Annual output, in Figure 4a, tripled from 10 to 30 megaton between 1884 and 1902. This was both due to an increase in the average mine size and to an increase in the number of mines from 700 to 900. Daily output per worker, in Figure 4b, increased from 2 to 3.3 tons for hand mines, and from 2.3 to 4.1 tons for machine mines.[23] Until 1898, this growth in output and productivity did not translate into higher wages: the daily wage of skilled labor remained around $ 1.8 until 1898, as can be seen in Figure 4c. After the large strikes in 1897-1898 and the introduction of centralized wage bargaining, wages rose. Coal prices per ton fell from $1.2 to $0.9 between 1884-1898, after which they increased again.

**Fact 2** *Mechanized mines used less skilled per unskilled worker.*

As was shown in Figure 4b, output per worker was higher in machine mines. The composition of labor was also different: in Figure 4d, I plot the ratio of the total number of skilled labor-days over the number of unskilled worker-days in per year. Mines without cutting machines used between 3 and 4 skilled labor-days per unskilled labor-days throughout the sample period, compared to 2 to 3 skilled labor-day per unskilled worker-day for machine mines. In every year, except 1894, machine mines used less skilled per unskilled worker. The skilled-unskilled labor ratio was on average 16.5% lower for machine mines compared to hand mines, and between 11% and 22% lower at a certainty of 90%. However, this difference is not necessarily a causal effect of cutting machines on skill-augmenting productivity: mines with higher productivity levels were probably more likely to adopt cutting machines. For the causal effects of cutting machines on total factor and factor-augmenting productivity levels, I refer to the empirical model in the next section. Anecdotal evidence suggests that cutting machines led to the substitution of skilled for unskilled workers. In his 1888 report, the Illinois Coal Mines Inspector asserts:

> "Herein lies the chief value of the [cutting] machine to the mine owner. It relieves him for the most part of skilled labor [...] it opens to him the whole labor market from which to recruit his forces." (Lord, 1888, 340)

Along the same lines, the State Inpector of Mines of Illinois wrote:

> "The mining machine is in fact the natural enemy of the coal miner; it destroys the value of his skill and experience, and reduces him to the rank of a common laborer." (Lord, 1888, 339)

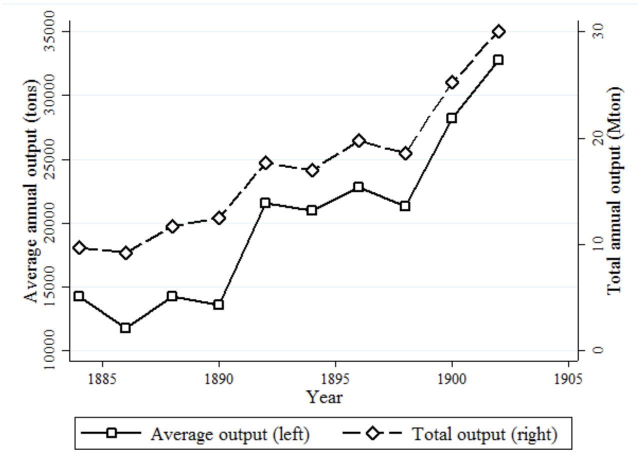
**Fact 3** *Skilled labor wages varied seasonally, unskilled labor wages did not.*

Coal demand was seasonal: during the cold winter months, energy demand increased compared to the warm summer months. As can be seen in Figure 5a, which plots average monthly skilled

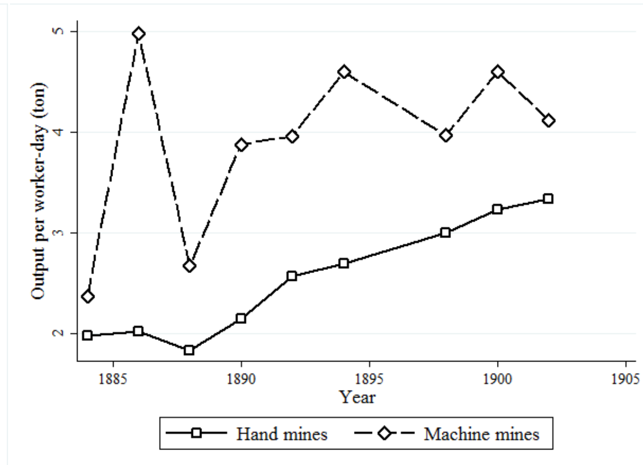
[23] This series is adjusted for the reduction of hours per working day in 1898, as explained in Appendix A.

**Figure 4: Aggregate quantities and prices**

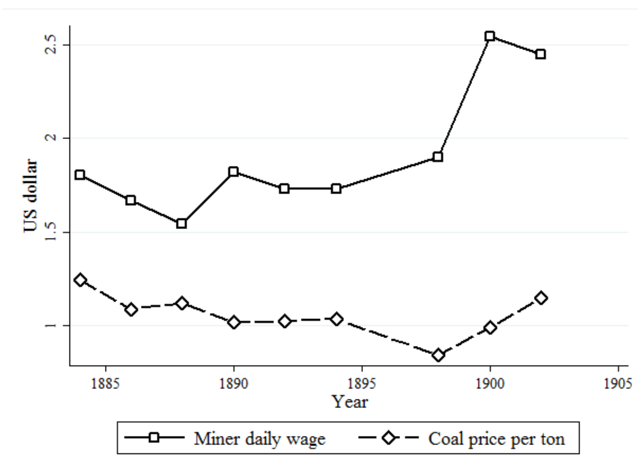
**(a) Output**



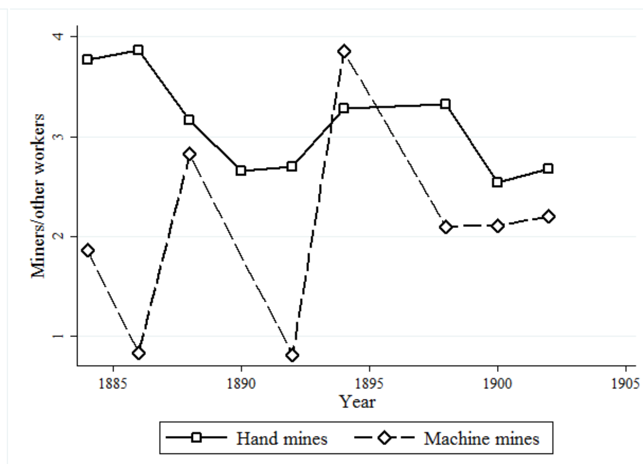

**(b) Output per worker**




**(c) Wages and prices**




**(d) Skilled/unskilled labor ratio**




**Notes:** Panel (a) plots average output per mine-year and total mine output in Illinois over time. Panel (b) plots the ratio of total output over total days worked at mines that used cutting machines ('machine mines') and mines that did not ('hand mines'). Panel (c) reports the aggregate skilled labor daily wage, defined as the total wage bill spend on skilled labor over the total number of skilled labor-days, and the aggregate price, defined as total revenue over total output, in the Illinois coal mining industry. The reduction in working hours in 1898 is taken into account. The coal price per ton is the mine-gate price. Panel (d) plots the ratio of total skilled worker-days over total unskilled worker-days at hand and machine mines. 1890 is omitted for machine mines in 1890 due to employment being unobserved for most machine mines in that year.

labor-days in 1890, skilled employment follows the coal demand cycle.[24] From August to February, employment is high, as coal for the cold winter months is extracted. Given that transporting coal to the final market took some time, coal demand already increased around August. Panel 5b shows that skilled wages followed this coal demand cycle: during summer, skilled wages fell compared to the

[24]This monthly data is based on a sample of mines selected by the Illinois Bureau of Labor Statistics across 5 counties in 1890, which covers 16% of skilled employment and 9% of unskilled employment.

other months. There is a lag between the wage and employment cycle of around a month, which might be due to the fact that wages are paid with a lag. In contrast unskilled worker wages did not co-vary with product and labor demand throughout the year.[25] Panel 5c also shows this by plotting monthly wages for both skilled and unskilled workers against the monthly number of worker-days of each type at the mine-month level throughout 1890. Skilled wages were positively correlated with monthly skilled employment, whereas the unskilled worker wage-employment schedule is flat. Moreover, there was a large variation in skilled wages across mines and months, but very little variation in unskilled wages

Skilled workers were paid piece rates, whereas unskilled workers were paid daily rates. If skilled workers were more productive during months of peak demand, this could be the reason that skilled wages co-vary with employment, rather than other explanations, such as monopsony power. However, as can be seen in Figure 5d, this is not the case. During the winter months, output per skilled worker-day was on average 2.53 ton, whereas it was 2.61 ton during summer. There is also a significant difference in skilled labor piece rates (wage per ton mined) between summer and winter: the wage per ton earned by a skiller worker was on average \$0.780 during summer and \$0.817 during winter, so it was 4.7% higher during the winter. This difference is significantly large than zero: the difference between summer and winter piece rates lies between 0.026 and 0.049 with a probability of 90%.

# 4 Empirical model

## 4.1 Model

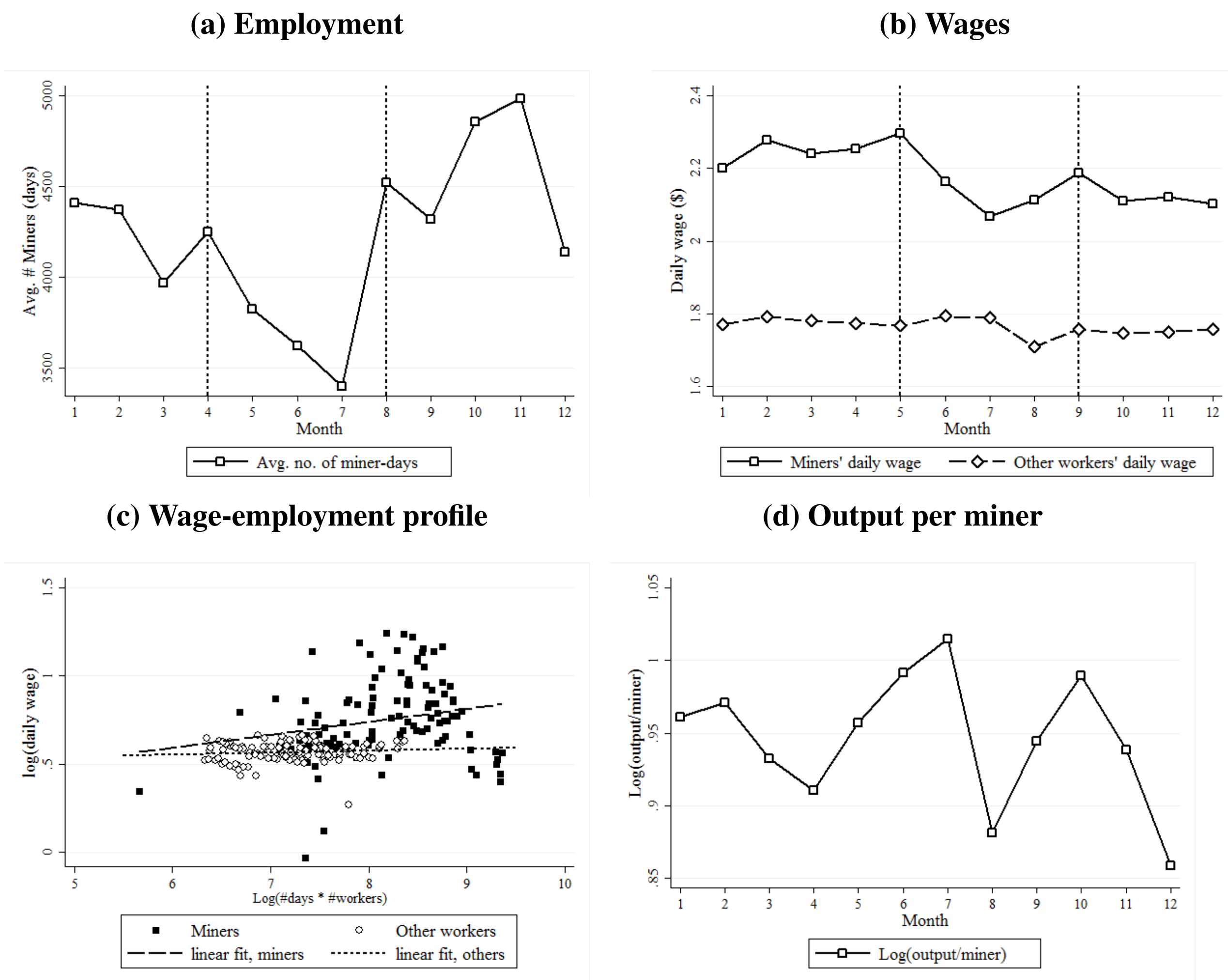
In this section, I model labor demand and supply in the coal mining industry by implementing an empirical version of the model in Section 2 with a concrete model of competition on both input and product markets. This serves as an input to the counterfactual exercise of understanding the effects of monopsony power on innovation in section 4.4.

### A Coal extraction

Let $f$ index firms and $t$ bi-yearly intervals. Biennial coal extraction is $Q_{ft}$ tons, the amount of skilled labor (in days worked) is $H_{ft}$, and unskilled labor-days is $U_{ft}$. Cutting machine usage is denoted $K_{ft} \in \{0, 1\}$. The model is written at the level of the firms $f$, which are observed in the data. The production function in logs is given by Equation (4a), denoting logarithms of variables in lowercases. I use a Cobb-Douglas production function in both labor types, but allow for the output elasticity of skilled labor $\beta_{ft}$ to vary flexibly across mines and years. The scale parameter $\nu$ is equal to the sum of the output elasticities of skilled and unskilled workers, and is assumed to be a constant. The

[25] As for coal prices, FOB coal prices in the Harbor of New York did not seem to fluctuate between months, but mine-gate coal prices in Illinois might have fluctuated. I have no data on monthly coal prices in Illinois.

**Figure 5: Seasonality in employment and wages**




**Notes:** Panel (a) plots average monthly skilled employment per month for a sample of 11 mines in 1890. Panel (b) does the same for average daily wages of skilled and unskilled employees. Panel (c) plots mine-level monthly employment against daily wages for both worker types across mines in 1890. Panel (d) plots total output per skilled worker per month during 1890.

Hicks-neutral productivity residual in logs is denoted $\omega_{ft}$.

$$q_{ft} = \beta_{ft} h_{ft} + (\nu - \beta_{ft}) l_{ft} + \omega_{ft} \tag{4a}$$

Besides labor, mines also use cutting machines, the usage of which is indicated by a dummy variable $K_{ft} \in \{0, 1\}$. Both the output elasticity of skilled workers $\beta_{ft}$ and the productivity residual $\omega_{ft}$ are assumed to be AR(1) processes, Equations (4b) and (4c), with serial correlations $\sigma^{\beta}$ and $\sigma^{\omega}$. This specification does not allow for some forms of cost dynamics in which current productivity is a function of the total amount of output produced in the past.[26] Both the output elasticity and Hicks-

[26] I refer to Appendix C.2 for a motivation and discussion of this assumption.

neutral productivity level are assumed to be linear functions of current machine usage $K_{ft}$ and a vector of other control variables $\mathbf{X}_{ft}$. I include a linear time trend, a constant, and the quantity of black powder used to this controls vector: both Hicks-neutral productivity and the output elasticity of workers could differ depending on how much black powder was used to blast the coal veins. The effects of using cutting machines on the output elasticity of skilled labor is parametrized by the coefficient $\alpha^{\beta}$, their effect on Hicks-neutral productivity is $\alpha^{\omega}$. The residual shocks to the skilled labor output elasticity and Hicks-neutral productivity are denoted $\gamma_{ft}^{\beta}$ and $\gamma_{ft}^{\omega}$. By using these parametric specifications, I assume that there is no heterogeneity across mines or time in the Hicks-neutral and factor-biased effects of cutting machines.

$$\beta_{ft} = \alpha^{\beta} K_{ft} + \sigma^{\beta} \mathbf{X}_{ft} + \rho^{\beta} \beta_{ft-1} + \gamma_{ft}^{\beta} \tag{4b}$$

$$\omega_{ft} = \alpha^{\omega} K_{ft} + \sigma^{\omega} \mathbf{X}_{ft} + \rho^{\omega} \omega_{ft-1} + \gamma_{ft}^{\omega} \tag{4c}$$

Although the assumption of unitary substitution between both inputs in the Cobb-Douglas model is strong, and goes against the canonical models in the labor literature, the model does better than usual Cobb-Douglas formulations in empirical work by allowing for unobserved heterogeneity in output elasticities of inputs across mines and time, and by allowing these output elasticities to be conditional on technology usage, which is crucial when studying non-Hicks neutral technological change.

I assume mines do not face a binding capacity constraint. This is consistent with the data: in 1898, the only year for which capacities are observed, merely 1.4% of the mines operated at full capacity, and they were responsible for 1.1% of industry sales. The entire distribution of capacity utilization rates is shown in Figure A2.

## B Coal demand

In contrast to the general model in Section 2, most coal firms were not monopolists on the coal market. Coal is assumed to be a homogeneous product.[27] Each firm operates on a single coal market, indexed by $m$ with a market share $s_{ft}^{q} \equiv \frac{Q_{ft}}{Q_{mt}}$ and market-level output $Q_{mt} \equiv \sum_{f \in m} Q_{ft}$. Coal markets will be defined in Section 4.2. The market-level coal demand curve is given by Equation (5), with a market-level mine-gate coal price $P_{mt}$, an inverse demand elasticity $\eta$, and a residual $\zeta_{mt}$ which reflects differences in coal prices across markets due to variation in local demand conditions, transport costs, etc. In the baseline model, I assume that all markets face the same coal demand elasticity.

$$P_{mt} = (Q_{mt})^{\eta} \exp(\zeta_{mt}) \tag{5}$$

[27]There is some differentiation between coal types in terms of heat rates and ash content, but all mines in the data set produce bituminous coal, which is assumed to be a homogeneous product.

## C Input supply

Each firm operates on exactly one labor market $n$ with skilled and unskilled labor market shares $s^h_{ft}$ and $s^l_{ft}$. More information on how labor markets are defined is in Section 4.2. Skilled labor in a market $n$ earns a daily wage $W^h_{nt}$, unskilled labor earns a daily wage $W^l_{nt}$. I convert the piece rates paid to skilled workers into daily wages in order to be comparable to the unskilled worker wages. Firms are assumed not to wage-discriminate in terms of skilled labor piece rates. Mine-employee-level wage data from the 1890 report show indeed that there was very little heterogeneity in both piece rates and daily wages across miners within firms at a certain point in time. A firm $f$ has an input market share $s^h_{ft} \equiv \frac{H_{ft}}{H_{mt}}$ on the market for $H$ and input market share $s^l_{ft} \equiv \frac{L_{ft}}{L_{mt}}$ on the market for $L$, with market-level employment $H_{mt} \equiv \sum_{f \in m} H_{ft}$ and $L_{mt} \equiv \sum_{f \in m} L_{ft}$. The market-level supply curve for both types of workers is given by equation (6). The inverse wage elasticity of skilled labor is $\psi^h_{nt} = \frac{\partial W^v_{nt}}{\partial H_{nt}} \frac{H_{nt}}{W^h_{nt}} + 1$ and for unskilled workers as $\psi^l_{nt} = \frac{\partial W^l_{nt}}{\partial L_{nt}} \frac{L_{nt}}{W^l_{nt}} + 1$. The error terms $\xi^h_{nt}$, $\xi^l_{nt}$ explains variation in wages across markets that cannot be explained by market size, which includes the outside options available to the workers in each market.

$$\begin{cases} W^h_{nt} &= {H_{nt}}^{\psi^h_{nt}-1} \exp(\xi^h_{nt}) \\ W^l_{nt} &= {L_{nt}}^{\psi^l_{nt}-1} \exp(\xi^l_{nt}) \end{cases} \tag{6}$$

Figure 5d revealed that unskilled worker wages were much less dispersed compared to skilled wages, and did not change in response to seasonal labor demand shocks. Therefore, I assume that unskilled labor supply is perfectly elastic, meaning that $\psi^l_{nt} = 1 \forall f$. There are, of course, other possible explanations for the fact that wages did not react to labor demand shocks, such as behavioral reasons, as shown in (Kaur, 2019). The key thing to note here is, however, that monthly wage profiles were only flat for unskilled labor, not for skilled labor. Although wage contracts differed between skilled and unskilled labor because skilled labor received a piece rate rather than a daily wage, both of these contracts were limited to monthly durations or less; it is hence not the case that unskilled wages did not respond to seasonal demand shocks because they were pre-negotiated for the entire year. In contrast, I allow for the elasticity of skilled labor supply, $\psi^h_{nt}$, to be above one. Although the log-linearity of Equation (6) imposes a strong functional form assumption, I allow the slope $\psi^h_{nt}$ to vary flexibly across markets and time, as local labor market conditions vary. I assume that cutting machines $K$ are sold on competitive markets, and that their prices are exogenous to each individual mine.

I assume that employers are homogeneous 'products' from the point of view of the workers: when choosing which firm to work for within a town, miners only care about the wage rate, not about other firm characteristics. The motivation for this assumption is that there is very little dispersion in wages within towns in a given year: town and year dummies explain 86% of the variation in skilled wages.

I do not formally model how employees gather their skills, and whether employees can move from being unskilled to skilled worker types. I do assume that firms cannot invest to turn unskilled workers into skilled workers - this would imply a dynamic input demand problem that does not fit the static input demand conditions that are outlined below.

## D Firm behavior

Using the terminology of Ackerberg et al. (2015), I assume that skilled and unskilled workers are both variable and static inputs. They are *variable* because they can be flexibly adjusted: as shown earlier, employment was adjusted throughout the year on a monthly basis, and wages were determined in short-term contracts until 1898.[28] Both labor types are also *static* because current labor choices do not affect future profits, i.e. there are no hiring or firing costs. Cutting machines are, in contrast, a *fixed* input. Firms need to make their cutting machine adoption decision one period in advance. Let the capital accumulation equation be given by the following equation, with machine acquisitions being denoted as $A_{ft-1} \in \{0,1\}$. Depreciation $\delta \in \{0,1\}$ takes the value of either zero or one. If there is no depreciation, meaning that $\delta = 1$, mines can only acquire a cutting machine if they do not already own one, and such an acquisition is permanent. If $\delta = 0$, machines fully depreciate within two years, and firms re-make the capital adoption decision every time period.

$$K_{ft} = \delta K_{ft-1} + A_{ft-1}(1 - \delta K_{ft-1}) \tag{7}$$

Cutting machines have both a common fixed cost component $\Phi$, which is the capital cost of acquiring the machine, and a common variable cost component $W^k$, due to the usage of electricity. Mine-level variable profits are denoted $\Pi_{ft} \equiv P_{ft}Q_{ft} - W^h_{mt}H_{ft} - W^l_{mt}L_{ft} - W^k K_{ft}$. Other intermediate input expenditure is not part of the mine's profit function, as these inputs were purchased and brought by the miners. I assume zero sunk costs of cutting machine usage, as the adoption model will be static. Hence, total mine profits are defined as $\Pi^{tot}_{ft} \equiv \Pi_{ft} - \Phi K_{ft}$.

I assume that firms make their input decisions in two phases. At time $t-1$, before the productivity residuals $\omega_{ft}$ and $\omega^h_{ft}$ are observed, firms simultaneously choose their cutting machine usage for the next period, $K_{ft}$. At time $t$, after the productivity residuals $\omega_{ft}$ and $\omega^h_{ft}$ are observed, firms simultaneously choose their optimal amounts of both labor types conditional on their capital technology, which was chosen earlier.

The second stage of the decision problem, the labor demand problem, is given by Equation (8). Taking capital usage as given, each mine $f$ independently chooses the amount of skilled and unskilled labor that maximizes its current variable profits. By choosing the amount used of both labor types, firms also choose their output $Q_{ft}$.

$$\max_{H_{ft}, L_{ft}} (\Pi_{ft}) \tag{8}$$

In the first stage, firms choose their capital investment $A_{ft} \in \{0,1\}$ that maximizes discounted total profits, with a common discount factor $\delta$. In the application, I will assume full depreciation of capital

[28] As explained further below, I will only consider the period 1884-1894 when estimating the structural model.

within two years, meaning that $\delta = 0$, in order to keep the cutting machine adoption problem static. This means that firms re-choose their capital stock in every two-year period, and do so by maximizing the profits in the next period.

$$\max_{A_{ft}} \Pi^{tot}_{f,t+1} + \mathbb{E}_t \sum_{r=2}^{\infty} (\delta^{r-1} \Pi^{tot}_{f,t+r}) \tag{9}$$

### E Equilibrium

By solving the first order conditions for the profit maximization problem in (8), the equilibrium expressions for all endogeneous static variables $(Q, P, H, L, W^h, W^l)$ can be solved for. These equilibrium expressions can be found in Appendix B.5. The skilled labor wage markdown charged by the firm is equal to $1 + s^h_{ft}(\psi^h_{nt} - 1)$:

$$\frac{\frac{\partial(P_{mt}Q_{ft})}{H_{ft}}}{W^h_{mt}} = 1 + (\psi^h_{mt} - 1)s^h_{ft}$$

The markdown parameter $\mu_f$ from the theoretical model hence corresponds to the markdown $1 + (\psi^h_{mt} - 1)s^h_{ft}$ in the empirical model. If the labor market share of the firm is equal to one, the actual markdown is equal to the monopsonistic markdown $\psi^h_{nt}$. If the firm is atomistically small, the markdown goes toward 1 in the limit, meaning that skilled laborers earn their marginal product of labor.

## 4.2 Identification and estimation

I now turn to the identification and estimation of the model. Five latent variables need to be identified: the entire distribution of output elasticities of skilled labor $\beta^h_{ft}$, in Equation (4a), the market-level inverse elasticity of skilled labor supply $\psi^h_{mt}$, in Equation (6), the inverse elasticity of coal demand $\eta$, in Equation (5), and the effect of cutting machines on the output elasticity of skilled labor and on Hicks-neutral productivity, in Equations (4b-4c). Although the model is specified at the firm-bi-year level, the dataset comes at the mine-bi-year level. I aggregate all the relevant variables from the mine- to the firm-level.[29] I restrict the panel to the time period 1884-1894 when estimating the model and conducting the counterfactual exercises, because wage and price data are missing in 1896, and because annual collective bargaining over wages between unions and coal firms was instituted in 1898, which does not fit the unilateral oligopsony framework of the model.

[29] Details on how I aggregate to the firm-level are in Appendix A.2.

## A Labor supply

**Identification** I start with the identification of the skilled labor supply function. Taking the logarithm of Equation (6) for skilled labor, and denoting logs as lowercases, gives equation (10).

$$w_{nt}^{h} = (\psi_{nt}^{h} - 1)h_{nt} + \xi_{nt}^{h} \tag{10}$$

The supply elasticity $\psi_{nt}^{h}$ cannot be recovered by simply regressing skilled labor wages on employment because of the latent outside options $\xi_{nt}^{h}$. Mines in labor markets with an unattractive outside option $\xi_{nt}^{h}$ can offer a lower wage to attract the same number of skilled laborers. In order to identify the slope of the skilled labor supply curve, a shock to labor demand that is excluded from skilled labor utility is necessary. I rely on the seasonal character of coal demand as a source of labor demand variation. As explained in section 3.1, coal demand rises during the fall and winter due to low temperatures. Denote skilled employment in town $n$ during winter and summer months as $H_{nt}^{WIN}$ and $H_{nt}^{SUM}$, and the corresponding daily skilled wages as $W_{nt}^{h,WIN}$ and $W_{nt}^{h,SUM}$. The supply residuals during winter and summer are $\xi_{nt}^{h,WIN}$ and $\xi_{nt}^{h,SUM}$. I assume that the outside option $\xi^{h}$ is the same during winters and summers: $\xi_{nt}^{h,WIN} = \xi_{nt}^{h,SUM}$. Under these assumptions, the slope of the skilled labor supply curve can then be calculated using equation (11):

$$\psi_{nt}^{h} = \frac{w_{nt}^{h,WIN} - w_{nt}^{h,SUM}}{h_{nt}^{WIN} - h_{nt}^{SUM}} + 1 \tag{11}$$

The main argument in favor of these two assumptions is that the monthly wage profile of unskilled workers did not fluctuate between the different seasons, as shown in Figure 5. It could be that outside options varied seasonally, for instance due to increased agricultural labor demand during the harvest season. This would, however, be consistent with higher wages during the summer, while lower summer wages are observed. Also, we would expect unskilled labor wages to fluctuate seasonally as well, which they did to a much lesser extent than skilled labor wages. Working conditions, such as mine safety, could vary seasonally, but were less easily adjustable than wages. Anecdotal sources mention that most skilled workers were (partially) unemployed during the summer months in Northern Illinois coalfields (Joyce, 2009), which is consistent with increased monopsony power over these workers during the summer.

**Labor market definition** Workers did not own cars yet, and railroads were only used for freight cargo except between large cities. Miners could hence only work in their own mining town or commute by foot to another town. Of the 448 towns reported in the data set, 75% were located more than 3 miles in a straight line from their closest mining town (town with at least one mine), and the average town was 5.6 miles away from the closest mining town. Given that miners had to bring their own equipment to the mines and that until 1898, they often worked 10 hours per day, it seems safe to assume that any town further than 3 miles apart is not a viable commuting option, as it would imply

2h30 of daily commuting time by foot.[30] In order to ensure isolated labor markets, I merge the towns that are closer than 3 miles from each other.[31] This results in 350 labor markets that lied on average 6.4 miles from the closest other town.

**Estimation** I calculate the slope of the skilled labor supply curve for each town using equation (11). Skilled wages are reported separately for winters and summers between 1884-1894.[32] The reported wage rates are piece rates, in wages per ton. Equation (11) was, however, written using daily wages per worker and days of employment, because workers care only about their daily wage, not their wage per ton of coal mined. I transform the piece rates that are observed in the data into daily wages by multiplying by the ton of coal mined per skilled labor-day at each mine. Next, I aggregate employment and daily wages to the town-year-level in order to estimate the town-level inverse skilled labor supply elasticity using Equation (11). This results in a skilled labor supply elasticity that can flexibly vary both across towns and over time.

## B Coal demand

**Identification** Taking logarithms of the coal demand function (5) results in $p_{mt} = \eta q_{mt} + \xi_{mt}$. As firms with attractive features $\xi_{mt}$, such as a convenient location, will set higher coal prices, this equation cannot be identified by regressing coal prices on quantities. I rely on the thickness of the coal vein as a cost shifter: whereas the vein thickness affects the marginal cost of mining, consumers do not care about it as it does not affect coal characteristics. Vein thickness was the result of geological variation, and hence plausibly exogenous to coal firms conditional on their location.

**Coal market definition** Coal firms either sold their output locally near the mine, or sold it to railroad firms who either transported it to final markets, or used it themselves to power their locomotives. I define coal markets $m$ as follows. If a mining town was not located on a railroad line, I infer that coal was sold locally, and define the coal market similarly to labor markets, being the town unless towns are located less than 3 miles from each other. If towns were connected to the railroad network, I let the railroad line be the market: as railroad firms were the main coal buyers, coal firms presumably competed against each other on the same railroad line, but did not compete against coal firms operating on different railroad lines.[33] Defining coal markets in this way results in 249 coal markets, of which 26 railroad lines and 223 local markets. Coal firms on markets not connected to the railroad network

[30]Taking a 10% sample of the town pairs to google maps shows that 3 miles of bird's eye distance corresponds on average to 3.9 miles by today's roads, and 77 minutes of walking (without equipment) one-way.

[31]More information is in Appendix A.2

[32]Summer and winter skilled employment is reported only up to 1890. However, the maximum number of workers is reported next to the average in 1892 and 1894. For these two years, I assume that the maximum number of employees coincides with the amount of workers during winter, and that the average number of workers throughout the year is a simple average between the summer and winter employment averages. This allows me to back out summer and winter worker averages for 1892 and 1894 as well.

[33]Details are in Appendix A.2.

have an average coal market share of 38%, compared to 5.8% for firms selling through the railroad network.

**Estimation** I estimate Equation (5) in logs by 2SLS using the log average vein thickness in the town and the log average mine depth in each year as instruments for the total coal quantity sold. I include the distance to Chicago and St. Louis, a dummy of whether a town was located on a railroad and whether it was located on a crossing of railroads, and a linear time trend as observable market-specific coal demand shifters. I define coal markets at the county-level: towns are too small for defining the coal market, as they were connected through the freight railroad system. On the other hand, I cannot assume too large coal markets, as there was a non-trivial transportation cost.[34]

## C Output elasticities of labor

**Identification** Both labor inputs were assumed to be variable, static inputs. Working out the input demand conditions from Equation (8), the output elasticity of skilled labor is equal to the product of its revenue share, its wage markdown, and its coal price markup.

$$\beta^h_{ft} = \frac{W^h_{nt} H_{ft}((\psi^h_{nt} - 1)s^h_{ft} + 1)}{P_{mt} Q_{ft}(1 + \eta s^q_{ft})} \tag{12}$$

The output elasticity of unskilled labor is then known up to the scale returns constant $\nu$: $\beta^l_{ft} = \nu - \beta^h_{ft}$. Relying on the first order conditions from the labor demand problem to identify the output elasticities of skilled and unskilled labor follows Hall (1988); Foster et al. (2008); Hsieh and Klenow (2009), with the difference that I allow for endogenous input prices. This approach has the benefit of not having to identify the production function, and of allowing for flexible heterogeneity in the output elasticities of skilled labor across mines and time, in contrast to 'input inversion' approaches that rely on a scalar, Hicks-neutral productivity residual. However, this comes at the cost of having to impose a fixed parameter for the degree of returns to scale. The assumption that coal markets are perfectly competitive can be relaxed.[35]

**Estimation** The full distribution of output elasticities of skilled and unskilled labor across mines and time can be readily computed from Equation (12) given that the revenue share of skilled labor is observed, and that the inverse skilled labor elasticity $\psi^h_{ft}$ was estimated earlier. However, doing so requires a calibration of the degree of scale returns $\nu$. I assume that there are decreasing returns to scale in coal mining extraction, because of three reasons. First, nearly all the mines produced far

[34] Only controlling for the state and year dummies explains merely 4% of coal price variation, whereas controlling for county dummies increases this explanatory power to an $R^2$ of 0.70.

[35] In principle, flexible markup variation can be allowed for by comparing cost share variation to revenue share variation. However, in this application, unskilled labor costs are unobserved, which rules out this approach. In Appendix C.1, I use an alternative production model which allows for market power on coal markets, but at the cost of allowing for less heterogeneity in the output elasticities across mines.

below their full capacity, despite coal markets being perfectly competitive. If there would be constant or increasing returns to scale and perfect competition downstream, mines without monopsony power on labor markets should produce at full capacity. Whereas half of the mines have a horizontal skilled labor supply function, and hence no monopsony power, merely 2% of mines produce at full capacity, and 90% of mines us less than four fifths of their capacity. Second, the monthly production data, which were discussed in Section 3, show that aggregate output per worker and output are negatively correlated across months within a year. Output per worker was 3% lower during winters compared to summers. This is consistent with decreasing returns to scale, as output was higher during winter. Third, in Appendix C.1, I specify an extension to the model in which I estimate degree of returns to scale while imposing more structure on the distribution of output elasticities of inputs across mines, which yields an estimated scale parameter 0.908. I calibrate the scale parameter to be $\nu = 0.9$, but conduct robustness checks with different values for $\nu$ in Appendix C.4.

## D Factor-biased and Hicks-neutral effects of cutting machines

**Identification** Finally, the effects of cutting machines on both the output elasticity of skilled labor and on Hicks-neutral productivity, Equations (4b) and (4c), need to be identified. Simply regressing the output elasticity of skilled labor $\beta^h_{ft}$ or Hicks-neutral productivity $\Omega_{ft}$ on cutting machine usage is subject to simultaneity bias, as both Hicks-neutral and factor-augmenting productivity affect input demand, an argument also made by Doraszelski and Jaumandreu (2017). I follow the production function identification literature by relying on timing assumptions to identify the cutting machine effects $\beta_1$ and $\beta_2$ (Olley & Pakes, 1996; Ackerberg et al., 2015). Following Blundell and Bond (2000) I take $\rho$-differences of Equation (4c), such that the skilled labor productivity shock can be written as $\gamma^{\beta}_{ft} = \alpha^{\beta}(K_{ft} - \rho^{\beta} K_{ft-1}) + \sigma^{\beta}(\mathbf{X}_{ft} - \rho^{\beta}\mathbf{X}_{ft-1})$, and the Hicks-neutral productivity shock as $\gamma^{\omega}_{ft} = \alpha^{\omega}(K_{ft} - \rho^{\omega} K_{ft-1}) + \sigma^{\omega}(\mathbf{X}_{ft} - \rho^{\omega}\mathbf{X}_{ft-1})$. Given that cutting machines are assumed to be a dynamic and fixed input, I assume that mines decide on cutting machine usage prior to the realization of both productivity shocks $\gamma^1_{ft}$ and $\gamma^2_{ft}$, which allows to identify the coefficients $\beta$, $\rho$, and $c$ by imposing that current and lagged capital usage are orthogonal to $\gamma^{\beta}_{ft}$ and $\gamma^{\omega}_{ft}$. As both labor inputs and black powder are variable inputs, they are chosen after the productivity shocks $\gamma^{\beta}_{ft}$ and $\gamma^{\omega}_{ft}$ are observed, but their lagged values are orthogonal these shocks. Hence, the moment conditions are:

$$\mathbb{E}\Big[\gamma^{\beta}_{ft}(\rho^{\beta}, \alpha^{\beta}, \sigma^{\beta}) | \begin{pmatrix} K_{ft} \\ K_{ft-1} \\ \mathbf{X}_{ft-1} \\ h_{ft-1} \\ l_{ft-1} \end{pmatrix} \Big] = 0 \qquad \mathbb{E}\Big[\gamma^{\omega}_{ft}(\rho^{\omega}, \alpha^{\omega}, \sigma^{\omega}) | \begin{pmatrix} K_{ft} \\ K_{ft-1} \\ \mathbf{X}_{ft-1} \\ h_{ft-1} \\ l_{ft-1} \end{pmatrix} \Big] = 0$$

**Estimation** I estimate Equations (4b) and (4c) using GMM with the moment conditions above. In the vector of controls $\mathbf{X}$, I include a constant, a linear time trend, and the logarithm of the number of powder kegs used by the firm, adding one within the logarithm to include mines that did not use any powder at all.

### E Fixed costs of cutting machines

As mentioned before, I assume a discount rate of $\delta = 0$, which implies that machines fully depreciate after two years, in order to make the adoption problem static. I assume fixed machine costs $\Phi$ are common across firms and time, and estimate $\Phi$ by matching the average observed machine usage rate to the predicted machine usage rate under this fixed cost. I re-estimate fixed costs in every bootstrapping iteration.

### F Bootstrapping

The entire estimation procedure that has been described in this section happens sequentially. First, I estimate the firm-bi-year-level inverse skilled labor supply elasticities $\psi^h_{ft}$. Next, I estimate the market-level coal demand elasticity $\eta$. Thirdly, I estimate the mine-level output elasticities $\beta^h_{ft}$, which requires knowledge of both $\psi^h_{ft}$ and $\eta$. Fourthly, I estimate the transition equations for the output elasticity of skilled labor $\beta$ and for Hicks-neutral productivity, $\omega$, in order to obtain the cutting machine effects $\alpha^\beta$ and $\alpha^\omega$. Finally, I estimate the level of fixed machine costs $\Phi$. In order to obtain the correct standard errors, I block-bootstrap this entire estimation procedure while resampling within firms over time, with 200 iterations.

## 4.3 Results

A summary of the key model estimates are in Table 1.[36] The skilled labor supply estimates are in Table 1a. The number of observations is 1,116 because the skilled wage markdown is estimated at the labor market-bi-yearly level on the subset of the panel for which seasonal wages are observed (1884-1894). The mean town-level inverse skilled labor supply elasticity $\psi^h_{nt}$ is 1.164, which implies that a monopsonist would set the marginal product of skilled laborers at 16.4% above their wage. A duopsonist with 50% market share would set the marginal product at 8.2% above the skilled wage. The average inverse skilled labor supply elasticity lies between 1.136 and 1.177 with a probability of 90%.[37]

The coal demand elasticities are in Table 1b. The number of observations is lower, at 484, because there are less coal markets than labor markets and because vein thickness (the instrument) is not observed in 1888 and 1890. The inverse demand elasticity is -0.263, so an increase in county-level coal output of 10% results in a drop in the county-level coal price by 2.6%. The inverse demand elasticity

[36] I refer to Appendix Table A4 for the full list of coefficient estimates.

[37] Appendix C.3 discusses how the markdown estimates are correlated with town and county characteristics. The distribution of firm-level markdown ratios, as opposed to the market-level skilled labor supply elasticity, is plotted in Figure A4a.

**Table 1: Model estimates**

| *(a) Miner supply (town-level)* | | Estimate | CI05 | CI95 |
|---|---|---|---|---|
| Inverse elasticity of miner supply | $\psi^h$ | 1.164 | 1.136 | 1.177 |
| Observations | | | 1116 | |
| *(b) Coal demand (county-level)* | | | | |
| Coal demand elasticity | $\eta$ | -0.263 | -0.307 | -0.233 |
| Observations | | | 484 | |
| R-squared | | | .202 | |
| *(c) Output elasticities* | | | | |
| Output elasticity of miners (avg.) | $\beta^h$ | 0.688 | 0.654 | 0.714 |
| Observations | | | 3723 | |
| *(d) Factor-biased productivity transition* | | | | |
| 1(Cutting machine) | $\alpha^\beta$ | -0.132 | -0.189 | -0.008 |
| Observations | | | 1133 | |
| R-squared | | | .006 | |
| *(e) Hicks-neutral productivity transition* | | | | |
| 1(Cutting machine) | $\alpha^\omega$ | 0.249 | -0.163 | 0.425 |
| Observations | | | 1050 | |
| R-squared | | | .238 | |
| *(f) Fixed machine costs* | | | | |
| Fixed machine cost (USD) | $\Phi$ | 8234.753 | 0.000 | 39735.496 |

**Notes:** Panel (a) reports the estimates of the labor supply function, Equation (11). Panel (b) reports the estimates of the coal demand function, Equation (5). Panel (c) reports the output elasticity of skilled labor, using Equation (12). Panels (d)-(e) report the estimated transition equations for the output elasticity of skilled labor and for Hicks-neutral productivity, Equations (4b)-(4c). Panel (f) reports estimates fixed cutting machine costs. Standard errors are block-bootstrapped with 250 iterations.

lies between -0.307 and -0.233 with a probability of 90%. The remaining coefficient estimates are in Table A4. The first-stage regression of the coal quantity on vein thickness has an F-statistic of 51.3. Coal demand is higher in markets that are connected to the railroad network and located on railroad crossings, and decreases with the distance to both St. Louis and Chicago. The distribution of firm-level markup ratios is plotted in Figure A4a.

Table 1c contains the estimated output elasticities of skilled and unskilled labor. The number of observations is 3,723, given that these output elasticities are estimated at the firm-bi-yearly level. The output elasticity of skilled labor is on average 0.688, and lies between 0.654 and 0.714 with a probability of 90%. The average output elasticity of unskilled labor is, mechanically, 0.222. The distribution of output elasticities across firms and time is plotted in Figure A4b.

The factor-biased effects of cutting machines are in Table 1d. Although this model is estimated at the firm-year level too, the number of observations is lower, at 1133, because lagged values of all variables are needed. The output elasticity of skilled labor is estimated to fall by 0.163 units due to the usage of cutting machines, which is a relative drop of 24% on average. This effect lies between -0.189 and -0.008 with a probability of 90%, and is thus significantly negative at the 95% confidence level. The findings that cutting machines were unskill-biased is consistent with the anecdotal historical evidence presented earlier, in Section 3. The effect of cutting machines on Hicks-neutral productivity is in Table 1e. The point estimate implies that cutting machines increased Hicks-neutral productivity by 28%, but this effect is very imprecisely estimated: cutting machines could have lowered Hicks-neutral productivity by 16% or increased it by 53%, all within a probability of 90%.

Finally, the fixed machine cost is in Table 1f, and is estimated to be \$8235, which is 3.1% higher than the yearly profit of the average firm.

## 4.4 Counterfactuals

### A Computation of the equilibrium

Let $X \in \{Q, H, L, W^h, W^l, P, \Pi, \Pi^{tot}\}$ be an endogenous variable in the model. I denote $X_{ft}^K(s_{ft}^h)$ as the equilibrium value of variable $X$ for usage of the technology $K \in \{0,1\}$ and for a certain labor market share $s_{ft}^h$. The equilibrium values for all endogenous variables $X$ can be computed using the expressions in Section B.5. For instance, $Q_{ft}^1(s_{ft}^h = 0.5)$ denotes the equilibrium output of firm $f$ in year $t$ when using cutting machines, and having a labor market share of 50%, and can be computed using Equation (13a).

In order to compute the equilibrium values of all endogenous variables $X$, I need to know the values of the output elasticity of skilled labor and the Hicks-neutral productivity level both when using cutting machines and when not doing so, $\beta_{ft}(K_{ft})$ and $\Omega_{ft}(K_{ft})$. If the mine does not use cutting machines, I calculate the counterfactual output elasticity if it would use cutting machines as $\beta_{ft}^h(K_{ft} = 1) = \beta_{ft}^h + \beta_1$, using Equation (4b). Similarly if the mine is already using cutting machines, the counterfactual output elasticity when not doing is $\beta_{ft}^h(K_{ft} = 0) = \beta_{ft}^h - \beta_1$. The counterfactual Hicks-neutral productivity levels are computed in the same way.

The market-level demand shifter $\zeta_{mt}$ is computed as the residual of the coal demand function.

Similarly, the labor supply residual $\xi^h_{nt}$ is the residual of the estimated labor supply function (6):

$$\xi^h_{nt} = \frac{W^h_{nt}}{\left(\frac{H_{ft}}{s^h_{ft}}\right)^{\psi^h_{nt}-1}}$$

## B Labor market competition and technology usage

Using the equilibrium variables $X^K_{ft}(s^h_{ft})$, I examine how three key outcomes of interest change in function of labor market competition $s^h_{ft}$. First, I am interested in the *variable profit return* to machine usage, $\frac{\Pi^1_{ft}(s^h_{ft})-\Pi^0_{ft}s^h_{ft}}{\Pi^0_{ft}s^h_{ft}}$. Secondly, I calculate capital usage in the absence of fixed machine costs, $\mathcal{I}[\Pi^1_{ft}(s^h_{ft}) - \Pi^0_{ft}(s^h_{ft})]$. This is an interesting metric because it tells us how much of the machine would be used if machine costs would be variable, rather than fixed. Thirdly, I calculate equilibrium machine usage as $\mathcal{I}[\Pi^1_{ft}(s^h_{ft}) - \Pi^0_{ft}(s^h_{ft}) - \Phi]$. The key counterfactual exercise of the paper is how these three metrics change in function of the level of labor market competition, as measured by the labor market share $s^h_{ft}$.

A number of assumptions need to be explained at this point. First, the labor supply and coal demand residuals $\xi^h_{nt}$, $\xi^l_{nt}$, $\zeta_{mt}$ are assumed to be invariant to both labor market structure and machine usage: both labor market structure and machine usage are assumed to affect worker and consumer preferences only through equilibrium wages and prices, but not directly. Secondly, I assume that unskilled worker characteristics, which are equal to unskilled worker wages, are the same across mines in a given year $\xi^l_{nt} = \xi^l_t$. This assumption is motivated by the evidence in Figure 5c, which showed that there is very little cross-sectional variation in unskilled wages. The residual $\xi^l_t$ is equal to the daily unskilled wage, which is unobserved. However, it can be backed out under the assumption of competitive unskilled labor markets. Writing out Equation (12) for both unskilled and skilled labor gives a system of equations (the variable input demand first order conditions) that can be solved for unskilled wages. The resulting unskilled wage expression is $W^l_{ft} = \frac{\beta^l_{ft}}{\beta^h_{ft}} \frac{P_{ft}Q_{ft}}{L_{ft}}(\psi^h_{nt} - 1)s^h_{ft}$. I take the yearly average of this imputed wage to be the unskilled wage $W^l_t$, which is equal to the unskilled labor supply residual $\xi^l_t$. Thirdly, when considering the effects of changing labor market structure on machine returns and machine usage, I do not let *coal* market structure vary simultaneously: the focus is to isolate the effects of labor market competition on technology returns and adoption, rather than the joint effect of labor and product market competition on these outcomes. Finally, fixed machine costs $\Phi$ are assumed to be invariant to the level of labor and product market competition.

## C Counterfactual technologies

In order to understand how the directed and Hicks-neutral effects of technologies shape the relationship between labor market competition and technology adoption, I carry out the counterfactual exercise from the previous section under three different technologies. Firstly, I consider the actual technology, the cutting machine, which was both unskill-biased and Hicks-neutral productivity-enhancing:

$\frac{\partial \beta^h}{\partial K} < 0$ and $\frac{\partial \Omega}{\partial K} > 0$. Secondly, I consider a skill-biased technology, such as hauling locomotives, for which $\frac{\partial \beta^h}{\partial K} > 0$ and $\frac{\partial \Omega}{\partial K} > 0$. I consider a technology with the exact opposite effect as cutting machines, meaning that its skilled labor output elasticity $\beta'$ is equal to $\beta'(K = 1) = \beta(K = 0)$ and $\beta'(K = 0) = \beta(K = 1)$. Finally, I do the same analysis for a counterfactual technology that would only be unskill-biased but not increase Hicks-neutral productivity, $\frac{\partial \beta^h}{\partial K} < 0$ and $\frac{\partial \Omega}{\partial K} = 0$. I re-estimate the fixed cost that rationalizes observed machine usage for every counterfactual technology.

## D Results

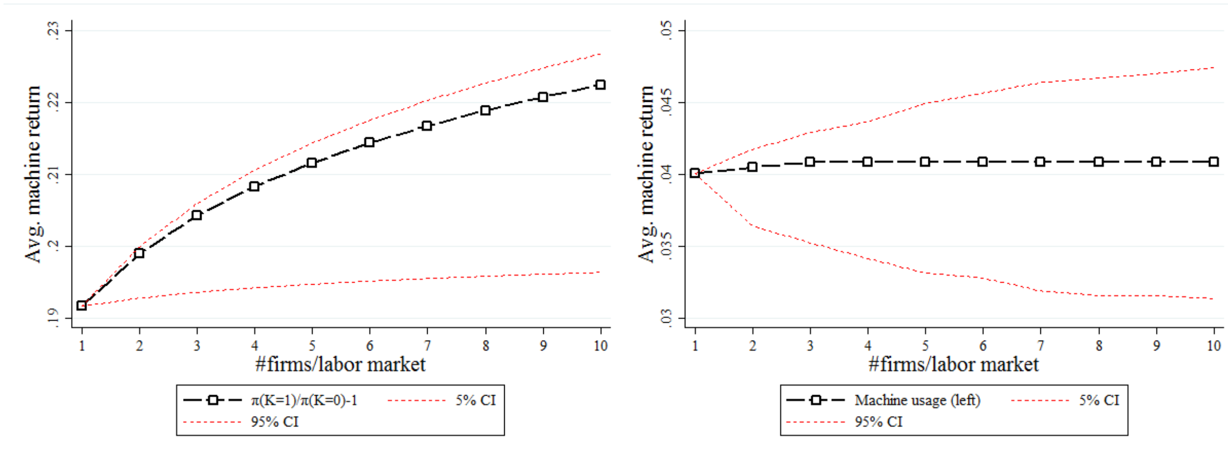
**Unskill-biased technology** The results of the counterfactual exercise for cutting machines, the skill-biased technology, is in Figure 6. In line with theorem 1, the variable profit return of cutting machines $\frac{\Pi(K=1)-\Pi(K=0)}{\Pi(K=0)}$ increases with the number of firms in the labor market. Whereas adopting a cutting machine increases variable profits by 19.2% on average under monopsonistic labor markets, this return is 20.0% under a symmetric duopsony, and 22.2% if there are 10 equally-sized firms on each labor market. Table 2 shows the bootstrapped confidence intervals on this change. The increase in returns to mechanization is statistically significant: when moving from 1 to 10 firms per labor market, the 90% confidence interval for the change in the cutting machine return is [0.2,3.7].

Although the average *relative* returns to mechanization would increase with labor market competition, machine *usage* would barely change: it would increase by 0.1 p.p., and this change is not significantly different from zero. Theorem 2 already said that the effect of skilled labor market competition on the absolute returns to technology adoption are ambiguous for an unskill-biased technology. A part of the explanation for this is that increasing labor market competition decreases profits, even if it increases the relative profit difference between using and not using technology. In order to recover fixed costs of technology usage, some degree of profitability is needed. A second reason for the zero net effect of competition on technology adoption is that cutting machines increase total factor productivity. Even if cutting machines decrease the usage of skilled workers relatively to unskilled workers, the absolute change in skilled labor demand is much less negative because cutting machines increase Hicks-neutral productivity.

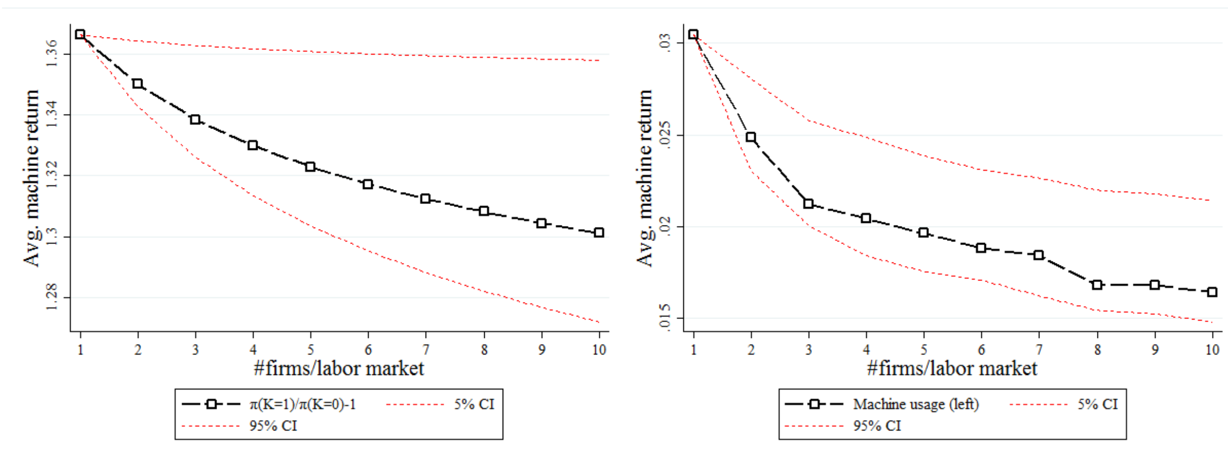
**Skill-biased technology** Next, consider a technology that is biased towards skilled labor, such as the mining locomotive. As is shown in Figure 6b, increasing skilled labor market competition would now decrease, rather than increase, the returns to such a technology. This is in line with theorem 1: increased competition for skilled labor decreases the markdown extracted from skilled workers, which decreases the incentive to adopt a technology that switches input usage towards these workers. The average return to the skill-biased technology drops from 137% to 130% on average when moving from a monopsonistic labor market to one with ten equally-size employers per market. As shown in Table 2, the usage rate of a skill-biased technology would drop from 3.0% to 1.6% when moving from 1 to 10 firms per labor market, a relative of nearly 50%. This decrease in adoption of a skill-biased technology has a 90% confidence interval of [-0.8 p.p.,-1.7 p.p.] and is hence statistically significant.

Figure 6: Counterfactuals

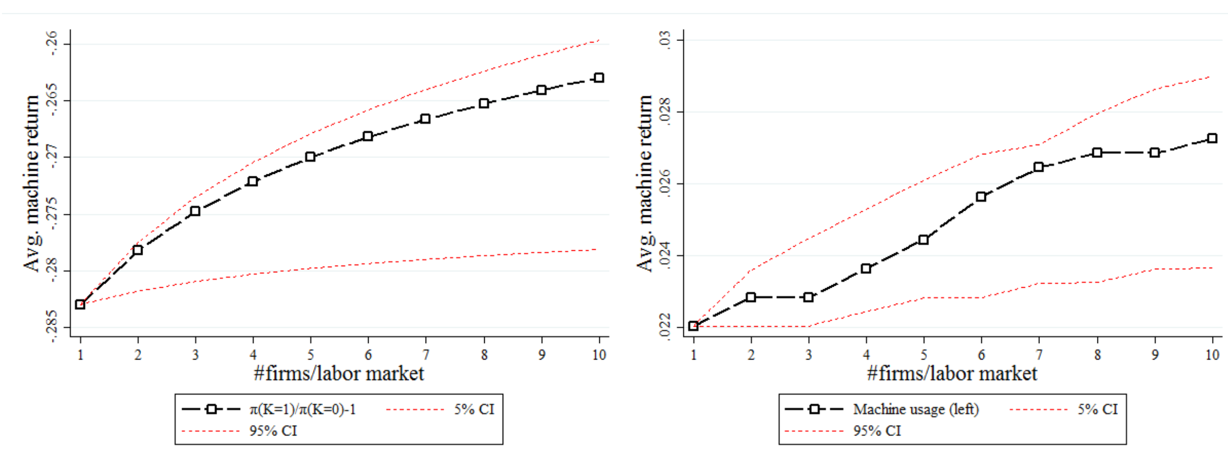
(a) Unskill-biased technology




(b) Skill-biased technology




(c) Unskill-biased technology without Hicks-neutral productivity effect




**Unskill-biased technology without Hicks-neutral productivity effect** Finally, suppose cutting machines would only change the output elasticity of skilled labor, but not Hicks-neutral productiv-

**Table 2: Counterfactuals: confidence intervals**

| *(a) Unskill-biased technology* | 1 firm | 10 firms | Dif. | CI05 | CI95 |
|---|---|---|---|---|---|
| Avg. relative return to technology | 0.192 | 0.222 | 0.031 | 0.002 | 0.043 |
| Avg. technology usage | 0.040 | 0.041 | 0.001 | -0.010 | 0.010 |
| *(b) Skill-biased technology* | 1 firm | 10 firms | Dif. | CI05 | CI95 |
| Avg. relative return to technology | 1.367 | 1.301 | -0.065 | -0.111 | -0.002 |
| Avg. technology usage | 0.030 | 0.016 | -0.014 | -0.017 | -0.008 |
| *(c) Purely unskill-biased technology* | 1 firm | 10 firms | Dif. | CI05 | CI95 |
| Avg. relative return to technology | -0.283 | -0.263 | 0.020 | 0.002 | 0.025 |
| Avg. technology usage | 0.022 | 0.027 | 0.005 | 0.001 | 0.008 |

ity. The relative returns from adoption would now be negative on average, but would increase from -0.283% to -0.263% on average when moving from 1 to 10 firms per labor market. As shown in Figure 6c, machine usage now *increases* with labor market competition, from 2.2% to 2.7% of firms, and this increase is significantly above zero. The more positive effect of labor market competition on machine usage compared to the baseline scenario is consistent with Lemma 2, and is due to the fact that without Hicks-neutral productivity change, skilled labor usage always falls when the output elasticity of skilled labor falls. Firms are less willing to adopt a technology that reduces their usage of skilled labor if their markdown earned over skilled labor is higher. Table 2 shows that the increase in relative returns to machine adoption is significantly positive, as is the change in cutting machine usage.

**Sizing the results** The counterfactual analysis shows that changes in labor market structure had little effects on cutting machine adoption, due to the fact that the factor-biased and Hicks-neutral effects of these machines counteract each other, and due to high fixed costs. However, under a different direction of technical change and/or different Hicks-neutral productivity changes, labor market power would have much larger effects on technology returns and adoption. This is remarkable given that the degree of oligopsony power inferred from the model is quite modest. The average firm sets wages at 3.5% below the marginal product of labor, and even under a pure monopsony, this markdown would be merely 14%. The current literature on oligopsony power usually finds much higher markdowns.

For instance, Azar et al. (2017) find a median markdown of 17%.[38] With higher markdown levels, the effects of oligopsony power on technology usage would be even more pronounced.

**Adoption vs. invention** Throughout the paper, I took the invention of new technologies, and their directionality, as given, and investigated how the adoption of such technologies varied with the degree of oligopsony power. Given that invention is likely impacted by the demand for new technologies, it is conceivable that labor market power does not only affect the usage of new technologies, but also their invention. The direction of newly invented technologies could hence be endogenous to the (aggregate) degree of oligopsony power on the various input markets, which would be an extension of Acemoglu (2002) and its ensuing literature to a setting with oligopsonistic input markets.

### E Current-day implications

Although the application in this paper is historical, the results have several important current-day implications. The model shows that in order to understand the effects of oligopsony power on technological change today, one needs to know (i) the direction of technological change, and (ii) the relative degrees of monopsony/oligopsony power over different types of inputs. These two primitives will most likely differ between industries. Across the board, the consensus seems that automation has been mainly skill-biased throughout the last couple of decades. If firms mainly exert market power over unskilled workers, then such market power is reducing the returns to automation. Little is known, however, about the relative degrees of oligopsony power across the skill and income distribution. Moreover, the model is not restricted to the study of labor markets and automation. Many technologies today are energy-saving. The model could potentially also be used to understand how oligopsony power on energy markets, which could exist locally by large energy-intensive industries, affects the incentives to adopt energy-saving technologies.

## 5 Conclusion

In this paper, I investigate how oligopsony power by firms affects the adoption of new production technologies. Using a theoretical model of log-linear labor supply and labor demand, I show that oligopsony power could either increase or decrease technological change, depending on the direction of technical change and its Hicks-neutral productivity effects. In an application, I implement an empirical version of this model to understand how oligopsony power over skilled coal miners affected the mechanization of the late 19th century Illinois coal mining industry. I find that the returns to unskill-biased technologies, such as cutting machines, increased with labor market competition, whereas the returns to skill-biased technologies, such as underground locomotives, decreased with labor market

[38] These higher markdown estimates are likely the consequence of differentiation between firms and jobs from the worker's point of view.

competition. In terms of technology usage, I find that changes in labor market competition had limited effects on cutting machine adoption. If cutting machines would have been purely unskill-biased, without changing Hicks-neutral productivity, their adoption would have increased with increased labor market competition. In contrast, increased labor market competition would have decreased the adoption of skill-biased technologies. These findings show that in order to understand the consequences of oligopsony power on technological change and productivity growth, it is crucial to know the direction of technological change and the relative magnitude of oligopsony power on the various factor markets. Both of these are likely to differ between markets, industries, and even firms.

# References


Acemoglu, D. (2002). Directed Technical Change. *The Review of Economic Studies*, *69*(4), 781–809.

Acemoglu, D. (2003). Labor- and Capital-Augmenting Technical Change. *Journal of the European Economic Association*, *1*(1), 1-37.

Ackerberg, D., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*(6), 2411-2451.

Ackermann, A. (1902). *Coal-cutting by machinery in America* (No. 16). The Colliery Guardian Company Limited.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: an inverted U-relationship. *Quarterly Journal of Economics*.

Aguirregabiria, V., & Luengo, A. (2017). *A microeconometric dynamic structural model of copper mining decisions.* (Mimeo)

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press.

Antras, P. (2004). Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. *Contributions in Macroeconomics*, *4*(1), 1-36.

Ashenfelter, O. C., Farber, H., & Ransom, M. R. (2010). Labor market monopsony. *Journal of Labor Economics*, *28*(2), 203-210.

Asker, J., Collard-Wexler, A., & De Loecker, J. (2019). (Mis)allocation, market power, and global oil extraction. *American Economic Review*, *109*(4), 1568–1615.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the US labor market. *American Economic Review*, *96*(2), 189–194.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279-1333.

Azar, J., Marinescu, I., & Steinbaum, M. (2017). *Labor market concentration* (Working Paper No. 24147). National Bureau of Economic Research.

Benkard, C. L. (2000). Learning and forgetting: the dynamics of aircraft production. *American Economic Review*, *90*(4), 1034-1054.

Berger, D., Herkenhoff, K., & Mongey, S. (2019). *Labor market power* (Tech. Rep.). Mimeo.

Bloch, L. (1922). *The Coal Miners' Insecurity: Facts about Irregularity of Employment in the Bituminous Coal Industry in the United States* (No. 7). Russell Sage Foundation.

Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: An application to production functions. *Econometric reviews*, *19*(3), 321–340.

Boal, W. M. (1995). Testing for employer monopsony in turn-of-the-century coal mining. *RAND Journal of Economics*, *26*(3), 519-536.

Boal, W. M. (2017). What did Unions do? The case of Illinois Coal Mining in the 1880s. *Journal of Labor Research*, *38*(4).

Brown, W. (1889). *Transactions of the federated institution of mining engineers* (Tech. Rep. No. 1). Federated Institution of Mining Engineers.

Card, D., & Krueger, A. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *American Economic Review*, *84*(4), 772-793.

Collard-Wexler, A., & De Loecker, J. (2015). Reallocation and technology: Evidence from the US steel industry. *American Economic Review*, *105*(1), 131-171.

Dechezleprêtre, A., Hémous, D., Olsen, M., & Zanella, C. (2019). *Automating labor: Evidence from firm-level patent data.*

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, *135*(2), 561–644.

Demirer, M. (2020). *Production function estimation with factor-augmenting technology: An application to markups* (Tech. Rep.). MIT working paper.

Doraszelski, U., & Jaumandreu, J. (2017). Measuring the bias of technological change. *Journal of Political Economy*, *126*(3), 1027-1084.

Edmond, C., Midrigan, V., & Xu, D. Y. (2018). *How costly are markups?* (Tech. Rep.). National Bureau of Economic Research.

Fishback, P. V. (1992). *Soft coal, hard choices: The economic welfare of bituminous coal miners, 1890-1930*. Oxford University Press.

Fishback, P. V., Holmes, R., & Allen, S. (2009). Lifting the curse of dimensionality: measures of states' labor legislation climate in the United States during the progressive era. *Labor History*, *50*(3), 313-346.

Foster, L., Haltiwanger, J., & Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, *98*(1), 394–425.

Goldin, C. D., & Katz, L. F. (2009). *The Race between Education and Technology*. Harvard University Press.

Goolsbee, A., & Syverson, C. (2019). *Monopsony power in higher education: a tale of two tracks* (Tech. Rep.). National Bureau of Economic Research.

Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics*, *89*(1), 118–133.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, *104*(8), 2509–26.

Graebner, W. (1974). Great expectations: The search for order in bituminous coal, 1890-1917. *The Business History Review*, 49–72.

Haanwinckel, D. (2018). Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution. *Unpublished manuscript*.

Hall, R. E. (1988). The relation between price and marginal cost in US industry. *Journal of Political Economy*, *96*(5), 921–947.

Hanlon, W. W. (2015). Necessity is the mother of invention: Input supplies and directed technical change. *Econometrica*, *83*(1), 67–100.

Harberger, A. C. (1954). The welfare loss from monopoly. *American Economic Review*, *44*(2), 77–87.

Hashmi, A. R., & Van Biesebroeck, J. (2016). The relationship between market structure and innovation in industry equilibrium: A case study of the global automobile industry. *The Review of Economics and Statistics*, *98*(1), 192-208.

Hicks, J. (1932). *The Theory of Wages*. London: Macmillan.

Hoover, E. D. (1960). Retail prices after 1850. In *Trends in the American economy in the nineteenth century* (pp. 141–190). Princeton University Press.

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, *124*(4), 1403-1488.

Huang, S.-Y., & Sexton, R. J. (1996). Measuring returns to an innovation in an imperfectly competitive market: Application to mechanical harvesting of processing tomatoes in Taiwan. *American Journal of Agricultural Economics*, *78*(3), 558–571.

Humphries, J. (2013). The lure of aggregates and the pitfalls of the patriarchal perspective: A critique of the high wage economy interpretation of the British industrial revolution. *The Economic History Review*, *66*(3), 693–714.

Igami, M., & Uetake, K. (2017). *Mergers, innovation and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996-2016* (Tech. Rep.).

Inderst, R., & Wey, C. (2003). Bargaining, mergers, and technology choice in bilaterally oligopolistic industries. *RAND Journal of Economics*, 1–19.

James, J. A., & Skinner, J. S. (1985). The resolution of the labor-scarcity paradox. *The Journal of Economic History*, *45*(3), 513–540.

Joyce, R. (2009). *The Early Days of Coal Mining in Northern Illinois.* Illinois Labor History Society. (Accessed through http://www.illinoislaborhistory.org/labor-history-articles/early-days-of-coal-mining-in-northern-illinois)

Just, R. E., & Chern, W. S. (1980). Tomatoes, technology, and oligopsony. *The Bell Journal of Economics*, 584–602.

Katz, L. F., & Margo, R. A. (2014). Technical change and the relative demand for skilled labor: The

United States in historical perspective. In *Human capital in history: The American record* (pp. 15–57). University of Chicago Press.

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*, *109*(10), 3585–3616.

Köhler, C., & Rammer, C. (2012). Buyer power and suppliers' incentives to innovate. *ZEW-Centre for European Economic Research Discussion Paper*.

Kroft, K., Luo, Y., Mogstad, M., & Setzler, B. (2020). *Imperfect competition and rents in labor and product markets* (Tech. Rep.). National Bureau of Economic Research.

Lamadon, T., Mogstad, M., & Setzler, B. (2019). *Imperfect competition, compensating differentials and rent sharing in the us labor market* (Tech. Rep.). National Bureau of Economic Research.

Lindner, A., Murakozy, B., Reizer, B., & Schreiner, R. (2019). *Technological change and skill demand in non-competitive labor markets* (Tech. Rep.). Mimeo.

Loertscher, S., & Marx, L. M. (2020). Countervailing power, integration, and investment under incomplete information.

Lord, J. S. (1883). *Statistics of Coal Production in Illinois, 1883: A Supplemental Report of the State Bureau of Labor Statistics*.

Lord, J. S. (1884). *Coal in illinois, 1884* (Biennial Report No. 3). State Bureau of Labor Statistics of Illinois.

Lord, J. S. (1886). *Coal in illinois, 1886* (Biennial Report No. 4). State Bureau of Labor Statistics of Illinois.

Lord, J. S. (1888). *Coal in illinois, 1888* (Biennial Report No. 5). State Bureau of Labor Statistics of Illinois.

Lord, J. S. (1890). *Coal in illinois, 1890* (Biennial Report No. 6). State Bureau of Labor Statistics of Illinois.

Lord, J. S. (1892). *Coal in illinois, 1892* (Biennial Report No. 7). State Bureau of Labor Statistics of Illinois.

Manning, A. (2011). Imperfect competition in the labor market. In *Handbook of labor economics* (Vol. 4, pp. 973–1041). Elsevier.

Manning, A. (2013). *Monopsony in motion*. Princeton University Press.

Mokyr, J. (1990). *Twenty five centuries of technological change: An historical survey* (Vol. 35). Taylor & Francis.

Morlacco, M. (2017). *Market power in input markets: theory and evidence from French manufacturing* (Tech. Rep.).

Naidu, S., Nyarko, Y., & Wang, S.-Y. (2016). Monopsony power in migrant labor markets. *Journal of Political Economy*, *124*(6), 1735-1792.

Naidu, S., & Yuchtman, N. (2017). Labor market institutions in the gilded age of american history. *The Oxford Handbook of American Economic History*, *Forthcoming*.

National Bureau of Economic Research, t. (n.d.). *Wholesale Price of Bituminous Coal, Georges*

*Creek, F.O.B. New York Harbor for New York, NY.* Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/M04046US35620M294NNBR, October 13, 2020.

Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, *64*(6), 1263-1297.

Parra, A., & Marshall, G. (2021). *Monopsony power and upstream innovation* (Tech. Rep.).

Reid, A. (1876). *Transactions of the North of England Institute of Mining and Mechanical Engineers*. North of England Institute of Mining and Mechanical Engineers.

Ross, D. (1898). *Coal in illinois, 1898* (Annual Report No. 17). State Bureau of Labor Statistics of Illinois.

Ross, D. (1900). *Coal in illinois, 1900* (Annual Report No. 19). State Bureau of Labor Statistics of Illinois.

Ross, D. (1902). *Coal in illinois, 1902* (Annual Report No. 21). State Bureau of Labor Statistics of Illinois.

Rubens, M. (2020). *Market structure, oligopsony, and productivity* (Mimeo). KU Leuven.

Salter, W. E. G. (1966). *Productivity and technical change* (Vol. 960). Cambridge University Press Cambridge.

Schilling, G. A. (1894). *Coal in illinois, 1894* (Annual Report No. 13). State Bureau of Labor Statistics of Illinois.

Schilling, G. A. (1896). *Coal in illinois, 1896* (Annual Report No. 15). State Bureau of Labor Statistics of Illinois.

Schubert, G., Stansbury, A., & Taska, B. (2020). Monopsony and outside options. *Available at SSRN*.

Schumpeter, J. (1942). *Capitalism, socialism and democracy*. New York: Harper.

Starr, E., Prescott, J. J., & Bishara, N. (2020). Noncompete agreements in the US labor force. *Journal of Law and Economics, forthcoming*.

Whitcomb, G. D. (1882, 08 8). *Coal mining machine* (No. 262225).

# Appendices

## A Data

### A.1 Sources

**Mine Inspector Reports** The main data source is the biennial report of the Bureau of Labor Statistics of Illinois between 1884-1902 (Lord, 1884, 1886, 1888, 1890, 1892; Schilling, 1894, 1896; Ross, 1898, 1900, 1902).

Each report contains a list of all mines in each county, and reports the name of the mine owner, which I take to be the firm, the town nearest the mine, and a selection of variables which varies across the volumes. An overview of all variables (including unused ones), and the years in which they are observed, is in Tables A8 and A9. Output quantities, the number of miners and other employees, mine-gate coal prices, and information about the usage of cutting machines are reported in every volume. Miner wages and the number of days worked are reported in every volume except 1896. The other variables, which includes information about the mine type, hauling technology, other technical characteristics, and other inputs, are reported in a subset of years.

**Census of Population, Agriculture, and Manufacturing** I use the 1880 population census to have information on county population sizes, demographic compositions, and areas. I also observe the county-level capital stock and employment in manufacturing industries from the 1880 census of manufacturing, and the number of farms and improved farmland area from the 1880 census of agriculture.

**Monthly data** In 1888, I observe monthly production data for a selection of 11 mines in Illinois, across 6 counties. I observe the monthly number of days worked and number of skilled and unskilled workers. I also observe the net earnings for all skilled and unskilled workers per mine per month, and the number of tons mined per worker per month. This allows me to compute the daily earnings of skilled and unskilled workers per month. I also obtain monthly coal price in the harbor of New York City from the Federal Reserve Economic Data: the *Wholesale Price of Bituminous Coal, George Creek, F.O.B. New York Harbor for New York, NY, Dollars per Short Ton*.[39]

### A.2 Data cleaning

**Employment** In every year, except for 1896, workers are divided in two categories, 'miners' and 'other employees'. In 1896, the distinction is made between 'underground workers' and 'aboveground workers', which is not the same distinction: all miners were underground workers, but some underground workers were not miners (e.g. doorboys, mule drivers, etc.). The employment data are

[39] Accessed through https://fred.stlouisfed.org

hence unobserved in 1896. From 1888-1896, boys are reported as a separate working category. Given that miners (cutters) were adults, I include these boys into the ‘other employee’ category.

The number of days worked is observed for all years. The average number of other employees per mine throughout the year is observed in every year bar 1896; in 1898 it is subdivided into underground other workers and above-ground other workers, which I add up into a single category.

The quantity of skilled and unskilled labor is calculated by multiplying the number of days worked with the average number of workers in each category throughout the year. Up to and including 1890, the average number of miners is reported separately for winters and summers. I calculate the average number of workers during the year by taking the simple average of summers and winters. If mines closed down during winters or, more likely, summers, I calculate the annual amount of labor-days by multiplying the average number of workers during the observed season with the total number of days worked during the year.

**Wages** Only miner wages are consistently reported over time at the mine level. The piece rate for miners is reported. Up to 1894, miner wages per ton of coal are reported separately for summers and winters. I weight these seasonal piece rates wages using the number of workers employed in each season for the years 1884-1890. In 1892 and 1894, seasonal employment is not reported, so I take simple averages of the seasonal wage rates. In 1896, wages are unobserved. From 1898 onwards, wages are reported at a yearly level, because wages were negotiated biennally after the large strikes of 1897-1898. For these years, wages are reported separately for hand and machine miners. In the mines that employed both hand and machine miners, I take the average of these two piece rates, weighted by the amount of coal cut by hand and cutting machines.

**Output** The total amount of coal mined is reported in every year, in short tons (2000 lbs). Up to and including 1890, the total quantity of coal extraction is reported, without distinguishing different sizes of coal pieces. After 1890, coal output is reported separately between ‘lump’ coal (large pieces) and smaller pieces, which I sum in order to ensure consistency in the output definition.

**Coal prices** Prices are normally given on average for all coal sizes, except in 1894 and 1896, where they are only given for ‘lump’ coal (the larger chunks of coal). I take the lump price to be the average coal price for all coal sizes in these two years. There does not seem to be any discontinuity in the time series of average or median prices between 1892-1894 or 1896-1898 after doing this, which I see as evidence for this assumption.

**Cutting machine usage** Between 1884 and 1890, the number of cutting machines used in each mine is observed. In between 1892 and 1896, a dummy is observed for whether coal was mined by hand, using cutting machines, or both. I categorize mines using both hand mining and cutting machines as mines using cutting machines. In 1898, I infer cutting machine usage by looking at which mines paid 'machine wages' and 'hand wages' (or both). In 1888, the number of cutting machines is reported by type of cutting machine as well. Finally, in 1900 and 1902, the output cut by machines and by hand is reported separately for each mine, on the basis of which I again know which mines used cutting machines, and which did not.

**Deflators** I deflate all monetary variables using the consumer price index from the *Handbook of Labor Statistics* of the U.S. Department of Labor, as reported by the Minneapolis Federal Reserve Bank website.[40]

**Hours worked** In 1898, eight-hour days were enforced by law, which means that the 'number of days' measure changes in unit between 1898 and 1900. As the inspector report from 1886 shows that ten-hour days were the standard, I multiply the number of working days after 1898 by 80% in order to ensure consistency in the meaning of a 'workday', i.e. to ensure that in terms of total number of hours worked, the labor quantity definition does not change after 1898.

**Mine and firm identifiers** The raw dataset reports mine names, which are not necessarily consistent over time. Based on the mine names, it is often possible to infer the firm name as well, in the case of multi-mine firms. For instance, the Illinois Valley Coal Company No. 1 and Illinois Valley Coal Company No. 2 mines clearly belong to the same company. For single-mine firms, the operator is usually mentioned as the mine name, (e.g. 'Floyd Bussard'). For the multi-mine firms, mine names were made consistent over time as much as possible. For the individual mine operators, it is impossible to link mines over time when the operator changes. There will hence be a lot of false exits and entries. The dataset is hence not very suitable for panel-data analysis when used at the mine-level.

**Town identifiers and labor market definitions** The raw data report town names. I link these names to geographical coordinates using Google Maps. I calculate the shortest distance between every town in the data. For towns that are located less than 3 miles from each other, I merge them and assign them randomly the coordinates of either of the two mines. This reduces the number of towns in the dataset from 448 to 350. The resulting labor markets lie at least 3 miles from the nearest labor market.

**Coal market definitions** Using the 1883 Inspector Report, I link every coal mining town to a railroad line, if any. Some towns are located at the intersection of multiple lines, in which case I assign the town to the first line mentioned. I make a dummy variable that indicates whether a railroad is located on a crossroad of multiple railroad lines. Towns not located on railroads are assumed to be isolated coal markets. For the connected towns, the market is defined as the railroad line on which they are located, of which there are 26. Given that data from 1883 is used, expansion of the railroad

[40]https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1800-

network after 1883 is not taken into account. However, the Illinois railroad network was already very dense by 1883.

**Aggregation from mine- to firm-level** I aggregate labor from the mine-bi-year- to firm-bi-year level by taking sums of the number of labor-days and labor expenses for both types of workers, both per year and per season. I calculate the wage rates for both types per workers by dividing firm-level labor expenditure on the firm-level number of labor-days. I also sum powder usage, coal output and revenue to the firm-level and calculate the firm-level coal price by dividing total firm revenue by total firm output. I aggregate mine depth and vein thickness by taking averages across the different mines of the same firm. I define the cutting machine dummy at the firm-level as the presence of at least one cutting machine in one of the mines owned by the firm. I define 'firm' as the combination of the firm name in the dataset and its town (the merged towns that are used to define labor markets), as firms are assumed to optimize input usage on a town-by-town basis.

# B Theory

## B.1 Proof of theorem 1

**To prove:** $\frac{\partial^2(\Pi_f)}{\partial\mu_f^h\partial\beta_f^h} \geq 0$

**Proof:** Omit subscripts $f$ for simplicity. Denote $\beta \equiv \beta^h$. Assume $\eta = 0$, without loss of generality. I let the firm be a monopsonist on market $L$, $s^l = 1$ and consider the effect of changes in the market share $s^h$. The proof is analogous when keeping $s^h = 1$ and considering variation in $s^l$. To simplify notation, denote $y = 1 + s^h(\psi^h - 1)$: I examine the sign of $\frac{\partial^2(\Pi)}{\partial y\partial\beta^h}$, which is the same as the sign of $\frac{\partial^2(\Pi)}{\partial s^h\partial\beta^h}$. Using Equations (2b)-(2a), variable profits are given by:

$$\Pi = [(\frac{\beta}{y})^{\frac{\beta}{\psi^h}}(\frac{\nu-\beta}{\psi^l})^{\frac{\nu-\beta}{\psi^h}}\Omega]^{\frac{1}{1-\frac{\beta}{\psi^h}-\frac{(\nu-\beta)}{\psi^l}}}(1-\frac{\beta}{y}-\frac{(\nu-\beta)}{\psi^l})$$

with $1 \leq y \leq \psi^h$; $\psi^h \geq 1$;$\psi^l \geq 1$; $\nu \leq 1$; $0 \leq \beta \leq \nu$.

Define $\pi \equiv \ln(\Pi)$. I prove that $\frac{\partial^2(y)}{\partial\psi^h\partial\beta^h} \geq 0$. Variable profits are weakly positive due to the economic restrictions on the parameter values. Given that $\Pi \geq 0$, and that $\pi(.)$ is twice differentiable, $\frac{\partial^2(\pi)}{\partial\beta^h\partial y} \geq 0 \Leftrightarrow \frac{\partial^2(\Pi)}{\partial y\partial\beta^h} \geq 0$.

The effect of monopsony power on profits is equal to:

$$\frac{\partial\pi}{\partial y} = \frac{-\frac{\beta}{\psi^h y}}{1-\frac{\beta}{\psi^h}-\frac{(\nu-\beta)}{\psi^l}} + \frac{\frac{\beta}{(y^2}}{1-\frac{\beta}{y}-\frac{(\nu-\beta)}{\psi^l}}$$

Taking second order derivatives w.r.t. the output elasticity of $H$, $\beta$, gives:

$$\frac{\partial}{\partial\beta}(\frac{\partial\pi}{\partial y}) = \frac{\frac{(\nu-\psi^l)y}{\psi^h\psi^l(y)^2}(1-\frac{\beta}{y}-\frac{(\nu-\beta)}{\psi^l})^2 + \frac{(\psi^l-\nu)\psi^h}{\psi^h\psi^l(y)^2}(1-\frac{\beta}{\psi^h}-\frac{(\nu-\beta)}{\psi^l})^2}{(1-\frac{\beta}{\psi^h}-\frac{(\nu-\beta)}{\psi^l})^2(1-\frac{\beta}{y}-\frac{(\nu-\beta)}{\psi^l})^2}$$

Working this out delivers the following expression, which is weakly positive given that $\psi^l \geq 1$, $\nu \leq 1$, and $y \leq \psi^h$:

$$\frac{\partial}{\partial\beta}(\frac{\partial\pi}{\partial y}) = \frac{(\psi^l-\nu)(\psi^h-y^h)[1-(\frac{\nu-\beta}{\psi^l})]^2}{(1-\frac{\beta}{\psi^h}-\frac{(\nu-\beta)}{\psi^l})^2(1-\frac{\beta}{y}-\frac{(\nu-\beta)}{\psi^l})^2} \geq 0 \qquad \square$$

## B.2 Proof of lemma 1

**To prove:** $\frac{\partial\Pi}{\partial\mu_f^h} \geq 0$.

**Proof:** Taking the first derivative of variable profits with respect to the markdown $\mu^h$ (analogously for $\mu^l$) gives:

$$\frac{\partial \Pi}{\partial \mu^h} = \frac{\partial Q}{\partial \mu^h}\Big(1 - \frac{\beta^h}{\mu^h} - \frac{\beta^l}{\mu^l}\Big) + Q\Big(\frac{\beta^h}{(\mu^h)^2}\Big)$$

$$= \frac{Q\beta^h}{(\mu^h)^2\psi^h}\Big(\psi^h - \mu^h\frac{(1-\frac{\beta^h}{\mu^h}-\frac{\beta^l}{\mu^l})}{(1-\frac{\beta^h}{\psi^h}-\frac{\beta^l}{\psi^l})}\Big)$$

This last expression is weakly positive because $\mu^h \leq \psi^h$. □

## B.3 Proof of lemma 2

**Proof:** Denote the variable profit margin as $m_f \equiv \Big(1 - \frac{\beta_f^h(1+\eta)}{\mu_f^h} - \frac{(\nu-\beta_f^h)(1+\eta)}{\mu_f^l}\Big)$

$$\frac{\partial \Pi_f}{\partial \Omega_f} = \frac{\partial Q_f}{\partial \Omega_f}\frac{\partial \Omega_f}{\partial K_f}m_f + \frac{\partial Q_f}{\partial \beta_f}\frac{\partial \beta_f}{\partial K_f}m_f + \frac{\partial m_f}{\partial \beta_f}Q_f + \underbrace{\frac{\partial m_f}{\partial \Omega}}_{=0}Q_f$$

Under the assumptions made, the variable profit margin $m_f$ is positive. It is easy to see that Hicks-neutral productivity increases output, $\frac{\partial Q_f}{\partial \Omega_f} > 0$. Hence, the higher the effect of the technology on Hicks-neutral productivity $\frac{\partial \Omega_f}{\partial K_f}$, the higher its effect on profits $\frac{\partial \Pi_f}{\partial \Omega_f}$. From Equation (3), it follows that a higher increase in profits due technology adoption also increases the effect of the markdown on this profit increase.

## B.4 Proof of lemma 3

**To prove:** $\frac{\partial^2(\Pi_f)}{\partial \mu_f^h \partial \Omega_f} \geq 0$

**Proof:**

$$\frac{\partial \Pi}{\Omega} = \frac{Q}{\Omega}(1 - \frac{\beta}{\mu} - \frac{\nu-\beta}{\psi^l})$$

$$\Rightarrow \frac{\partial}{\partial \mu^h}(\frac{\partial \Pi}{\partial \Omega}) = \frac{Q\beta}{\Omega \mu^2} + \frac{\partial Q}{\partial \mu}\frac{1}{\Omega}(1 - \frac{\beta}{x} - \frac{\nu-\beta}{\psi^l})$$

Working this out gives:

$$\frac{\partial}{\partial \mu^h}(\frac{\partial \Pi}{\partial \Omega}) = \frac{Q\beta}{\Omega\mu}\left(\frac{(\psi^h-\mu)(1-\frac{\nu-\beta}{\psi^l})}{\mu\psi^h(1-\frac{\beta}{\psi^h}-\frac{\nu-\beta}{\psi^l})}\right) \geq 0$$ □

The numerator of this expression is weakly positive, because $\mu \leq \psi^h$, $\beta \leq \nu$, and $\psi^h \geq 1, \psi^l \geq 1$. The denominator is weakly positive (strictly positive if $\nu < 1$ or $\psi^h > 0$ or $\psi^l > 0$), because of the same reason.

## B.5 Equilibrium expressions for empirical model

The equilibrium output of a mine $f$ at time $t$ is denoted $Q^*_{ft}$. It can be solved for by computing the first order conditions of the profit maximization problem, (8), and using Equations (4a), (5), and (6), which are respectively the production, coal demand, and labor supply functions. The resulting equilibrium output expression is in Equation (13a), which is the empirical analogue of Equation (2a) with Cournot competition upstream and downstream. When assuming that the firm is a monopolist and monopsonist (all market shares become one, and $f = m = n$), and there are no latent differences between coal and labor markets (no $\xi^l_{nt} = \xi^h_{nt} = \zeta_{mt} = 1$), Equation (13a) simplifies to Equation (2a).

$$Q^*_{ft} = \Bigg[ \Big( \frac{\beta^h_{ft} (s^h_{ft})^{\psi^h_{nt}-1} (1 + s^q_{ft}\eta) (\frac{1}{s^q_{ft}})^\eta \exp(\zeta_{ft})}{((\psi^h_{nt} - 1) s^h_{ft} + 1) \exp(\xi^h_{nt})} \Big)^{\frac{\beta^h_{ft}}{\psi^h_{nt}}} \Big( \frac{\beta^l_{ft} (s^l_{ft})^{\psi^l_{nt}-1} (1 + s^q_{ft}\eta) (\frac{1}{s^q_{ft}})^\eta \exp(\zeta_{ft})}{((\psi^l_{nt} - 1) s^l_{ft} + 1) \exp(\xi^l_{nt})} \Big)^{\frac{\beta^l_{ft}}{\psi^l_{nt}}} \Omega_{ft} \Bigg]^{\frac{1}{1 - \frac{\beta^h_{ft}(\eta+1)}{\psi^h_{nt}} - \frac{\beta^l_{ft}(\eta+1)}{\psi^l_{nt}}}} \tag{13a}$$

The equilibrium coal price is $P^*_{mt} = {Q^*_{mt}}^\eta \zeta_{mt}$. The equilibrium quantities of both labor types are then given by Equation (13b):

$$\begin{cases} H^*_{ft} & = \left( \frac{\beta^h_{ft} Q^*_{ft} P^*_{mt} (1 + s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^\eta}{((\psi^h_{mt} - 1) s^h_{ft} + 1) \xi^h_{mt}} \right)^{\frac{1}{\psi^h_{mt}}} \left( s^h_{ft} \right)^{\frac{\psi^h_{mt}-1}{\psi^h_{mt}}} \\ L^*_{ft} & = \left( \frac{\beta^l_{ft} Q^*_{ft} P^*_{mt} (1 + s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^\eta}{((\psi^l_{nt} - 1) s^l_{ft} + 1) \xi^l_{nt}} \right)^{\frac{1}{\psi^l_{nt}}} \left( s^l_{ft} \right)^{\frac{\psi^l_{nt}-1}{\psi^l_{nt}}} \end{cases} \tag{13b}$$

Substituting the equilibrium labor quantities from (13b) into the labor supply functions in (6) gives the expression for equilibrium wages, Equation (13c).

$$\begin{cases} W^{h*}_{mt} & = \left( \frac{\beta^h_{ft} P^*_{ft} Q^*_{ft} (1+\eta)}{((\psi^h_{mt} - 1) s^h_{ft} + 1) s^h_{ft}} \right)^{\frac{\psi^h_{mt}-1}{\psi^h_{mt}}} (\exp(\xi^h_{mt}))^{\frac{1}{\psi^h_{mt}}} \\ W^{l*}_{nt} & = \left( \frac{\beta^l_{ft} P^*_{ft} Q^*_{ft} (1+\eta)}{((\psi^l_{nt} - 1) s^l_{ft} + 1) s^l_{ft}} \right)^{\frac{\psi^l_{nt}-1}{\psi^l_{nt}}} (\exp(\xi^l_{nt}))^{\frac{1}{\psi^l_{nt}}} \end{cases} \tag{13c}$$

# C Empirical analysis

## C.1 Alternative production model

In the main text, I assumed that the scale parameter $\nu$ was equal to 0.9 and imposed a homogeneous goods Cournot model on the coal market to estimate markups. In this section, I use an alternative model in which I estimate the scale parameter and do not impose a demand model on the coal market, but which does not allow for unobserved heterogeneity in the output elasticities across firms and time.

**Production** In Equation (14), I impose a Cobb-Douglas production function in skilled and unskilled labor with each output elasticity shifting linearly with the usage of cutting machines, which is measured by the interaction effects $\beta^{hk}$ and $\beta^{lk}$.

$$q_{ft} = \beta^h h_{ft} + \beta^l l_{ft} + \beta^{hk} h_{ft} K_{ft} + \beta^{lk} l_{ft} K_{ft} + \beta^k K_{ft} + \omega_{ft} \tag{14}$$

I assume that cutting machines do not change the degree of returns to scale in both labor inputs, which implies that $\beta^{hk} = -\beta^{lk}$. As usual, I rely on timing assumptions on the input demand problem of the firm to identify the production function coefficients (Olley & Pakes, 1996; Ackerberg et al., 2015). As in Blundell and Bond (2000), I assume that total factor productivity evolves following an AR(1) process, with shock $\varepsilon_{ft}$:

$$\omega_{ft} = \rho\omega_{ft-1} + \varepsilon_{ft}$$

I keep the assumptions from the baseline model that both labor types are static, variable inputs, whereas cutting machines are a dynamic, fixed input. Hence, capital choices at time $t$ and any past time period are orthogonal to the productivity shock $\varepsilon_{ft}$ at time $t$, while labor choices at time $t-1$ are orthogonal to productivity shocks at time $t$:

$$\mathbb{E}\Big[\varepsilon_{ft} \Big| \begin{Bmatrix} h_{f\theta-1} \\ l_{f\theta-1} \\ K_{f\theta} \end{Bmatrix}_{\theta=1}^{t} \Big] = 0$$

The markup $\mu^q_{ft}$ can be expressed as the ratio of the output elasticity of miners over the product of its revenue share and markdown:[41]

$$\mu^q_{ft} = \frac{\beta^h + \beta^{hk} K_{ft}}{\frac{W^l_{ft} H_{ft}}{P_{ft} Q_{ft}} \psi^h_{ft}}$$

[41] Alternatively, the markup could be estimated using unskilled labor as well, but unskilled labor costs are latent.

**Estimation** As in the main text, I proceed by aggregating the data set to the town-year level, by summing output and labor and by defining the capital dummy at the town-year level. The estimable production function becomes Equation (15). I again denote towns as $f$, assuming that each town consists of either one firm or by multiple firms that are perfectly colluding:

$$q_{jt} = \beta^h h_{jt} + \beta^l l_{jt} + \beta^{hk}(h_{jt} - l_{jt})K_{jt} + \beta^k K_{jt} + \omega_{jt} \quad (15)$$

Similarly to the main text, I estimate the integrated model using a block-bootstrapping procedure that resamples within towns, using 200 iterations.

**Results** The results of this alternative production model are in Table A5. Coal cutting machines are still unskill-biased: the output elasticity miners is estimated to fall by 0.353 points when adopting a cutting machine, coming from 0.687. In the baseline model, this was a smaller drop of 0.110 points, down from 0.688. The scale parameter, $\nu$ is equal to be 0.768, whereas it was assumed to be 0.9 in the main text. Thus, the assumption of decreasing returns to scale is confirmed. The average markup ratio $\mu^q$ is estimated at 1.126, which implies that the coal price is 12.6% above marginal costs. This estiamate does not impose any model of competition on the coal market. In contrast, the homogeneous goods Cournot model in the baseline model delivered an average markup ratio of 1.067, or a markup of 6.7%, which is around half of the markup with the production model.

## C.2 Cost dynamics

There are multiple sources of cost dynamics that would invalidate the productivity transition equation. If it becomes increasingly costly to operate deeper mines, for instance, productivity would depend on past *cumulative* output, as Aguirregabiria and Luengo (2017) find for copper mining. Such dependence could also exist due to learning by doing, as in Benkard (2000), but productivity would then increase with cumulative output, rather than fall. I test this by regressing the logarithms of the productivity residual $\omega_{ft}$ on log cumulative output. The estimated coefficients in Table A2. If not including mine fixed effects, lagged cumulative output is associated with higher total factor productivity. This could, however, be due to selection: more productive mines are more likely to have extracted and sold more coal. Once I include mine fixed effects to track how productivity co-varies with cumulative output within each mine over time, the coefficient on lagged cumulative output becomes small and insignificant.

## C.3 Inverse miner supply elasticity: correlations

Appendix Table A3 regresses the log town-level inverse miner supply elasticity $\psi^h_{mt}$ on a number of town and county characteristics. The town-level inverse supply elasticity is equal to the profit-maximizing wage markdown of a monopsonist. Labor supply is more inelastic (implying higher wage markdowns) if the share of total coal employment over the town population is higher: in towns that are coal mining towns with few outside work opportunities, miner supply is more inelastic. A second regressor is the log of the ratio of the total farmed area in a county divided by the county's surface. Miner supply is more inelastic in areas with less farming (for instance, because of rugged geography), presumably because there are less outside work opportunities to switch to. Thirdly, the population share of African Americans in the county does not correlate significantly with the miner supply elasticity. Fourthly, towns with a higher share of firms connected to the railroad network have slightly more inelastic miner supply. This is in line with historical evidence that railroads were not used to transport workers, and confirms the assumption of isolated mining towns.[42] Finally, the average wage in manufacturing industries in the same county does not correlate significantly with the miner supply elasticity, which suggests that the outside option was mainly to work in agriculture, rather than in manufacturing industries, which were in any case scarce in rural Illinois.

## C.4 Robustness checks

### A Alternative values for the scale returns parameter

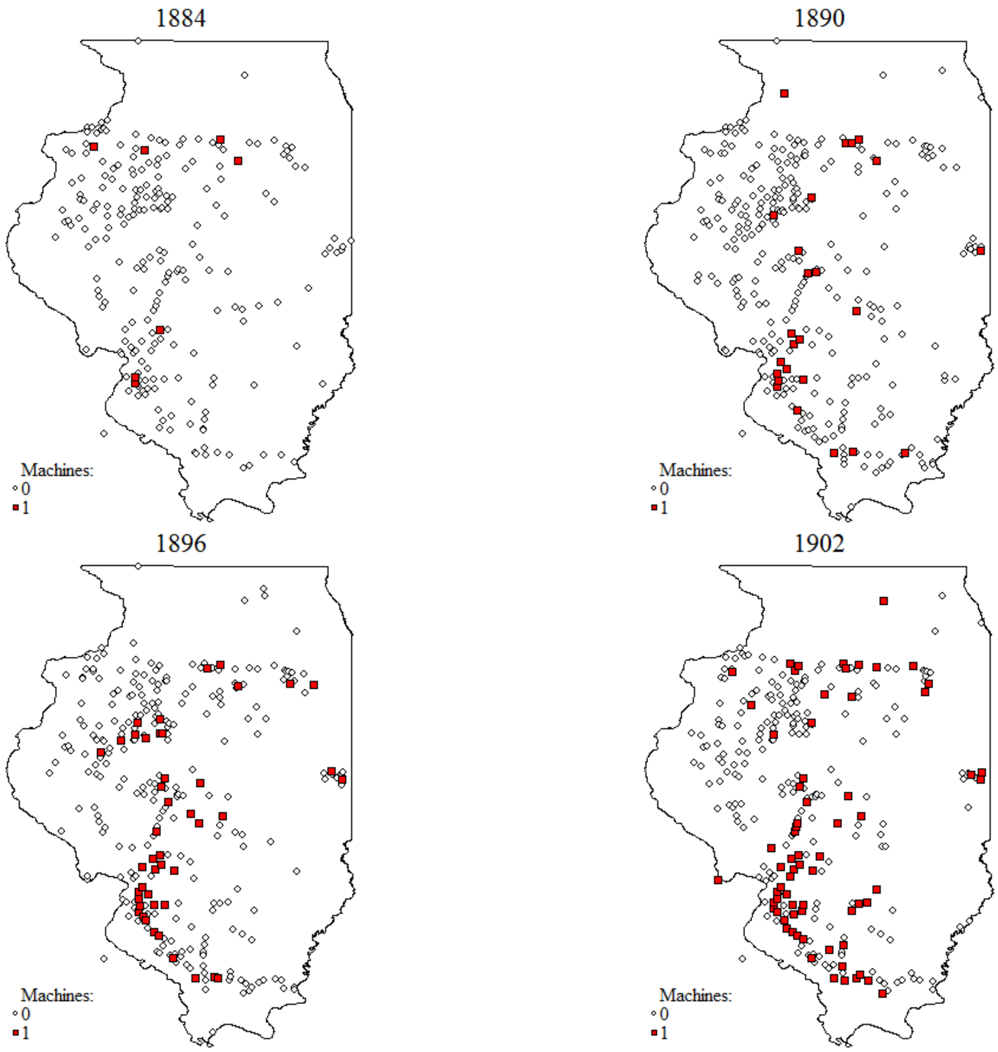
In the baseline analysis, I calibrated the degree of returns to scale $\nu$ to be 0.9, and motivated why decreasing returns to scale is an appropriate assumption in the historical coal mining industry setting. In this robustness check, I re-calibrate the returns parameter to be, alternatively, 0.85 and 0.95. The results are in Figure A5 and Figure A5 for $\nu = 0.85$ and $\nu = 0.95$, respectively. The direction of all counterfactual effects is the same as in the baseline analysis with $\nu = 0.90$. The size of the changes in technology usage in response to variation in labor market structure increases with the degree of returns to scale.

### B Wage screens

The usage of wage screens cause measurement error in the data, because miners were not remunerated on the total (reported) output at the mine, but based on a lower coal output that consisted only of larger pieces of coal. In 1898, I observe which firms paid screened wages and which did not. As a robustness check, I re-run the analysis while excluding counties for which at least one firm paid a screened wage in 1898, which was the case for 13.9% of observations. The estimated counterfactuals for this selected sample of firms is in Figure A7, and look very similar to those in the baseline analysis containing both firms that pay screened wages and those that do not.

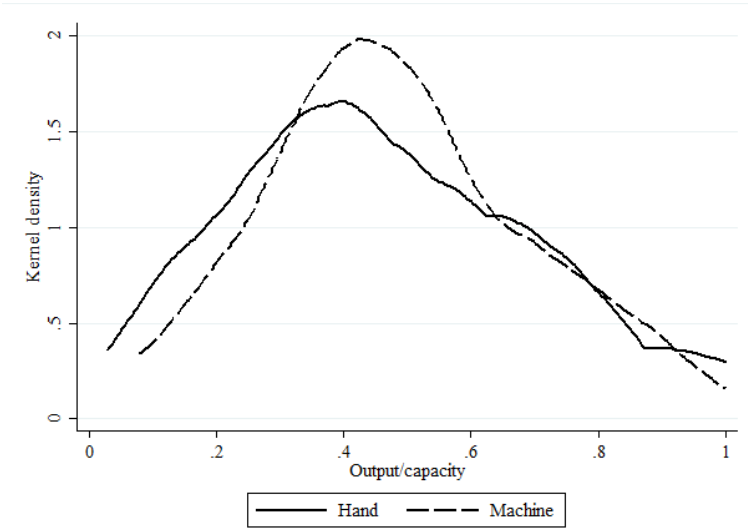
[42] If mining towns would not be isolated due to workers commuting by train, being connected to the railroad network should result in more elastic labor supply.

**Figure A1: Geographical spread of cutting machines**




**Notes:**The dots represent mining towns, each of which can contain multiple mines. Villages with squares contain at least one machine mine.

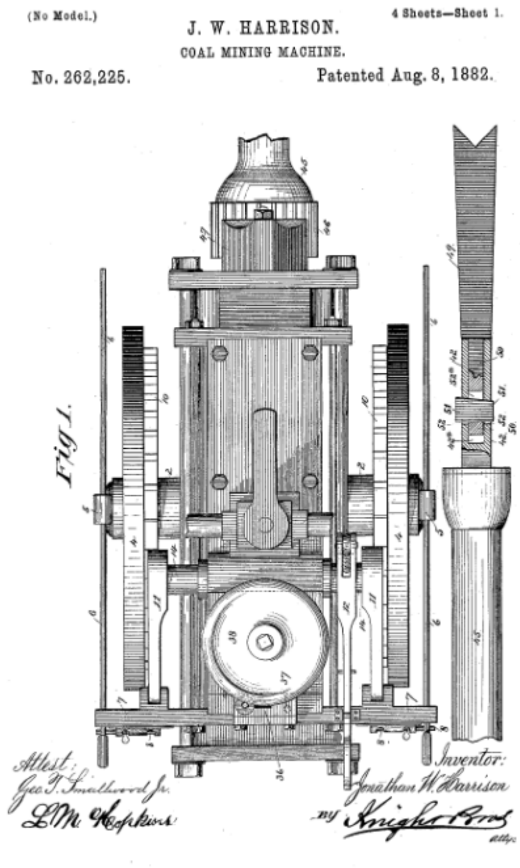
## Figure A2: Capacity utilization



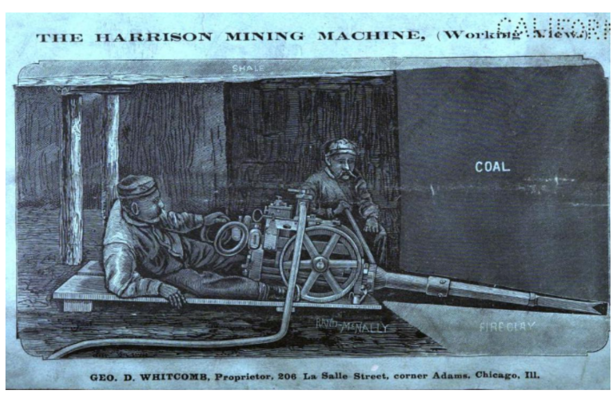

**Notes:** This graph plots the distribution of capacity utilization, defined as annual mine output over annual mine capacity, across mines in 1898. A distinction is made between hand mines, which did not use cutting machines, and machine mines, which did.

## Figure A3: Harrison Cutting Machine

### (a) Patent



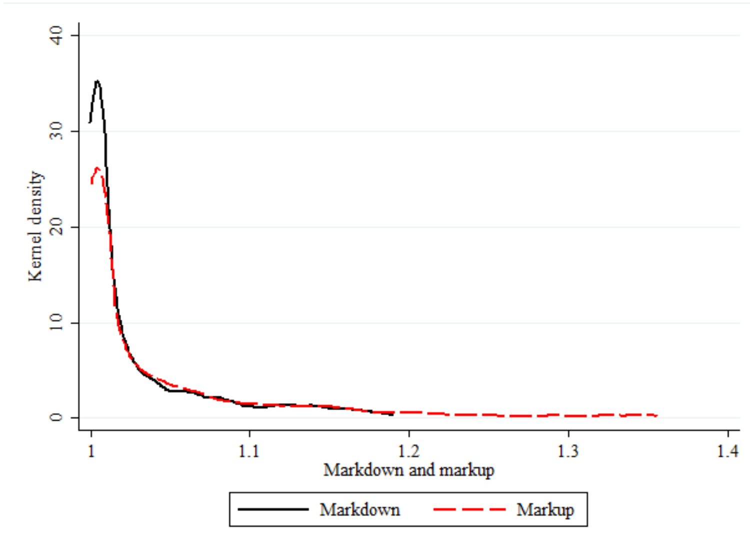

### (b) Illustration



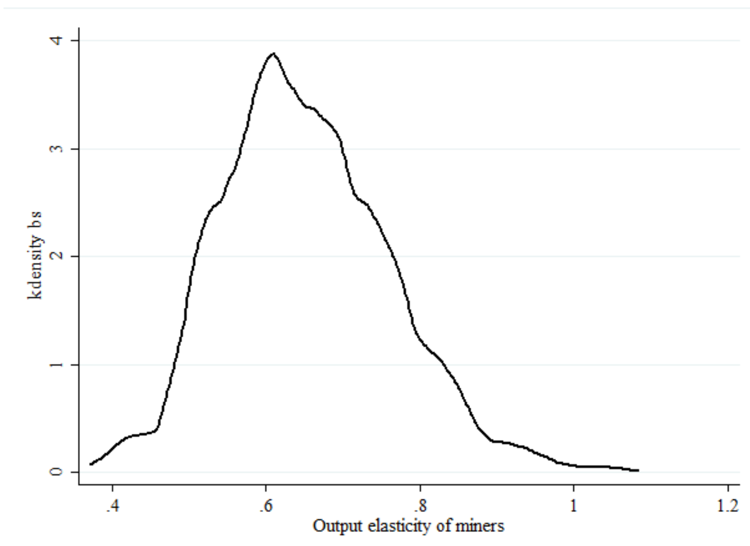

**Notes:** U.S.A. patent of the 1882 Improved Harrison Coal Cutting Machine (Whitcomb, 1882). This was the most frequently used coal cutting machine in the data set.

**Figure A4: Distributions of latent variables**

**(a) Markdowns and markups**



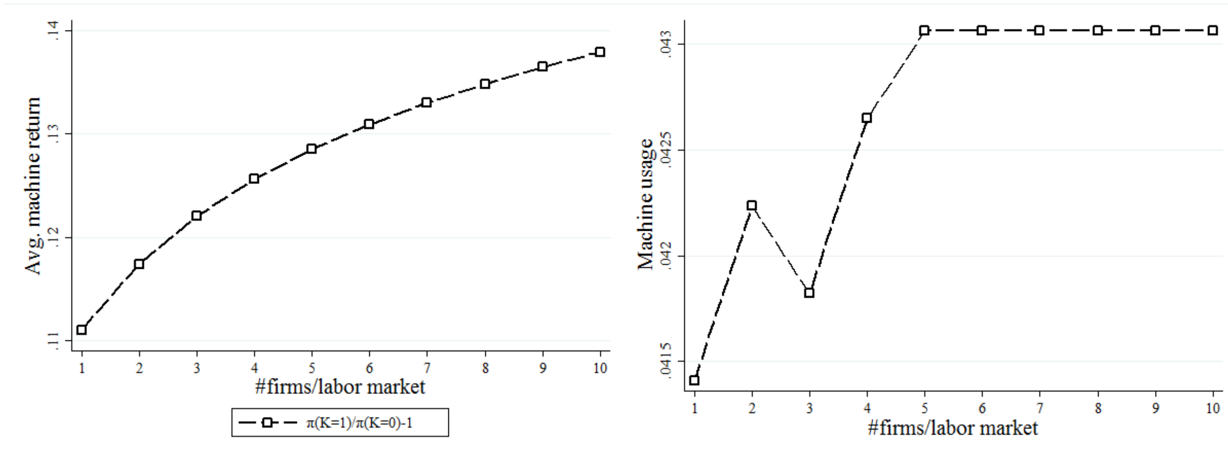

**(b) Skilled labor output elasticity**



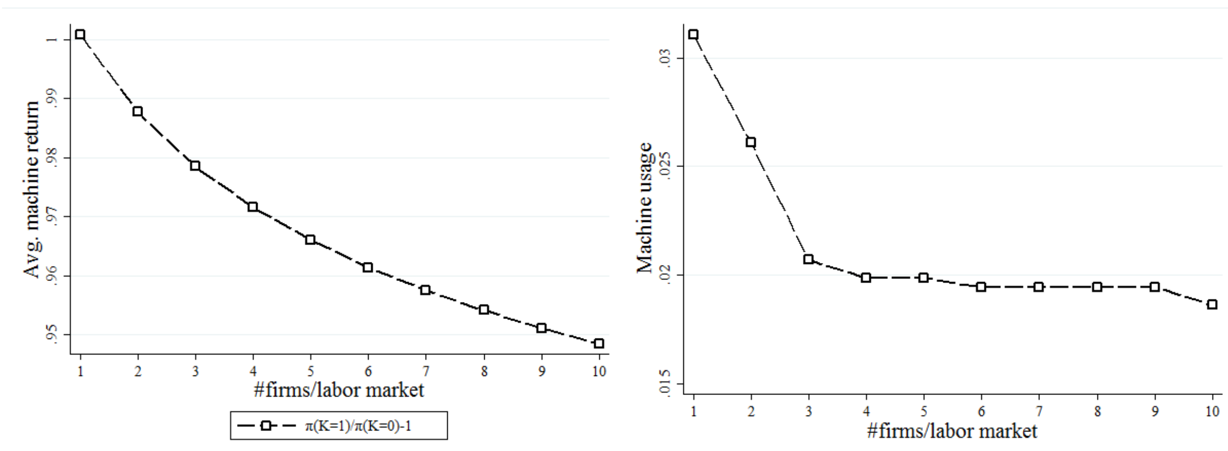

**Notes:** Distribution of the inverse miner supply elasticity across mines between 1884-1894. Each distribution censored at its 5th and 95th percentile.

**Figure A5: Counterfactuals with $\nu = 0.85$**

**(a) Unskill-biased technology**




**(b) Skill-biased technology**




**(c) Unskill-biased technology without Hicks-neutral productivity effect**

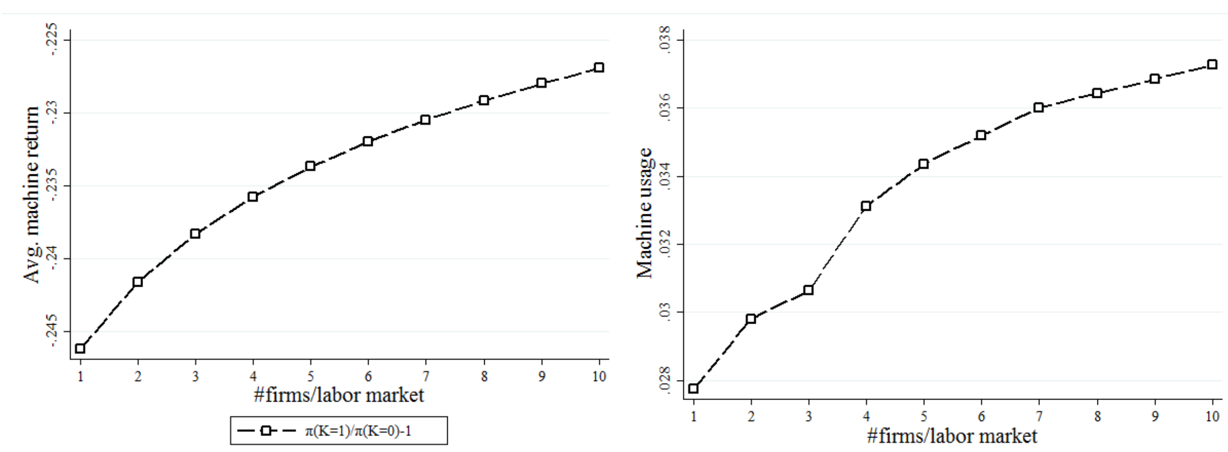
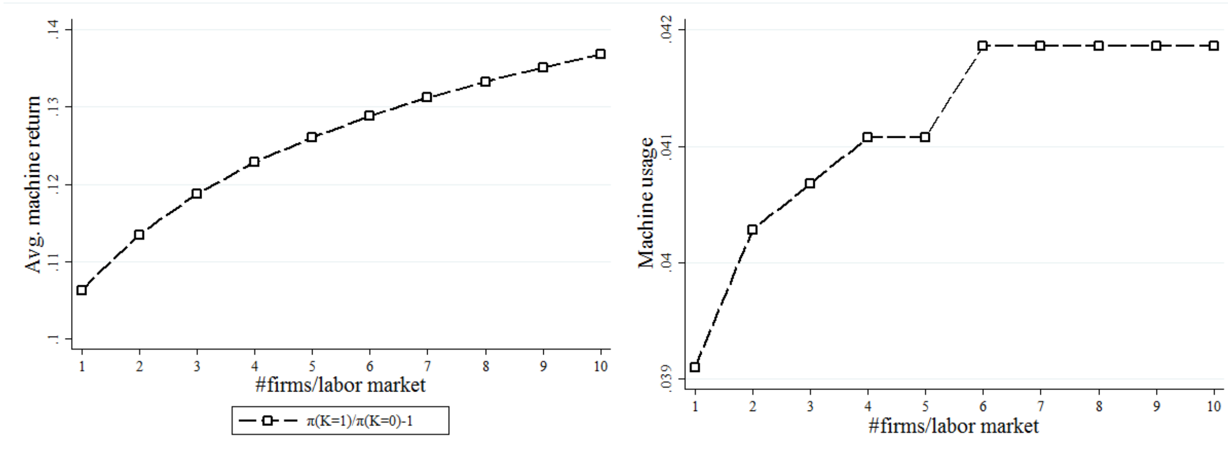
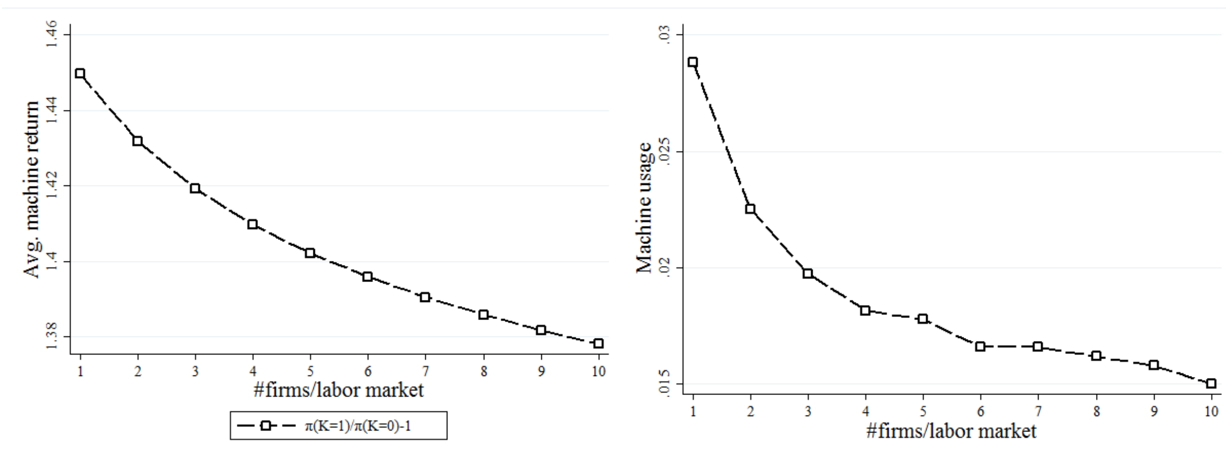


**Figure A6: Counterfactuals with $\nu = 0.95$**

**(a) Unskill-biased technology**




**(b) Skill-biased technology**




**(c) Unskill-biased technology without Hicks-neutral productivity effect**

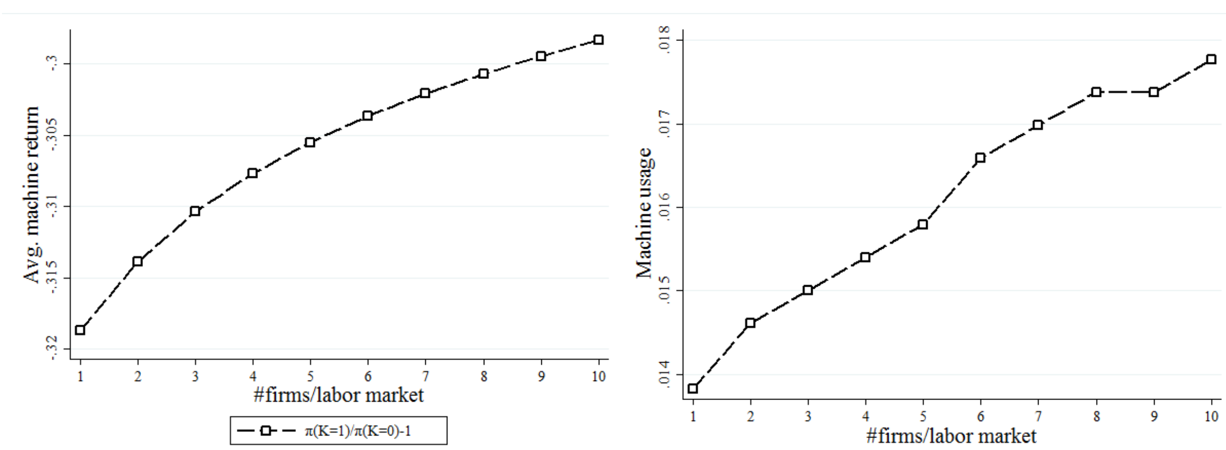
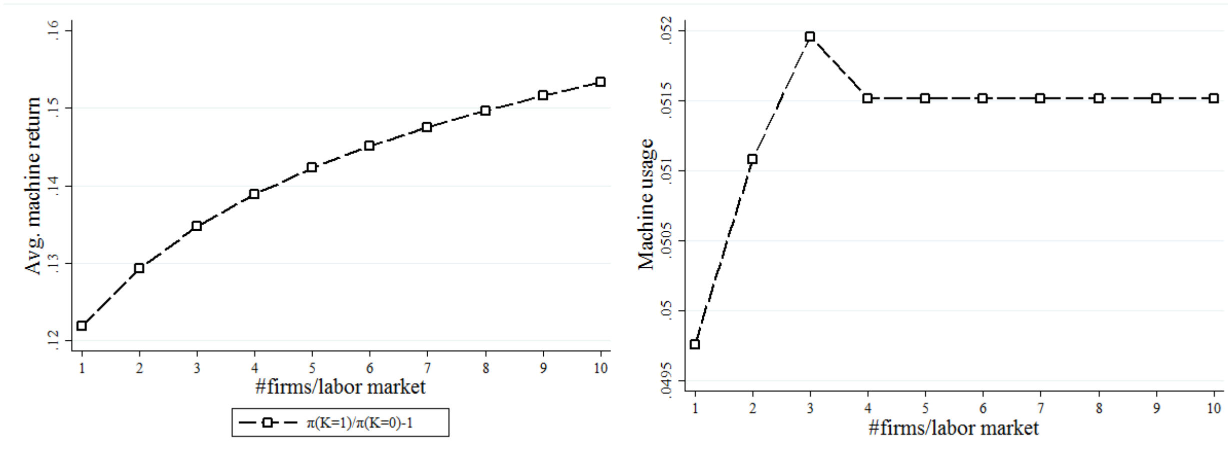
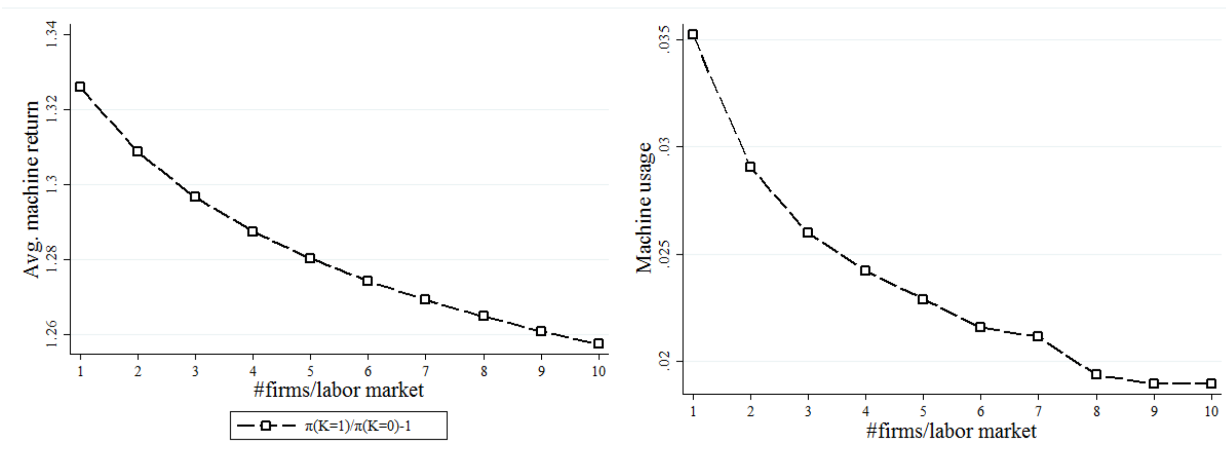


**Figure A7: Counterfactuals accounting for wage screens**

**(a) Unskill-biased technology**




**(b) Skill-biased technology**




**(c) Unskill-biased technology without Hicks-neutral productivity effect**

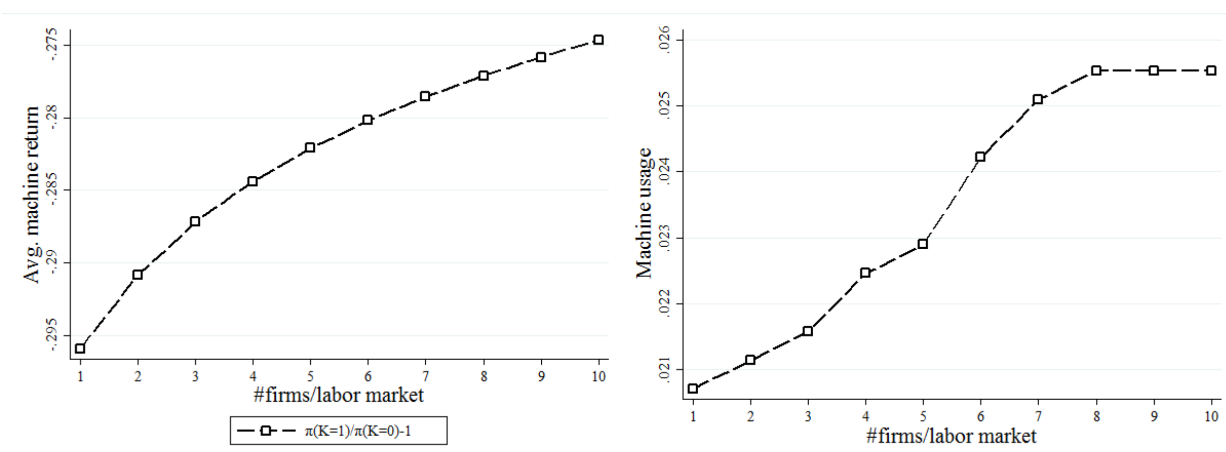


**Table A1: Occupations and wages**

| | Daily wage (USD) | Employment share (%) |
|---|---|---|
| Miner | 2.267 | 61.5 |
| Laborers | 1.76 | 14.30 |
| Drivers | 1.83 | 5.91 |
| Loaders | 1.74 | 3.63 |
| Trappers | 0.80 | 1.86 |
| Timbermen | 2.02 | 1.68 |
| Roadmen | 2.36 | 1.46 |
| Helpers | 1.70 | 0.92 |
| Brusher | 2.06 | 0.75 |
| Cagers | 1.87 | 0.70 |
| Engineer | 2.11 | 0.61 |
| Firemen | 1.60 | 0.57 |
| Entrymen | 2.01 | 0.56 |
| Pit boss | 2.70 | 0.56 |
| Carpenter | 2.09 | 0.53 |
| Blacksmith | 2.08 | 0.46 |
| Trimmers | 1.50 | 0.36 |
| Dumper | 1.68 | 0.36 |
| Mule tender | 1.65 | 0.31 |
| Weighmen | 1.95 | 0.29 |

**Notes:** Occupation-level data for the top-20 occupations by employment share in the 1890 sample of 11 mines in Illinois. The 20 occupations with highest employment shares together cover 97% of coal mining workers in the sample.

**Table A2: Cost dynamics**

| | log(Output/(labor-days)) | | | |
|---|---|---|---|---|
| | Estimate | SE | Estimate | SE |
| log(Cum. output) | 0.124 | 0.003 | -0.011 | 0.017 |
| Mine FE | No | | Yes | |
| Observations | 3766 | | 3766 | |
| R-squared | .326 | | .810 | |

**Notes:** Regression of log output per worker-day against log cumulative output (lagged by one time period) at the mine-year level. Sample only includes mines for which lagged output is observed.

**Table A3: Markdown correlations**

| | log(Markdown) | |
|---|---|---|
| | Estimate | SE |
| log(Coal employment share) | 0.022 | 0.004 |
| log(Farmland/Total Area) | -0.136 | 0.066 |
| log(African Americans / Population) | -0.001 | 0.004 |
| Share of firms connected to railroad | 0.029 | 0.010 |
| log(Manufacturing wage) | 0.021 | 0.022 |
| Observations | 876 | |
| R-squared | 0.342 | |

**Notes:** Regression of log miner wage markdown on mine and county characteristics. Standard errors clustered at the county level.

**Table A4: Coal demand and production estimates: all coefficients**

| *(a) Coal demand (county-level)* | log(Coal price) | | |
|---|---|---|---|
| | Est. | CI05 | CI95 |
| log(Quantity) | -0.263 | -0.307 | -0.233 |
| 1(Railroad connection) | 0.433 | 0.330 | 0.599 |
| 1(Railroad crossing) | 1.164 | 0.676 | 1.377 |
| log(Dist. to St. Louis) | -0.073 | -0.152 | -0.016 |
| log(Dist. to Chicago) | -0.287 | -0.447 | -0.144 |
| Observations | | 484 | |
| F-stat 1st stage | | 51.3 | |
| R-squared | | .202 | |
| *(b) Output elasticity transition* | log(Output elasticity of skilled miners) | | |
| | Est. | CI05 | CI95 |
| 1(Cutting machine) | -0.132 | -0.189 | -0.008 |
| log(Materials) | 0.012 | -0.022 | 0.006 |
| Year | -0.011 | -10.189 | 41.916 |
| Constant | 20.481 | -0.014 | 0.023 |
| Observations | | 1133 | |
| R-squared | | .006 | |
| *(c) Hicks-neutral productivity transition* | log(Hicks-neutral productivity) | | |
| | Est. | CI05 | CI95 |
| 1(Cutting machine) | 0.249 | -0.163 | 0.425 |
| log(Materials) | 0.124 | -0.270 | 0.142 |
| Year | -0.011 | | |
| Constant | -60.954 | -0.209 | 0.201 |
| Year | | -0.883 | 0.466 |
| Observations | | 1050 | |
| R-squared | | .238 | |

**Notes:**

## Table A5: Alternative production model

| | | | |
|---|---|---|---|
| *(a) Production function* | | | |
| log(Skilled labor) | 0.687 | 0.574 | 1.509 |
| log(Skilled labor/Unskilled labor)*1(Cutting machine) | -0.353 | -0.798 | -0.240 |
| log(Unskilled labor) | 0.081 | -0.433 | 0.430 |
| 1(Cutting machine) | 0.551 | 0.433 | 1.088 |
| Constant | 1.950 | -2.834 | 7.283 |
| Observations | | | |
| R-squared | | | |
| *(b) Markup and returns to scale* | | | |
| Returns to scale | 0.768 | 0.577 | 1.424 |
| Markup | 1.126 | 0.946 | 2.487 |

**Notes:** Alternative production function that estimates markup and degrees to scale, as specified in Appendix C.1. Standard errors are block-bootstrapped with 200 iterations.

**Table A6: Counterfactual with different returns to scale parameter $\nu$**

$\nu = 0.85$

| *(a) Competitive equilibrium* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
|---|---|---|---|---|---|
| Returns to machine adoption | 0.009 | 0.024 | 0.015 | 0.007 | 0.027 |
| Cutting machine usage | 0.051 | 0.057 | 0.006 | -0.003 | 0.027 |
| *(b) Elastic labor supply* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
| Returns to machine adoption | 0.009 | -0.012 | -0.021 | -0.041 | -0.009 |
| Cutting machine usage | 0.051 | 0.054 | 0.002 | -0.019 | 0.010 |

$\nu = 0.95$

| *(a) Competitive equilibrium* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
|---|---|---|---|---|---|
| Returns to machine adoption | 0.000 | 0.015 | 0.015 | 0.006 | 0.027 |
| Cutting machine usage | 0.051 | 0.056 | 0.004 | -0.003 | 0.034 |
| *(b) Elastic labor supply* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
| Returns to machine adoption | 0.000 | -0.020 | -0.021 | -0.040 | -0.009 |
| Cutting machine usage | 0.051 | 0.050 | -0.001 | -0.024 | 0.010 |

**Notes:** Same counterfactual exercise as in Table 2 but with scale return parameters $\nu = 0.85$ and $\nu = 0.95$.

**Table A7: Counterfactual dropping counties with screened wages.**

| *(a) Competitive equilibrium* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
|---|---|---|---|---|---|
| Returns to machine adoption | -0.003 | 0.004 | 0.007 | 0.007 | 0.027 |
| Cutting machine usage | 0.040 | 0.050 | 0.011 | -0.003 | 0.024 |
| *(b) Elastic labor supply* | Reality | CF | Dif. | Dif. (5%) | Dif (95%) |
| Returns to machine adoption | -0.003 | -0.021 | -0.018 | -0.041 | -0.009 |
| Cutting machine usage | 0.040 | 0.032 | -0.007 | -0.025 | 0.008 |

**Notes:** Same counterfactual exercise as in Table 2 but dropping counties in which 'wage screens' were used.

**Table A8: All variables per year**

| Year | 1884 | ’86 | ’88 | ’90 | ’92 | ’94 | ’96 | ’98 | ’00 | ’02 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Output quantities** | | | | | | | | | | |
| Total | X | X | X | X | X | X | X | X | X | X |
| Lump | | | | | X | X | X | X | X | X |
| Mine run | | | | | | | | | X | X |
| Egg | | | | | | | | | X | X |
| Pea | | | | | | | | | X | X |
| Slack | | | | | | | | | X | X |
| Shipping or local mine | | | | | X | X | X | | | |
| Shipping quantities | | | | | | | | | | X |
| **Input quantities** | | | | | | | | | | |
| Miners, winter | X | X | X | X | | | | | | |
| Miners, summer | X | X | X | X | | | | | | |
| Miners, avg entire year | | | | | X | X | | X | X | X |
| Miners, max entire year | | | | | X | X | | | | |
| Other employees | X | X | X | X | X | X | | X | X | X |
| Other employees, underground | | | | | | | | X | | |
| Other employees, above ground | | | | | | | | X | | |
| Other employees winter | | | | | | | X | | | |
| Other employees summer | | | | | | | X | | | |
| Boys employed underground | | | X | X | X | X | X | | | |
| Mules | | X | | | | | | | | |
| Days worked | X | X | X | X | X | X | | X | X | X |
| Kegs powder | X | X | X | X | X | X | | X | | X |
| Men killed | X | X | X | X | X | X | | X | | X |
| Men injured | X | X | X | X | X | X | | X | | X |
| Capital (in dollar) | X | | | | | | | | | |

**Table A9: All variables per year (cont.)**

| Year | 1884 | '86 | '88 | '90 | '92 | '94 | '96 | '98 | '00 | '02 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Output price** | | | | | | | | | | |
| Price/ton at mine | X | X | X | X | X | | | X | X | X |
| Price/ton at mine, lump | | | | | | X | X | X | | |
| **Input prices** | | | | | | | | | | |
| Miner piece rate (summer) | X | X | X | X | X | X | | | | |
| Miner piece rate (winter) | X | X | X | X | X | X | | | | |
| Miner piece rate (hand) | | | | | | | | X | X | X |
| Miner piece rate (machines) | | | | | | | | X | X | |
| Piece rate dummy | | | | | X | | | | | |
| Payment frequency | | | | | | X | X | X | X | |
| Net/gross wage | | | | | | | X | | | |
| Oil price | | | | | | | X | | | |
| **Technicals** | | | | | | | | | | |
| Type (drift, shaft, slope) | X | X | | | X | X | X | X | | |
| Hauling technology | X | X | | | X | X | | X | | |
| Depth | X | X | | | X | X | X | X | X | |
| Thickness | X | X | | | X | X | X | X | X | |
| Geological seam type | X | X | | | X | X | | X | | |
| Longwall or PR method | X | X | | | X | X | X | | X | |
| Number egress places | X | X | | | | | | | | |
| Ventilation type | X | X | | | | | | | | |
| New/old mine | | | | | X | X | | | | |
| # Acres | | | | | X | X | X | | | |
| Mine capacity | | | | | | | | X | | |
| Mined or blasted | | | | | | | | X | | |
| **Cutting machine usage** | | | | | | | | | | |
| Cutting machine dummy | | | | | X | X | X | X | | |
| # Cutting machines | X | X | X | X | | | | | | |
| # Tons cut by machines | | | | | | | | | X | X |
| # Cutting machines, by type | | | X | | | | | | | |