Oligopsony Power and Factor-Biased Technology Adoption

Michael Rubens*

January 19, 2023

Abstract

I show that buyer power of firms could either increase or decrease their technology adoption, depending on the direction of technical change and on which input markets are imperfectly competitive. I examine this relationship empirically in a setting that features both concentrated labor markets and a large technology shock: the introduction of mechanical cutters in the 19th century Illinois coal mining industry. Using a model of production and labor supply which is estimated with mine-level data, I find that oligopsony power over skilled miners reduced the usage of cutting machines, an unskill-biased technology. However, it would have increased the usage of counterfactual skill-biased and Hicks-neutral technologies.

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

JEL codes: L11, L13, J42, N51

*UCLA, Department of Economics. Contact: rubens@econ.ucla.edu.

This paper previously circulated under the title 'Monopsony Power and Factor-Biased Technology Adoption'. I thank Jan De Loecker, Frank Verboven, Jo Van Biesebroeck, Chad Syverson, Otto Toivanen, Mert Demirer, Scott Stern, Penny Goldberg, Steve Berry, Suresh Naidu, Cuicui Chen, and Daniel Haanwinckel for feedback and discussion, and participants at various seminars and conferences. This paper is based on 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 buyer power across various industries, countries, and types of factor markets.[1] When studying the welfare consequences of such buyer power, prior research has typically assumed that buyer 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 buyer 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.

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 sets the price of each input at a markdown below its marginal revenue product, in function of the input supply elasticity and of the degree of competition on each market. I examine how these input price markdowns affect the adoption of a new technology, which can have both a factor-biased effect, rotating the production isoquant, and a Hicks-neutral effect, shifting the isoquant. I show that the effect of markdowns on technology usage is ambiguous, except in a limited number of special cases, such as for technologies that are biased away from perfectly elastic or monopsonistic inputs.

In order to study the effects of buyer power on technology adoption outside of these special cases, I turn to an empirical application. I study how the mechanization of the Illinois coal mining industry between 1884 and 1902 was affected by market power held by firms over their miners. There are three reasons why this provides a unique setting to study the relationship between buyer power and innovation. First, 19th century Illinois coal mining towns are a textbook example of oligopsonistic markets, as local labor markets were geographically isolated and highly concentrated. Up to 1898, wages were set unilaterally by firms, without bargaining with trade unions. Following large strikes in 1897-1898, the wage-setting process changed to wage bargaining between employers and trade union representatives, which provides a within-sample shock to employers' buyer power. Second, the introduction of coal cutting machines in the U.S. in 1882 provides a large observed technological shock. The data set tracks the usage of these cutting machines and other

[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. (2022); Rubens (2020b); Morlacco (2017); Lamadon et al. (2022); Kroft et al. (2020).

[2] Examples include, among many others, Schumpeter (1942), Aghion, Bloom, Blundell, Griffith, and Howitt (2005) and Igami and Uetake (2020).

technologies over time, together with input and output quantities, wages and coal prices, all at the mine level. Third, bituminous coal firms are single-product firms producing a nearly homogeneous product, which facilitates the empirical analysis.

The key question of the empirical application is how the degree of labor market competition affected cutting machine adoption, and how this would have changed if cutting machines would have had different directed effects. I start with two pieces of descriptive evidence to motivate this analysis. First, firms in concentrated markets adopted less cutting machines, a technology that saved primarily on skilled workers, but more underground mining locomotives, a technology that saved on unskilled workers. Second, the introduction of wage bargaining in 1898 led to increased cutting machine adoption in markets where workers bargained succesfully for higher wages. Taken together, these facts suggest that market power over miners decreased cutting machine usage, but increased locomotive usage.

In order to explain these descriptive facts and replicate them 'out-of-sample', I construct an empirical model of input supply and demand in the coal mining industry. The model has three components. First, I specify a production function for coal with three factors: skilled miners who cut coal, unskilled other workers who did a variety of tasks such as driving mules and sorting coal, and capital, in the form of cutting machines and locomotives. I rely on a Cobb-Douglas production function in both labor types, but with output elasticities that are a function of cutting machine and locomotive usage, and that vary flexibly across firms and over time.[3] 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 (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 that their output is undifferentiated conditional on their location. Third, I specify a model of oligopsonistic competition on each labor market, as these were concentrated: the median mining town contained merely 2 coal firms. For each labor type, I use a log-linear supply curve of which the elasticity varies flexibly across labor markets and over time. I assume that firms are homogeneous from the employees' point of view and abstract from search and adjustment frictions, which implies a static Cournot employment-setting game played by the employers.

I estimate the production model with mine-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, whereas locomotives were

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

skill-biased, in line with the anecdotal evidence. Both technologies increased Hicks-neutral productivity.[4] The coal demand model is identified by exploiting geological variation in coal vein thickness, a cost shifter that does not affect consumer utility. Finally, the labor supply model is estimated using labor-market level data on wages and employment, and is identified by comparing the wage responses to seasonal weather variation across labor types. The labor supply estimates reveal a moderate degree of oligopsony power over skilled workers, but no oligopsony power over unskilled workers.[5]

I combine the estimated labor supply and demand model to compute joint labor and product market equilibrium under the observed labor market structure. Then, using the estimated model, I compute how changes in labor market structure would affect technology adoption. I carry out this exercise for the two observed technologies, cutting machines and mining locomotives, and for two counterfactual technologies: one technology that has identical effects as cutting machines, except for an opposite direction, and a second technology that is Hicks-neutral. I find that moving from one to ten employers per labor market would increase cutting machine usage by 18%, but would decrease mining locomotive usage by 10%. This finding is in line with the descriptive evidence provided at the start of the paper. Finally, the usage of counterfactual cutting machine technologies with the same Hicks-neutral effect and fixed cost, but that would have been skill-biased or Hicks-neutral, rather than unskill-biased, would have decreased, rather than increased, with additional labor market competition.

Although the empirical setting of the paper is historical, these findings have important current-day implications. The model shows that in order to know whether market power held by employers over their workers increases or decreases automation incentives today, it is crucial to know two things: first the direction of technological change, and second, the different levels of wage markdowns across the worker skill distribution. While there is much evidence on the direction of technological change,[6] the relative degree of market power across the skill distribution is mostly still an open empirical question. The labor literature has mainly focused on monopsony and oligopsony power over low-skilled workers, such as Card and Krueger (1994), for instance due to a lack of outside options of workers

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

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

[6]See, among many others, Autor et al. (2006); Machin and Van Reenen (1998); Goos et al. (2014); Katz and Margo (2014).

(Caldwell & Danieli, 2022; Schubert, Stansbury, & Taska, 2020). Non-compete clauses are, however, most frequent among high-skilled jobs in the U.S (Starr, Prescott, & Bishara, 2021). Also, Prager and Schmitt (2021) find that mergers only affect wages of workers with industry-specific skills. 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 increase energy efficiency.

This paper makes three main contributions. First, it contributes to the literature on competition and innovation (Schumpeter, 1942; Aghion et al., 2005; Collard-Wexler & De Loecker, 2015; Bloom, Schankerman, & Van Reenen, 2013; Igami & Uetake, 2020) by studying the effect of *factor* market power on innovation, rather than *product* market power. In a closely related paper, Goolsbee and Syverson (2019) find that monopsony power over tenure-track faculty induces universities to substitute these workers for adjunct faculty members. In contrast, I endogenize the choice of the production technology: buyer power does not just let firms move along the input demand curves, but also leads to changes of the input demand curves, due to different technology choices. Méndez-Chacón and Van Patten (2022) examine investment by a historical monopsonist, the United Fruit Company. This paper differs by offering a different mechanism for why buyer power is related to investment, by showing that buyer power could either increase or decrease investment depending on the direction of technological change, and by presenting a model of oligopsony, rather than pure monopsony. A number of theoretical papers study innovation and investment by firms with buyer power, such as Inderst and Wey (2003) and Loertscher and Marx (2022), which differ from this paper by relying on a bargaining setting with imperfect information, whereas I study unilateral oligopsony power in a perfect information setting, but allowing for directed techological change. Finally, there is a literature on the effects of buyer power on technology choices of their suppliers (Just & Chern, 1980; Huang & Sexton, 1996; Köhler & Rammer, 2012; Parra & Marshall, 2021), whereas I focus on technology adoption of the buyers themselves.

Second, this paper contributes to the literature on directed technological change and factor bias. The seminal models of directed technical change, such as Autor et al. (2003); Acemoglu (2002, 2003) and Antras (2004), assume that input markets are perfectly competitive. Contemporaneous work by Haanwinckel (2018) and Lindner et al. (2019) examines the effects of skill-biased technologies on skill demand and wage inequality with imperfectly competitive labor markets. This paper contributes to this literature by showing that factor-biased technology choices are themselves endogenous to the degree of buyer power. This also relates to the 'induced innovation' hypothesis of Hicks (1932), which posits that labor-saving technological change is more likely if wages are high, because cost savings are then higher as well.[7] This hypothesis has been empirically studied in a variety of settings, including Hanlon (2015) and Dechezleprêtre, Hémous, Olsen, and Zanella (2019). As noted in Acemoglu (2002), a critique of the induced innovation hypothesis is that the notion 'expensive input' is inconsistent with settings in which factor prices are equal to marginal products (Salter, 1966). In contrast, I do allow for a wedge between factor prices and marginal products.

Third, I contribute to the literature on monopsony and oligopsony power. An increasing volume of research studies the distributional and efficiency effects of labor market power (Manning, 2013; Berger et al., 2022; Lamadon et al., 2022) but this is done keeping technology choices fixed. In contrast, I show that endogenous technology choices present an additional channel through which input market power shapes aggregate outcomes. Finally, by studying labor market power in a historical setting, this paper is also related to a body of work on labor market power during the late 19th century, such as Boal (1995), Naidu and Yuchtman (2017), and Delabastita and Rubens (2022).

The remainder of this paper is structured as follows. Section 2 contains the theoretical model. Section 3 presents the industry background and stylized facts. Section 4 discusses the empirical model, results, and counterfactuals. Section 5 concludes.

2 Theory

2.1 Environment

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 (1). The output elasticity of input $V \in \{H, L\}$ at firm f is denoted β^v_f. Scale returns are parametrized as $\nu = \beta^h_f + \beta^l_f$, which is below, above or equal to one if returns to scale are decreasing, increasing, or constant. Total factor productivity is denoted Ω_f. Firms choose whether to use a technology $K_f \in \{0, 1\}$ or not, with the

[7] This theory has been forwarded as a reason why Britain was the first country to experience an industrial revolution (Allen, 2009).This hypothesis that has in turn been criticized by, among others, Humphries (2013).

technology $K_f = 1$ having a common fixed cost Φ.

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

Technology usage can affect both the output elasticities and the Hicks-neutral productivity residual. I call the technology K 'H-biased' if $\frac{\partial \beta_f^h}{\partial K_f} > 0$, because K then increases the marginal rate of technical substitution of H for L, keeping factor proportions constant.[8] Conversely, K is termed 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 scale parameter ν.

Using a Cobb-Douglas production function with technology-specific output elasticities is a first-order approximation of the canonical models on technical change, which usually rely on a constant elasticity of substitution (CES) production function. Although imposing a unitary elasticity of substitution between different types of workers is 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 assumption is that it permits analytical expressions for joint input and output market equilibrium in the presence of both oligopsonistic and oligopolistic competition.

B Markets

A firm f pays its input suppliers prices W_f^h and W_f^l, and cannot price discriminate between different suppliers of the same input. The firm faces the input supply functions in Equation (2), with inverse supply elasticity $(\psi^h - 1)$ for input H and $(\psi^l - 1)$ for input L. I assume that the supply functions are weakly upward-sloping, $\psi^h \geq 1$ and $\psi^l \geq 1$.

$$\begin{cases} W_f^h = {H_f}^{\psi^h - 1} \\ W_f^l = {L_f}^{\psi^l - 1} \end{cases} \tag{2}$$

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 η, in Equation (3). I assume that the demand

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

curve is either horizontal or downward-sloping, which implies that $\eta \leq 0$.

$$P_f = Q_f^{\eta} \tag{3}$$

2.2 Behavior and equilibrium

A Behavior

Variable profits are defined as $\Pi_f \equiv P_f Q_f - W_f^h H_f - W_n^l L_f$, whereas 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 current variable profits, taking the technology K as given. Depending on how competitive the input markets are, the firm sets the price of each input at a markdown below its marginal revenue product, as parametrized by the markdowns $\mu_f^h \in [1, \psi^h]$ and $\mu_f^l \in [1, \psi^l]$. If the firm is a monopsonist on the market for H, the profit-maximizing markdown is equal to the inverse supply elasticity, $\mu_f^h = \psi^h$, and similarly for the other input.

$$\max_{H_f, L_f} (P_f Q_f - W_f^h H_f - W_f^l L_f)$$

Solving the first order conditions of this optimization problem results in the input demand functions in Equation (4):

$$\begin{cases} H_f &= \frac{P_f Q_f \beta_f^h (1+\eta)}{W_f^h \mu_f^h} \\ L_f &= \frac{P_f Q_f \beta_f^l (1+\eta)}{W_f^l \mu_f^l} \end{cases} \tag{4}$$

Denote the marginal product of input suppliers H as $MR_f^h \equiv \frac{\partial(P_f Q_f)}{\partial H_f} = \beta_f^h P_f Q_f (1+\eta)$. From Equation (4), one can see that the markdown parameters μ^h and μ^l are equal to the ratio of the marginal product of an input over its price: $\mu_f^h = \frac{MR_f^h}{W_f^h}$ and $\mu_f^l = \frac{MR_f^l}{W_f^l}$.

B Equilibrium

Solving Equations (1)-(4) yields an equilibrium expression for output in Equation (5a), at which both the goods and input markets are in equilibrium.

$$Q_f = \left[\left(\frac{\beta_f^h (1+\eta)}{\mu_f^h} \right)^{\frac{\beta_f^h}{\psi_f^h}} \left(\frac{\beta_f^l (1+\eta)}{\mu_f^l} \right)^{\frac{\beta_f^l}{\psi_f^l}} \Omega_f \right]^{\frac{1}{1 - \frac{\beta_f^h (1+\eta)}{\psi_f^h} - \frac{\beta_f^l (1+\eta)}{\psi_f^l}}} \tag{5a}$$

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 are equal to the product of equilibrium revenues and a variable profit margin.

$$\Pi_f = \underbrace{{Q_f}^{(1+\eta)}}_{\text{revenue}} \underbrace{\left(1 - \frac{\beta_f^h(1+\eta)}{\mu_f^h} - \frac{(\nu - \beta_f^h)(1+\eta)}{\mu_f^l}\right)}_{\text{variable profit margin}} \tag{5b}$$

2.3 The returns to technology adoption

With these equilibrium expressions at hand, I now consider how the effect of technology usage K on variable profits Π depends on the level of competition on each input market.

A Relative profit return

I start by examining the relative profit return to technology adoption, $\frac{\Pi(K=1)-\Pi(K=0)}{\Pi(K=0)}$. This relative profit change is approximated by the derivative of log variable profits with respect to technology usage, $\frac{\partial \ln(\Pi)}{\partial K}$. Lemma 1 states that if one input market is monopsonistic, the input price markdown over the other input weakly increases the relative variable profit return to a technology if it is biased towards the non-monopsonistic input, but weakly decreases the returns to a technology if it is biased towards the monopsonistic input.

Lemma 1 *Consider a market equilibrium characterized by Equations (1)-(4), with one input market being monopsonistic. Then, the input price markdown of the other input weakly increases the relative variable profit return to a technology if it is biased towards that other input, but weakly decreases the returns to a technology if it is biased towards the monopsonistic input.* ***Proof:*** *see Appendix A-A.*

$$\textit{For } \mu_f^l = \psi_f^l : \qquad \frac{\partial \beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\ln(\Pi_f))}{\partial \mu_f^h \partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$$

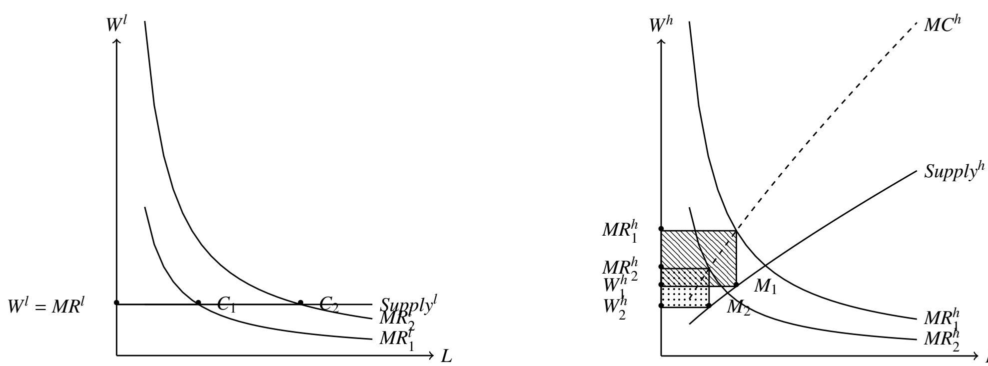
Theorem 1 generalizes the result from Lemma 1 any market structure, as long as the supply of one of the inputs is perfectly elastic.

Theorem 1 *Consider a market equilibrium characterized by Equations (1)-(4), with the supply of one input being perfectly elastic. The input price markdown of the inelastic input weakly increases the relative variable profit return to a technology if it is biased towards* *the inelastic input, but weakly decreases the returns to a technology if it is biased towards the elastic input.* ***Proof:*** *see Appendix A-B.*

$$\text{For } \psi_f^l = 1: \qquad \frac{\partial \beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \ \Rightarrow \frac{\partial^2(\ln(\Pi_f))}{\partial \mu_f^h \partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$$

The intuition behind Theorem 1 becomes clear from Figure 1. Given that the supply of L is perfectly elastic, the firm extracts no profit from input L: the price of L, W^L is equal to its marginal revenue product. In contrast, a markdown is charged to the inelastically supplied input H: the price of H lies below its marginal revenue product. Adopting an L-biased technology leads to lower usage of H, but to higher usage of L: the equilibria move from point C_1 to C_2 on the market for L, and from M_1 to M_2 on the market for H. The technology hence decreases variable profits, as it shifts input usage from the input from which the firm extracts rents, H, towards the input from which it extracts zero rents, L. Hence, the firm wants to use the technology that shifts input usage towards the input from which it extracts a markdown, which is H.

Figure 1: Input price markdowns and technology choices



More in general, it could be that both inputs are supplied inelastically and both are not sold on a monopsonistic market. Proposition 1 states that in that case, the effect of markdowns on the relative profit change from a technology is ambiguous.

Proposition 1 *If both input markets are oligopsonistic and supplied inelastically, the price markdown of an input could either increase or decrease the relative variable profit return* *to a technology that is biased towards that input.* ***Proof:*** *see Appendix A-C.*

$$\text{For } \mu_f^l < \psi_f^l \text{ and } \psi_f^l \neq 1: \qquad \frac{\partial \beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\ln(\Pi_f))}{\partial \mu_f^h \partial K_f} \Big\{ \geq \; or \; \leq \Big\} 0$$

B Absolute profit return

Theorem 1 and Lemma 1 provided results on 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 effects of markdowns on the *absolute* change in variable profits after machine adoption, $\Pi(K = 1) - \Pi(K = 0)$, and compare it against the costs of machine usage, Φ. Proposition 2 states that the results from Theorem 1 and Lemma 1 generalize to the absolute profit effect of a technology that is biased towards the inelastic or non-monopsonistic input, but not necessarily for a technology that is biased towards the perfectly elastic or monopsonistic input.

Proposition 2 *If an input is monopsonistic of perfectly elastic, the markdown of the other input increases the absolute return to a technology that is biased towards that other input. It can increase or decrease the absolute return to a technology that is biased towards the monopsonistic or perfectly elastic input.* ***Proof:*** *see Appendix A-D.*

$$\text{For } \mu_f^l = \psi_f^l \text{ or } \psi_f^l = 1: \qquad \frac{\partial \beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\Pi_f)}{\partial \mu_f^h \partial K_f} \begin{Bmatrix} \geq \\ \geq \; or \; \leq \end{Bmatrix} 0$$

The intuition behind Proposition 2 is as follows. The main reason why the result in Theorem 1 does not necessarily translate to the absolute profit effect of a technology relates to Hicks-neutral productivity changes. Even if the technology that lowers β^h reduces the relative demand for H compared to L, the firm could still end up using more units of input H in absolute terms due to the Hicks-neutral productivity shock. In that case, a higher markdown over H could increase the absolute profit return from the technology, even if it is L-biased.

C Hicks-neutral vs. factor-biased technology effects

Suppose the markdown decreases the relative return to technology adoption. What then determines whether the effect of the markdown on the absolute return to adoption will be negative or positive? Proposition 3 says that the higher the Hicks-neutral productivity effect of the technology is, the more likely it becomes that markdowns increase the absolute returns to technology adoption. The reason for this is that higher (more positive) Hicks-neutral productivity effects increase the likelihood that the absolute usage of all inputs increases with technology adoption.

Proposition 3 *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 A-E*

In the limiting case of a neutral technology that only increases Hicks-neutral productivity but not the output elasticities, the markdown on any input market increases the absolute return from technology adoption, as is stated in Proposition 4. An increase in Hicks-neutral productivity results in higher equilibrium output produced by the firm. The higher input price markdowns are, the higher profits are, and hence the higher the absolute change in the profit level after technology adoption.

Proposition 4 *The absolute profit effect of a technology that weakly increases Hicks-neutral productivity but does not change the output elasticity of any input weakly increases with either input price markdown.* ***Proof:*** *see Appendix A-F.*

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

3 Coal mining in Illinois, 1884-1902

The theoretical model in the previous section showed that the effect of buyer power on technology usage is ambiguous, except in a limited number of special cases, and depends both on the Hicks-neutral and directed effects of new technologies. Hence, in order to know how buyer power affects technology usage in a more general setting, empirical analysis is required. Prior to setting up the empirical model, I discuss the industry background and present two motivating facts.

3.1 Industry background

The setting of the empirical application is the Illinois coal mining industry between 1884 and 1902. Throughout this time period, this industry grew rapidly: annual output tripled from 10 to 30 megatons 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 680 to 898.

A Extraction process

The coal extraction process consisted of three consecutive steps. First, the coal vein had to be accessed, as it lied below the surface for 98.0% of the mines and 99.4% of output. Second, upon reaching the vein, the coal wall was 'undercut', traditionally by hand, 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 (Fishback, 1992).Third, coal had to be transported back to the surface and sorted from impurities. The hauling was done using mules or underground locomotives. Mines used two types of intermediate inputs. First, black powder was used to blast the coal wall. This powder and other materials, such as picks, was purchased and brought by the miners. Second, coal itself was used to power steam engines, electricity generators, and air compressors.

Figure 2(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'). Daily output per worker increased from 2 to 3.3 tons for hand mines, and from 2.3 to 4.1 tons for machine mines.[9]

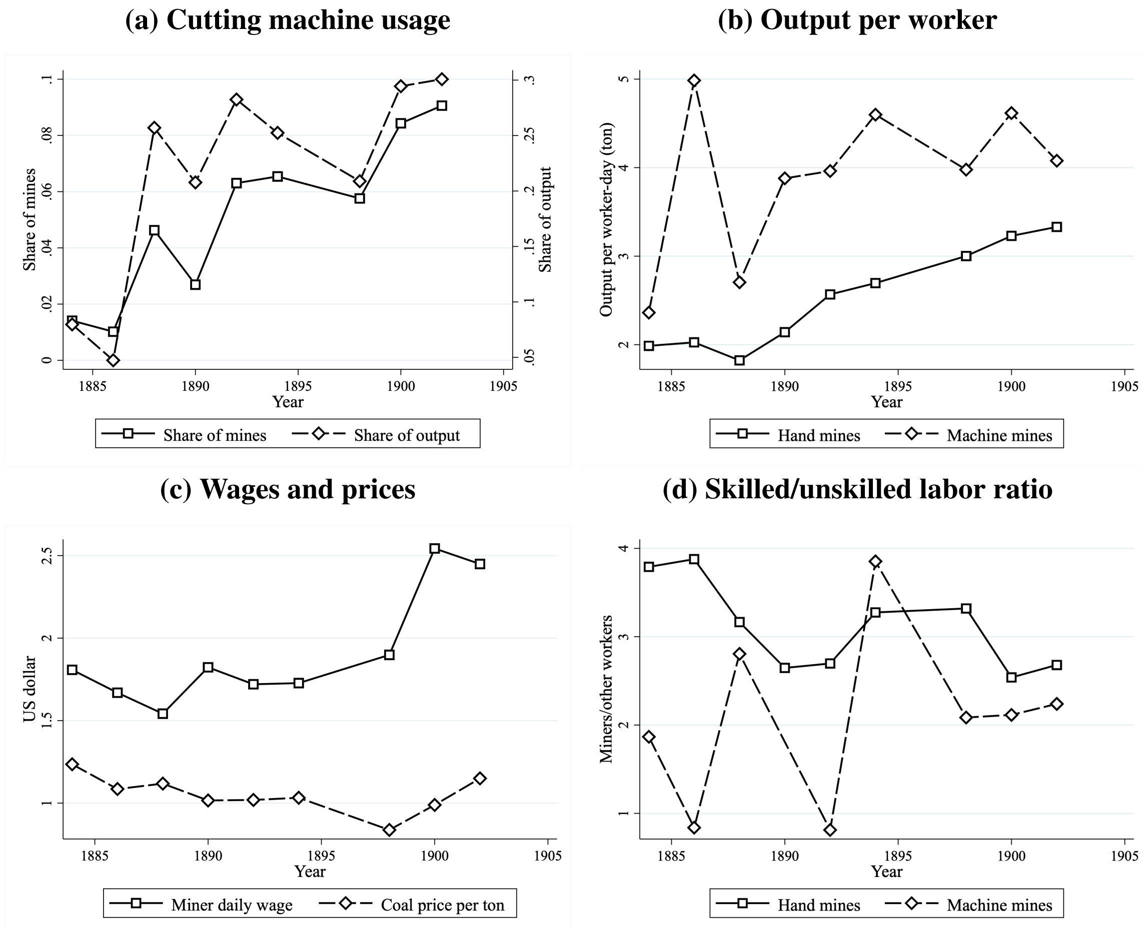
B Occupations

Coal mining involved a variety of occupational 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 undercutting coal walls.

The 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

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

Figure 2: Output, inputs, and prices



flooding. From this point onward, I classify workers into two types: miners, which I will denote as 'skilled labor', and all other employees, which are called 'unskilled labor'. This follows the categorization of labor provided in the data set.

C Technological change

The first mechanical coal cutter in the U.S.A. was invented by J.W. Harrisson in 1877, but it was merely a prototype.[10] The Harrisson patent was acquired and adapted by Chicago industrialist George Whitcomb, whose 'Improved Harrison Cutting Machine' was released in 1882.[11] As shown in Figure 2a, the share of Illinois coal mines using a cutting machine

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

[11]A picture of the patent is in Appendix Figure A5. The spatial diffusion of cutting machines is shown in Appendix Figure A1.

increased from below 2% to 9% between 1884 and 1902. Mechanized mines were larger: their share of output increased from 7 to 30% over the same time period. The mechanization of the hauling process, which replaced mules with underground locomotives, was another source of technical change, and started during the 1870s. By the start of the panel in 1884, mining locomotives were already widely used in Illinois: the share of output mined in locomotive mines was around 90%.

As was shown in Figure 2(b), output per worker was higher in cutting machine mines. The composition of labor was also different: in Figure 2(d), I plot the ratio of the total number of skilled labor-days over the number of unskilled worker-days per year.[12] 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 this difference is statistically significant. 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 estimates of the causal effect 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 unskilled for skilled 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 [...] 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, 1892).

In contrast, underground mining locomotives had very different effects: rather than saving on the skilled miners, locomotives replaced mules and some of the unskilled workers involved in the hauling process, such as mule tenders.

D Labor markets

Skilled workers received a piece rate per ton of coal mined, whereas unskilled workers were paid a daily wage.[13] Converting the piece rates to daily wages, the net salary of skilled labor

[12] 1890 is omitted for machine mines in 1890 due to employment being unobserved for most machine mines in that year.

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

was on average 22% higher compared to unskilled labor. 'Net salary' means net of material costs and other work-related expenses. Rural Illinois was, and still is, sparsely populated: the median and average population sizes of the towns in the dataset were 845 and 1706 inhabitants. In the average town, 16% of the population was employed in a coal mine. Considering that women and children under the age of 12 did not work in the mines, this implies that a large share of the local working population was employed in coal mining. Of all the villages, 42% had just one coal firm, and 75% had three or less coal firms. 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 next closest 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 were not yet introduced. In order to switch employers, miners had to migrate to another town.[14]

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 through 'yellow-dog' labor contracts that forced employees not to join a trade union.[15] The first successful trade 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 employers after 1898, which took place in January (Bloch, 1922). Wages were therefore set by employers until 1898, but were the results of bargaining afterwards. Wages were bargained over in a tiered negotiation procedure: first, a general agreement was made at the state-industry level, afterwards mine owners individually negotiated wages with miner representatives (Bloch, 1922). 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.

Figure 2(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. The fast growth in labor productivity did not translate into higher wages until 1898, as daily miner wages remained around $1.8. After the subsequent introduction of wage bargaining, wages rose.

[14]Some more evidence supporting the isolated mining towns assumption is in Appendix C.5.

[15]These contracts were criminalized in Illinois in 1893, with fines of $100 USD, which was equivalent to on average six months of a miner's wage. (Fishback, Holmes, & Allen, 2009).

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 92% of the mines' coal output was either sold to railroad firms or transported by train to final markets. The remaining 8% was sold to local consumers. The main coal destination markets for Illinois mines were St. Louis and, to a lesser extent, Chicago.[16] Railway firms acted as an intermediary between coal firms and consumers, and were also major coal consumers themselves. Historical evidence points to intense 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 80 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. Figure 2(c) shows that the mine-gate coal price per ton, weighted by output shares, fell from \$1.2 to \$0.9 between 1884-1898, after which it increased again.

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 is obtained from the *Biennial Report of the Inspector of Mines of Illinois*. 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 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 to 1896. 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.

[16]Chicago mainly sourced its coal from fields in Ohio, Pennsylvania, and West Virginia using lake steamers (Graebner, 1974).

However, the data set reports the usage of wage screens in 1898, and shows that they were used in mines representing merely 2% of total employment. 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 utilize different other datasets. First, the inspection report from 1890 contains monthly data on wages and employment for both types of workers, and monthly production quantities for a sample of 11 mines that covers 15% of skilled and 9% of unskilled workers. Second, town- and county-level information from the 1880 and 1900 population census and the censuses of agriculture and manufacturing are collected as well. Third I collect information on coal cutting machine costs from Brown (1889). I refer to Appendix B for more details regarding the data sources and cleaning procedures.

3.3 Facts to be explained

The empirical model in the next section will examine how coal firms' technology choices were affected by how much market power they possessed over the miners. Prior to estimating the model, I present two motivating facts on technology choices.

Fact 1 *Less cutting machine, but more locomotive usage in concentrated labor markets.*

First, I compare the usage of cutting machines and mining locomotives across towns with different labor market structures. I regress average technology usage at the town level on variables indicating whether there are one, two, or three coal firms in the town. I control for the log average vein thickness and depth in each town, the log distances to Chicago and St. Louis, and a linear time trend, as these could explain both labor market structure and machine usage. The resulting estimates are in Table 1(a), and are very different for both technologies. Whereas cutting machine usage is 4.3 percentage points lower in duopsonistic towns compared to towns with four or more firms, the locomotive usage rate is 8 percentage points higher in monopsonistic towns compared to towns with four or more firms.

Fact 2 *Cutting machine usage increased in response to successful wage bargaining by miners.*

Table 1: Facts to be explained

(a) Market structure and technology usage	1(Cutting machine)		1(Locomotive)	
	Est.	S.E.	Est.	S.E.
1(One firm in town)	-0.003	0.020	0.080	0.042
1(Two firms in town)	-0.043	0.020	0.029	0.042
1(Three firms in town)	-0.006	0.028	0.034	0.040
Mine FE	No		Yes	
R-squared	.124		.598	
Observations	960		728	
(b) Unionization and technology usage	1(Cutting machine)			
	Est.	S.E.	Est.	S.E.
1(Successful wage bargaining)*1(t$\geq$ 1898)	0.091	0.025	0.054	0.025
1(Successful wage bargaining)	0.059	0.020	-0.008	0.013
1(t$\geq$ 1898)	0.012	0.011	0.007	0.013
Mine FE	No		Yes	
R-squared	.024		.755	
Observations	4310		4310	

Notes: Controls in panel (a) are log vein thickness and depth, log distances to Chicago and to St. Louis, and a linear time trend. Standard errors in panel (a) are clustered at the town level. Standard errors in panel (b) are clustered at the mine level.

Second, I examine how the move to wage bargaining after the 1897 strike affected cutting machine usage. I cannot carry out the same exercise for mining locomotives, as these are not recorded in the data set after 1898. The introduction of wage bargaining between labor unions and mine owners led to higher wages on average, and hence lower wage markdowns compared to the pre-1898 period. However, it did not affect wages at all mines in the same way. At mines where miners were already paid their marginal product prior to 1898, wage bargaining presumably did not lead to large wage increases, as this would have led to exit, compared to firms that did charge high wage markdowns prior to 1898. Hence, I compare machine usage between two sets of mines i over time t: those at which wage negotiations led to a wage increase immediately after the 1897 strikes, denoted as $WI_{it} = 1$ compared to the mines at which wages were not raised after these strikes, $WI_{it} = 0$. To do so, I estimate the following difference-in-differences equation for the six years before and after the introduction of wage bargaining.

$$K_{it} = a_0 + a_1 WI_{it} + a_2 WI_{it} * I(t > 1897) + a_3 I(t > 1897) + u_{it}$$

The results are in Table 1(b). I find that when not controlling for mine fixed effects, cutting machine usage increased by 9.1 percentage points at mines with successful wage bargaining after 1898 compared to the other mines. When controlling for mine fixed effects, the relative increase is 5.4 percentage points. Appendix Figure A2 visualizes the results, and also shows that cutting machine usage was already higher in the mines that saw successful wage bargaining, but that the pre-trends were similar.

Summing up, cutting machine and locomotive usage have the opposite correlation with labor market structure, and the drop in wage markdowns after 1898 seems to have led to an increase in cutting machine usage. The empirical model in the next model will be used to replicate and explain these descriptive facts.

4 Empirical model

4.1 Model

I implement an empirical version of the labor supply and demand model from Section 2, tailored to the coal mining industry setting. The goal of this model is to answer the counterfactual question of how changes in labor market competition would affect technology usage, and how this relationship depends on the directed and Hicks-neutral effects of the technologies in question.

A Coal extraction

Let f index firms and t even years. The model is set up at the firm level, because it is more plausible that firm owners optimize at this level, rather than at each mine independently.[17] Annual 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 $K_{ft}^{cut} \in \{0, 1\}$, locomotive usage is $K_{ft}^{loc} \in \{0, 1\}$, together they form the capital vector $\mathbf{K}_{ft}$. The production function in logs is given by Equation (6a), 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 β_{ft} to vary flexibly across firms and years. The scale parameter ν is equal

[17] For the few firms that operate in multiple labor markets, I assume that firms make decisions independently across labor markets, given that labor markets will be assumed to be isolated.

to the sum of the output elasticities of skilled and unskilled workers, and is assumed to be a constant. The Hicks-neutral productivity residual in logs is denoted ω_{ft}.

$$q_{ft} = \beta_{ft} h_{ft} + (\nu - \beta_{ft}) l_{ft} + \omega_{ft} \tag{6a}$$

Both the output elasticity of skilled workers β_{ft} and the productivity residual ω_{ft} are assumed to be AR(1) processes, Equations (6b) and (6c), with serial correlations ρ^{β} and ρ^{ω}. 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.[18] Both the output elasticity and Hicks-neutral productivity level are assumed to be linear functions of current technology usage $\mathbf{K}_{ft}$ 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 miners 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 $\alpha^{\beta,cut}$, their effect on Hicks-neutral productivity is $\alpha^{\omega,cut}$. Similarly, the effects of mining locomotives on the output elasticity of skilled labor and on Hicks-neutral productivity are $\alpha^{\beta,loc}$ and $\alpha^{\omega,loc}$. The residual shocks to the skilled labor output elasticity and Hicks-neutral productivity are denoted γ^{β}_{ft} and γ^{ω}_{ft}. By using these parametric specifications, I assume that there is no heterogeneity across firms and time in the Hicks-neutral and factor-biased effects of cutting machines and mining locomotives.

$$\beta_{ft} = \alpha^{\beta,cut} K^{cut}_{ft} + \alpha^{\beta,loc} K^{loc}_{ft} + \sigma^{\beta} \mathbf{X}_{ft} + \rho^{\beta} \beta_{ft-1} + \gamma^{\beta}_{ft} \tag{6b}$$
$$\omega_{ft} = \alpha^{\omega,cut} K^{cut}_{ft} + \alpha^{\omega,loc} K^{loc}_{ft} + \sigma^{\omega} \mathbf{X}_{ft} + \rho^{\omega} \omega_{ft-1} + \gamma^{\omega}_{ft} \tag{6c}$$

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.[19]

B Coal demand

Each firm operates on a single coal market, indexed by m with a market share $s^{q}_{ft} \equiv \frac{Q_{ft}}{Q_{mt}}$ 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 (7), with a market-level mine-gate coal price P_{mt}, an inverse demand elasticity η, and a market-level residual ζ_{mt}

[18] I refer to Appendix C.4 for a motivation and discussion of this assumption.
[19] The entire distribution of capacity utilization rates is shown in Figure A4.

that 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.[20]

$$P_{mt} = (Q_{mt})^{\eta} \exp(\zeta_{mt}) \tag{7}$$

Coal is assumed to be a homogeneous product, conditional on the market and year: there is no within-market, within-year variation in coal quality. Although different coal types exist, the mines in the data set all extract bituminous coal. There might be minor quality differences even within this coal type due to variation in sulfur content, ash yield, and calorific value (Affolter & Hatch, 2002). Most of this variation is, however, dependent on the mine's geographical location. The coal homogeneity assumption will be defended using a variance decomposition of prices in Section 4.2.

C Input supply

Each firm operates on exactly one labor market n, which will be defined 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 reveals that there was almost no within-firm wage variation for workers of either skill type at a given point in time. Firms have a skilled labor market share $s^h_{ft} \equiv \frac{H_{ft}}{H_{nt}}$ and unskilled labor market share $s^l_{ft} \equiv \frac{L_{ft}}{L_{nt}}$, with market-level employment $H_{nt} \equiv \sum_{f \in n} H_{ft}$ and $L_{nt} \equiv \sum_{f \in n} L_{ft}$. The market-level supply curve for both types of workers is given by Equation (8). The inverse wage elasticities are $\psi^h_{nt} = \frac{\partial W^v_{nt}}{\partial H_{nt}} \frac{H_{nt}}{W^h_{nt}} + 1$ and $\psi^l_{nt} = \frac{\partial W^l_{nt}}{\partial L_{nt}} \frac{L_{nt}}{W^l_{nt}} + 1$. The error terms ξ^h_{nt}, ξ^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.

$$\begin{cases} W^h_{nt} &= {H_{nt}}^{\psi^h_{nt}-1} \exp(\xi^h_{nt}) \\ W^l_{nt} &= {L_{nt}}^{\psi^l_{nt}-1} \exp(\xi^l_{nt}) \end{cases} \tag{8}$$

Coal demand was seasonal: during the cold winter months, energy demand increased compared to the warm summer months. Mining firms could not arbitrage across seasons because coal storage often led to deteriorating coal quality, and was expensive and dangerous (Stoek, Hippard, & Langtry, 1920). As Williams (1901) asserts, "The product of a mine can

[20]This assumption is relaxed in Appendix C.2.

be stored with economy only in the mine itself [...] Coal must be sold, therefore, as fast as it is mined". This is confirmed in Figure 3(a), that plots the distribution of average skilled employment (in days) during winter and summer: more workers were hired during winter compared to the summer months.[21] Figure 3(b) shows that skilled wages followed this coal demand cycle: they were higher during winters than during summers. However, this pattern held only for skilled wages, not for unskilled wages. Although the seasonal demand shocks increased both skilled and unskilled labor demand, only skilled wages increased during winter. This is shown in Figure 3(c), that plots how average daily wages for both skilled and unskilled workers in 1890 change with the monthly number of worker-days of each type at the mine-month level throughout 1890.[22] Skilled wages were positively correlated with monthly skilled employment, whereas the unskilled worker wage-employment schedule is flat. Moreover, there was a lot of variation in skilled wages across mines and months, but very little cross-sectional and intertemporal variation in unskilled wages.

Given that Figure 3(c) reveals that unskilled worker wages were much less dispersed compared to skilled wages, and did not change in response to seasonal labor demand shocks, I assume that unskilled labor supply is perfectly elastic, meaning that $\psi^l_{nt} = 1$, but allow for inelastic skilled labor supply, $\psi^h_{nt} \geq 1$. This is in line with recent evidence on the wage effects of mergers (Prager & Schmitt, 2021). There are, of course, other possible explanations for the fact that wages did not react to labor demand shocks, such as behavioral reasons (Kaur, 2019). Key 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.

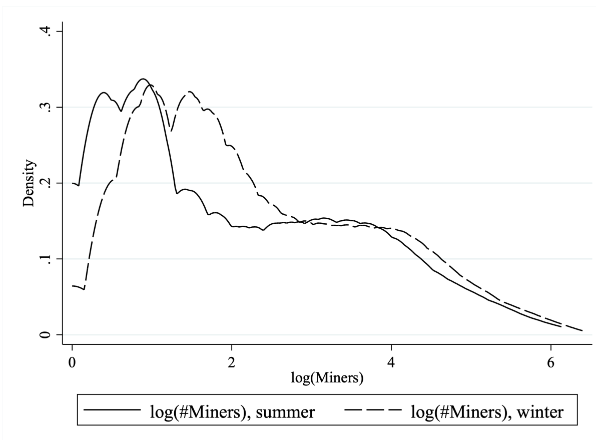
In contrast to unskilled labor, I allow for the elasticity of skilled labor supply, ψ^h_{nt}, to be above one. Although the log-linearity of Equation (8) imposes a strong functional form assumption, I allow the slope ψ^h_{nt} to vary flexibly across markets and time, as local labor market conditions may vary. I assume that cutting machines are sold on competitive markets, and that their prices are exogenous to each individual mine. I assume that employers

[21] This is consistent with anecdotal evidence: Joyce (2009) mentions that miners were (partially) unemployed during the summer months.

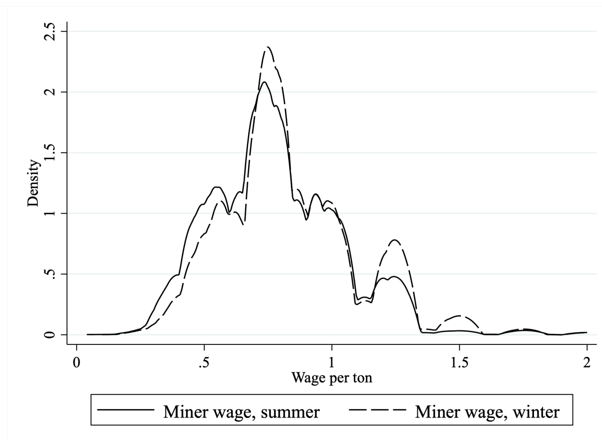
[22] Unlike skilled wages and employment, unskilled wages and employment are not broken down by season in the entire dataset. However, monthly wage and employment data is available for a sample of mines selected by the Illinois Bureau of Labor Statistics across 5 counties in 1890, which covers 16% of skilled employment and 9% of unskilled employment.

Figure 3: Seasonality in employment and wages

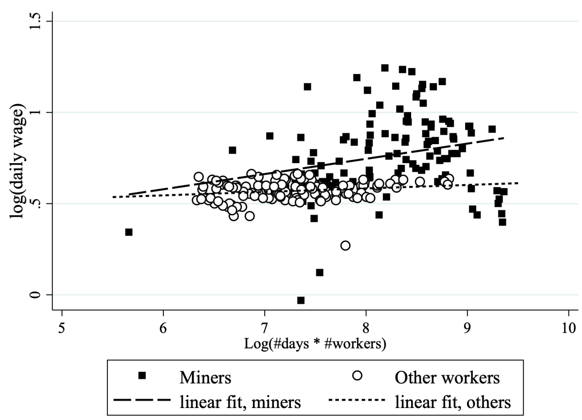
(a) Skilled employment



(b) Skilled wages



(c) Wage-employment profile by skill type



are homogeneous 'products' from the point of view of the workers, conditional on their location and the time period in question. A similar assumption was made in Arnold (2019). 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 firm-level 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, and wages were determined using short-term contracts until 1898.[23] Both labor types are also static because current labor choices do not affect future profits, i.e. there are no hiring or firing costs. This fits the description in Fishback (1992) of miners as being more similar to contractors than to employees. Cutting machines and locomotives are, in contrast, fixed inputs. Firms need to make their technology adoption decisions one period in advance. Let capital accumulation for each technology $\tau \in \{loc, cut\}$ be given by the following equation, with machine acquisitions being denoted as $A^{\tau}_{ft} \in \{0, 1\}$. Depreciation $\delta \in \{0, 1\}$ takes the value of either zero or one. If there is no depreciation, meaning that $\delta = 1$, firms can only acquire a cutting machine or locomotive if they do not already own one, and such an acquisition is permanent. If $\delta = 0$, machines fully depreciate within two years, and firms repeat their capital adoption decision in every time period.

$$K^{\tau}_{ft} = \delta K^{\tau}_{ft-1} + A^{\tau}_{ft-1}(1 - \delta K^{\tau}_{ft-1}) \tag{9}$$

Observed entry and exit of machine usage is frequent: whereas there are 109 observed instances of cutting machine installation, there are 62 observed instances in which an installed cutting machine is scrapped again. This suggests the existence of an aftermarket for capital. Because of this, I assume that firms re-optimize their capital choices every time period, for the next time period, which implies a depreciation rate of $\delta = 0$. Each technology has a common fixed cost component Φ^{τ}, with variable technology costs being normalized to zero. Firm-level variable profits are denoted $\Pi_{ft} \equiv P_{ft}Q_{ft} - W^{h}_{mt}H_{ft} - W^{l}_{mt}L_{ft}$. Intermediate input expenditure is not part of the mine's costs, as black powder and materials were purchased and brought by the miners. Total profits are defined as $\Pi^{tot}_{ft} \equiv \Pi_{ft} - \Phi^{cut}K^{cut}_{ft} - \Phi^{loc}K^{loc}_{ft}$.

I assume that firms make their input decisions in two phases. At time $t-1$, before the productivity shocks γ^{ω}_{ft} and γ^{β}_{ft} are observed, firms simultaneously choose their technology usage for the next period, $\mathbf{K}_{ft}$, through the capital adoption decisions $\mathbf{A}_{f,t-1} = (A^{loc}_{f,t-1}, A^{cut}_{f,t-1})$. Firms pick the capital investments that maximize total profits in the next

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

period.

$$\max_{\mathbf{A}_{f,t-1}} (\Pi_{ft}^{tot}) \tag{10}$$

At time t, after the productivity shocks are observed, firms simultaneously choose their optimal amounts of both labor types conditional on their capital technology, which was chosen earlier. This decision follows Equation (11): firms independently choose the amounts of skilled and unskilled labor that maximize their current variable profits. By choosing the amount used of both labor types, firms also choose their output Q_{ft}.

$$\max_{H_{ft}, L_{ft}} (\Pi_{ft}) \tag{11}$$

In contrast to labor choices, which are strategic, capital investment is assumed to be a single-agent problem, similarly to Olley and Pakes (1996). Firms do not take into account that their technology choices affect wages and markdowns at other firms in the same market, and hence technology choices of these other firms.

E Equilibrium

Solving the first order conditions for the profit maximization problem in (11) gives equilibrium expressions for all endogeneous static variables (Q, P, H, L, W^h, W^l), which can be found in Appendix C.1. The skilled labor wage markdown is equal to the market-level inverse supply elasticity, weighted by the labor market share:

$$\frac{\frac{\partial(P_{mt}Q_{ft})}{H_{ft}}}{W_{nt}^h} = 1 + (\psi_{nt}^h - 1)s_{ft}^h$$

The markdown parameter μ_f^h from the theoretical model hence corresponds to the markdown $1 + (\psi_{nt}^h - 1)s_{ft}^h$ 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 ψ_{nt}^h. If the firm is atomistically small, the markdown goes toward one in the limit, meaning that skilled laborers earn their marginal product of labor. Due to the Cournot coal market assumption, the markup is equal to $\mu_{ft} = (1 + s_{ft}^q \eta)^{-1}$.

4.2 Identification and estimation

I now turn to the identification and estimation of the model. The following latent variables need to be identified: the market-level inverse elasticity of skilled labor supply ψ_{nt}^h, in

Equation (8), the inverse elasticity of coal demand η, in Equation (7), the entire distribution of output elasticities of skilled labor β_{ft}, in Equation (6a), the effects of cutting machines and locomotives on the output elasticity of skilled labor and on Hicks-neutral productivity, in Equations (6b)-(6c), and the fixed costs of cutting machines and locomotives. Although the model is specified at the firm-bi-year level, the dataset comes at the mine-year level. I aggregate all the relevant variables from the mine- to the firm-year-level.[24] Given that labor markets are isolated, I assume that firms make their input decisions independently across towns. 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 wage bargaining between unions and coal firms was instituted in 1898, which does not fit the unilateral oligopsony framework of the model.

A Labor supply

Identification I start with the identification of the skilled labor supply function. Taking the logarithm of Equation (8) for skilled labor, and denoting logs as lowercases, gives Equation (12).

$$w_{nt}^{h} = (\psi_{nt}^{h} - 1)h_{nt} + \xi_{nt}^{h} \tag{12}$$

The supply elasticity ψ_{nt}^{h} cannot be recovered by simply regressing skilled labor wages on employment because of the latent outside options ξ_{nt}^{h}. Firms in labor markets with an unattractive outside option ξ_{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 $\xi_{nt}^{h,SUM}$. The slope of the skilled labor supply curve can be expressed as Equation (13).

$$\psi_{nt}^{h} = \frac{(w_{nt}^{h,WIN} - \xi_{nt}^{h,WIN}) - (w_{nt}^{h,SUM} - \xi_{nt}^{h,SUM})}{h_{nt}^{WIN} - h_{nt}^{SUM}} + 1 \tag{13}$$

[24] Firms are assumed to be mine owners. Details on how I aggregate to the firm-level are in Appendix B.2.

The problem is that not just labor demand could vary seasonally, but also labor supply, as measured by the latent variables $\xi_{nt}^{h,SUM}, \xi_{nt}^{h,WIN}$. One example for seasonal variation in miner supply would be seasonal wage variation in other industries, such as in agriculture due to the harvesting season. I proceed by assuming that the labor supply residual is identical across skill groups, $\xi_{nt}^{h} = \xi_{nt}^{l} = \xi_{nt}$. This implies that both skilled and unskilled miners can obtain the same wage when moving to other industries than coal mining: coal miner skills are industry-specific, which is consistent with Fishback (1992). Given that unskilled labor supply is perfectly elastic, $\psi_{ft}^{l} = 1$, the outside option ξ_{nt} is equal to the unskilled wage. The seasonal outside option can hence be substituted for seasonal unskilled wages. As shown in Figure 3(c), unskilled wages barely varied across seasons, which implies that the labor supply residuals must have been stable throughout the year too. Hence, $\xi_{nt}^{h,WIN} = \xi_{nt}^{h,SUM}$, and the outside options cancel out in Equation (13).

In Appendix C.3, I propose an alternative identification strategy that relies on cross-sectional variation in labor demand due to railroad connections. This avoids having to assume stable outside options throughout the year. Moreover, this also addresses another issue: the above identification strategy likely measures a short-run labor supply elasticity, but labor supply is likely more elastic on the longer term due to miner migration. The alternative labor supply estimates in Appendix C.3 indeed suggest more elastic labor demand, but deliver counterfactual results that are very similar to those derived from the baseline model.

Labor market definition Miners could only work in their own mining town or commute by foot to another town, as railroads were only used for freight cargo outside of the large cities. Of the 448 towns 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.[25] In order to ensure isolated labor markets, I merge the towns that are closer than 3 miles from each other.[26] This results in 374 labor markets that lie on average 6.4 miles from the next nearest town.

[25] 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.

[26] More information is in Appendix B.2

Estimation I calculate the slope of the skilled labor supply curve for each town using Equation (13). Skilled wages are reported separately for winters and summers between 1884-1894.The reported wage rates are piece rates, in wages per ton. Equation (13) was, however, written using daily wages per worker and days of employment, because workers care only about their daily wage, not their wage per tons 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. 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 (13). For the bottom percentile of the distribution, the inverse supply elasticity is negative, because either winter employment or wages are below summer employment or wages. Given that the equilibrium model gives negative production quantities with a downward-sloping supply curve, I restrict these elasticities to be zero instead, which implies $\psi^h_{nt} = 1$.

B Coal demand

Identification Taking logarithms of the coal demand function, (7), results in $p_{mt} = \eta q_{mt} + \zeta_{mt}$. I decompose the market-level residual ζ_{mt} into an unobservable component $\bar{\zeta}_{mt}$ and an observable component $\boldsymbol{\eta}^z \mathbf{Z}_{mt}$. Hence, I estimate the following equation:

$$p_{mt} = \eta q_{mt} + \boldsymbol{\eta}^z \mathbf{Z}_{mt} + \bar{\zeta}_{mt}$$

As firms in coal markets with attractive features ζ_{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 quality (Affolter & Hatch, 2002). 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 to coincide with the labor market. 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.[27] 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 have an average coal market share of 38%, compared to 5.8% for firms selling through the railroad network.

Estimation I estimate Equation (7) in logs by 2SLS using the log average vein thickness in the town as an instrument for coal output. In the observed covariates vector $\mathbf{Z}_{mt}$, I include the following market-level demand shifters: the log 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, year dummies, and log average mine depth in the market, as this affected coal quality (Affolter & Hatch, 2002). Regressing firm-level coal prices on market and year fixed effects yields an R-squared of 0.69, adding the observed market-level demand and cost shifters increases this to 0.83. Hence, there is just 17% unobserved variation in coal prices within markets between firms, which supports the homogeneous product assumption.

C Output elasticities of labor

Identification Working out the variable profit maximization problem in Equation (11) yields Equation (14), which equates the output elasticity of skilled labor to the product of its revenue share, the wage markdown, and the coal price markup. This expression follows the tradition of Hall (1988), Foster et al. (2008), and Hsieh and Klenow (2009), with the difference that I allow for endogenous input prices.

$$\beta_{ft} = \frac{W^h_{nt} H_{ft}((\psi^h_{nt} - 1)s^h_{ft} + 1)}{P_{mt} Q_{ft}(1 + \eta s^q_{ft})} \tag{14}$$

Given that markups and markdowns were estimated earlier, the output elasticity of unskilled labor is known up to the scale returns parameter ν, as it is $(\nu - \beta_{ft})$. The intuition behind (14) is that after netting out any markup and markdown variation, the residual variation in revenue shares should be due to variation in output elasticities. Although this approach comes at the cost of having to impose a fixed parameter for the degree of returns to scale and a model of competition both upstream and downstream,[28] the benefit is that it allows for unobserved heterogeneity in the output elasticity of skilled labor, both across

[27]Details are in Appendix B.2.

[28]In principle, imposing a model of competition downstream would not be required in any case, as one could net out markdowns from *cost* share variation across firms, rather than from *revenue* share variation across firms. However, in this paper, unskilled labor costs are unobserved, which rules out this approach.

firms and time. Moreover, in order to solve for market equilibrium and carry out the counterfactual exercises, imposing a model of competition upstream and downstream is required in any case.

Estimation In order to recover the entire distribution of output elasticities of skilled and unskilled labor using Equation (14), a value for the degree of scale returns ν needs to be calibrated. I calibrate the scale parameter to be $\nu = 0.9$, because of two reasons. First, assuming decreasing returns to scale makes sense because nearly all the mines produced below their full capacity, despite coal markets being perfectly competitive. If there would be constant or increasing returns to scale and perfect competition downstream, firms without monopsony power on labor markets should produce at full capacity. Whereas half of the firms have a horizontal skilled labor supply function, and hence no monopsony power, merely 2% of firms produce at full capacity, and 90% of firms use less than four fifths of their capacity. Second, in Appendix C.2, I present an alternative production model that also estimates scale returns, but that does not allow for unobserved heterogeneity in the output elasticities. This results in a returns to scale parameter of 0.856, which is close to the calibrated value of 0.9.

D Factor-biased and Hicks-neutral productivity transitions

Identification The effects of cutting machines and locomotives on both the output elasticity of skilled labor and on Hicks-neutral productivity, Equations (6b) and (6c), need to be identified. Simply regressing the output elasticity of skilled labor β_{ft} or Hicks-neutral productivity Ω_{ft} on technology usage is subject to simultaneity bias, as both Hicks-neutral and factor-augmenting productivity affect input demand (Doraszelski & Jaumandreu, 2017). I follow the production function identification literature by relying on timing assumptions to identify the technology effects $\boldsymbol{\alpha}^{\beta}$ and $\boldsymbol{\alpha}^{\omega}$ (Olley & Pakes, 1996; Ackerberg et al., 2015). Following Blundell and Bond (2000) I take ρ-differences of Equation (6c), such that the skilled labor productivity shock can be written as $\gamma^{\beta}_{ft} = \boldsymbol{\alpha}^{\beta}(\mathbf{K}_{ft} - \rho^{\beta}\mathbf{K}_{ft-1}) + \boldsymbol{\sigma}^{\beta}(\mathbf{X}_{ft} - \rho^{\beta}\mathbf{X}_{ft-1})$, and the Hicks-neutral productivity shock as $\gamma^{\omega}_{ft} = \boldsymbol{\alpha}^{\omega}(\mathbf{K}_{ft} - \rho^{\omega}\mathbf{K}_{ft-1}) + \boldsymbol{\sigma}^{\omega}(\mathbf{X}_{ft} - \rho^{\omega}\mathbf{X}_{ft-1})$. As was explained earlier, cutting machines and locomotives are chosen prior to the realization of both productivity shocks γ^{ω}_{ft} and γ^{β}_{ft}, which allows to identify the production function coefficients by imposing that current and lagged capital usage are orthogonal to γ^{β}_{ft} and γ^{ω}_{ft}. As both labor inputs and black powder are variable inputs, they are chosen after the productivity shocks γ^{β}_{ft} and γ^{ω}_{ft} are observed, but their lagged values are orthogonal to these shocks. Hence, the moment conditions are:

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

Estimation I estimate Equations (6b) and (6c) 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 one plus the number of powder kegs used by the firm.

E Fixed technology costs

As mentioned before, I assume a discount rate of $\delta = 0$ in order to make the adoption problems static. I assume fixed costs Φ^{cut} and Φ^{loc} are both common across firms and time. For both technologies, I estimate the fixed cost level that rationalizes the observed technology usage rate. Solving for labor and product market equilibrium, as explained in Section 4.4A, results in a variable profit gain for each technology τ for each firm-year combination, $\Delta\Pi^{\tau}_{ft} \equiv \Pi_{ft}(K^{\tau}_{ft}=1) - \Pi_{ft}(K^{\tau}_{ft}=0)$. I estimate fixed costs by minimizing the distance between the observed technology usage rate and the predicted technology usage rate:

$$\min_{\Phi^{\tau}} \Big(\sum_{f,t} K^{\tau}_{ft} - \sum_{f,t} I(\Delta\Pi^{\tau}_{ft} > \Phi^{\tau})\Big)^2$$

The entire estimation procedure that has been described in this section is implemented sequentially. First, I estimate the inverse skilled labor supply elasticities ψ^{h}_{ft}. Next, I estimate the inverse coal demand elasticity η. Third, I estimate the firm-level output elasticities β_{ft}, which requires knowledge of both ψ^{h}_{ft} and η. Fourth, I estimate the transition equations for the output elasticity of skilled labor β and for Hicks-neutral productivity, ω, in order to obtain the cutting machine and locomotive effects $\boldsymbol{\alpha}^{\beta}$ and $\boldsymbol{\alpha}^{\omega}$. Finally, I estimate the level of fixed machine costs $\boldsymbol{\Phi}$. 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

Miner supply The market-level skilled labor supply elasticity is in Table 2(a). The number of observations is 1153 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 ψ^h_{nt} is 1.157. This implies that a monopsonist would set the marginal product of skilled laborers at 15.7% above their wage, but a firm with a labor market share of s^h_{ft} would set the marginal product at s^h_{ft}*15.7% above the skilled wage. The average firm charges a markdown of 4.5%.[29] The inverse skilled labor supply elasticity has a standard error of 0.023, and hence lies significantly above one. Appendix C.5 discusses how the miner supply elasticity is correlated with town and county characteristics.

Coal demand The market-level coal demand elasticity is in Table 2(b). The number of observations is lower, at 453, because there are fewer coal markets than labor markets and because vein thickness (the instrument) is not observed in 1888 and 1890. The inverse demand elasticity is estimated to be -0.465, with a standard error of 0.101. I refer to Appendix Table A4 for the other coefficient estimates. The first-stage regression of the coal quantity on vein thickness has an F-statistic of 70.1. Coal demand is higher in markets that are connected to the railroad network and are located on railroad crossings, and decreases with the distance to both St. Louis and Chicago, although this distance effect is not statistically significant. The average firm charges a price of 14.8% above its marginal cost, but the median firm only charges a price that is 2.2% above its marginal cost. Downstream market power is hence skewed considerably towards large firms. As a validation exercise, I re-estimate markups using an alternative model that does not impose a model of coal market competition, but instead imposes that there is no unobserved heterogeneity in output elasticities across firms and time. This model, which is in Appendix C.2, estimates the average markup at 13.9%, reasonably close to the Cournot markup of 14.8%.

Production Table 2(c) contains the estimated output elasticities of skilled labor, the other coefficients are again in Appendix Table A4. The number of observations is 3800, given that this output elasticity is estimated at the firm-bi-yearly level. The output elasticity of skilled labor is on average 0.734, with a standard error of 0.033. The distribution of output elasticities across firms and time is plotted in Appendix Figure A6(b).

The factor-biased effects of cutting machines are in Table 2(d). Although this model is estimated at the firm-year level too, the number of observations is lower, at 1149, because lagged values of all variables are needed to estimate the equation of motion for the output elasticities. The output elasticity of skilled labor is estimated to fall by 0.160 units due to

[29] The distribution of firm-level markup and markdown ratios is plotted in Appendix Figure A6(a).

Table 2: Model estimates

(a) Miner supply (town-level)		Est.		S.E.
Inverse elasticity of miner supply	ψ^h	1.157		0.023
Observations			1153	
(b) Coal demand (county-level)				
Coal demand elasticity	η	-0.465		0.101
Observations			453	
R-squared			.191	
(c) Output elasticities				
Output elasticity of miners (avg.)	β	0.734		0.033
Observations			3800	
(d) Factor-biased productivity transition				
1(Cutting machine)	$\alpha^{\beta,cut}$	-0.160		0.068
1(Locomotive)	$\alpha^{\beta,loc}$	0.101		0.029
Observations			1149	
R-squared			.008	
(e) Hicks-neutral productivity transition				
1(Cutting machine)	$\alpha^{\omega,cut}$	0.218		0.157
1(Locomotive)	$\alpha^{\omega,loc}$	0.277		0.182
Observations			1066	
R-squared			.225	
(f) Fixed machine costs				
Fixed cutting machine cost (USD)	Φ^{cut}	5705.000		3127.542
Fixed locomotive cost (USD)	Φ^{loc}	1601.250		367.739

Notes: Standard errors are block-bootstrapped with 200 iterations.

the usage of cutting machines, which is a relative drop of 22% on average. The standard error on this coeffient is 0.068 so cutting machines have a significant negative effect on the output elasticity of skilled labor. In contrast, mining locomotives increase the output elasticity of skilled labor, and hence decrease the output elasticity of unskilled labor, by 0.101 points, and this change is significant as well. The finding that cutting machines saved on skilled miners, whereas mining locomotives saved on unskilled labor, is consistent with the fact that cutting machines automated the cutting process, whereas mining locomotives automated the hauling process. The effect of cutting machines on Hicks-neutral productivity is in Table 2(e). The point estimate of 0.218 implies that cutting machines increased Hicks-neutral productivity by 24%, but this effect is not statistically significant. Locomotives are estimated to increase productivity by 31%, but this is again not statistically significant. The estimated productivity effects of locomotives are similar in size to the estimates for coal-hauling steam locomotives in Pennsylvanian mines, in Rubens (2020a).

Fixed costs Finally, fixed technology costs are in Table 2(f), and are estimated to be \$5705 for cutting machines, and \$ 1601 for locomotives. In comparison, the average annual variable profit of a coal firm was \$3456, and the median firm's variable profit barely \$423. External cost information for cutting machines is obtained from Brown (1889), which reports a cost of \$8000 for eight cutting machines. The average firm in my data set also used 8 cutting machines, so the estimated fixed cost is below the external cost estimate. One reason for the smaller fixed cost estimate could be a non-zero scrap value of cutting machines after two years of usage.

4.4 Counterfactuals

Using the estimated model, I now examine how technology usage would change under different levels of labor market competition, for different types of technologies.

A Computation of the equilibrium

Let $(Q, H, L, W^h, W^l, P, \Pi, \Pi^{tot})$ be the endogenous variables of the model, other than capital. I denote a variable with a tilde as the equilibrium value of that variable for a given set of technology choices and for a certain labor market share s^h_{ft}. For instance, $\tilde{Q}_{ft}(K^{cut}_{ft} = 1, K^{loc}_{ft} = 0, s^h_{ft} = 0.5)$ denotes the equilibrium output of firm f in year t when using cutting machines, not using locomotives, and having a labor market share of 50%. These values can be computed using the equilibrium expressions in Appendix C.1.

In order to compute the equilibrium values under different technology choices, 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, $\beta_{ft}(K^{cut}_{ft})$ and $\Omega_{ft}(K^{cut}_{ft})$, and similarly for locomotives. If the mine does not use cutting machines, I calculate the counterfactual output elasticity if it would use cutting machines as $\beta_{ft}(K^{cut}_{ft} = 1) = \beta_{ft} + \alpha^{\beta,cut}$, using Equation (6b). Similarly if the mine is already using cutting machines, the counterfactual output elasticity when not doing is $\beta_{ft}(K^{cut}_{ft} = 0) = \beta_{ft} - \alpha^{\beta,cut}$. The counterfactual Hicks-neutral productivity levels are computed in the same way: $\omega_{ft}(K^{cut}_{ft} =$

$0) = \omega_{ft} - \alpha^{\omega,cut}$ is the counterfactual productivity of not using cutting machines, calculated if the firm is already using cutting machines, and $\omega_{ft}(K^{cut}_{ft} = 1) = \omega_{ft} + \alpha^{\omega,cut}$ is the counterfactual productivity of using cutting machines, calculated if the firm is not currently using cutting machines. I proceed analogously for locomotives.

The market-level demand shifter ζ_{mt} is computed as the residual of the coal demand function. Similarly, the labor supply residual ξ^h_{nt} is the residual of the estimated labor supply function, Equation (8):

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

B Labor market competition and technology usage

The equilibrium outcomes $(\tilde{Q}, \tilde{H}, \tilde{L}, \tilde{W}^h, \tilde{W}^l, \tilde{P}, \tilde{\Pi}, \tilde{\Pi}^{tot})$ can be evaluated under any labor market structure, factual or counterfactual. Accordingly, equilibrium usage of any technology can be calculated as an indicator function $\mathcal{I}(.)$ of whether the technology increases total profits at the firm or not, $\mathcal{I}[\tilde{\Pi}_{ft}(1, K^{loc}_{ft}, s^h_{ft}) - \tilde{\Pi}_{ft}(0, K^{loc}_{ft}, s^h_{ft}) - \Phi^{cut} > 0]$ for cutting machines, and $\mathcal{I}[\tilde{\Pi}_{ft}(K^{cut}_{ft}, 1, s^h_{ft}) - \tilde{\Pi}_{ft}(K^{cut}_{ft}, 0, s^h_{ft}) - \Phi^{loc} > 0]$ for locomotives. Averaging across firms allows to compare the usage rate of each technology at each labor market structure s^h_{ft}.

A number of assumptions needs to be explained at this point. First, the labor supply and coal demand residuals ξ^h_{nt}, ξ^l_{nt}, ζ_{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, not in any other way. Second, I assume that unskilled worker characteristics, which are equal to unskilled worker wages, are the same across firms in a given year $\xi^l_{nt} = \xi^l_t$. This assumption is motivated by the evidence in Figure 3(c), which showed that there is very little cross-sectional variation in unskilled wages. The residual ξ^l_t is equal to the daily unskilled wage, which is unobserved for most years. However, it can be backed out under the assumption of competitive unskilled labor markets. Writing out Equation (14) 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 $W^l_{nt} = \frac{(\nu - \beta_{ft}) P_{mt} Q_{ft} \mu_{ft}}{L_{ft}}$. 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 ξ^l_t. Third, 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 $\mathbf{\Phi}$ are assumed to be invariant to the level of labor and product market competition.

C Observed and counterfactual technologies

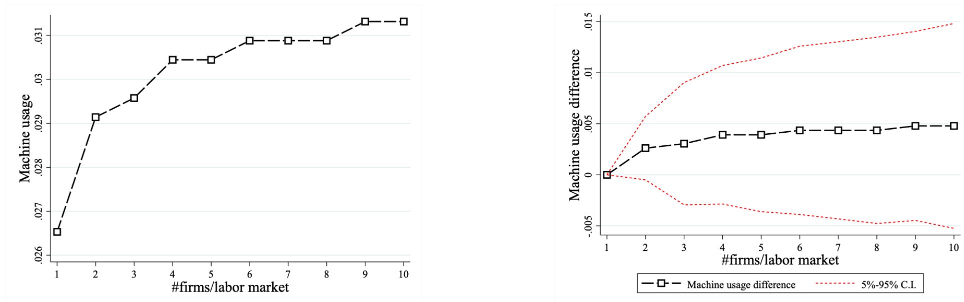
I simulate equilibrium technology usage under the various counterfactual labor market structures for four technologies: two observed and two counterfactual technologies. First, I consider cutting machines, which were unskill-biased, Hicks-neutral productivity-enhancing, and had a high fixed cost. Second, I consider hauling locomotives, which were skill-biased, Hicks-neutral productivity-increasing, and had a relatively low fixed cost. Third, I consider a counterfactual cutting machine with the exact same parametrization and fixed cost, except that it has the opposite direction of technological change, $-\alpha^{\beta,cut}$ rather than $\alpha^{\beta,cut}$. Fourth, I consider a counterfactual technology with the same fixed cost as cutting machines, but which is purely Hicks-neutral, meaning that $\alpha^{\beta,cut} = 0$. In order to compare only the effect of changes in the direction of technological change, I keep the Hicks-neutral effects $\boldsymbol{\alpha}^{\omega}$ of each technology constant across bootstrap iterations.

D Results

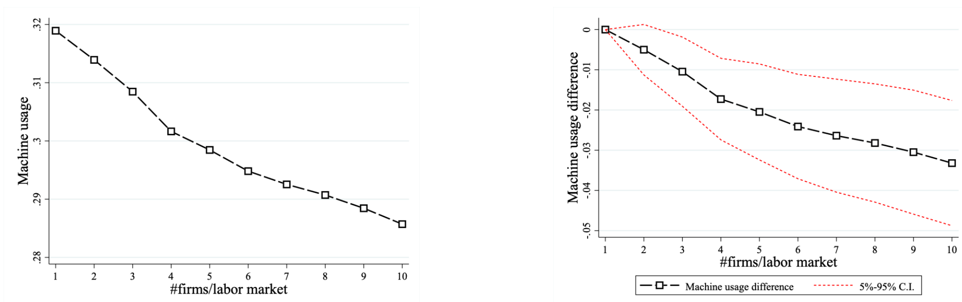
Observed technologies The results of the counterfactual exercise for all technologies are in Figure 4. The left figures plot technology usage across labor market structures, the right figures plot the difference in technology usage compared to monopsonistic labor markets for each labor market structure, with 5%-95% confidence intervals based on the bootstrapped standard errors. The result for the first observed technology, cutting machines, is in Figure 4(a). Cutting machine usage is estimated to increase with the number of firms per labor market: the usage rate increases from 2.65% of firms in monopsony to 3.13% if markets have 10 equally sized firms. This difference in usage rate between these two labor market structures is not statistically significant, which is in part due to the uncertainty over the value of the output elasticity effects $\boldsymbol{\alpha}^{\beta}$. The negative effect of labor market concentration on cutting machine adoption is the net result of the Hicks-neutral and directed effects of cutting machines, which counteract each other, as was explained in Proposition 3. Whereas the effect of labor market concentration on cutting machine usage is not statistically significant, its effect on the relative returns to cutting machine usage, $\frac{\tilde{\Pi}(1,K_{ft}^{loc},s_{ft}^{h})-\tilde{\Pi}(0,K_{ft}^{loc},s_{ft}^{h})}{\tilde{\Pi}(0,K_{ft}^{loc},s_{ft}^{h})}$, is statistically significant, as shown in Appendix Figure A8. This negative effect is in line with Theorem 1. Whereas adopting a cutting machine increases variable profits by 2.73% on average under monopsonistic labor markets, this return is 5.21% if there are 10 equally-sized firms on each labor market.

Figure 4: Technology usage and labor market structure

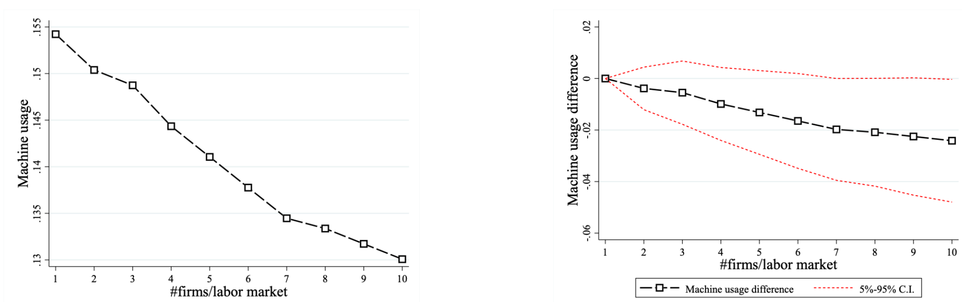
(a) Cutting machines



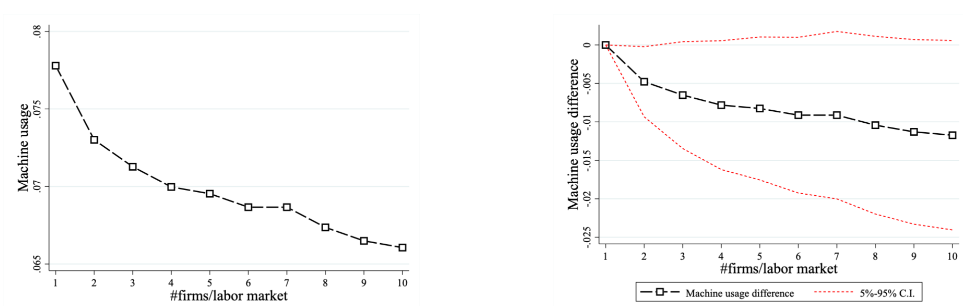
(b) Locomotives



(c) Skill-biased cutting machines



(d) Hicks-neutral cutting machines



Second, consider mining locomotives, which were biased towards skilled labor, rather than towards unskilled labor. Applying 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, such as mining locomotives. Figure 4(b) confirms this intuition: the usage rate of locomotives drops from 31.9% in a monopsonistic labor market to 28.6% with 10 firms per labor market, and this change is statistically significant. As shown in Appendix Figure A8, the variable profit return to locomotive adoption decreases substantially with the number of firms per labor market.

These counterfactual exercises explain the two motivating facts from Section 3.3. The first fact showed that cutting machine was lower in more concentrated labor markets, but locomotive usage higher. Given that unskilled labor markets were competitive, labor market structure only drives skilled miner wage markdowns; this is also consistent with the very low geographical variation in unskilled wages documented earlier. More concentrated labor markets were hence markets with higher skilled wage markdowns, and the counterfactual exercise above shows that this indeed led to lower cutting machine usage, but higher locomotive usage. The second descriptive fact showed that the usage of cutting machines increased after the introduction of wage bargaining in 1898 for the mines at which this bargaining led to a decrease in wage markdowns. Given that unskilled labor markets were already competitive prior to 1898, the move from oligopsonistic wage-setting to wage bargaining must have affected mainly skilled wage markdowns. The fact that a drop in these markdowns after 1898 led to increased cutting machine usage is, again, consistent with the results of the counterfactual exercise.

Counterfactual technologies What would happen if cutting machines would have had a different direction of technological change? Figure 4(c) shows that the usage rate of a counterfactual skill-biased cutting machine would be 5 to 6 times higher compared to the observed, unskill-biased, cutting machines. Moreover, the usage of such skill-biased cutting machines would fall with the number of firms per labor market, which is the opposite effect as for the observed cutting machines. The much higher usage rate for skill-biased technologies helps explaining why mining locomotives, which were introduced earlier than cutting machines, were adopted four times more rapidly than cutting machines, as is shown in Appendix Figure A3.

Finally, I consider a counterfactual Hicks-neutral cutting machine. In line with the theory, the relative profit effect of such a technology is invariant to labor market structure, as shown in Appendix Figure A8. However, the usage rate of this Hicks-neutral technology still falls with the number of firms per labor market, as shown in Figure 4(d), although this drop is merely borderline significant. The absolute profit gain from technology usage falls with additional firms per labor market: in order to recover the fixed technology costs, some degree of market power is required.

4.5 Discussion

A Effect sizes

The first two counterfactual exercises above explain and replicate the descriptive facts about the correlation between the usage of different technologies, market structure, and wage-setting conduct. The last two counterfactual exercises show that the direction of technological change plays a crucial role in shaping the effect of labor market power on technology adoption incentives: keeping all technology characteristics fixed, but muting or inverting its directional effects, results in an opposite effect of labor market concentration on technology usage.

It is remarkable that the effects on technology usage are relatively large, considering the modest degree of oligopsony power inferred from the model. At the average firm, the marginal revenue product of miners is 4.5% above their wage, and even under a pure monopsony, this markdown ratio would still be merely 15.7%. The current literature on oligopsony power usually finds higher markdowns. For instance, Azar et al. (2022) finds a markdown ratio of 21% using current U.S. data. With higher markdown levels, the effects of oligopsony power on technology usage would be even more pronounced.

B Adoption vs. invention

Throughout the paper, I took the invention of new technologies and their directionality as given, in contrast to, for instance, Acemoglu (2002). Given that invention is likely impacted by the demand for new technologies, which depends on usage rates, it is likely 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.

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 demand, I show that the effects of factor market power on technology usage are ambiguous, and depend on the direction of technical change, Hicks-neutral productivity effects, and which inputs firms have market power over. 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 competition. In terms of technology usage, I find that oligopsony power on labor markets had a negative effect on cutting machine adoption, but this effect would have been positive if cutting machines would have been either skill-biased or Hicks-neutral.

Although the application in this paper is historical, the results have several important current-day implications for industries with imperfectly competitive factor markets. The model shows that in order to understand the effects of labor market power on technological change today, it is crucial 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, factor markets, and, as the application in this paper showed, across technologies. In the case of labor markets, the consensus seems to be that automation has been mainly skill-biased throughout the last couple of decades. If firms mainly exert market power over unskilled workers, then it is likely that such market power is reducing the returns to automation. Much less is known, however, about the relative degrees of employer market power across the skill and income distribution, which is a crucial element in order to understand how labor market power shapes technology choices today.

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.

Affolter, R. H., & Hatch, J. R. (2002). Characterization of Quality of Coals from the Illinois Basin. In R. H. Affolter & J. R. Hatch (Eds.), (Vol. 1625, p. 1-227). U.S. Geological Survey.

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

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.

Arnold, D. (2019). Mergers and acquisitions, local labor market concentration, and worker outcomes. *Local Labor Market Concentration, and Worker Outcomes (October 27, 2019)*.

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

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. A., Berry, S. T., & Marinescu, I. (2022). *Estimating labor market power* (Tech. Rep.). 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. (2022). Labor market power. *American Economic Review*, *112*(4), 1147–93.

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.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347-1393.

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.

Caldwell, S., & Danieli, O. (2022). Outside options in the labor market.

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.*

Delabastita, V., & Rubens, M. (2022). Monopsony and industrialization. *Available at SSRN*.

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

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., & 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.

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. (2020). Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996–2016. *The Review of Economic Studies*, *87*(6), 2672–2702.

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

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. (2022). Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, *112*(1), 169–212.

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. (2022). Incomplete information bargaining with applications to mergers, investment, and vertical integration. *American Economic Review*, *112*(2), 616–49.

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

Lord, J. S. (1892). *Coal in Illinois, 1884-1892* (Biennial Reports No. 2-7).

Machin, S., & Van Reenen, J. (1998). Technology and changes in skill structure: Evidence from seven OECD countries. *The Quarterly Journal of Economics*, *113*(4), 1215–1244.

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.

Méndez-Chacón, E., & Van Patten, D. (2022). Multinationals, monopsony and local development: Evidence from the United Fruit Company. *Econometrica*(Forthcoming).

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*.

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.).

Prager, E., & Schmitt, M. (2021). Employer consolidation and wages: Evidence from hospitals. *American Economic Review*, *111*(2), 397–427.

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

Engineers. North of England Institute of Mining and Mechanical Engineers.

Rubens, M. (2020a). Managerial knowledge and technology choice: Evidence from US mining schools. *FEB Research Report Department of Economics*.

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

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

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. (2021). Noncompetes in the us labor force. *Journal of Law and Economics*.

Stoek, H., Hippard, C., & Langtry, W. (1920). Bituminous coal storage practice. *Bulletin of the University of Illinois Engineering Experiment Station*, *116*, 1-142.

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

Williams, T. (1901). The anthracite coal crisis. *The Atlantic*, *April*.

Appendices

A Appendix: Proofs

A Proof of Lemma 1

To prove: For $\mu^l_f = \psi^l_f$: $\frac{\partial \beta^h_f}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\; \Rightarrow \frac{\partial^2(\ln(\Pi_f))}{\partial \mu^h_f \partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$

Proof: Omit firm subscripts f for simplicity, and denote $\beta \equiv \beta^h_f$, and $\tilde{\eta} \equiv 1 + \eta$. Let $\mu^l = \psi^l$ and consider a change in the markdown over input H, μ^h. The proof is analogous when assuming $\mu^h = \psi^h$ and considering changes in μ^l. Note that $1 \leq \mu^h \leq \psi^h$; $\psi^h \geq 1$, $0 \leq \beta \leq \nu$, and $0 \leq \tilde{\eta} \leq 1$. Using Equations (5a)-(5b), variable profits are given by:

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

Define $\pi \equiv \ln(\Pi)$. Variable profits are weakly positive due to the economic restrictions on the parameter values. The first derivative of log profits with respect to the markdown is:

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

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

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

Given that $\Pi \geq 0$ and that $\pi(.)$ is twice differentiable, $\frac{\partial^2(\pi)}{\partial\beta\partial\mu^h} \geq 0 \Leftrightarrow \frac{\partial^2(\pi)}{\partial\mu^h\partial\beta} \geq 0$. The denominator of this expression is always positive. The numerator is weakly positive if expression (15) holds.

$$(\frac{\mu^l-\nu\tilde{\eta}}{\psi^l-\nu\tilde{\eta}})\frac{\psi^l\psi^h}{\mu^l\mu^h} \geq \frac{(1-\frac{\beta\tilde{\eta}}{\mu^h}-\frac{(\nu-\beta)\tilde{\eta}}{\mu^l})^2}{(1-\frac{\beta\tilde{\eta}}{\psi^h}-\frac{(\nu-\beta)\tilde{\eta}}{\psi^l})^2} \tag{15}$$

The right-hand side of (15) is weakly smaller than one, because $\mu^h \leq \psi^h$ and $\mu^l =$

$\psi^l = 1$. For the same reason, the left-hand side is weakly larger than one, so it holds that $\frac{\partial}{\partial\beta}(\frac{\partial\pi}{\partial\mu^h}) \geq 0$.

Suppose that $\frac{\partial\beta_f^h}{\partial K_f} \geq 0$. Then, $\frac{\partial^2(\pi)}{\partial\mu^h\partial\beta}$ has the same sign as $\frac{\partial^2(\pi)}{\partial\mu^h\partial K}$, so it follows that $\frac{\partial^2(\pi)}{\partial\mu^h\partial K} \geq 0$. In contrast, if $\frac{\partial\beta_f^h}{\partial K_f} \leq 0$, then $\frac{\partial^2(\pi)}{\partial\mu^h\partial\beta}$ has the opposite sign of $\frac{\partial^2(\pi)}{\partial\mu^h\partial K}$. It follows that $\frac{\partial^2(\pi)}{\partial\mu^h\partial K} \leq 0$ □

B Proof of Theorem 1

To prove: For $\psi_f^l = 1$: $\frac{\partial\beta_f}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\ln(\Pi_f))}{\partial\mu_f^h\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$

Proof: Theorem 1 follows immediately from Lemma 1: if $\psi^l = 1$, this implies that $\mu^l = \psi^l$, because $\mu^l \geq 1$: input suppliers are never paid more than their marginal revenue product. □

C Proof of Proposition 1

To prove: For $\mu_f^l < \psi_f^l$and $\psi_f^l \neq 1$:
$\frac{\partial\beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\ln(\Pi_f))}{\partial\mu_f^h\partial K_f} \begin{Bmatrix} \geq \text{ or } \leq \end{Bmatrix} 0$

Proof: This proposition can be proven by a numerical example. Consider the case in which $\frac{\partial\beta_f^h}{\partial K_f} \geq 0$. Take $\psi^l = \psi^h = 2$, $\beta = 0.1$, $\mu^l = 1$, $\mu^h = \psi^h$, $\sigma = 0.9$. Then, subtracting the right-hand side of expression (15) from its left-hand sign gives a positive number, 0.0715, meaning that $\frac{\partial^2(\pi)}{\partial\mu^h\partial K} \geq 0$. But taking another value for β, $\beta = 0.8$, gives a negative number for this subtraction, -0.429, meaning that $\frac{\partial^2(\pi)}{\partial\mu^h\partial K} \leq 0$. Hence, depending on the size of the markdowns and of the output elasticity, $\frac{\partial^2(\pi)}{\partial\mu^h\partial K}$ can have any sign, except if $\psi^l \geq 1$ or $\mu^l \leq \psi^l$, in which case Lemma 1 and Theorem 1 apply. □

D Proof of Proposition 2

To prove: For $\mu_f^l = \psi_f^l$ or $\psi_f^l = 1$:
$\frac{\partial\beta_f^h}{\partial K_f} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0 \;\Rightarrow\; \frac{\partial^2(\Pi_f)}{\partial\mu_f^h\partial K_f} \begin{Bmatrix} \geq \\ \geq \text{ or } \leq \end{Bmatrix} 0$

Proof: The first derivative of log profits to capital usage can be written as

$$\frac{\partial\ln(\Pi)}{\partial K} = \frac{1}{\Pi}\frac{\partial\Pi}{\partial K}$$

This implies that the effect of markdowns on the absolute profit effect of the technology can be decomposed as in equation (16).

$$\frac{\partial}{\partial \mu^h}(\frac{\partial \Pi}{\partial K}) = \underbrace{\Pi \frac{\partial}{\partial \mu^h}(\frac{\partial \ln(\Pi)}{\partial K})}_{(I)} + \underbrace{\frac{\partial \Pi}{\partial \mu^h}\frac{\partial \ln(\Pi)}{\partial K}}_{(II)} \tag{16}$$

Given that the firm is operating, variable profits are positive: $\Pi \geq 0$. The term $\frac{\partial \ln(\Pi)}{\partial K}$ is positive too, because otherwise the technology would decrease variable profits, which is not a relevant case because such a technology would never be adopted. The effect of markdowns on variable profits is positive: $\frac{\partial \Pi}{\partial \mu_f^h} \geq 0$. To see this, take the first derivative of variable profits with respect to the markdown μ^h, which gives:

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

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

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

It follows that term (II) is always positive. If $\frac{\partial \beta_f^h}{\partial K_f} \geq 0$, then term (I) is weakly positive, due to Theorem 1 and Lemma 1. Hence, the markdown increases the absolute profit return to the technology, $\frac{\partial}{\partial \mu^h}(\frac{\partial \Pi}{\partial K}) \geq 0$. In contrast, if $\frac{\partial \beta_f^h}{\partial K_f} \leq 0$, it follows that term (II) is weakly negative, again due to Theorem 1 and Lemma 1. The sign of $\frac{\partial}{\partial \mu^h}(\frac{\partial \Pi}{\partial K})$ is now ambiguous, depending on the relative size of terms (I) and (II). □

E Proof of Proposition 3

To prove: The higher $\frac{\partial \Omega}{\partial K}$, the more likely that $\frac{\partial}{\partial \mu^h}\frac{\partial \Pi}{\partial K} \geq 0$.

Proof: Denote the variable profit margin as $m \equiv \Big(1 - \frac{\beta\tilde{\eta}}{\mu^h} - \frac{(\nu-\beta)\tilde{\eta}}{\mu^l}\Big)$. Variable profits can hence be written as $\Pi = Q^{\tilde{\eta}} m$. The effect of technology usage on variable profits is:

$$\frac{\partial \Pi}{\partial K} = \tilde{\eta} Q^{\tilde{\eta}-1} \frac{\partial Q}{\partial \Omega}\frac{\partial \Omega}{\partial K} m + \tilde{\eta} Q^{\tilde{\eta}-1} \frac{\partial Q}{\partial \beta}\frac{\partial \beta}{\partial K} m + \frac{\partial m}{\partial \beta} Q^{\tilde{\eta}} + \underbrace{\frac{\partial m}{\partial \Omega}}_{=0} Q^{\tilde{\eta}}$$

Under the assumptions made, the variable profit margin m is positive. It is easy to see that Hicks-neutral productivity increases output, $\frac{\partial Q}{\partial \Omega} > 0$. Hence, the higher the effect of the technology on Hicks-neutral productivity $\frac{\partial\Omega}{\partial K}$, the higher its effect on profits $\frac{\partial\Pi}{\partial K}$. This also implies higher effects on log profits, $\frac{\partial\pi}{\partial K}$. Because markdowns increase variable profits, $\frac{\partial\Pi}{\partial\mu^h} \geq 0$, as was shown in the proof of Proposition 2, it follows from Equation (16) that a higher value of $\frac{\partial\Pi}{\partial K}$ makes the effect of the markdown on this absolute profit change more positive. $\square$

F Proof of Proposition 4

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

Proof: Given that $\frac{\partial\beta^h}{\partial K} = 0$, the variable profit effect of the technology is equal to:

$$\frac{\partial\Pi}{\partial K} = \frac{\partial\Pi}{\partial\Omega}\frac{\partial\Omega}{\partial K}$$

The first derivative of variable profits with respect to Hicks-neutral productivity is:

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

Taking the second derivative with respect to the markdown μ^h gives:

$$\frac{\partial}{\partial\mu^h}(\frac{\partial\Pi}{\partial\Omega}) = \frac{Q^{\tilde{\eta}}\beta\tilde{\eta}^2}{\Omega(\mu^h)^2\psi^h}\left(\psi^h - \tilde{\eta}\mu^h\frac{(1 - \frac{\beta\tilde{\eta}}{\mu^h} - \frac{(\nu-\beta)\tilde{\eta}}{\mu^l})}{(1 - \frac{\beta\tilde{\eta}}{\psi^h} - \frac{(\nu-\beta)\tilde{\eta}}{\psi^l})}\right)$$

This expression is weakly positive, because $1 \leq \mu^h \leq \psi^h$, $1 \leq \mu^l \leq \psi^l$, and $0 \leq \tilde{\eta} \leq 1$. This implies that

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

Similarly, the second derivative with respect to the other markdown, μ^l , is

$$\frac{\partial}{\partial\mu^l}(\frac{\partial\Pi}{\partial\Omega}) = \frac{Q^{\tilde{\eta}}(\nu-\beta)\tilde{\eta}^2}{\Omega(\mu^l)^2\psi^l}\left(\psi^l - \tilde{\eta}\mu^l\frac{(1 - \frac{\beta\tilde{\eta}}{\mu^h} - \frac{(\nu-\beta)\tilde{\eta}}{\mu^l})}{(1 - \frac{\beta\tilde{\eta}}{\psi^h} - \frac{(\nu-\beta)\tilde{\eta}}{\psi^l})}\right)$$

Again, this expression is weakly positive, which implies:

$$\frac{\partial\Omega}{\partial K_f}\begin{Bmatrix}\geq\\\leq\end{Bmatrix}0 \Rightarrow \frac{\partial^2(\Pi_f)}{\partial\mu_f^l\partial K_f}\begin{Bmatrix}\geq\\\leq\end{Bmatrix}0 \qquad \square$$

B Appendix: Data

B.1 Sources

Mine Inspector Reports The main data source is the biennial report of the Bureau of Labor Statistics of Illinois, of which I collected the volumes between 1884 and 1902. Each report contains a list of all mines in each county, and reports the name of the mine owner, the town in which the mine is located, and a selection of variables that varies across the volumes. An overview of all the variables (including unused ones), and the years in which they are observed, is in Tables A7 and A8. Output quantities, the number of miners and other employees, mine-gate coal prices, and information on 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 The 1888 report contains monthly production data for a selection of 11 mines in Illinois, across 6 counties. I observe the monthly number of days worked and the 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.

B.2 Data cleaning

Employment In every year except 1896, workers are divided into two categories, 'miners' and 'other employees'. In 1896, a different d istinction is made, between 'underground workers' and 'above-ground workers'. This does not correspond to the miner-others categorization because all miners were underground workers, but some underground workers were not miners (e.g. doorboys, mule drivers, etc.). Hence, I do not use the 1896 data. From 1888 to 1896, boys are reported as a separate working category. Given that miners (cutters) were adults, I include these boys in 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 except 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 no longer reported seasonally, because wages were negotiated biennially from that year onwards. 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 quantity and price 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. Mine-gate 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 motivating 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.[30]

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 the total number of hours worked, the labor quantity definition does not change after 1898. Given that the model is estimated on the pre-1898 period, this does not affect the model estimates, only the descriptive evidence.

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

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

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.

Town identifiers and labor market definitions The data set contains 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 374. 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 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 worker 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.

C Appendix: Empirics

C.1 Equilibrium expressions for the 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, (11), and using Equations (6a), (7), and (8), which are respectively the production, coal demand, and labor supply functions. The resulting equilibrium output expression is in Equation (17a), which is the empirical analogue of Equation (5a) 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 (17a) simplifies to Equation (5a).

$$Q^*_{ft} = \left[\left(\frac{\beta_{ft}(s^h_{ft})^{\psi^h_{nt}-1}(1 + s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^{\eta} \exp(\zeta_{ft})}{((\psi^h_{nt} - 1)s^h_{ft} + 1) \exp(\xi^h_{nt})} \right)^{\frac{\beta_{ft}}{\psi^h_{nt}}} \left(\frac{\nu - \beta_{ft}(s^l_{ft})^{\psi^l_{nt}-1}(1 + s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^{\eta} \exp(\zeta_{ft})}{((\psi^l_{nt} - 1)s^l_{ft} + 1) \exp(\xi^l_{nt})} \right)^{\frac{\nu-\beta_{ft}}{\psi^l_{nt}}} \Omega_{ft} \right]^{\frac{1}{1-\frac{\beta_{ft}(\eta+1)}{\psi^h_{nt}}-\frac{\nu-\beta_{ft}(\eta+1)}{\psi^l_{nt}}}} \tag{17a}$$

The equilibrium coal price is $P^*_{mt} = {Q^*_{mt}}^{\eta}\zeta_{mt}$. The equilibrium quantities of both labor types are then given by Equation (17b):

$$\begin{cases} H^*_{ft} &= \left(\frac{\beta_{ft}Q^*_{ft}P^*_{mt}(1+s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^{\eta}}{((\psi^h_{mt}-1)s^h_{ft}+1)\xi^h_{mt}} \right)^{\frac{1}{\psi^h_{mt}}} (s^h_{ft})^{\frac{\psi^h_{mt}-1}{\psi^h_{mt}}} \\ L^*_{ft} &= \left(\frac{\nu-\beta_{ft}Q^*_{ft}P^*_{mt}(1+s^q_{ft}\eta)(\frac{1}{s^q_{ft}})^{\eta}}{((\psi^l_{nt}-1)s^l_{ft}+1)\xi^l_{nt}} \right)^{\frac{1}{\psi^l_{nt}}} (s^l_{ft})^{\frac{\psi^l_{nt}-1}{\psi^l_{nt}}} \end{cases} \tag{17b}$$

Substituting the equilibrium labor quantities from (17b) into the labor supply functions in (8) gives the expression for equilibrium wages, Equation (17c).

$$\begin{cases} W^{h*}_{mt} &= (\frac{\beta_{ft}P^*_{ft}Q^*_{ft}(1+\eta)}{((\psi^h_{mt}-1)s^h_{ft}+1)s^h_{ft}})^{\frac{\psi^h_{mt}-1}{\psi^h_{mt}}} (\exp(\xi^h_{mt}))^{\frac{1}{\psi^h_{mt}}} \\ W^{l*}_{nt} &= (\frac{\nu-\beta_{ft}P^*_{ft}Q^*_{ft}(1+\eta)}{((\psi^l_{nt}-1)s^l_{ft}+1)s^l_{ft}})^{\frac{\psi^l_{nt}-1}{\psi^l_{nt}}} (\exp(\xi^l_{nt}))^{\frac{1}{\psi^l_{nt}}} \end{cases} \tag{17c}$$

C.2 Alternative production model

In the main text, I assumed that the scale parameter ν was equal to 0.9 and imposed a homogeneous goods Cournot model on the coal market to estimate markups. As a robustness check, I use an alternative model in which I estimate the scale parameter and do not impose a demand model on the coal market. This comes at the cost of having to assume that there is no unobserved heterogeneity in output elasticities β_f across firms and time.

Production Equation (18) is an alternative production function in skilled, unskilled labor, materials (black powder), and capital, with interaction terms between each labor type and each capital technology.

$$q_{ft} = \beta^h h_{ft} + \beta^l l_{ft} + \beta^{hk}_{cut} h_{ft} K^{cut}_{ft} + \beta^{hk}_{loc} h_{ft} K^{loc}_{ft} + \beta^{lk}_{cut} l_{ft} K^{cut}_{ft} + \beta^{lk}_{loc} l_{ft} K^{loc}_{ft} + \beta^{k}_{cut} K^{cut}_{ft} + \beta^{k}_{loc} K^{loc}_{ft} + \beta^m m_{ft} + \omega_{ft} \quad (18)$$

I assume that cutting machines and locomotives do not change the degree of returns to scale in both labor inputs, which implies that $\beta^{hk}_{cut} = -\beta^{lk}_{cut}$ and $\beta^{hk}_{loc} = -\beta^{lk}_{loc}$. I keep the timing assumptions on the input demand problem from the main text, and impose an AR(1) process for total factor productivity with a productivity shock ε_{ft}:

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

As was explained in the main text, the input timing decisions correspond to the following moment conditions, which I estimate up to one time lag, again as in the main text.

$$\mathbb{E}\Big[\varepsilon_{ft} \Big| \left\{ \begin{matrix} h_{f\theta-1} \\ l_{f\theta-1} \\ m_{\theta-1t} \\ \mathbf{K}_{f\theta} \\ \mathbf{K}_{f\theta} h_{f\theta-1} \\ \mathbf{K}_{f\theta} l_{f\theta-1} \end{matrix} \right\}_{\theta=1}^{t} \Big] = 0$$

The markup μ_{ft} can be expressed as the ratio of the output elasticity of miners over the product of its revenue share and markdown:[31]

$$\mu_{ft} = \frac{\beta^h + \beta^{hk}_{cut} K^{cut}_{ft} + \beta^{hk}_{loc} K^{loc}_{ft}}{\frac{W^l_{ft} H_{ft}}{P_{ft} Q_{ft}} \psi^h_{ft}}$$

Results The results of this alternative production model are in Table A5. Coal cutting machines are still unskill-biased: the output elasticity of miners is estimated to fall by 0.346 points when adopting a cutting machine, coming from 0.695. In the baseline model, this was a smaller drop of 0.160 points, down from 0.734. In contrast, the usage of mining locomotives is estimated to barely change the output elasticity of skilled workers, in contrast to the main specification where it increased this output elasticity compared to unskilled workers. The scale parameter, ν, is estimated at 0.856, 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 μ is estimated at 1.139, which implies that the coal price is 13.9% above marginal costs. This estimate does not impose any model of competition on the coal market. The homogeneous goods Cournot model in the baseline model delivered a very similar average markup ratio of 1.148, which is close to the markup in the alternative model that does not impose Cournot competition.

C.3 Alternative labor supply model

In the main text, I estimated the skilled labor supply curve, Equation (8) by relying on intertemporal variation in labor demand. This results in a short-run labor supply elasticity. In this robustness check, I re-estimate the skilled labor supply curve using a cross-sectional demand shifter instead. Given that railroads were used for coal transport, but not for passenger services, being located on a railroad line should shift coal demand, but not miner supply. Therefore, I re-estimate the inverse skilled labor supply function using dummies for railroad connections and railroad crossings as instruments for the amount of employees in each town. I control for the minimum log distance to either Chicago and St. Louis, to control for town locations within Illinois, and for year fixed effects. The results are in Table A6. The resulting inverse supply elasticity is 0.061, compred to 0.157 in the main version of the model. This difference is intuitive: labor supply should be more elastic on the long run compared to the short run. This alternative model comes at the cost, however, of assuming identical labor supply elasticities across towns and over time, whereas the model

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

in the main text allowed supply elasticities to vary flexibly across towns and over time. In Figure A7, I compare the counterfactual exercises for technology usage between the baseline model (on the left) and the alternative labor supply model (on the right). The resulting changes in machine usage in response to changes in labor market structure are generally a bit smaller with the alternative labor supply model, but the differences are minimal.

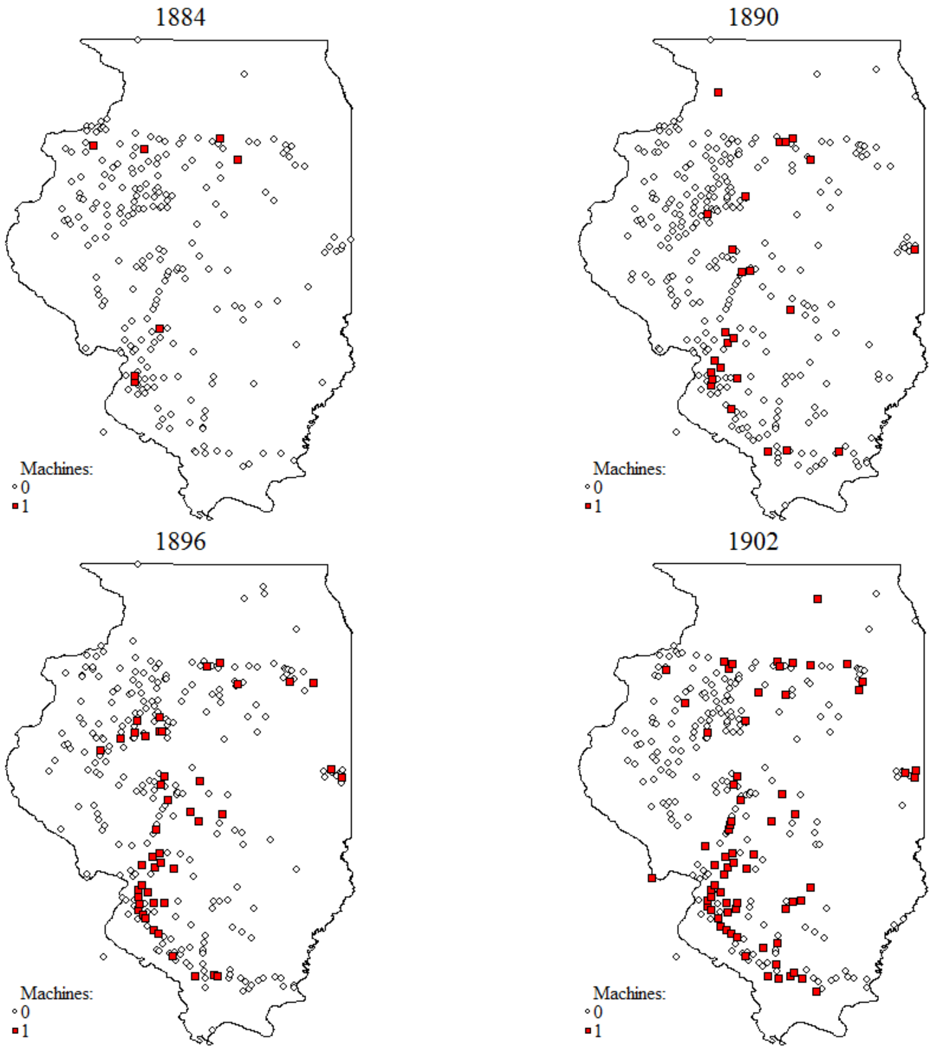
C.4 Cost dynamics

Cost dynamics would invalidate the assumed AR(1) productivity transition, Equation (6c). For instance, if it becomes increasingly costly to operate deeper mines, productivity would fall with past cumulative output. Another violation of the AR(1) process could be due to learning by doing, as in Benkard (2000), but productivity would then increase with cumulative output. I test for cost dynamics by regressing the logarithm of the Hicks-neutral productivity residual ω_{ft} on log cumulative output. The estimated coefficients are in Table A2. If not including mine fixed effects, lagged cumulative output is associated with higher total factor productivity. However, this could be due to selection: more productive mines are more likely to have extracted and sold more coal in the past. 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, which defends the assumption of no cost dynamics made in Equation (6c).

C.5 Inverse miner supply elasticity: correlations

I regress the log town-level inverse miner supply elasticity $\ln(\psi^h_{nt})$ on observed town and county characteristics. A higher town-level inverse supply elasticity implies more inelastic miner supply. The results are in Appendix Table A3. First, miner supply is more inelastic if total coal employment as a share of the town population is higher, which implies fewer outside work opportunities. 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 fewer outside work opportunities to switch to. Third, the population share of African Americans in the county does not correlate significantly with the miner supply elasticity. Fourth, the miner supply elasticity does not differ with the share of firms connected to the railroad network, which is in line with historical evidence that railroads were not used to transport workers.[32] Finally, the average wage in manufacturing industries in the same county correlates positively, but not significantly, with the inverse 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.

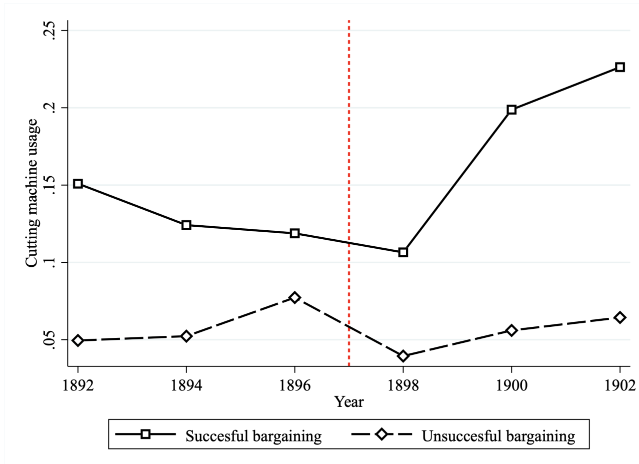
Figure A1: Geographical spread of cutting machines



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

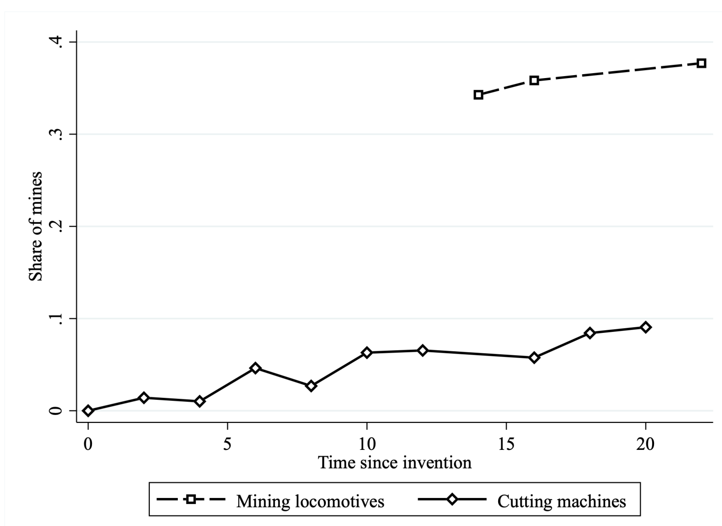
[32] If mining towns would not be isolated due to workers commuting by train, being connected to the railroad network should result in more elastic labor supply.

Figure A2: Cutting machine usage pre- and post-unionization



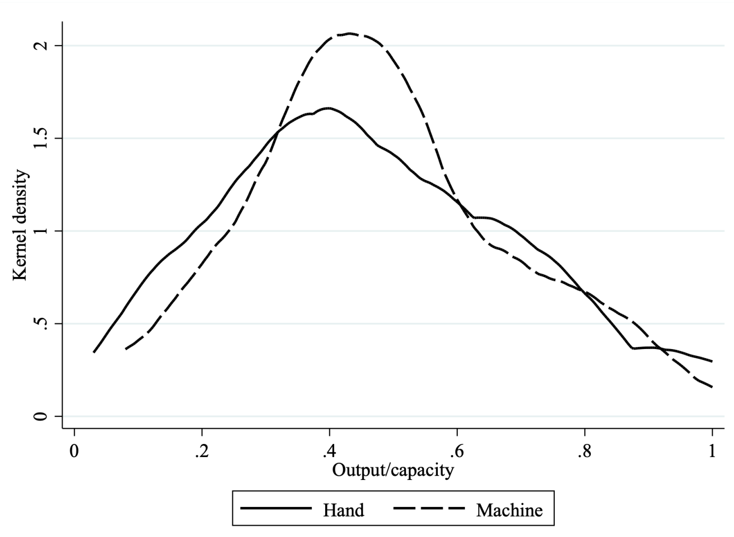
Notes: This figure compares average cutting machine usage between mines at which the introduction of wage bargaining in 1898 resulted in increased miner wages, and mines at which wages remained unchanged. The dotted line indicates the large Illinois coal strikes of 1897, which led to the introduction of wage bargaining.

Figure A3: Cutting machine vs. locomotive adoption



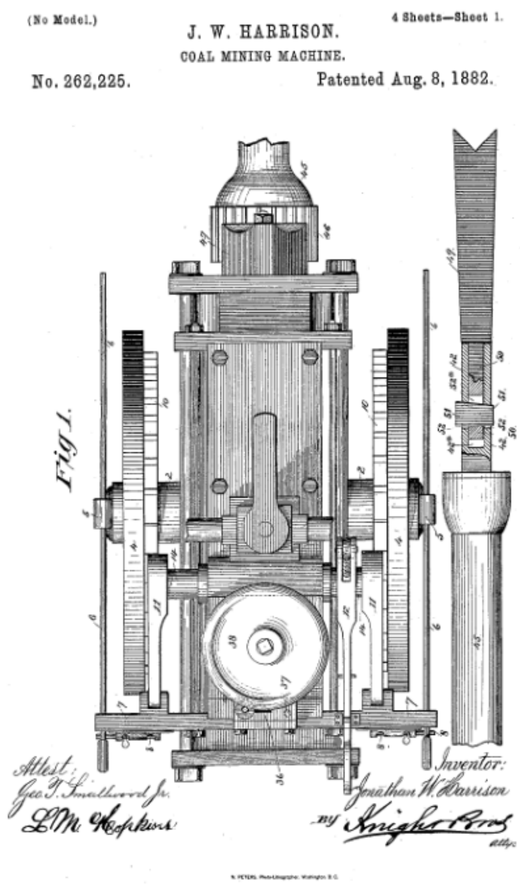
Notes: This figure compares the share of Illinois mines using locomotives and cutting machines against how long these technologies have been invented.

Figure A4: 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 so.

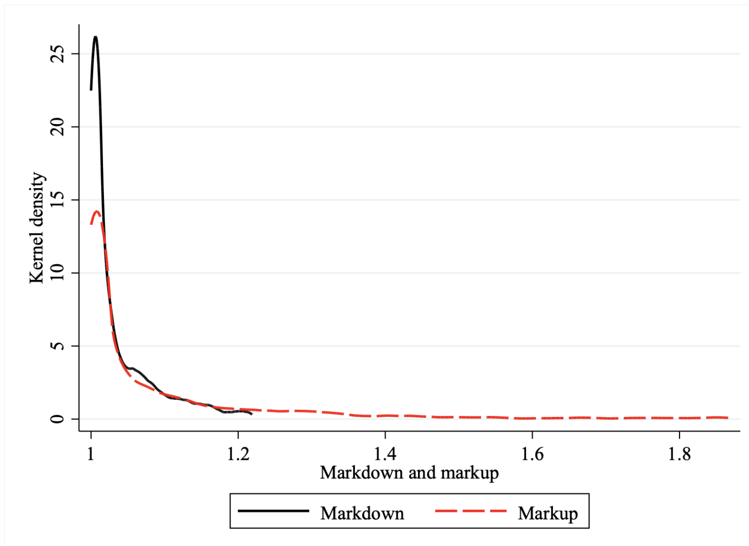
Figure A5: Patent of the Harrison Cutting Machine



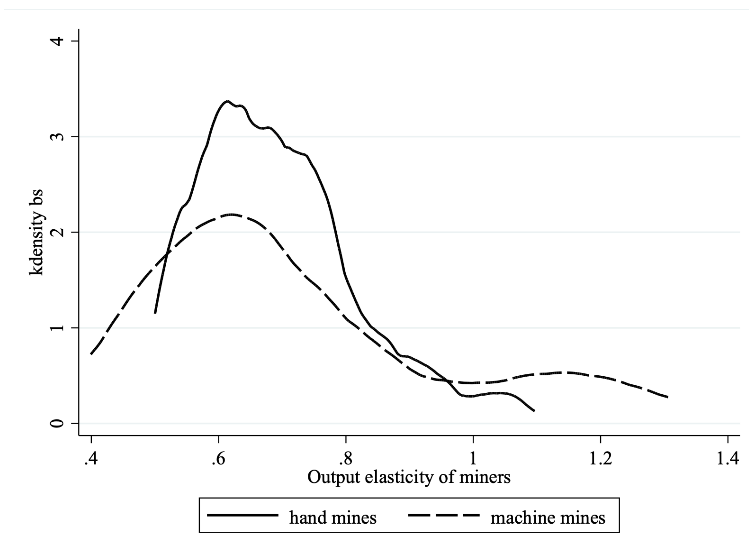
Notes: U.S. patent of the 1882 Improved Harrison Coal Cutting Machine (Whitcomb, 1882). This was the most frequently used coal cutting machine in the data set.

Figure A6: Distributions of latent variables

(a) Markdowns and markups



(b) Skilled labor output elasticity

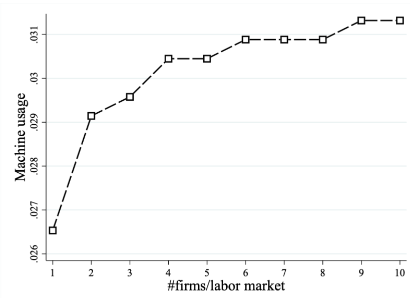


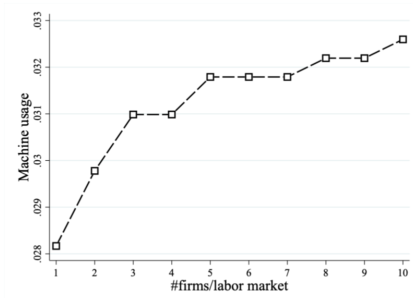
Notes: Panel (a) shows the distribution of the miner wage markdown and coal price markup across mines between 1884-1894. Panel (b) shows the distribution of miner output elasticities for firms using at least one cutting machines ('machine mines' and firms that did not use any cutting machines ('hand mines'). All distributions are censored at their 5th and 95th percentile.

Figure A7: Technology usage and labor market structure - alternative supply model

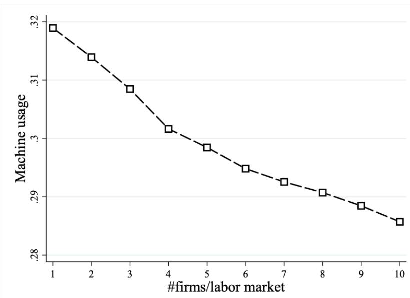
Main model: **Alternative model:**

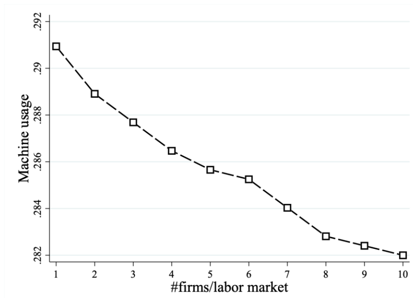
(a) Cutting machines



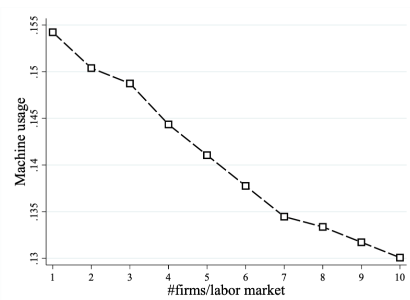


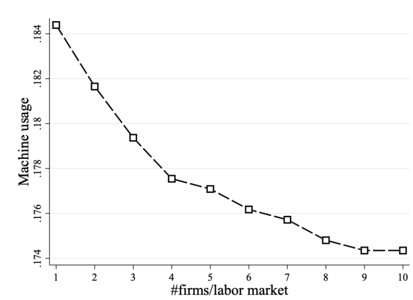
(b) Locomotives



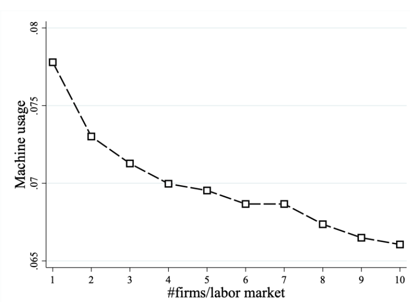


(c) Skill-biased cutting machines





(d) Hicks-neutral cutting machines



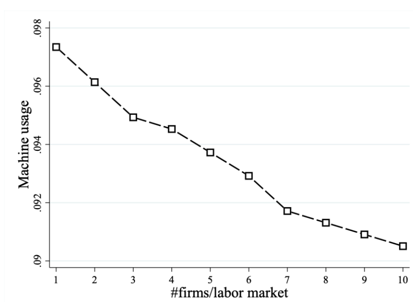
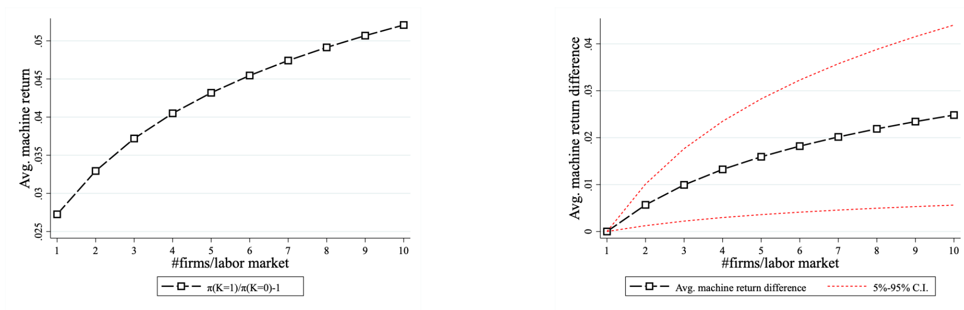
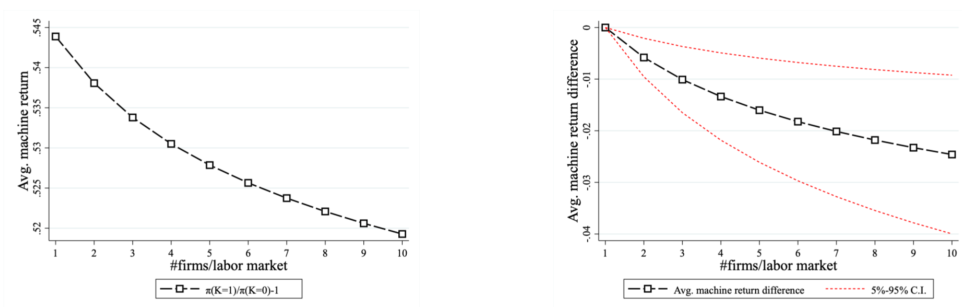


Figure A8: Variable profit returns and labor market structure

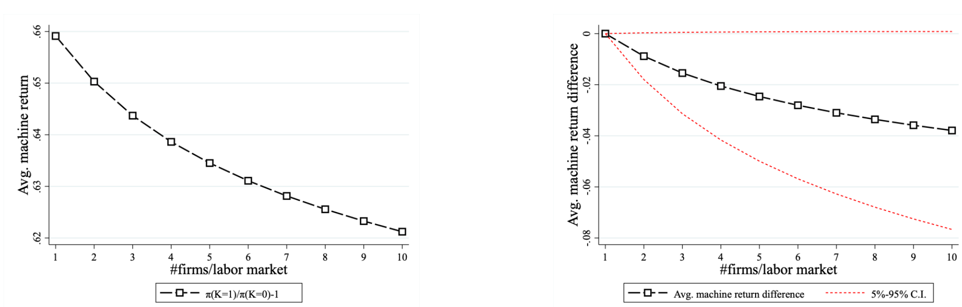
(a) Cutting machines



(b) Locomotives



(c) Skill-biased cutting machines



(d) Hicks-neutral cutting machines

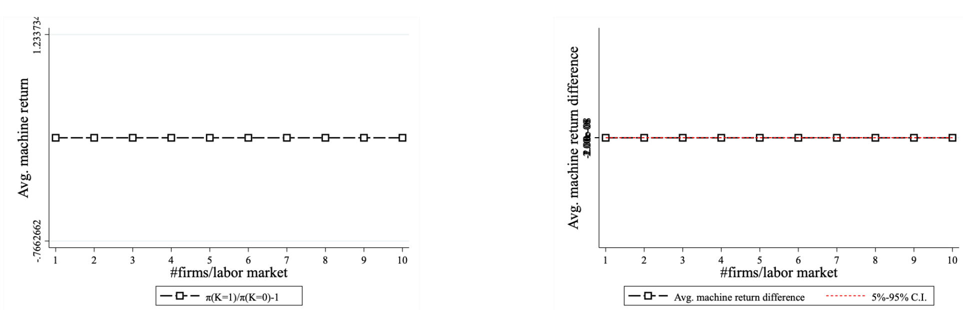


Table A1: Occupations and wages

	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))			
	Est.	S.E.	Est.	S.E.
log(Cum. output)	0.126	0.004	0.010	0.016
Mine FE	No		Yes	
Observations	3520		3520	
R-squared	.336		.816	

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(Inverse miner supply elasticity+1)	
	Est.	S.E.
log(Coal employment share)	0.015	0.005
log(Farmland/Total Area)	-0.113	0.068
log(African Americans / Population)	-0.003	0.004
Share of firms connected to railroad	-0.003	0.012
log(Manufacturing wage)	0.045	0.026
R-squared	.113	
Observations	831	

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.	S.E.
Coal demand elasticity	-0.465	0.101
1(Railroad connection)	0.353	0.123
1(Railroad crossing)	0.689	0.209
log(Dist. to St. Louis)	-0.069	0.071
log(Dist. to Chicago)	-0.044	0.070
Observations	453	
F-stat 1st stage	70.1	
R-squared	.191	

(b) Output elasticity transition	log(Output elasticity of skilled miners)	
	Est.	S.E.
1(Cutting machine)	-0.160	0.068
1(Locomotive)	0.101	0.029
log(Materials)	0.005	0.017
Year	-0.011	0.013
Constant	21.111	25.417
Observations	1149	
R-squared	.008	

(c) Hicks-neutral productivity transition	log(Hicks-neutral productivity)	
	Est.	S.E.
1(Cutting machine)	0.218	0.157
1(Locomotive)	0.277	0.182
log(Materials)	0.105	0.130
Year	0.028	0.197
Constant	-51.688	.
Observations	1066	
R-squared	.225	

Table A5: Alternative production model

(a) Production function	log(Output) Est.	S.E.
log(Skilled labor)	0.695	0.345
log(Unskilled labor)	0.161	0.373
log(Skilled labor/Unskilled labor)*1(Cutting machine)	-0.346	0.317
log(Skilled labor/Unskilled labor)*1(Locomotive)	0.021	0.246
1(Cutting machine)	0.401	0.297
1(Locomotive)	-0.109	0.374
R-squared		
Observations		
(b) Markup and returns to scale		
Returns to scale	0.856	0.168
Markup	1.139	0.520

Notes: Alternative production function that estimates markup and degrees to scale, as specified in Appendix C.2. Standard errors are block-bootstrapped with 200 iterations.

Table A6: Alternative labor supply model

	log(Employment) Estimate	S.E.
log(Wage)	0.061	0.022
log(Min. distance to Chicago or St. Louis)	0.111	0.031
R-squared	.010	
Observations	1097	

Table A7: 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 A8: 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 vein 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							