Labor-Market Power, Deadweight Loss, and Technology Adoption

Michael Rubens*

July 23, 2025

Abstract

Buyer power can induce deadweight loss, but it can also incentivize buyers' technology adoption by reducing investment holdup. In this paper, I construct a structural model that incorporates these two opposing forces, and I use it to quantify the net welfare effects of employers' market power over their workers. Applying the model to the late-19th-century Illinois coal mining industry, a textbook monopsony example that experienced a large technological shock due to the invention of mechanical cutting machines, I find that an increase in employer power leads to substantially higher mechanization rates. Assuming exogenous capital investment would lead to overestimating the consumer and labor welfare losses from employer power by 13% and 7%.

Keywords: Monopsony, Buyer Power, Investment, Technological Change

JEL codes: L11, L13, J42, N51

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

This paper previously circulated under the titles "Monopsony Power and Factor-Biased Technology Adoption" and "Oligopsony 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, and Daniel Haanwinckel for feedback and discussion, as well as participants at various seminars and conferences. Financial support from the Belgian American Educational Foundation, the Fulbright program, the Research Foundation Flanders (FWO grant 11A4618N), the European Research Council (ERC grant 816638), the Terry and Susan Kramer Family Foundation, and the NBER Program on Economics of Innovation in the Energy Sector is gratefully acknowledged.

1 Introduction

There is growing empirical evidence for the existence of monopsony power across various industries, countries, and types of factor markets.[1] An increasingly large literature has examined the aggregate welfare consequences of monopsony and oligopsony power (Berger et al., 2022; Lamadon et al., 2022). This literature has typically assumed firms' technology choices and investments to be exogenous to the degree of buyer power. However, buyer power could increase investment in either human or physical capital by mitigating investment holdup problems (Williamson, 1971; Joskow, 1987). This trade-off between anticompetitive distortions and endogenous investment plays an important role in various debates around labor-market power, such as regulation of noncompete agreements (Starr, Prescott, & Bishara, 2021; Shi, 2023), and the role of buyer power in horizontal merger control (Hemphill & Rose, 2018; Loertscher & Marx, 2019).

In this paper, I theoretically and empirically examine the welfare effects of employer power while allowing for these two countervailing forces. I construct a model in which employer power both leads to monopsony-induced deadweight loss, as employers cut back on input usage in order to push down input prices, but also incentivizes the adoption of new productivity-enhancing technologies, as it lets employers appropriate more of the rents from innovation. Although the model is written in terms of employers and employees, it applies more generally to vertical relationships between buyers and sellers.

I construct a bargaining model of wage negotiations between workers and employers with upward-sloping labor supply that nests the classical monopsony wage-posting model as a limit case in which all bargaining power is on the employer side. I conduct comparative statics in terms of '"employer power," which I define as the employer's bargaining ability. Employers combine different labor types to produce output using a constant elasticity of substitution (CES) production function. Employers choose whether to adopt a new technology, which increases both Hicks-neutral and factor-specific productivity levels but requires incurring a fixed cost. Using the model, I show that the net effect of employer power on output, consumer surplus, and total welfare is ambiguous, as the output-employer power relationship assumes an inverted U-shape. The output-

[1] See literature reviews by Ashenfelter et al. (2010) and Manning (2011), and recent papers by, among many others, Naidu et al. (2016); Berger et al. (2022); Morlacco (2017); Lamadon et al. (2022); Kroft et al. (2020); Rubens (2023); Chambolle et al. (2023).

maximizing level of employer power depends on the relative size of (i) the elasticity of labor supply and (ii) the productivity effects of the technology.

Given that the net effect of employer power on output is theoretically ambiguous, I empirically implement the model in the context of the Illinois coal mining industry between 1884 and 1902. There are three main reasons why this provides a unique and interesting setting to study the relationship between buyer power and innovation. First, the introduction of coal-cutting machines in the United States in 1882 provides a large technological shock, and I observe the gradual mechanization of this industry. Second, 19th-century Illinois coal mining towns are textbook examples of monopsonistic labor markets, with geographically isolated local labor markets. Prior work has found evidence for substantial monopsony power during the second industrial revolution (Boal, 1995; Naidu & Yuchtman, 2017; Delabastita & Rubens, 2024). Third, a series of large strikes by nascent miner unions in 1898 provides in-sample variation in employer bargaining power, which I use to validate the model predictions.

I estimate the model using a novel, uniquely rich archival dataset on mine-level production, coal prices, input quantities and prices, technology usage, and geological data. I rely on observed variation in the thickness of coal veins as cost shifters to estimate coal demand, and on international coal-price shocks to estimate labor supply. Identifying the production function relies, as usual, on timing assumptions on input choices as a function of both Hicks-neutral and labor-augmenting productivity shocks. In line with anecdotal historical evidence, I find that cutting machines were unskill-biased, similarly to many other technologies developed throughout the 19th century (Mokyr, 1990; Goldin & Katz, 2009).

Using the estimated production, labor supply, and coal demand models, I estimate the relative bargaining weights of the employers and the coal miner unions. I find that employers and unions had roughly equal bargaining weights on average, although their relative bargaining power fluctuated over time. I find that a series of large strikes in 1898 led to a relative increase of union bargaining power at striking mines. This led to increased output but decreased machine adoption, which validates the model's predictions. There is also substantially higher observed cutting-machine investment at mines for which I estimate higher employer power, which is again in line with the theoretical model. Finally, I estimate the fixed costs of cutting-machine usage by comparing the variable profit gains from machine adoption to the observed machine-usage rates.

I use the estimated model for two counterfactual exercises. First, I assess the effects of a 5% increase in employer power for the observed technology, cutting machines. I compute how quantities, prices, investment, and welfare would have changed under two scenarios: keeping machine usage fixed, and allowing machine usage to be endogenous to the degree of employer power. I find that the increase in employer power would have increased cutting-machine usage by 45%. However, output would have decreased on average by 14%, as the deadweight losses from employer power dominate the holdup reduction effect. Consequently, increased employer power would induce a reduction in consumer and labor surplus by 11% and 22%, respectively. Assuming exogenous capital investment leads to overestimating these losses to consumer and labor surplus by 13% and 7%. Employer power increases producer surplus by 62%, and this compensates the consumer and worker losses, as total surplus increases by 0.7%. In contrast, a model with exogenous capital investment would predict a total welfare reduction of 1.7%.

In a second counterfactual exercise, I examine the same increase in employer power for a counterfactual technology that has larger productivity effects. I find that in this case, labor surplus decreases by 17%, whereas consumer welfare barely changes, by 1.5%. The directedness of the technology implies that employer power strongly decreases employment, thereby harming workers, but only has limited effects on output. These asymmetric effects of employer power on output and employment stand in contrast to the usual monopsony model, in which employer power decreases both output and employment at similar levels, thereby harming both workers and consumers.

This paper contributes to four distinct literatures. First, it contributes to prior work on the welfare consequences of monopsony power. Existing work on ”neoclassical” monopsony and oligopsony power usually focuses on deadweight loss and/or on misallocation (Morlacco, 2017; Berger et al., 2022; Lamadon et al., 2022; Rubens, 2023). Whereas this literature keeps technology choices fixed when conducting welfare counterfactuals, I show that endogenous technology choices present an additional channel through which input market power affects welfare. I contribute to this literature by examining the welfare effects of monopsony power in the presence of both deadweight loss and investment holdup.[2]

[2]Although there is work on holdup in a different class of monopsony models in the labor-search literature (Acemoglu & Shimer, 1999; Shi, 2023), these models do not feature monopsony-induced deadweight loss.

Second, this paper contributes to the empirical literature on technology adoption and market power. Most of this literature has focused on product market power while assuming competitive input markets (Collard-Wexler & De Loecker, 2015; Miller, Osborne, Sheu, & Sileo, 2023). Two notable exceptions stand out. One, Goolsbee and Syverson (2023) show that monopsony power of higher education institutions leads to substitution from tenure-track toward adjunct faculty, which is a technological change. My model differs by allowing monopsony to induce not only movements *along* the production isoquant but also shifts *of* the isoquant itself. Two, Lindner, Muraközy, Reizer, and Schreiner (2022) study the effects of directed technological change in monopsonistic markets. In contrast, I examine the effects of monopsony power on those technological changes.

Third, this paper builds on the vertical relations literature. Relative to existing work on holdup (Williamson, 1971; Joskow, 1987; Zahur, 2022), I incorporate monopsony distortions and upward-sloping marginal-cost curves, and I also use a model with multiple substitutable inputs. In contrast to the literature that studies the effects of buyer power on technology choices of suppliers (Just & Chern, 1980; Huang & Sexton, 1996; Köhler & Rammer, 2012; Parra & Marshall, 2024), I focus on its effects on the technology choices of the buyers.

Finally, this paper relates to the literature on the effects of labor unions. One one hand, labor unions can countervail monopsony power (Dodini, Salvanes, & Willén, 2022; Azkarate-Askasua & Zerecero, 2024; Angerhofer, Collard-Wexler, & Weinberg, 2025; Demirer & Rubens, 2025); on the other hand, they can decrease innovation incentives by capturing innovation rents (Grout, 1984; Menezes-Filho & Van Reenen, 2003). I contribute by constructing and estimating a model that contains both of these counteracting effects of unionization.

The rest of this paper is structured as follows. Section 2 contains the theoretical model. Section 3 discusses the data, the industry background, and the empirical model. Section 4 covers the estimation of the model. Section 5 contains the counterfactual simulations. Section 6 concludes.

2 Theory: Buyer Power and Investment

I start with a theoretical framework to examine the welfare effects of buyer power while allowing for both deadweight losses and endogenous investment.

2.1 Primitives

Firms f produce output Q_f using two variable inputs. The model can be applied to any buyer of multiple factors, but given the empirical application I consider the firm to be an employer of high- and low-skilled labor H_f and L_f. I rely on the CES production function (1), in which both inputs are substitutable at a constant elasticity σ. To keep the model tractable, I assume constant returns to scale, but the results are robust to relaxing this assumption.[3] Skill-augmenting productivity is denoted as A_f, Hicks-neutral productivity as Ω_f, and the low-skilled-labor coefficient as β^l. The capital stock K enters the production function in two ways: it changes skill-augmenting productivity $A_f(K_f)$ and Hicks-neutral productivity $\Omega_f(K_f)$.

$$Q_f = \left((A_f(K_f)H_f)^{\frac{\sigma-1}{\sigma}} + \beta^l L_f^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \Omega_f(K_f) \tag{1}$$

Firms sell their output at a price P_f. Consumer demand for the good is given by a standard horizontal differentiation demand system, with an average industry-level price P_0, unobserved firm characteristics ξ_f, and a constant demand elasticity η:

$$Q_f = \left(\frac{P_f}{P_0} \right)^{\eta} \xi_f \tag{2}$$

High-skilled workers have an outside option Z_f, which they can earn when choosing not to be employed at firm f. I allow this outside option Z_f to be an upward-sloping curve, with constant inverse elasticity ψ, as shown in Equation (3). Firms are differentiated from the worker's perspective through an amenity term ζ_f. The average industry wage is equal to W_0. In contrast to models with a constant outside-option value, such as in Abowd and Lemieux (1993), an increasing outside option generates an upward-sloping labor supply curve, which allows for the possibility of monopsonistic behavior by employers. An increasing outside-option curve can be rationalized by the fact that workers are heterogeneous in terms of their outside options, and that firms cannot wage-discriminate as a function of these worker-specific outside options. Hence, the labor supply curve to each firm is upward-sloping: hiring an extra worker requires a higher wage to compensate

[3] See Appendix C.2.1.

for the higher outside option of the marginal worker:

$$Z_f = \frac{W_0}{1+\psi}\Big(\frac{H_f}{\zeta_f}\Big)^{\psi} \tag{3}$$

In contrast, the outside option of low-skilled labor is assumed a constant, V. This implies that firms pay low-skilled workers a uniform wage V and that low-skilled labor supply is perfectly elastic.

High-skilled workers are unionized at the firm level. The utility of the labor union at firm f is denoted as Π_f^u, which is defined as the integral of all differences between wages and the outside options for the high-skilled workers. Hence, I assume that the labor union aims to maximize total labor surplus:

$$\Pi_f^u = \int_0^H (W_f - Z_{if})di = (W_f - \frac{Z_f}{1+\psi})H_f$$

Employer variable profits are denoted as Π_f^d:

$$\Pi_f^d = P_f Q_f - W_f H_f - V L_f$$

Investing in capital induces a fixed cost ϕK_f on the employers, with ϕ denoting the cost per unit of capital. Therefore, total employer profits $\overline{\Pi}_f^d$ are given by $\overline{\Pi}_f^d \equiv \Pi_f^d - \phi K_f$.

2.2 Behavior

Decisions take place in three stages. First, employers choose their level of capital investment. Second, workers and employers bargain over a linear wage contract. Third, workers decide how much labor to supply.[4] I discuss these decisions in reverse chronological order.

In the third stage of the model, the labor union decides how much labor it is willing to supply given the negotiated wage level, in order to maximize union profits Π^u, which leads to the following upward-sloping high-skilled labor supply curve:

$$W_f = \frac{W_0}{1+\psi}\Big(\frac{H_f}{\zeta_f}\Big)^{\psi} \tag{4}$$

[4]The linear wage contract assumption will be motivated in the empirical application in Section 3. An alternative model that features efficient bargaining is in Appendix B.2.

At the same time, low-skilled workers are chosen by the employers to maximize their profits, $\max_{L_f}(\Pi_f^d)$. Taking the first-order condition leads to the following low-skilled-labor demand curve:

$$P_0\Big(\frac{1+\eta}{\eta}\Big)Q_f^{\frac{1}{\eta}}\Big(\frac{Q_f}{L_f}\Big)^{\frac{1}{\sigma}}(\Omega_f)^{\frac{\sigma-1}{\sigma}}\beta_l = V \tag{5}$$

In the second stage of the model, wages are bargained over between the labor union and the employers, with $\gamma_f \in [0, 1]$ denoting the relative bargaining ability of the union. Crucially, workers and employers cannot contract directly on machine adoption, and they bargain over joint variable profits, rather than total profits. Capital is chosen by the employers prior to the wage-bargaining process, so unless the wage-bargaining phase includes negotiations about capital investment as well, unions will appropriate a part of the rents generated by capital investment during the wage-bargaining phase without internalizing the effects of this rent-sharing on employer investment incentives. The inability to contract on capital investment is the source of the investment-holdup problem; this occurs in reality, as it is hard to write complete contracts on investment decisions, as opposed to wage contracts. The Hollywood writers strikes are an example of the tension between salary negotiations and negotiations about technology adoption (Kinder, 2024).

$$\max_{W_f}(\Pi_f^u)^{\gamma_f}(\Pi_f^d)^{1-\gamma_f}$$

Taking first-order conditions for the bargaining problem results in Equation (6). Workers and firms anticipate labor supply that will form in the second stage, and therefore internalize the partial derivative $\frac{\partial H_f}{\partial W_f}$:

$$\begin{aligned}\gamma_f\Big((1-\frac{1}{1+\psi}\frac{\partial Z_f}{\partial W_f})H_f + \frac{\partial H_f}{\partial W_f}(W_f - \frac{Z_f}{1+\psi})\Big)(P_f Q_f - W_f H_f - V L_f)\\ + (1-\gamma_f)(W_f - \frac{Z_f}{1+\psi})H_f(P_f\frac{\partial Q_f}{\partial H_f}\frac{\partial H_f}{\partial W_f} - H_f - W_f\frac{\partial H_f}{\partial W_f}) = 0\end{aligned} \tag{6}$$

By attributing the ”right to manage” to the workers rather than to the employers, I assume that the equilibrium lies on the labor supply curve, rather than on the labor demand curve. I motivate this assumption in the context of the empirical application in Section 4.5. This model collapses to the classical monopsony model in the case of perfect buyer power ($\gamma = 0$).[5]

Optimal quantities and prices $(Q_f^*, P_f^*, W_f^*, H_f^*, L_f^*)$ are the solution of Equations (1), (2), (4), (5), and (6), which are the production, goods demand, high-skilled labor supply, low-skilled labor demand, and wage-bargaining equations.

2.3 Effects of Employer Power: Investment Holdup vs. Deadweight Loss

I define "employer power" as the employer's bargaining weight $(1-\gamma_f)$. Employer power has two countervailing effects on output. First, holding capital K_f fixed, an increase in employer power leads to decreased output:

Proposition 1. *Holding capital K_f constant, output Q_f decreases in employer power ($1-\gamma_f$).*

As employer power increases, the equilibrium moves along the labor supply curve further toward the monopsonistic equilibrium, resulting in a lower level of output. This is the classical deadweight loss from monopsony power: employers push down wages by buying fewer inputs, which in turn decreases output as well.

However, capital investment is not invariant to employer power: employer power increases the share of the rents from capital investment that accrue to the employer. As employers weigh these rents against the fixed costs of capital investment, which are borne by the employers, higher rents imply higher capital investment.

Proposition 2. *Let production be increasing and strictly concave in capital, product demand be strictly concave and marginal labor costs be strictly convex. Under these assumptions, capital usage K_f increases in employer power* $(1-\gamma_f)$.

The functional form assumptions of concave demand, convex labor costs, and concavity of production in capital are satisfied for specifid functional forms imposed in the model. The intuition behind Proposition 2 is the following. On one hand, keeping total surplus fixed, increasing employer power increases the share of total profits that goes to the employer, so it also increases the absolute profit change upon adopting new capital. As employers face a trade-off between this profit change and the fixed cost to adopt capital, higher employer power incentivizes capital investment.

[5] I refer to Demirer and Rubens (2025), who present a model in which vertical conduct is endogeneous, rather than imposed by assumption.

On the other hand, increasing employer power shrinks the total surplus, as linear-price contracts are inefficient. However, under the assumptions made, the total profit effect for employers of gaining bargaining power is still positive, so the increased share of surplus dominates the change in total surplus. The formal proof of Proposition 2 is in Appendix B.1.

Net Effect of Employer Power on Output

Combining Propositions 1 and 2 reveals that the net effect of employer power on output is ambiguous. On one hand, employer power decreases output through the monopsony distortion. On the other hand, it can increase technology usage, thereby reducing marginal costs and increasing output. Which of these effects dominates depends on the relative magnitude of the deadweight loss and the endogenous investment mechanism. In the empirical application, I will quantify the relative size of these effects to examine how counterfactual changes in employer power affect output, producer surplus, consumer surplus, and worker welfare.

Simulation

To illustrate the importance of endogenous technology usage for the welfare effects of labor-market power, I simulate a parametrized version of the model.[6] I calibrate the goods demand elasticity at $\eta = -7$ and the inverse labor supply elasticity at $\psi = 0.25$, following the estimates for U.S. construction workers in Kroft et al. (2020). I consider a new technology that increases H-augmenting productivity (A) by 5% and increases TFP (Ω) by 20%.[7]

Figure 1a plots optimal technology usage K against employer power $(1 - \gamma_f)$. The solid red line depicts the model in which technology usage is allowed to change as a function of employee bargaining power. In line with the theoretical model, technology usage increases with the level of employer power. By comparison, the dashed blue line depicts the model in which technology usage is exogenous to the degree of employer power. In this model, technology usage is fixed equal to average technology usage in the endogenous adoption model.

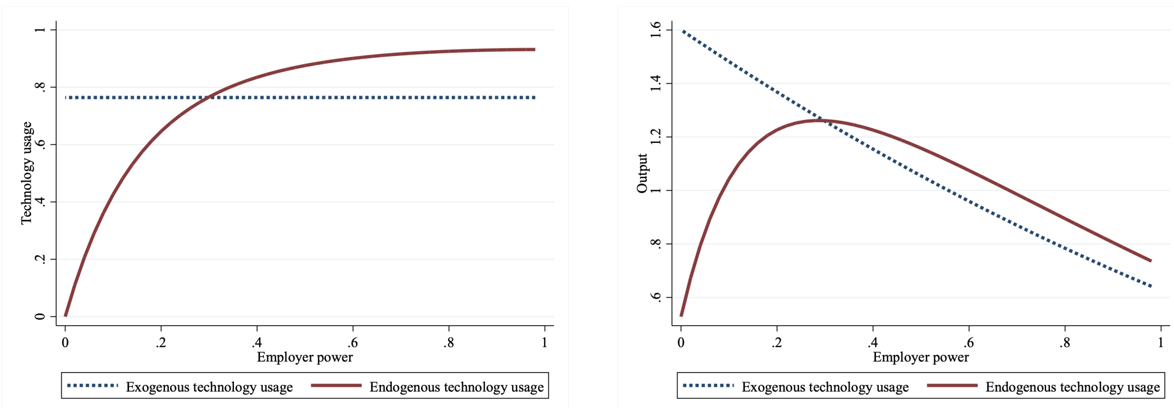
Figure 1b shows output Q as a function of employer power $(1 - \gamma_f)$. Under the assumption of exogenous technology usage, the blue solid line, output monotonically decreases with employer power. This is due to deadweight loss induced by the employer's monopsony power. The wage

[6] The parametrization is specified in Appendix B.3.
[7] In Appendix B.3.2, I show that this pattern is robust to various alternative parametrizations.

markdown set by the employer shrinks to zero as employer bargaining power goes to zero, reducing deadweight loss to zero. However, allowing for endogenous capital usage turns the output–bargaining power relationship into an inverted-U shape. At low levels of employer power, the output decrease due to monopsony power is countered by the reduction in marginal costs due to increased technology usage. Under the parametrization of the model, the positive output effect of increased technology usage outweighs deadweight loss until the bargaining weight of the employer is around one-third. Hence, output is maximized at this level of employer power.

Figure 1: Bargaining Power, Capital Investment, and Output

(a) Capital Investment **(b) Output**



Notes: Panel (a) shows how capital investment changes with the degree of employer power in the simulated model. Panel (b) shows output as a function of the employer's bargaining parameter, both when assuming exogenous investment and when letting investment vary with employer power.

3 Empirical Application: Coal Mining in Illinois

The model in Section 2 reveals that the relationship between employer power and output is ambiguous—it depends on the relative magnitude of the monopsony-induced deadweight loss and endogenous technology-adoption mechanisms. Therefore, empirical analysis is needed.

In this section, I empirically implement the model to study the net welfare effects of employer power in the context of the 19th-century Illinois coal mining industry. I observe the gradual mechanization of this industry through the introduction of coal-cutting machines, which provides an interesting case to study investment holdup.

3.1 Data

The main dataset is derived from the *Biennial Report of the Inspector of Mines of Illinois*. I observe every bituminous coal mine in Illinois between 1884 and 1902 at two-year intervals, resulting in 7,996 observations. The dataset contains the name of the mine, the mine owner's name, yearly coal extraction, average employee counts for both skilled and unskilled workers, days worked, and a dummy for cutting-machine usage per year. 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 these variables are used in the analysis, given that some are observed in a small subset of years.[8]

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, meaning that skilled workers were paid only for their output of large coal pieces, rather than for their total output. This introduces some measurement error in labor costs. However, according to the dataset, wage screens were used for merely 2% of total employment in 1898. Skilled wages and employment are separately reported for the summer and winter months from 1884 to 1894. For some years, I observe additional variables such as mine capacities, the value of the total capital stock, and a breakdown 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 other datasets. First, the Inspection Report of 1890, which contains monthly data on wages and employment for both types of workers, and monthly production quantities for a sample of 11 mines covering 15% of skilled and 9% of unskilled workers. Second, town- and county-level information are derived from the 1880 and 1900 population censuses and the censuses of agriculture and manufacturing. Third coal-cutting-machine costs are obtained from Brown (1889). Appendix A contains more details on my data sources and cleaning procedures.

[8]Appendix Table A7 contains a full list of observed variables and the years in which they are observed.

3.2 Industry Background

The Illinois coal mining industry grew rapidly throughout the sample period: annual output tripled from 8.7 to 28.1 megatons from 1884 to 1902. This was due to both an increase in the average mine size and because the number of mines grew from 643 to 859 units.

Coal-Extraction Process

The coal-extraction process consisted of three main steps. First, the coal vein had to be accessed, as it lay below the surface for 98.75% of the mines and 99.3% of output. Second, after miners reached the vein, the coal wall was "undercut," traditionally by hand, but from 1882 onward also with coal-cutting machines. Mechanization of the cutting process is considered to be the most significant technological advancement during this period (Fishback, 1992). Third, the coal had to be transported to the surface and separated 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, were purchased and brought to the mine 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 2.0 to 3.4 tons for hand mines, and from 2.2 to 4.0 tons for machine mines.[9]

Although different coal types exist, the mines in the dataset all extracted bituminous coal. There might have been 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 and, hence, not a choice variable of the firms conditional on operating in a certain location.

Occupations

Coal mining involved numerous occupational tasks. The Inspector Report from 1890 reports wages at the occupation level; I report this subdivision in Appendix Table A2 for the 20 occupations with

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

Figure 2: Output, Inputs, and Prices

Notes: Panel (a) plots the evolution of cutting-machine usage, both as a share of firms and weighted by output. Panel (b) documents the evolution of output per worker at mines where coal was cut manually, and at mines where cutting machines were used. Panel (c) shows the evolution of daily skilled wages and of the coal price per ton in Illinois, weighted by employment and output, respectively. Panel (d) shows the evolution of the aggregate ratio of skilled to unskilled workers in Illinois for both hand and machine mines.

the highest employment shares, together covering 97% of employment. Three of five workers were miners; they did the actual coal cutting. This required significant skill: to determine the thickness of the pillars, miners faced a trade-off, lower output vs. risk of collapse. The other 40% of workers did various 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"), and opening doors and elevators ("trappers"). The skills required to carry out these tasks were usually less complex than those of the miners; moreover they were not specific to coal mining: tending mules and loading carts were general-purpose tasks, in contrast to mining-specific tasks such as undercutting coal walls.

The difference in industry-specific skills is reflected in daily wages: miners earned higher daily wages than almost all other mining employee types.[10] The higher wages of miners cannot be explained as a risk premium—most of the other occupations were performed below the surface as well, and were hence subject to the same risks from mine collapse or flooding. From this point onward, I follow the skill categorization in the Inspector Reports by classifying workers into two types: miners, which I denote as "skilled labor," and all other employees, which I denote as "unskilled labor."

Technological Change

The first prototype of a mechanical coal cutter in the United States was invented by J.W. Harrison in 1877.[11] The Harrison patent was acquired and adapted by Chicago industrialist George Whitcomb, whose "Improved Harrison Cutting Machine" was released on the market in 1882.[12] As shown in Figure 2a, the share of Illinois coal mines using a cutting machine increased from 1% to 9% between 1884 and 1902. Mechanized mines were larger: their share of output increased from 7% to 30% over the same time period. Mechanization of the hauling process, which replaced mules with underground locomotives, was another source of technical advancement that started during the 1870s. By the start of the panel, in 1884, mining locomotives were already widely used in Illinois: the share of output of mines that operated locomotives was around 80% in 1884 and 90% in 1886.

As 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-labor days.[13] Mines without cutting machines used between three and four skilled labor-days per unskilled labor-day throughout the sample period, compared to two to three skilled labor-days per unskilled labor-day for machine mines. In every year except 1894, machine mines had a lower skilled-to-unskilled labor ratio than the other mines. On average, the skilled-to-unskilled labor ratio was 16% lower for machine mines compared to hand

[10] The only exception being "pit bosses" (middle managers), and "roadmen," who maintained and repaired mine tracks. These two categories of workers made up barely 2% of the workforce.

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

[12] Appendix Figure A5 depicts the patent. The spatial diffusion of cutting machines is shown in Appendix Figure A3.

[13] 1890 is omitted for machine mines due to employment being unobserved for most machine mines in that year.

mines, and this difference is statistically significant.[14] 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 TFP and factor-augmenting productivity levels, I refer to the empirical model in Section 4. 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).

Labor Markets

Skilled workers received a piece rate per ton of coal mined, which is a classical linear price contract, whereas unskilled workers were paid a daily wage.[15] Converting the piece rates to daily wages, the net salary of skilled labor was on average 16% higher compared to unskilled labor. "Net salary" means net of material costs and other work-related expenses. Rural Illinois was sparsely populated: the median and average populations of the towns in the dataset were 872 and 1,697 inhabitants. In the average town, 13.6% of the population was employed in a coal mine. Women and children under the age of 12 did not work in the mines, which implies that a large share of the local working population was employed in coal mining. Of all the towns, 43% had just one coal firm, and 75% had three or fewer. Towns with three or fewer coal firms accounted for 62% of total mining employment. Although most of the towns in the dataset were connected by railroad, these were exclusively used for freight: passenger lines operated only between major cities (Fishback, 1992). Given that the average village was 6.6 miles from the next village, and that skilled workers had to bring their own supplies to the mine, commuting between towns was not an option and the mining towns can be considered as isolated local labor markets. Most roads were unpaved and

[14] Regressing the log skilled-to-total labor ratio on a cutting-machine usage indicator results in a coefficient of -0.170 with a standard error of 0.037.

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

automobiles had not yet been introduced. To switch employers, miners had to migrate to another town.

The first attempts to unionize the Illinois coal miners started around 1860, but without much success (Boal, 2017). Although Illinois coal miners were unionized, for instance through the United Mine Workers of America and the Knights of Labor, union power was constrained by the use of "yellow-dog" labor contracts, which forced employees not to join a trade union.[16] A major strike in 1897–1898 led to a modest increase in wages, to a reduction of working hours, and to the introduction of annual wage negotiations, which took place each January (Bloch, 1922). Nevertheless, industrial relations remained tense for the ensuing years (Bloch, 1922).

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 dataset did not pay for company housing of the miners (Lord, 1883, p. 75), which would otherwise have been a labor cost in addition to miner wages.

Figure 2(c) reports the aggregate skilled-labor daily wage, defined as the total wage bill spent 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; daily miner wages remained around $1.80. After the strikes, wages rose sharply to around $2.50 per day.

Coal Markets

Coal was sold at the mine gate, and there was no vertical integration with postsales coal treatment, such as coking. On average, 92% of the mines' coal output was either sold to railroad firms or transported by train or barge to final markets. The importance of waterways decreased sharply in the favor of railroads during the early 1880s (of Natural Resources, 2025) The remaining 8% was sold to local consumers. The main coal-destination markets for Illinois mines were St. Louis and Chicago. Railway firms acted as an intermediary between coal firms and consumers; they were also major coal consumers themselves.

Historical evidence points to intense competition in coal markets during the last two decades of the 19th century, before the large consolidation wave in the early 1900s (Graebner, 1974).

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

Nevertheless, coal was still costly to transport, which means that coal markets were likely not perfectly integrated: coal prices varied considerably across Illinois. In 1886, for instance, the coal price varied between 91 cents per short ton at the 10th percentile of the price distribution to 1.75 dollars per 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 $1.25 to $0.84 between 1884 and 1898, after which it increased again.

Descriptive Evidence for Upward-Sloping Labor Supply

Coal demand was seasonal: demand for energy was higher in winter than in summer. Coal storage costs meant that firms could not fully arbitrage between winters and summers, and, hence, needed to hire more workers during the winter. Joyce (2009) mentions that miners were (partially) unemployed during the summer months. This cyclical pattern provides useful variation to compare wage responses of skilled and unskilled workers to coal demand shocks. In Figure 3(a), I show 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 demand for both skilled and unskilled labor, only skilled wages increased during winter. This is also shown in Figure 3(b), which plots how average daily wages for both skilled and unskilled workers in 1890 changed with the monthly number of worker-days of each type at the mine-month level throughout 1890.[17] Skilled wages were positively correlated with monthly skilled employment, whereas the unskilled worker wage-employment schedule was flat. Moreover, skilled wages varied greatly across mines and months, but there was very little cross-sectional and intertemporal variation in unskilled wages.

The fact that skilled wages increased in response to coal demand shocks whereas unskilled wages did not, and the fact that unskilled wages were nearly uniform across Illinois whereas skilled wages were not, suggests an inelastic high-skilled labor supply curve and a fully elastic low-skilled labor supply curve. This implies that firms had the potential to exert monopsony power on the high-skilled labor market, but not on the low-skilled labor market. However, it could be that seasonal employment movements reflected not only labor demand variation but also labor supply shocks,

[17] 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 are available for a sample of mines selected by the Illinois Bureau of Labor Statistics across five counties in 1890, which cover 16% of skilled employment and 9% of unskilled employment.

for instance due to the harvesting season. Moreover, within-year demand shocks trace out a short-run supply curve, whereas labor supply could be more elastic in the longer run. Hence, in the structural model, I will instead rely on an instrumental-variable strategy that relies on international coal-price shocks to identify the high-skilled labor supply elasticity.

Figure 3: Wage-Employment Profile by Skill Type

(a) Wages **(b) Wage-Employment Profile**

Notes: Panel (a) shows how the wages of skilled miners and other mine employees evolved monthly during 1890 from the 1890 Inspector Report (Illinois Bureau of Labor Statistics, 1890). Panel (b) plots mine-month-level wages for both types of workers against monthly employment from the same source, again for both types of workers.

3.3 Empirical Model

Production Function

I implement an empirical model of the Illinois coal industry based on the general model outlined in Section 2. Let f index coal firms per town and let t index all even years between 1884 and 1902. The model is set up at the firm-town-year level: it is plausible that employers optimize at the firm level, rather than at each mine independently. However, I let firms optimize on a labor-market-by-labor-market basis: firms with mines in different labor markets do not internalize the cross-labor-market effects of their decisions. This is consistent with the model, given that it does not feature strategic interaction between firms on the labor market. Annual coal extraction in short tons is denoted as Q_{ft}, the number of days worked by high-skilled labor is denoted as H_{ft}, and the number of low-skilled labor-days is denoted as L_{ft}. In contrast to the theoretical model, capital investment is modeled as a binary variable: firms choose whether to use cutting machines or not, with usage being denoted as $K_{ft} \in \{0,1\}$. I abstract from other technologies, such as mining locomotives, because they were already widely adopted by the start of the panel, and because they are not observed in all years of the sample.

I maintain the CES production function from Equation (1), with an elasticity of input substitution σ and low-skilled-labor coefficient β^l. Firms differ in terms of their skill-augmenting productivity A_{ft} and in their Hicks-neutral productivity Ω_{ft}. In Appendix C.2, I estimate various extensions of the production model to allow for nonconstant returns to scale, the existence of intermediate inputs, and the possibility that cutting machines change scale returns. All these extensions lead to very similar production-function estimates. Taking the logs of Equation (1) and adding the time index leads to Equation (7), which is the production function I will estimate:

$$q_{ft} = \left(\frac{\sigma}{\sigma-1}\right) \log \left(\left(H_{ft} A_{ft}(K_{ft})\right)^{\frac{\sigma-1}{\sigma}} + \beta^l L_{ft}^{\frac{\sigma-1}{\sigma}} \right) + \omega_{ft}(K_{ft}) \tag{7}$$

Cutting-machine usage K_{ft} is allowed to affect both productivity terms Ω_{ft} and A_{ft}. The logarithms of both these productivity terms, a_{ft} and ω_{ft}, are assumed to evolve as AR(1) processes, as specified in Equations (8) and (9). The productivity terms have serial correlations ρ^a and ρ^ω and are assumed to be affected linearly by cutting-machine usage, as parametrized by the coefficients α^k and β^k for labor-augmenting and Hicks-neutral productivity, respectively.[18] Skill-augmenting and Hicks-neutral productivity shocks are denoted as e^a_{ft} and e^ω_{ft}:

$$a_{ft} = \alpha^k K_{ft} + \rho^a a_{ft-1} + e^a_{ft} \tag{8}$$

$$\omega_{ft} = \beta^k K_{ft} + \rho^\omega \omega_{ft-1} + e^\omega_{ft} \tag{9}$$

I assume that 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 1.9% of the mines operated at full capacity, and they were responsible for just 2.7% of industry sales.[19] I also abstract from

[18] Although these AR(1) specifications do not allow for richer models of cost dynamics in which current productivity is a function of the total amount of output produced in the past, they do have the benefit of not requiring inversion of the production function, thereby allowing for rich heterogeneity in both productivity terms, markdowns, and markups. See Appendix C.2.4 for a motivation and discussion of this assumption.

[19] Figure A4 depicts the entire distribution of capacity utilization rates.

stockpiling of coal, and I assume that coal must be sold immediately after extraction: coal storage usualy led to deteriorating coal quality; moreover it was expensive and dangerous (Stoek, Hippard, & Langtry, 1920). As Williams (1901) asserted:

> "The product of a mine can be stored with economy only in the mine itself [...] Coal must be sold, therefore, as fast as it is mined."

Labor Supply

Adding time subscripts to the inverse labor supply function (Equation 4), then inverting it, results in the labor supply equation (Equation 10a). The daily wage of high-skilled workers W_{ft} is computed as the piece rate multiplied by daily tonnage per worker. I measure W_{0t} as the average daily wage in year t.

I include observed firm characteristics $\mathbf{X}^h_{ft}$ next to the latent amenity term ζ_{ft}. Specifically, I include a linear time trend, county fixed effects, and the logarithm of the minimal distance of the firm to Chicago and St. Louis as observed characteristics, to account for proximity to the large regional population centers.

$$H_{ft} = \left(\frac{W_{ft}}{W_{0t}}\right)^{\frac{1}{\psi}} \exp(\mathbf{X}^h_{ft})^{\boldsymbol{\psi}^x} \zeta_{ft} \tag{10a}$$

I estimate the inverse labor supply equation in logs, which is given by Equation (10b). I estimate inverse labor supply, rather than labor supply, because this makes it easier to test whether firms are wage-takers, in which case $\psi = 0$, or have some market power over wages.

$$w_{ft} - w_{0t} = \psi h_{ft} + \frac{\psi}{\boldsymbol{\psi}^x} \mathbf{x}^h_{ft} - \psi \log(\zeta_{ft}) - \ln(1 + \psi) \tag{10b}$$

The amenity term ζ_{ft} captures firm differentiation from the miner's perspective. In contrast to Delabastita and Rubens (2024), who rely on a homogeneous employers model, I do allow for firm differentiation because skilled wages varied substantially across mines, even within the same labor markets.[20]

Similarly to the theoretical model, the market for low-skilled labor is assumed to be perfectly

[20] In Appendix Table A6, I report the R^2 of regressing log daily miner wages on (subsequently) year, county, town, and firm dummies. Town and year dummies explain only 29% of the variation in skilled miner wages.

competitive, and low-skilled workers are paid a uniform wage V, which is equal to their outside option. The main reason for this assumption lies in the fact that unskilled wages barely varied across Illinois, nor did they react to seasonal weather shocks, as shown in Figure 3.

Coal Demand

Coal produced in Illinois mines was a nearly homogeneous product. However, coal firms were differentiated by their locations, which resulted in price differences between coal firms. I again include the shortest distance to either Chicago or St. Louis, county fixed effects, and a linear time trend, because these variables likely affected coal demand:

$$Q_{ft} = \left(\frac{P_{ft}}{P_{0t}}\right)^{\eta} \exp(\mathbf{x}_{ft}^{q})^{\boldsymbol{\eta}^{x}} \xi_{ft} \tag{11}$$

Taking logarithms and inverting Equation (11) results in Equation (12), which is the demand model I estimate. I again estimate inverse demand, rather than demand for coal, to test whether firms are price setters or price takers on coal markets.

$$p_{ft} - p_{0t} = \frac{1}{\eta} q_{ft} - \frac{\boldsymbol{\eta}^{x}}{\eta} \mathbf{X}_{ft}^{q} - \frac{1}{\eta} \ln(\xi_{ft}) \tag{12}$$

Vertical Conduct

I make two "vertical conduct" assumptions. First, I assume that unions and firms bargain over linear wage contracts, rather than over nonlinear contracts or over wages and quantities, as they would under efficient bargaining. This is motivated by the observed nature of linear (piece-rate) wage contracts, and by the fact that employers and unions bargained over wages, not employment (Bloch, 1922). Second, I assume that employment is chosen by the union rather than by the employers, meaning that the equilibrium lies on the labor supply curve. This assumption is motivated by the fact that the 1898 coal strikes, which are a negative shock to employer power, led to increased output, as shown in Section 4.5. If employment would be chosen by the employer instead (being on the labor demand curve), output should increase instead, as employer power would decrease double marginalization.

Variable Input Decisions

I assume that firms make decisions in two stages. First, they choose whether or not to use cutting machines. Second, conditional on this choice, they bargain over high-skilled wages and they choose low-skilled employment. I describe the model in reverse chronological order, given that estimation proceeds in this order.

In year t, employers negotiate a high-skilled wage rate with the union according to the bargaining protocol specified in Equation (13), following the bargaining model that was described in Equation (6). Given the institutional background, I assume that employers and miner representatives bargain over wages at the firm level—this implies a passive-beliefs assumption that employers and unions take the bargaining outcomes and actions at all other firms as given (Horn & Wolinsky, 1988).

$$\gamma_{ft}\Big((1-\frac{1}{1+\psi}\frac{\partial Z_{ft}}{\partial W_{ft}})H_{ft}+\frac{\partial H_{ft}}{\partial W_{ft}}(W_{ft}-\frac{Z_{ft}}{1+\psi})\Big)(P_{ft}Q_{ft}-W_{ft}H_{ft}-VL_{ft}) + (1-\gamma_{ft})(W_{ft}-\frac{Z_{ft}}{1+\psi})H_{ft}(P_{ft}\frac{\partial Q_{ft}}{\partial H_{ft}}\frac{\partial H_{ft}}{\partial W_{ft}}-H_{ft}-W_{ft}\frac{\partial H_{ft}}{\partial W_{ft}})=0 \quad (13)$$

Following the labor supply curve, Equation (10b), this negotiated wage rate results in a certain amount of high-skilled labor being supplied at each coal firm. Coal firms simultaneously choose low-skilled labor, as specified in Equation (14). Although the miner unions also represented non-miner workers, I do not model the wage-formation process for low-skilled labor as a bargaining model, given that the wage-negotiation documents show only wage bargaining over skilled wages (Bloch, 1922) and that there was little variation in low-skilled wages to begin with, suggesting a competitive labor market. I assume that these variable input decisions happen after the productivity shocks e^{ω} and e^{a} are observed by the firm. The combination of low-skilled and high-skilled labor and capital (the decisions are specified below) results in coal output Q_f according to the production function, Equation (7). The coal demand curve, Equation (12), determines the price every firm can charge at that level of coal output.

$$P_{0t}\Big(\frac{1+\eta}{\eta}\Big)Q_{ft}^{\frac{1}{\eta}}\Big(\frac{Q_{ft}}{L_{ft}}\Big)^{\frac{1}{\sigma}}(\Omega_{ft})^{\frac{\sigma-1}{\sigma}}\beta_l=V \quad (14)$$

Fixed Input Choices

Cutting machines are chosen unilaterally by the employers rather than negotiated, given the evidence of the wage-bargaining process (Bloch, 1922). Employers trade off the variable profit gains form cutting machines with their fixed costs.

To compute the variable profit gains from capital investment, I obtain the optimal quantities and prices at every firm in every year by solving the system of Equations (7), (10a), (11), (13), and (14): the production function, high-skilled labor supply, coal demand, high-skilled labor demand, and low-skilled labor demand. This delivers certain outcomes $(Q^1_{ft}, P^1_{ft}, H^1_{ft}, L^1_{ft}, W^1_{ft})$ if the firm uses cutting machines, and different outcomes $(Q^0_{ft}, P^0_{ft}, H^0_{ft}, L^0_{ft}, W^0_{ft})$ if the firm does not use cutting machines. The variable profit gain of the employer from using cutting machines is denoted as $\Delta\Pi^d_{ft}$:

$$\Delta\Pi^d_{ft} \equiv (P^1_{ft}Q^1_{ft} - W^1_{ft}H^1_{ft} - V_tL^1_{ft}) - (P^0_{ft}Q^0_{ft} - W^0_{ft}H^0_{ft} - V_tL^0_{ft}) \tag{15}$$

The costs of technology usage are denoted ϕ_t, so total employer profits are equal to $\Pi^d_{ft} - \phi_t K_{ft}$. As in Peters, Roberts, Vuong, and Fryges (2017), I parametrize technology costs as an exponential distribution. I let the rate parameters (ϕ^0, ϕ^1, ϕ^2) evolve over time, with ϕ^0 measuring the time-invariant fixed cost of technology usage, ϕ^1 measuring the time-varying component of fixed machine costs, and ϕ^2 measuring the variable cost of using cutting machines. I allow machine costs to have a variable component, in addition to a fixed-cost component, because larger mines might require more cutting machines.

$$\phi \sim \exp(\phi^0 + \phi^1 t + \phi^2 q_{ft})$$

I assume that prior to observing the productivity shocks e^ω and e^a, firms independently and simultaneously choose whether or not they will use cutting machines. They make this decision by trading off the costs of machine adoption ϕ_t with the expected variable profit return $\Delta\Pi^d$. I assume that firms do not choose their cutting machines in a dynamic manner, but rather that they optimize their technology mix period by period. The main reason for this assumption is that observed entry *and* exit of machine usage is frequent: 106 cutting-machine installations, 60 of which were scrapped. This suggests the existence of an aftermarket for capital.

Using the exponential form of fixed costs, the probability that a firm uses cutting machines $p^k_{ft}(\phi)$ is equal to

$$p^k_{ft}(\phi^0, \phi^1, \phi^2) = 1 - \exp\left(\frac{-\Delta\Pi^d_{ft}}{\phi^0 + \phi^1 t + \phi^2 q_{ft}}\right) \tag{16}$$

4 Identification and Estimation

In this section, I turn to the identification and estimation of the model. I consecutively estimate miner supply, coal demand, the coal production function, relative bargaining ability of unions and coal firms, and cutting-machine costs. Table 1 summarizes the sources of identification in the model, which are explained in detail in the next subsections.

Table 1: Identification: Summary

	Equation	Parameters	Identification
Miner supply	(10b)	$\boldsymbol{\psi}$	International coal-price shocks
Coal demand	(12)	$\boldsymbol{\eta}$	Coal vein thickness
Production function	(7)	$\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\rho}$	Input timing assumptions
Bargaining weights	(22)	γ	Bargaining first-order conditions
Machine costs	(16)	$\boldsymbol{\phi}$	Comparing variable profit gains to observed cutting-machine usage

4.1 Labor Supply Estimation

Although the model is specified at the firm level, the dataset is observed at the mine level. Given that firms are assumed to have optimized at the firm-town level, I aggregate all the relevant variables to the firm-town-year level, as detailed in Appendix A.2.

I start with the identification of the inverse skilled labor supply curve, Equation (10b). The inverse labor supply elasticity ψ cannot be recovered by simply regressing high-skilled-labor wages on employment, because of the latent firm characteristics ζ_{ft}. Firms with a high ζ_{ft} knew they were attractive to miners, which permitted them to offer a lower wage to attract the same number of miners. 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 international coal-market price shocks for identification. I obtain the average annual coal price on international coal markets in Europe from Degrève (1982). I use as instruments both this coal price and its interaction term with an indicator for whether or not a mine shipped coal over the railroad network. These instruments imply three assumptions. One, individual Illinois coal mines were too small to affect coal prices on the European market. This makes sense given that Illinois produced only around 10% of the total U.S. output, and that U.S. bituminous coal mines exported only around 1.2% of their coal in 1898 (Graebner, 1974). Two, international coal-price shocks affected the demand for Illinois coal. Given that Chicago was one of the destination markets for Illinois coal, and that Chicago also sourced coal from both the East Coast and other coal fields by lake steamers, changes in nonlocal coal prices affected demand for Illinois coal. Three, international coal-price shocks affected coal demand more if coal mines were shipping their coal over the railroad network compared to coal mines that sold their coal only locally. This third assumption makes sense given that mines that sold only locally did not compete with coal fields outside of Illinois; neither these mines not their consumers were connected to the railroad network.

I compute the baseline wage level $w_{0i(f)t}$ as the average miner wage in Illinois. I estimate Equation (10b) with a two-stage least squares estimator using the European coal price and an interaction term of the European coal price and a shipping dummy as instruments for the log relative wage at each firm. I control for whether the firm was a shipping mine or a mine that sold only locally, and I include county fixed effects and a linear time trend.

For unskilled wages, I rely on the average daily wage for unskilled labor in the Illinois coal industry in every year. Given that I observe this wage from only 1884 through 1894, I linearly interpolate for the 1896–1902 period using a loglinear time trend.

The inverse skilled labor supply elasticity is estimated to be 0.258 with a standard error of 0.126, as reported in Table 2(a). This means that in the monopsony case, which corresponds to full employer power $\gamma_{ft} = 0$, the marginal product of skilled labor would surpass their wages by 26 %.[21]

[21] The wage markdown is equal to $\frac{MRPL-w}{w} = \psi$.

Table 2: Model Estimates

(a) Labor Supply		Est.	S.E.
Inverse labor supply elasticity	ψ	0.258	0.126
1(Shipping mine)		-0.395	0.283
Year		0.009	0.005
First-stage F-stat		12.4	
Observations		6315	
(b) Coal Demand			
Inverse coal demand elasticity	$\frac{1}{\eta}$	-0.187	0.017
Log(min. distance to big city)		-0.029	0.022
No. railroads		0.025	0.016
Year		-0.000	0.001
First-stage F-stat		773.1	
Observations		3127	
(c) Production Function			
Input substitution elasticity	σ	0.359	0.057
Skill-augmenting technology effect	α^k	0.192	0.095
Hicks-neutral technology effect	β^k	0.100	0.153
Low-skilled labor coefficient	β^l	0.006	0.003
Serial correlation Hicks-neutral productivity	ρ^ω	0.287	0.117
Serial correlation skill-augmenting productivity	ρ^ω	0.830	0.076
Observations		1626	
(d) Fixed Machine Costs			
Fixed machine cost in 1882 ('000 USD)	ϕ^0	9.230	2.909
Fixed machine cost time trend ('000 USD)	ϕ^1	-0.928	0.326
Variable machine cost (USD)	ϕ^2	0.108	0.169

Notes: Panel (a) reports the skilled labor supply estimates, Panel (b) reports the estimates of the coal demand function, Panel (c) contains the estimates of the production function, with block-bootstrapped standard errors over 200 iterations, and Panel (d) reports the cutting-machine fixed-cost estimates.

4.2 Coal Demand Estimation

I estimate the coal demand function in logarithms, Equation (12), using firm-level quantities and prices. I rely on the thickness of the coal vein as a cost shifter: whereas the vein thickness affects the marginal cost of mining, it does not enter consumer utility conditional on the coal price, because vein thickness does not affect coal quality (Affolter & Hatch, 2002). The thickness of the coal vein was the result of geological variation, and hence not a choice variable.

I estimate Equation (12) using a two-stage least squares estimator, with the log average vein thickness in the town as the instrument for coal output. In the observed covariates vector $\mathbf{X}^q_{ft}$, I include the following coal demand shifters: the logarithm of the minimal distance to either Chicago or St. Louis, the number of railroads passing through the mine's town, whether or not the mine was a shipping mine, and a linear time trend. I compute p_{0t} as the average coal price in each year.

Table 2(b) contains the coal demand estimates. The number of observations, 3,127, is lower than when estimating labor supply because the thickness of the coal veins is not observed in 1888 and 1890. The coal demand elasticity is estimated at -0.187 with a standard error of 0.017. This corresponds to a coal demand elasticity of $\eta = -5.347$, and implies that firms set coal prices at a markup of 23% above marginal costs.[22] The minimal distance from either Chicago or St. Louis had a negative but statistically insignificant effect on demand. A more important demand shifter seems to be the number of railroad lines passing through the mine's town. Coal demand was roughly stable throughout the sample period.

The finding that coal firms were not price takers on coal markets is somewhat surprising, given that coal is a homogeneous good and that there were many firms in this market. Prior work on coal mining in the same period describes the firms as price takers on their product markets (Fishback, 1992). In Appendix Table A4, I examine whether the demand estimates are driven by the large number of small mines that only sell coal locally, rather than exporting coal over the railroad network. I estimate the inverse coal demand equation separately for shipping and local mines. However, the inverse price elasticity is even higher (more negative) for shipping mines than for local mines, so even mines that sold their coal over the railroad network had price-setting power.

4.3 Production Function and Bargaining Weights Estimation

I estimate the production function in two steps. First, I estimate the elasticity of input substitution σ and the skill-augmenting effects of cutting machines, α^k. Second, I estimate all other production-

[22] The coal-price markup above marginal costs is $\frac{P_{ft}}{mc_{ft}} = \frac{\eta}{\eta+1}$.

function coefficients, β^l and β^k.

Elasticity of Substitution

The elasticity of substitution is usually estimated by taking the ratio of the input demand functions from the employer's profit-maximization first-order conditions, e.g., in Doraszelski and Jaumandreu (2018). In the bargaining model, however, the marginal-revenue product of high-skilled labor is not equal to its wage as long as $\gamma < 1$. Setting γ to zero in Equation (6), which implies perfect monopsony power, gives:

$$\frac{\partial R_{ft}}{\partial H_{ft}} = W_{ft}(1 + \psi)$$

Conversely, if γ becomes one, which implies that the labor union had all the bargaining power, the wage of high-skilled workers is equated to their marginal revenue product:

$$\frac{\partial R_{ft}}{\partial H_{ft}} = W_{ft}$$

These two first-order conditions for extremes of the bargaining parameter γ_f can be linearly interpolated using the bargaining parameter γ_{ft}, which results in a linear approximation of the first-order conditions:

$$\frac{\partial R_{ft}}{\partial H_{ft}} = W_{ft}(1 + (1 - \gamma_{ft})\psi) \tag{17}$$

Working out the first-order conditions (5) and (17), then dividing (5) by (17), results in Equation (18). This equation is a variant of the first-stage regression from Doraszelski and Jaumandreu (2018), except that the labor supply elasticity enters into the first-order conditions, as in Rubens, Wu, and Xu (2024):

$$l_{ft} - h_{ft} = \sigma(w_{ft} - v + \ln(1 + (1 - \gamma_{ft})\psi)) + \underbrace{\sigma \ln(\beta^l) + (1 - \sigma)a_{ft}}_{\equiv \tilde{a}_{ft}} \tag{18}$$

Given that Equation (8) specifies an AR(1) process for the factor-augmenting productivity term a_{ft}, the residual $\tilde{a}_{ft}$ also evolves as an AR(1). Hence, taking ρ^a differences of Equation (18) isolates the productivity shock e^a_{ft} as a function of the coefficients ρ^a, σ, and α^k. Using the previously stated assumptions that capital is chosen prior to observing the skill-augmenting productivity shock e^a_{ft}, but that variable inputs are chosen afterwards, the moment conditions, Equation (19), can be specified to estimate the elasticity of input substitution σ, the skill-augmenting productivity effect of cutting machines α^k, and the serial correlation in skill-augmenting productivity ρ^a:

$$\mathbb{E}\Big[e^a_{ft}(\rho^a, \alpha^k, \sigma)| \begin{pmatrix} K_{ft-r} \\ L_{ft-r-1} \\ H_{ft-r-1} \end{pmatrix} \Big]_{r=0}^{T-1} = 0 \tag{19}$$

Second-Stage Production-Function Coefficients

From Equation (18), the log factor-augmenting productivity residual a_{ft} can be written as a function of the estimated parameters σ and ψ, and the yet-to-be-estimated parameters β^l and β^k:

$$a_{ft} = \Big(\frac{l_{ft} - h_{ft}}{1-\sigma}\Big) - \frac{\sigma}{1-\sigma}\big(\ln(\beta^l)\big) - \frac{\sigma}{1-\sigma}\big(w_{ft} - v_t + \ln(1 + (1-\gamma_{ft})\psi)\big)$$

Substituting this factor-augmenting productivity term into the log production function gives:

$$q_{ft} = \frac{\sigma}{\sigma-1}\ln\Bigg(\Big(H_{ft}\exp\big(\frac{l_{ft} - h_{ft}}{1-\sigma} - \frac{\sigma}{1-\sigma}\ln(\beta^l) - \frac{\sigma}{1-\sigma}(w_{ft} - v_t + \ln(1+(1-\gamma_{ft})\psi))\big)\Big)^{\frac{\sigma-1}{\sigma}} + \beta^l L_{ft}^{\frac{\sigma-1}{\sigma}}\Bigg) + \omega_{ft} \tag{20}$$

I define the first linear term in the log production function as $B_{ft}(.)$:

$$B_{ft} \equiv \frac{\sigma}{\sigma-1}\ln\Bigg(\Big(H_{ft}\exp\big(\frac{l_{ft} - h_{ft}}{1-\sigma} - \frac{\sigma}{1-\sigma}\ln(\beta^l) - \frac{\sigma}{1-\sigma}(w_{ft} - v_t + \ln(1+(1-\gamma_{ft})\psi))\big)\Big)^{\frac{\sigma-1}{\sigma}} + \beta^l L_{ft}^{\frac{\sigma-1}{\sigma}}\Bigg)$$

Using the productivity transition in Equation (9), taking ρ^ω differences isolates the Hicks-neutral productivity shock e^{ω}_{ft} as a function of the parameters $(\rho^{\omega}, \beta^{k}, \beta^{l})$:

$$e^{\omega}_{ft} = q_{ft} - \rho^{\omega} q_{ft-1} - \beta^{k}(k_{ft} - \rho^{\omega} k_{ft-1}) - (B_{ft} - \rho^{\omega} B_{ft-1})$$

Using the timing assumption that employers chose capital prior to the realization of the Hicks-neutral productivity shock but chose low-skilled labor and bargained over wages *after* the realization of this shock leads to the moment conditions in Equation (21). I estimate this model using lags of up to one time period.

$$\mathbb{E}\Big[e^{\omega}_{ft}(\rho^{\omega}, \beta^{k}, \beta^{l})| \begin{pmatrix} K_{ft-r} \\ L_{ft-r-1} \\ H_{ft-r-1} \end{pmatrix} \Big]_{r=0}^{T-1} = 0 \tag{21}$$

Estimating Bargaining Weights

Adding time subscripts to the wage-bargaining first-order condition, Equation (13), and rearranging terms as a function of the union bargaining weights γ_{ft} leads to Equation (22). I estimate the bargaining parameters using this equation, whose variables are all either observed or have been estimated in the production, labor supply, and goods demand models.

$$\gamma_{ft} = \frac{(W_{ft} - Z_{ft})(\frac{\partial R_{ft}}{\partial H_{ft}}\frac{\partial H_{ft}}{\partial W_{ft}} - H_{ft} - W_{ft}\frac{\partial H_{ft}}{\partial W_{ft}})}{(W_{ft} - Z_{ft})(\frac{\partial R_{ft}}{\partial H_{ft}}\frac{\partial H_{ft}}{\partial W_{ft}} - H_{ft} - W_{ft}\frac{\partial H_{ft}}{\partial W_{ft}}) - (\frac{\psi H_{ft}}{1+\psi} + \frac{\partial H_{ft}}{\partial W_{ft}}(W_{ft} - Z_{ft}))\frac{\Pi^{d}_{ft}}{H_{ft}}} \tag{22}$$

To estimate both the first and second stages of the production-function estimation, Equations (18) and (20), the bargaining parameters γ_{ft} need to be known, as they enter into the first-order condition for high-skilled labor demand. However, estimating the bargaining parameters γ_{ft} requires knowing the production-function coefficients, as is clear from Equation (22). I proceed by estimating the production function and the bargaining parameter using a fixed-point algorithm. I start with an initial value of $\gamma_{ft} = 0.5$ to estimate the production function and the bargaining parameter. Then, I use the resulting bargaining parameter to reestimate the production function and the bargaining parameter. I continue this estimation loop until the production-function coefficients converge, up to a sensitivity level of 0.001. I find that the estimates quickly converge to a fixed point.

Results: Production

Table 2(c) reports the production-function estimates. The elasticity of substitution between skilled and unskilled miners is estimated at 0.359, which implies that these two types of workers are gross complements. It is not surprising that this elasticity is relatively low, given that skilled miners were employed for cutting coal whereas unskilled miners were employed mainly for hauling coal, two tasks that are complementary. Cutting machines are estimated to increase skill-augmenting productivity by 21%,[23] so cutting machines are a skill-augmenting technology. Given that skilled and unskilled labor are gross complements, this makes cutting machines an unskill-biased technology (Acemoglu, 2002), similarly to many other technologies developed throughout the 19th century that were also unskill-biased (Mokyr, 1990; Goldin & Katz, 2009). The finding that cutting machines were unskill-biased is consistent with the fact that cutting machines automated the cutting process, which was reliant on skilled miners, in contrast to the hauling process, which was mainly reliant on unskilled workers. Besides increasing skill-augmenting productivity, cutting machines also increased Hicks-neutral productivity by 11%, although this effect is not statistically significant. The low-skilled labor parameter β^l is estimated, imprecisely, at 0.006. Easier to interpret are the corresponding output elasticities of low- and high-skilled labor, which are estimated at 0.683 and 0.317, respectively. Finally, skill-augmenting and Hicks-neutral productivity have serial correlations of 0.287 and 0.830.

Results: Bargaining Weights

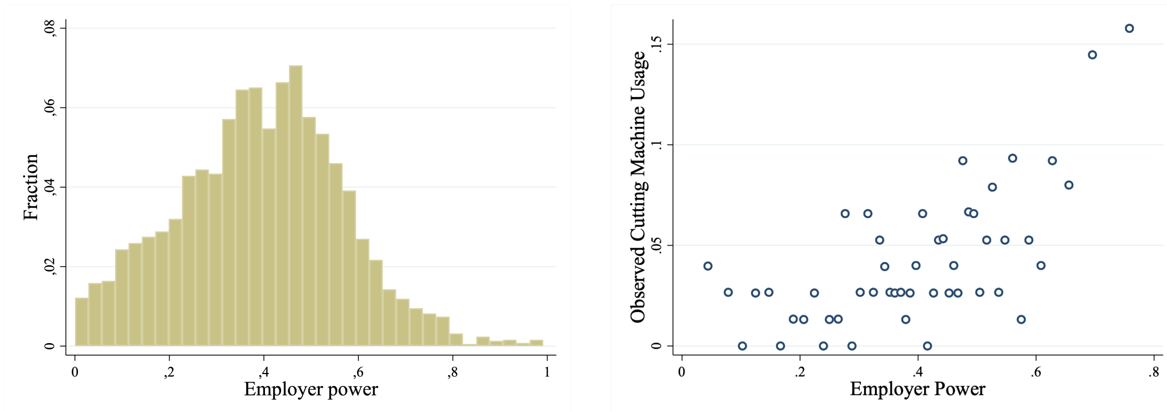
The average and median employer's bargaining weight are both 0.39, so bargaining power was balanced between mine owners and the miners' union, slightly favoring the union. I keep only the bargaining parameter values that range between zero and one, as values outside of this range are meaningless in the context of the bargaining model. This reduces the number of observations by 11%. Figure 4a shows the entire distribution of employer power.

Table 3 regresses the logarithm of the employer bargaining parameter over high-skilled workers ($\ln(1-\gamma_f)$) on firm and market characteristics. The first two columns do not include firm fixed

[23] $\exp(0.192) - 1$

Figure 4: Employer Power Estimates

(a) Distribution of Employer Power **(b) Machine Usage and Employer Power**



Notes: Panel (a) shows a histogram of the employer power estimates. The average is 0.39 and the median is 0.40. Panel (b) compares observed cutting-machine usage and estimated employer power, plotting average machine usage by ventiles of the employer power distribution.

effects, the last two do. In both specifications, employers with higher market shares on the high-skilled-worker market had more bargaining power. Moreover, employer power was increasing at a rate of 0.7% to 1.1% per year. Employers in markets with more immigrants and a higher population share of African Americans had more bargaining power. This is consistent with the lower bargaining ability of immigrants, who often did not speak much English, and with the use of African American"strike-breakers" who migrated from the southern U.S. states as a means to decrease the unions' bargaining ability. Finally, firms in counties with higher manufacturing wages had less bargaining power, due to the miners' more favorable negotiating position. As I document in Section 4.5, the 1897–1898 strikes led to a relative decrease in employer power at striking mines compared to nonstriking mines, although this did not reverse the overall rise in employer power.

4.4 Fixed Costs Estimation

Solving the Model Conditional on Machine Usage

Using the estimated model, I solve the system of Equations (7), (10a), (11), (13), and (14) for every firm in every year, both if using cutting machines and if not using cutting machines. Given that

Table 3: Covariates of Employer Power Over Skilled Labor

	Log(Employer Bargaining Power)			
	Est.	S.E.	Est.	S.E.
High-skilled-employment market share	0.280	0.022	0.186	0.039
Low-skilled-employment market share	-0.311	0.022	-0.173	0.045
Year	0.011	0.002	0.007	0.003
Pop. share foreigners	0.925	0.298		
Pop. share Afro-Americans	2.650	0.702		
Log(manufacturing wage)	-0.192	0.070		
Firm FE:	No		Yes	
R-squared	.069		.673	
Observations	3466		3466	

Notes: In this table, I regress $\ln(1-\gamma_f)$ on firm and market characteristics.

this system of equations is nonlinear and cannot be solved analytically, I solve for it numerically.[24] For every outcome variable $Y \in \{Q, P, H, L, W\}$, this yields an optimal outcome if the firms used cutting machines, denoted as Y^1_{ft}, and if not, denoted as Y^0_{ft}.

Estimation of Cutting-Machine Costs

Using the estimated model, the cutting-machine probabilities at each firm in each year, $p^k_{ft}(\phi)$, can be computed using Equation (16), up to the unknown fixed-cost parameters (ϕ^0, ϕ^1, ϕ^2). I estimate these fixed-cost parameters using a maximum-likelihood estimator. Using Equation (16), the log likelihood function of using cutting machines $\ln(\mathcal{L}_{ft}(\phi)$ can be written as:

$$\ln(\mathcal{L}_{ft}(\phi)) = \sum_{f,t}[K_{ft}\ln(p^k(\phi)) + (1-K_{ft})\ln(1-p^k_{ft}(\phi))]$$

I estimate the machine cost parameters (ϕ^0, ϕ^1, ϕ^2) by maximizing the log likelihood function $\ln(\mathcal{L}_{ft}(\phi))$. Because the number of observed capital-usage decisions is sparse, I do not rely on the observed capital usages K_{ft} in the raw data; I rather estimate a conditional choice probability $\tilde{K}_{ft}$ first by running a probit model of cutting-machine usage on log Hicks-neutral and labor-

[24] I use the Matlab optimizer `fsolve` with function tolerance 10^{-3}, 10^5 maximum iterations, and 600 maximum function evaluations.

augmenting productivity, the labor supply shifter, and the coal demand shifter. I estimate this model on the entire sample and obtain predicted usage rates of cutting machines for every firm in every year. Next, I use these predicted usage rates in the log likelihood function to estimate cutting-machine fixed and variable costs. The resulting estimates are in Table 2(d). The fixed cost of using a cutting machine is estimated to be \$9,230 in 1882, and is estimated to fall by \$928 every two years throughout the sample period. This decline in machine fixed cost is in line with the falling costs of many new technologies. Variable machine costs are estimated to be 0.108 USD per ton of coal extracted. External cost information for cutting machines in 1889 from Brown (1889), reported a total purchasing cost of \$8,000 for eight cutting machines, which is the average number of cutting machines used in the dataset. The estimated average machine cost in the data 1888 is of a similar magnitude, at \$6,446.

Solving for Optimal Machine Usage

Using Equation (16), I estimate optimal cutting-machine usage for every firm in every year. I compute the optimal values $\hat{Y}_{ft}$ for variables $Y \in \{Q, P, H, L, W\}$ as the weighted average of the value when the firm used cutting machines and when if did not, weighted by the probability of using cutting machines:

$$\hat{Y}_{ft} = Y^0_{ft} Pr(K_{ft} = 0) + Y^1_{ft} Pr(K_{ft} = 1) \tag{23}$$

4.5 Model Validation

Evidence for Investment Holdup

Figure 4b shows average observed cutting-machine usage for each 5th percentile of the employer-bargaining-parameter distribution. Firms with higher levels of employer power are on average more likely to adopt cutting machines. This correlation is consistent with the investment-holdup channel in the model.

Although this correlation is suggestive of investment holdup, it does not establish causality as it does not rely on an exogenous shock in employer power. Therefore, I complement this evidence with documenting how capital investment and output evolved in response to a large miners' strike in Illinois in 1897–1898, which became known as the "Illinois coal war." Miners went on strike at one-quarter of the coal mines in Illinois; wages subsequently increased at 92% of these striking coal mines. Although the strike itself was likely endogenous, the fact that the strike was successful at increasing wages provides in-sample variation in union bargaining power that is useful to test the model predictions.

I estimate a difference-in-differences model comparing striking mines to nonstriking mines for an outcome variable y_{ft}. I start by estimating how the miner union's bargaining ability changed in response to the strike, by including $\log(\gamma_{ft})$ as the left-hand-side variable y_{ft} in Equation (24). I compare its evolution between striking mines for which $I(strike)_f = 1$ and nonstriking mines before and after 1898, when the strike occurred. I include a linear time trend.

$$y_{ft} = a_0 + a_1 I(strike)_f + a_2 I(strike)_f I(t \geq 1898) + a_3 I(t \geq 1898) + a_4 t + e_{ft} \tag{24}$$

The estimates are in panel (c) of Table 4. When comparing mines at which the strike was successful (defined as a wage increase in response to the strike) to mines that did not strike and mines that did but were not granted a wage increase, union power increased by 20.3% on average. When comparing all mines that went on strike, successfully or not, to nonstriking mines, union power increased on average by 14.9%.

Next, we look how the output changed differentially at striking and nonstriking mines after the 1898 miners strike, excluding the strike year of 1898, as output mechanically decreased that year. The first column of Panel (a) in Table 4 shows that output increased on average by 40.8% at mines that went on strike and obtained a wage increase compared to nonstriking and unsuccessfully striking mines. When comparing all striking mines, independently of the strike outcome, the increase is smaller, at 33.4%. This output expansion is important because it confirms the model assumption that the equilibrium lies on the labor supply curve, which is implied by the assumption that the union chooses employment. If the employers would choose employment, the equilibrium would lie on the labor demand curve and the wage increase due to the strike should lead to lower rather than higher employment and output.

Panel (b) in Table 4 regresses capital investment, measured as an indicator of whether or not the mine acquired a cutting machine, on the difference-in-differences regressors using a linear proba-

Table 4: Effects of the 1898 Miners Strike

	(I) Succesful strikes:		(II) All strikes:	
(a) Output	Log(Output)			
	Est.	S.E.	Est.	S.E.
1(Strike)*1(year$\geq$ 1898)	0.342	0.101	0.288	0.096
Firm fixed effects	Yes		Yes	
R-squared	.945		.944	
Observations	7154		7154	
(b) Capital Investment	1(Acquired Machine)			
	Est.	S.E.	Est.	S.E.
1(Strike)*1(year$\geq$ 1898)	-0.020	0.021	-0.036	0.019
Firm fixed effects	No		No	
R-squared	.006		.008	
Observations	3358		3358	
(c) Union Power	Log(Union Bargaining Power)			
	Est.	S.E.	Est.	S.E.
1(Strike)*1(year $\geq$ 1898)	0.185	0.067	0.139	0.067
R-squared	.712		.710	
Observations	.710		3787	

Notes: Panel (a) reports the difference-in-difference estimates of how output changed at striking mines relatively to non-striking mines, excluding the strike year. Panel (b) reports the same effects for an indicator variable of whether the mine acquired a cutting machine or not, using a linear probability model. Panel (c) reports the same effects for our estimated level of employer power.

bility model. I do not include mine fixed effects, as there is almost no within-mine variation over time in the adoption variable. The estimates reveal that capital investment decreases by 0.20 points for succesfully-striking mines and by 0.36 points for all striking mines, which are large changes given the average investment rate of 0.49. The investment decrease is statistically significant when using all striking mines as the treatment group. This reduction in capital investment as a result of increased union bargaining power again confirms the investment-holdup mechanism.

Model Fit

Figure A2 shows average observed and predicted cutting-machine usage by year. The model-predicted machine usage rates fit both the level and the time trend in observed mechanization rates reasonably well. In Appendix Table A1, I also compare the model-predicted quantities and prices to their observed values, and I find a reasonable model fit despite not explicitly targeting any of these moments in our estimation procedure, except for the average cutting-machine usage rate.

5 Welfare Effects of Employer Power

To examine the effects of changes in employer power, I compute how all quantities, prices, and welfare change for counterfactual values of employer bargaining power $\tilde{\gamma}_{ft}$. I examine a 5% increase in the level of employer power and compare its effects both when assuming that capital investment is exogenous and endogenous to employer power.

5.1 Welfare Computation

In both the actual and the counterfactuals, I compute consumer surplus CS_{ft} as the area in between the demand curve and the optimal price:

$$CS_{ft} \equiv \int_0^{\hat{Q}_{ft}} (P_0\Big(\frac{Q_{ft}}{\xi_{ft}}\Big)^{\frac{1}{\eta}} - \hat{P}_{ft})dQ_{ft} = \Big(\frac{-1}{\eta+1}\Big)\frac{P_0}{\xi^{\frac{1}{\eta}}}(\hat{Q}_{ft})^{\frac{\eta+1}{\eta}}$$

Similarly, I compute labor surplus LS_{ft} as the area between the labor supply curve and the optimal wage $\hat{W}_{ft}$:

$$LS_{ft} \equiv \int_0^{\hat{H}_{ft}} (\hat{W}_{ft} - \frac{W_{0,i(f)t}}{1+\psi}\Big(\frac{H_{ft}}{\zeta_{ft}}\Big)^{\psi})dH_{ft} = \frac{\psi}{(\psi+1)^2}\frac{W_0}{\zeta_{ft}^{\psi}}(\hat{H}_{ft})^{\psi+1}$$

Finally, producer surplus is equal to variable employer profits:

$$PS_{ft} \equiv \hat{P}_{ft}\hat{Q}_{ft} - \hat{H}_{ft}\hat{W}_{ft} - \hat{L}_{ft}V_t$$

5.2 Counterfactual: Increased Employer Power

I conduct a counterfactual exercise in which employer power $(1-\gamma)$ increases by 5% at all firms. I resolve all equilibrium values $\hat{\mathbf{Y}}$ for the higher values of employer power. The results are in Panel (a) of Table 5. The first column reports the average changes in the selected variables if capital investment is assumed to be exogenous to employer power, whereas the second column allows cutting-machine usage to adjust. Without adjustment in capital investment, output would fall by by 15.7%, because the exertion of monopsony power induces deadweight loss. However, capital investment does not remain fixed: the increase in employer power results in an increase in the cutting-machine usage rate from 2.1% to 3.1%, a relative increase of 45%. This indicates that investment holdup is substantial. As a result, the output reduction is 12% smaller when taking into account endogenous cutting-machine usage: increased investment lowers marginal costs, which leads to output expansion.

Given that output falls, employer power reduces both consumer and labor welfare, by 11.2% and 22.1%, respectively. These losses would be overestimated by 13% and 7%, respectively, if capital investment were assumed to be exogenous. In contrast, producer surplus increases by 62.2%, as increased employer power allows them to capture a larger share of the rents. Adding up consumer, labor, and employer surplus, increased employer power leads to a total welfare gain of 0.7%, whereas assuming exogenous capital investment would predict a welfare *loss* of 1.7%.

5.3 Increased Employer Power for Counterfactual Technology

I reconsider the effects of employer power for a counterfactual technology that has larger productivity effects. I set both β^k and α^k to three times their estimated value and again compute the effects of a 5% increase in employer power. The results are in Panel (b) of Table 5. The increase in employer power now still decreases output, but by only 3%, which indicates that the holdup-reduction effect nearly outweighs the deadweight-loss channel. As a result, increased employer power now barely changes consumer surplus, whereas labor surplus still falls by 17.3%. Hence, this simulation shows that technologies exist under which employer power has very different effects on consumer and labor surplus. This stands in contrast with the usual monopsony models, in which both consumer and labor surplus suffer to similar extents as monopsony depresses both output and input quantities (Hemphill & Rose, 2018).[25] The cutting machines, because they are a (skilled) labor-saving technology, reduce miner employment relative to other inputs and production, despite increasing output through reduced marginal costs. This explains why output barely falls, and could even increase for technologies with even stronger effects, whereas wages—and possibly also employment—fall sharply, thereby harming workers.

Table 5: Effects of a 5% Increase in Employer Power

(a) Observed Technology	% Changes:	
Endogenous Investment?	No	Yes
Output	-15.741	-13.814
Cutting machine usage	0.000	44.832
Consumer surplus	-12.857	-11.214
Producer surplus	56.508	62.158
Worker surplus	-23.656	-22.074
Total surplus	-1.709	0.710
(b) Technology with Larger Productivity Effects	% Changes:	
Endogenous Investment?	No	Yes
Output	-20.745	-2.829
Cutting machine usage	0.000	77.535
Consumer surplus	-16.738	-1.531
Producer surplus	91.492	147.219
Worker surplus	-31.178	-17.261
Total surplus	3.867	27.682

Notes: Panel (a) reports average percentage changes for output, machine usage, and the different welfare metrics if employer power increases by 5%. The first column assumes cutting machine usage remains constant, the second column allows for endogenous cutting machine usage. Panel (a) reports the counterfactuals for the observed technology, Panel (b) for a technology with higher productivity effects.

[25] One exception is models of double marginalization, under which employer power also benefits consumers but harms workers; however, these models are not "monopsonistic," as equilibrium lies on the labor demand, rather the labor supply, curve.

6 Conclusion

In this paper, I investigate the welfare effects of employer power by studying the trade-off between monopsony distortions and endogenous investment. Using a model of production and labor supply that allows for monopsony power, wage bargaining, and imperfectly competitive goods markets, I find that an increase in employer power could either increase or decrease output and total welfare, depending on the relative size of the monopsony distortion, on the marginal-cost reduction due to endogenous investment, and on the initial level of employer power.

In the empirical context of the mechanization of the late-19th-century Illinois coal mining industry, I find that an increase in employer power lowered output, because the monopsony distortion dominated the marginal-cost reduction that was due to the adoption of additional coal-cutting machines. Although consumer and labor welfare declined in response to increased employer power, this decline is 7% to 13% smaller than one would find when holding capital investment fixed. The total welfare effects of employer power are underestimated to an even larger extent. Therefore, the results indicate that taking into account endogenous capital quantitatively matters for assessing the welfare effects of labor market power.

References

Abowd, J. A., & Lemieux, T. (1993). The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada. *The Quarterly Journal of Economics*, *108*(4), 983–1014.

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

Acemoglu, D., & Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, *40*(4), 827–849.

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.

Angerhofer, T., Collard-Wexler, A., & Weinberg, M. (2025). *Oligopsony and collective bargaining.*

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

Azkarate-Askasua, M., & Zerecero, M. (2024). Union and firm labor market power. *Available at SSRN 4323492*.

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.

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.

Chambolle, C., Christin, C., & Molina, H. (2023). Buyer Power and Exclusion: A Progress Report. *International Journal of Industrial Organization*, 102969.

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

Degrève, D. (1982). *Le commerce extérieur de la Belgique, 1830-1913-1939: Présentation critique des données statistiques* (Vol. 1b). Brussels: Palais des Académies.

Delabastita, V., & Rubens, M. (2024). Colluding Against Workers. *Available at SSRN*.

Demirer, M., & Rubens, M. (2025). *Welfare effects of buyer and seller power* (Tech. Rep.). National Bureau of Economic Research.

Dodini, S., Salvanes, K. G., & Willén, A. (2022). *The dynamics of power in labor markets: Monopolistic unions versus monopsonistic employers*. IZA-Institute of Labor Economics.

Doraszelski, U., & Jaumandreu, J. (2018). 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.

Goldin, C., & Katz, L. F. (2009). *The race between education and technology*. Harvard University Press.

Goolsbee, A., & Syverson, C. (2023). Monopsony power in higher education: A tale of two tracks. *Journal of Labor Economics*, *41*(S1), S257-S290.

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

Grout, P. A. (1984). Investment and wages in the absence of binding contracts: A nash bargaining approach. *Econometrica: Journal of the Econometric Society*, 449–460.

Hemphill, C. S., & Rose, N. L. (2018). Mergers that harm sellers. *Yale Law Journal*, *127*, 2078.

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

Horn, H., & Wolinsky, A. (1988). Bilateral monopolies and incentives for merger. *The RAND Journal of Economics*, 408–419.

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.

Illinois Bureau of Labor Statistics, . (1890). *Sixth biennal report of the bureau of labor statistics of illinois* (Tech. Rep.). Springfield, Illinois: Bureau of Labor Statistics of Illinois.

Joskow, P. L. (1987). Contract duration and relationship-specific investments: Empirical evidence from coal markets. *The American Economic Review*, 168–185.

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.

Kinder, M. (2024). Hollywood writers went on strike to protect their livelihoods from generative ai. their remarkable victory matters for all workers. *Brookings Institute*.

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., Muraközy, B., Reizer, B., & Schreiner, R. (2022). Firm-level technological change and skill demand.

Loertscher, S., & Marx, L. M. (2019). Merger review for markets with buyer power. *Journal of Political Economy*, *127*(6), 2967–3017.

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).

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

Menezes-Filho, N., & Van Reenen, J. (2003). Unions and innovation: a survey of the theory and empirical evidence. *International handbook of trade unions*, 293–334.

Miller, N., Osborne, M., Sheu, G., & Sileo, G. (2023). Technology and market power: The united states cement industry, 1974-2019. *Georgetown McDonough School of Business Research Paper*(4041168).

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*.

of Natural Resources, I. D. (2025). *The archaeology of 19th century canal boats.* (Accessed through https://dnr.illinois.gov/naturalresources/cultural/canalboatsp2.html)

Parra, A., & Marshall, G. (2024). Monopsony power and upstream innovation. *Journal of Industrial Economics*, *72*(2), 1005-1020.

Peters, B., Roberts, M., Vuong, V. A., & Fryges, H. (2017). Estimating Dynamic R&D Demand: An Analysis of Costs and Long-Run Benefits. *RAND Journal of Economics*, *48*(2), 409-437.

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. (2023). Market Structure, Oligopsony Power, and Productivity. *American Economic Review*, *113*(9), 2382–2410.

Rubens, M., Wu, Y., & Xu, M. (2024). *Exploiting or Augmenting Labor?* (Tech. Rep.).

Shi, L. (2023). Optimal regulation of noncompete contracts. *Econometrica*, *91*(2), 425–463.

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*.

Williamson, O. E. (1971). The vertical integration of production: market failure considerations. *The American Economic Review*, *61*(2), 112–123.

Zahur, N. B. (2022). Long-term contracts and efficiency in the liquefied natural gas industry. *Available at SSRN 4222408*.

Online Appendix

A Data Appendix

A.1 Sources

Mine Inspector Reports

My main data source is the biennial report of the Bureau of Labor Statistics of Illinois; I collected the volumes issued from 1884 to 1902. Each report contains a list of all mines in each county and contains 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, appears 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 include mine type, hauling technology, other technical characteristics, and other inputs—are reported in a subset of years.

Censuses of Population, Agriculture, and Manufacturing

I rely on three 1880 Illinois censuses—population, manufacturing, and agriculture—by digitizing copies of the original prints, which I accessed from the HathiTrust Digital Library[26]. The population census yields information on county population sizes, demographic compositions, and areas; the manufacturing census allows me to observe county-level capital stock and employment in manufacturing industries; and the agriculture census contains the number of farms and the amount of improved farmland area.

Monthly Data

The 1888 report contains monthly production data for a selection of 11 mines in Illinois, across six 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,

[26] https://www.hathitrust.org/.

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.

A.2 Data Cleaning

Employment

In every year except 1896, workers are divided into two categories: “miners” and “other employees.” In 1896, a different distinction is made: “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). 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 employees” 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

Miner wages are the only ones 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-rate wages using the number of workers employed in each season for the years 1884 to 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 onward, wages are no longer reported seasonally, because wages were negotiated biennially from that year onward. 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 (2,000 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 and 1894 or 1896 and 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. 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.[27]

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

Hours Worked

In 1898, eight-hour days were enforced by law for the first time, which means that the "number of days" measure changes in unit between 1898 and 1900. Because the Inspector Report from 1886 shows that ten-hour days were the standard, I multiply the number of working days after 1898 by 80% 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 multimine firms. For instance, the Illinois Valley Coal Company No. 1 and Illinois Valley Coal Company No. 2 mines clearly belong to the same company. For single-mine firms, the operator is usually mentioned as the mine name, (e.g. "Floyd Bussard"). For the multimine firms, I made mine names consistent over time as much as possible.

Town Identifiers and Labor-Market Definitions

The dataset 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 two 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 480 to 391. The resulting labor markets lie at least two 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 at the crossroads of multiple railroad lines. 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 the firm-bi-year level by taking sums of the total days worked 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 by 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 a "firm" as the combination of the firm name in the dataset and its town (the merged towns that are used to define labor markets), as firms are assumed to optimize input usage on a town-by-town basis.

B Theory Appendix

B.1 Proofs

B.1.1 Proof of Proposition 1

This proposition is equivalent to Lemma 1 in Demirer and Rubens (2025), and proven in their Appendix B.5.

B.1.2 Proof of Proposition 2

Proof. The downstream profit returns to capital are $B(\gamma_f, K_f) \equiv \frac{d\Pi_f^d}{dK_f}$. We prove that these returns decrease with union power, $\frac{dB(\gamma_f,K_f)}{d\gamma_f} < 0.$. Given that the per-unit costs of capital are invariant to union power, decreasing capital returns imply decreased capital usage.

We start by writing downstream profits as a share $\tilde{\gamma}$ of total profits $\Pi_f^{tot} = P_f Q_f - V_f L_f - \frac{Z_f H_f}{1+\psi}$:

$$\Pi_f^d = (1 - \tilde{\gamma}_f)\Pi_f^{tot}$$

In contrast to models of efficient bargaining, in which $\tilde{\gamma} = \gamma$, the upstream profit share is not necessarily identical to the bargaining weight $\tilde{\gamma}$ in models with linear price contracts. However, $\frac{\partial\tilde{\gamma}}{\partial\gamma} > 0$: a higher bargaining weight on upstream profits results in a higher share of total profits that accrue to the upstream party.

Denote the marginal return to capital as $MRK_f \equiv \frac{dQ_f}{dK_f}$. Taking the derivative of downstream profits to union power γ_f results in:

$$\frac{dB(\gamma_f, K_f)}{d\gamma_f} = -\underbrace{\frac{d\Pi_f^{tot}}{dK_f}}_{(I)}\underbrace{\frac{d\tilde{\gamma}}{d\gamma}}_{>0} + (1-\tilde{\gamma}_f)(\underbrace{\frac{d(\frac{d\Pi_f^{tot}}{dK_f})}{d\gamma_f}}_{(II)})\underbrace{\frac{d\tilde{\gamma}}{d\gamma}}_{>0}$$

The term $\frac{d\Pi^{tot}}{dK_f}$ is strictly positive, as capital would otherwise never be adopted, even with full employer power. Therefore, term (I) is negative. Higher union power reduces the downstream's share of total profits, so keeping total profits fixed, the profit return to employers decreases with union power.

The employer's profit weight $(1-\tilde{\gamma}_f) > 0$ is positive by construction. We now prove that term (II) is negative. First, write out the total profit returns to capital using the chain rule:

$$
\begin{aligned}
\frac{d}{d\gamma_f}\Big[\frac{d\Pi_f^{tot}}{dK_f}\Big] &= \frac{d}{dQ_f}\Big[\frac{d\Pi_f^{tot}}{dK_f}\Big]\frac{dQ_f}{d\gamma_f} \\
&= \frac{d}{dQ_f}\Big[\frac{d\Pi_f^{tot}}{dQ_f}\frac{dQ_f}{dK_f}\Big]\frac{dQ_f}{d\gamma_f} \\
&= \Big[\underbrace{\frac{\partial}{\partial Q_f}(\frac{\partial Q_f}{\partial K_f})}_{<0}\underbrace{\frac{\partial \Pi_f^{tot}}{\partial Q_f}}_{>0} + \underbrace{\frac{\partial}{\partial Q_f}(\frac{\partial \Pi_f}{\partial Q_f})}_{<0}\underbrace{\frac{\partial Q_f}{\partial K_f}}_{>0}\Big]\underbrace{\frac{dQ_f}{d\gamma_f}}_{>0} < 0
\end{aligned}
$$

Proposition 1 states that $\frac{dQ_f}{d\gamma_f} > 0$. Given that the production function is increasing and concave in K, this means that $\frac{dQ_f}{dK_f} > 0$ and $\frac{d}{dQ_f}(\frac{dQ_f}{dK_f}) < 0$. The term $\frac{\partial \Pi_f^{tot}}{\partial Q_f} > 0$ because the monopsonistic equilibrium results in an output level below joint profit maximization. Hence, increasing output locally results in increased total profits. Finally, total profits are concave, meaning that $\frac{\partial}{\partial Q_f}(\frac{\partial \Pi_f}{\partial Q_f}) < 0$. To see this, denote marginal costs of low- and high-skilled labor as $MCL_f \equiv \frac{d}{dQ_f}(V_f L_f)$ and $MCH_f \equiv \frac{d}{dQ_f}(\frac{Z_f H_f}{1+\psi})$:

$$
\frac{\partial}{\partial Q_f}(\frac{\partial \Pi_f}{\partial Q_f}) = \underbrace{\frac{d}{dQ_f}(MR_f)}_{<0} - \underbrace{\frac{d}{dQ_f}(MCH_f)}_{>0} - \underbrace{\frac{d}{dQ_f}(MCL_f)}_{>0} < 0
$$

The concavity of the product demand curve and convexity of the marginal cost curves for both labor types imply that total profits are concave in output.

Bringing these different terms together, it becomes clear that both $(I) < 0$ and $(II) < 0$. Therefore, $\frac{dB(\gamma_f, K_f)}{d\gamma_f} < 0$, which concludes the proof.

$\square$

B.2 Efficient Bargaining

B.2.1 Behavior

In the main text, it is assumed that firms and unions bargain over linear wage contracts. In this Section, I discuss an alternative model of vertical conduct: efficient bargaining between unions and firms over both wages *and* employment. This provides a useful benchmark against which to compare the results from the full model, because the strongly efficient model does not feature any monopsony distortions, only endogenous technology choices.

In the efficient bargaining model, employers and unions bargain over both employment and wages in a Nash bargaining protocol, with γ_f still indicating union bargaining power. Crucially, the unions and employers still do not bargain over technology adoption.

$$\max_{H_f, L_f, W_f} (\Pi_f^u)^{\gamma_f} (\Pi_f^d)^{1-\gamma_f}$$

The model implies that the union and employers jointly optimize joint profits and split the surplus according to the bargaining parameters γ_f. Taking the first-order condition for the high-skilled wage results in:

$$W_f = (1-\gamma_f)\frac{Z_f}{1+\psi} + \gamma_f\Big(\frac{P_f Q_f - V L_f}{H_f}\Big) \tag{25}$$

The first-order conditions for the labor inputs are given by:

$$P_0\Big(\frac{1+\eta}{\eta}\Big)Q_f^{\frac{1}{\eta}}\Big(\frac{Q_f}{H_f}\Big)^{\frac{1}{\sigma}}(\Omega_f A_f)^{\frac{\sigma-1}{\sigma}} = \frac{W_0}{1+\psi}\Big(\frac{H_f}{\zeta_f}\Big)^{\psi} \tag{26}$$

$$P_0\Big(\frac{1+\eta}{\eta}\Big)Q_f^{\frac{1}{\eta}}\Big(\frac{Q_f}{L_f}\Big)^{\frac{1}{\sigma}}(\Omega_f)^{\frac{\sigma-1}{\sigma}}\beta_l = V \tag{27}$$

Optimal quantities and prices $(P_f^*, Q_f^*, H_f^*, L_f^*)$ is the solution to the system of equations (1), (2), (26), and (5): the production function, the goods demand curve, and the two input demand equations. Wages are determined as a function of the bargaining parameter, as described in Equation (25); they do not have any effect on output, inputs, and goods prices as long as capital is held fixed.

An important difference between the model highlighted above and the model of Abowd and Lemieux (1993) is that the latter assume the outside option of workers to be a scalar, whereas I allow the outside option to be increasing. The analog of this feature in models of vertical relations would be that sellers face increasing marginal costs.

B.2.2 Effects of Employer Power: Endogenous Investment

Although employer power $(1-\gamma_f)$ does not affect output when holding the capital stock K_f fixed, employer power increases investment, which in turn affects marginal costs and, hence, output. Suppose firms need to pay a capital cost ϕ per unit of capital K_f, which is a fixed cost because it does not vary with production. I maintain the assumption that the technology increases variable profits. Proposition 3 says that under strongly efficient bargaining, employer power increases firms' technology adoption.

Proposition 3. *Under strongly efficient bargaining, buyer power increases capital investment:*

$$\frac{\partial K_f}{\partial(1-\gamma_f)} > 0$$

The proof of this theorem is straightforward. Denoting joint variable profits as $\Pi^j \equiv \Pi^d + \Pi^u$, the effect of capital on total employer profits $\overline{\Pi}^d = \Pi^d - \phi K$ is given by:

$$\frac{\partial \overline{\Pi}_f^d}{\partial K_f} = \frac{\partial \Pi_f^d}{\partial A_f}\frac{\partial A_f}{\partial K_f} - \phi = (1-\gamma_f)\frac{\partial \Pi_f^j}{\partial A_f}\frac{\partial A_f}{\partial K_f} - \phi$$

Taking the derivative with respect to employer power $(1-\gamma)$ gives:

$$\frac{\partial}{\partial(1-\gamma_f)}\Big(\frac{\partial \Pi_f^d}{\partial K_f}\Big) = \frac{\partial \Pi_f^j}{\partial A_f}\frac{\partial A_f}{\partial K_f}$$

This last term is positive under the assumption that the technology is variable enhancing variable profits.

The intuition behind Proposition 3 is that buyer power increases the share of the rents created by capital investment that flows to the buyer. Hence, this increases the incentive for the buyer to invest. This a reformulation of the well-known holdup mechanism from Williamson (1971), which hinges on the assumption that workers and firms can only write incomplete contracts that do not condition on investments by the employer. The wage contracts used in the Illinois coal mining industry are an example of such an incomplete contract.

Corrolary 1. *Under strongly efficient bargaining, buyer power increases output.*

It follows immediately from Proposition 3 that employer power increases output in the strongly efficient bargaining model. Given the strong efficiency assumption, employer power does not affect output conditional on technology adoption K_f. However, employer power increases technology adoption, hence, decreases marginal costs. This marginal-cost reduction results in increased output.

B.3 Simulating the Theoretical Model

B.3.1 Baseline Parametrization

In Section 2.3, I simulate the theoretical model with the following parameter values. I use the estimates from Kroft et al. (2020) for the U.S. construction industry to set the product-demand elasticity to $\eta = -7$ and the inverse labor-supply elasticity to $\psi = 0.25$. I calibrate the elasticity of substitution between high- and low-skilled labor at $\sigma = 0.7$. I normalize most parameters at one: $\xi = 1$, $\zeta = 1$, $w_0 = p_0 = v = 1$, $\omega = 1$, $a = 1$. I set the low-skilled production coefficient at 0.2: $\beta^l = 0.1$. I simulate a dataset with 50 observations, in which the bargaining parameter γ_f is distributed uniformly between 0 and 1. I let fixed technology costs be distributed as an exponential distribution with a mean of 0.05.

Under these parametrizations, I solve the system of equations (1), (2), (5), (4), (6) for (Q, P, W, H, L).

B.3.2 Alternative Parametrizations

In Figure A1, I compare the baseline calibration of the structural model to various alternative parametrizations. First, I let labor supply be more inelastic. Second, I increase the productivity effects of the new technology.

C Extensions and Robustness Checks

C.1 Model Fit

Table A1 compares the model-predicted outcomes against the observed outcomes in the data. I use medians for all variables except for cutting-machine usage, for which I report the average (as median machine usage is zero). Cutting-machine usage is almost identical in the model and the data. The model generates coal quantities and prices that are very similar to those observed in the data. Both unskiled and skilled labor days worked are overestimated, and skilled wages are overestimated compared to their true values. However, the estimation of the model does not target any of these moments, except for the capital-investment rate through the maximum-likelihood estimation of fixed costs. Considering that these moments are untargeted, the model fits the data reasonably, especially for the variables that are relevant to compute consumer surplus (output quantities and prices).

C.2 Alternative Production-Function Specifications

C.2.1 Nonconstant Returns to Scale

In the main text, the production function (1) relied on constant returns to scale. In contrast, Equation (30) allows for nonconstant returns to scale, as parametrized by ν:

$$Q_f = \Big((A_f(K_f)H_f)^{\frac{\sigma-1}{\sigma}} + \beta^l L_f^{\frac{\sigma-1}{\sigma}}\Big)^{\frac{\nu\sigma}{\sigma-1}} \Omega_f(K_f) \tag{28}$$

The first step of the production-function estimation procedure, the estimation of Equation (18), remains the same. However, the second step of the estimation procedure needs to estimate the scale parameter ν in addition to the other production-function coefficients ρ^ω, β^l, and β^k. Given that we have four instruments (lagged employment for both labor types, and current and lagged capital), the model is still identified.

$$q_{ft} = \frac{\nu\sigma}{\sigma-1} \ln\Bigg(\Big(\exp\Big((\frac{l_{ft}-h_{ft}}{1-\sigma}) - \frac{\sigma}{1-\sigma}(\ln(\beta^l)) - \frac{\sigma}{1-\sigma}(w_{ft} - v_t + \ln(1+(1-\gamma_{ft})\psi))\Big)H_{ft}\Big)^{\frac{\sigma-1}{\sigma}} + \beta^l L_{ft}^{\frac{\sigma}{\sigma-1}}\Bigg) + \omega_{ft}$$

The results are in the first column of Table A5. The scale parameter is estimated at 1.032, which indicates modestly increasing returns to scale, but is not significantly different from 1. Hence, the assumption of constant returns to scale cannot be rejected. The other production coefficients look very similar to the estimates in the main model, which assumes constant returns to scale.

C.2.2 Adding Materials

As a second robustness check, I add the materials to the production function as a third production input. I use the number of kegs of black powder to measure materials, as this is the main intermediate input that is measured in the dataset. This implies that a fifth coefficient, β^m, needs to be estimated. I assume that changing the stock of black powder requires adjustment costs: black powder is a durable good but needs to be safely stored. Hence, it is conceivable that there was an adjustment cost when increasing the stock of black powder, as additional storage space was needed. Conforming with this assumption, I include current and lagged materials as an additional instrument when estimating the production function:

$$Q_f = \left((A_f(K_f)H_f)^{\frac{\sigma-1}{\sigma}} + \beta^l L_f^{\frac{\sigma-1}{\sigma}} + \beta^m M_f^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\nu\sigma}{\sigma-1}} \Omega_f(K_f) \tag{29}$$

The estimates are in the second column of Table A5. The material coefficient is estimated to be very close to zero, which means that ignoring materials in the main production model does not matter much. The remaining production coefficient looks very similar to the previous ones, with the exception of the serial correlation in TFP, which increases to 0.516.

C.2.3 Capital and Returns to Scale

The degree of returns to scale may have changed when firms adopted cutting machines. To test this, I interact the returns-to-scale parameter with the cutting-machine indicator variable, thereby allowing returns to scale to differ between firms that do and do not use cutting machines. Now, an additional instrument is needed to identify all six parameters in the production function. I rely on nonfatal accident rates as shifters of labor supply, which should directly affect input usage but not productivity. I measure the probability of nonfatal accidents as the ratio of the number of such accidents over total employment at the mine, in days worked:

$$Q_f = \left((A_f(K_f)H_f)^{\frac{\sigma-1}{\sigma}} + \beta^l L_f^{\frac{\sigma-1}{\sigma}} + \beta^m M_f^{\frac{\sigma-1}{\sigma}}\right)^{\frac{(\nu_0+\nu_1 K_f)\sigma}{\sigma-1}} \Omega_f(K_f) \tag{30}$$

The estimates are in the third column of Table A5. The interaction effect between returns to scale and cutting machines is close to zero and not statistically significant. Hence, the null hypothesis that returns to scale are invariant to cutting-machine usage cannot be rejected.

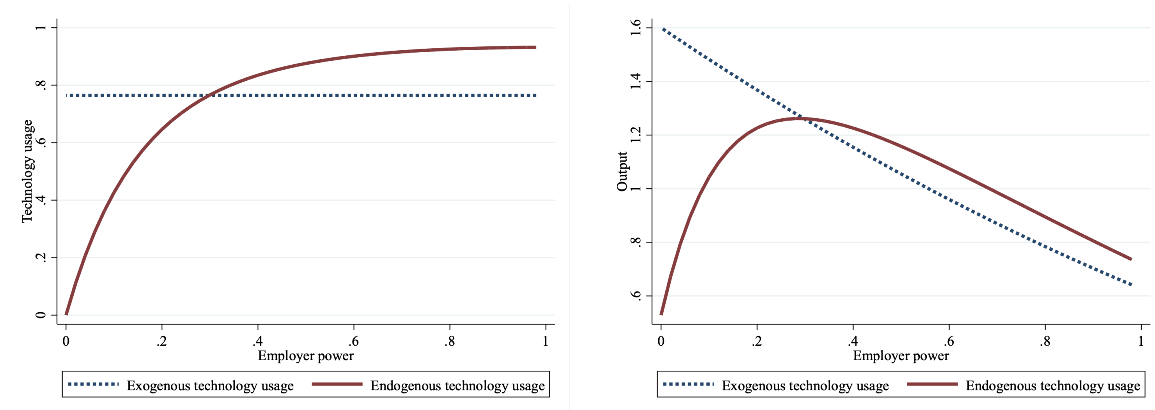
C.2.4 Cost Dynamics

In Table A3, in the spirit of Benkard (2000), I test for cost dynamics by regressing labor productivity, measured as output per labor-day, on log cumulative output. I find that when not taking mine fixed effects, cumulative past output correlates with higher productivity. However, this is likely due to a selection effect: more-productive mines exist longer and produce more. As soon as I include mine fixed effects and look at time-series variation in productivity within mines, the relationship between log cumulative output and labor productivity vanishes. This suggests that cost dynamics are not a key feature to be included in the model.

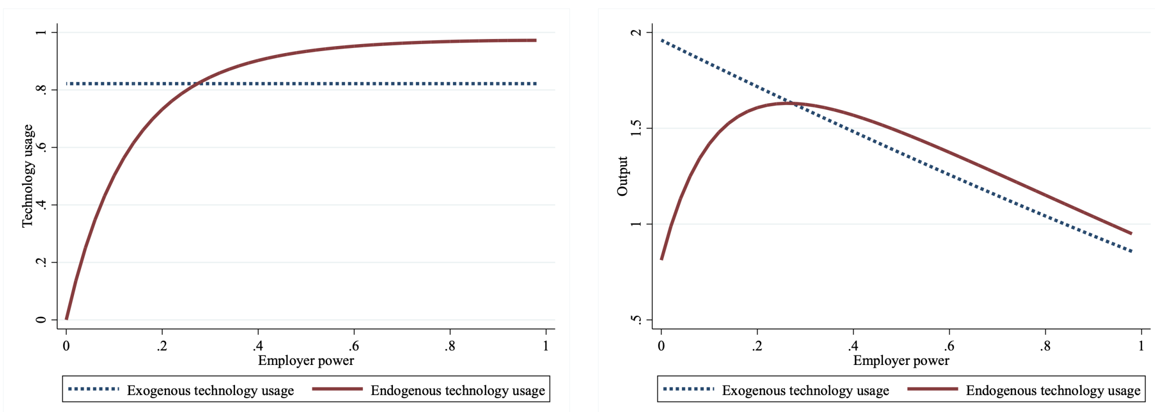
C.3 Appendix Tables and Figures

Figure A1: Simulations: Alternative Parametrization

(a) Baseline: $\psi = 0.25, \beta^k = 0.2$



(b) $\psi = 0.5, \beta^k = 0.2$



(c) $\psi = 0.25, \beta^k = 0.05$

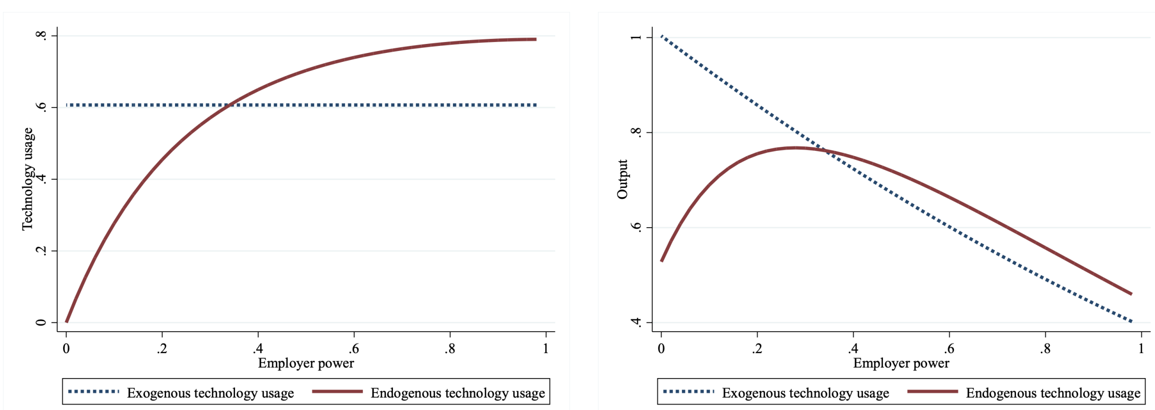
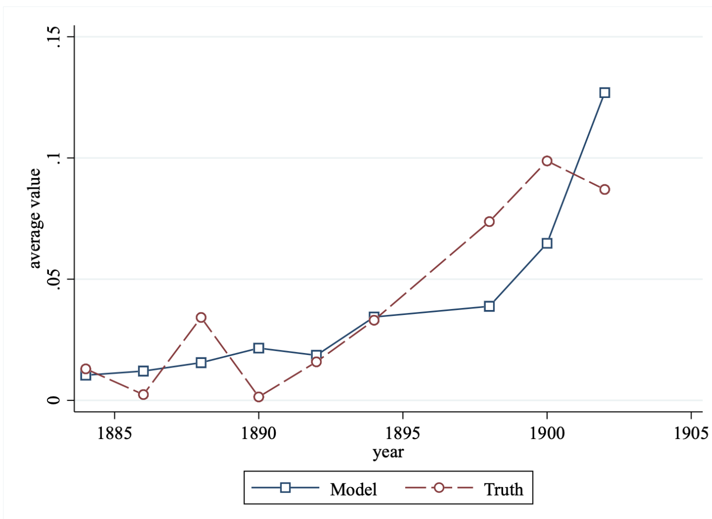
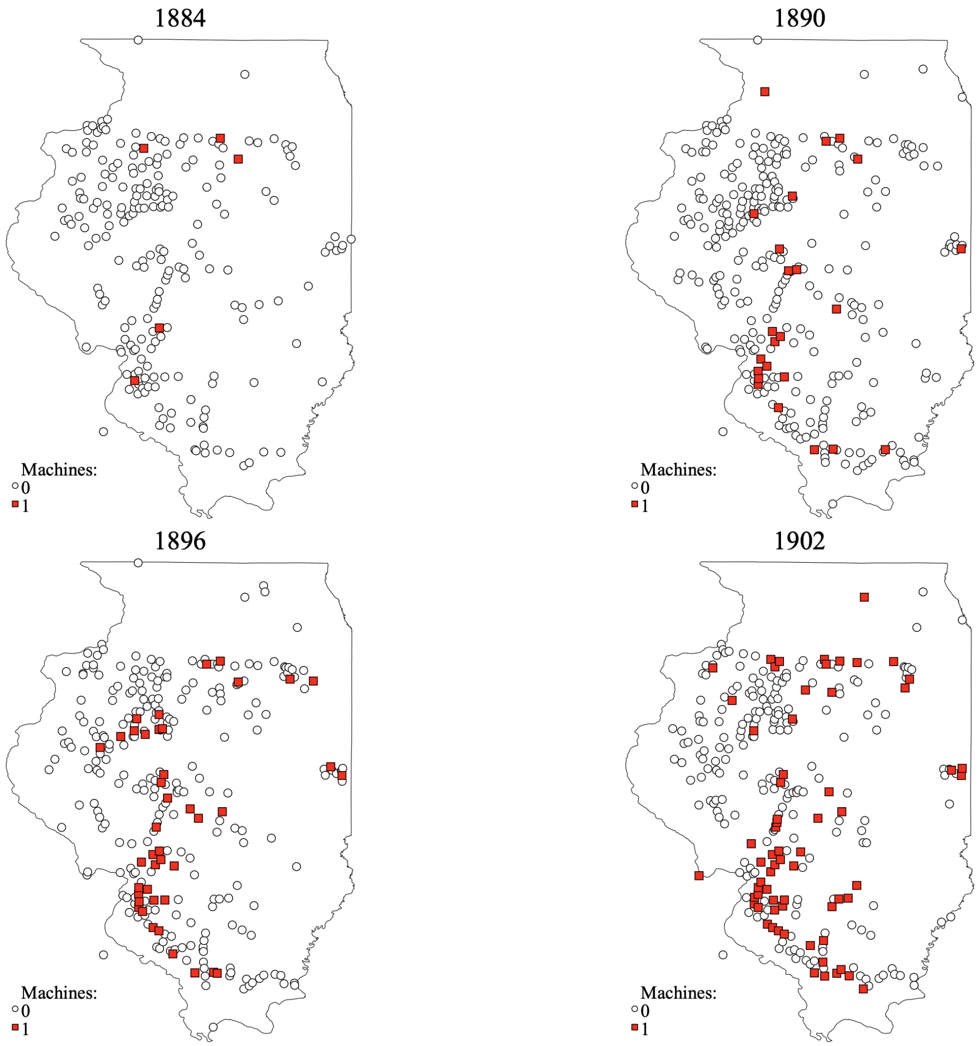


Figure A2: Predicted and Observed Evolution of Machine Usage



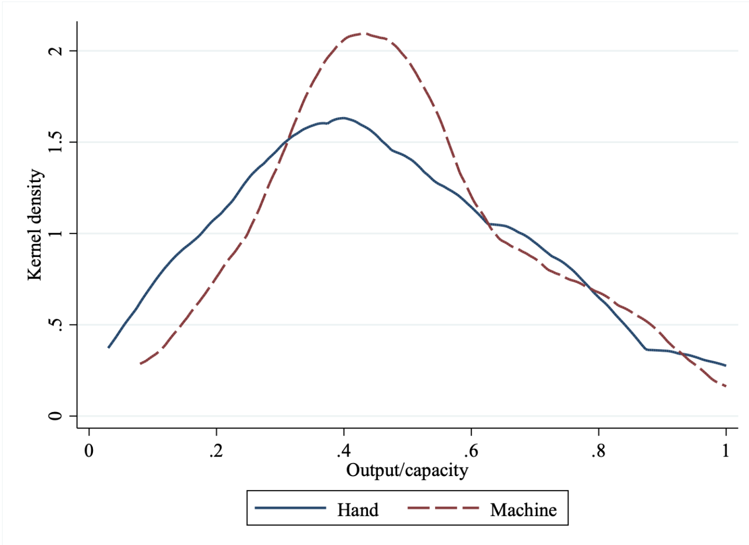
Notes: This figure compares annual average observed and model-predicted cutting-machine usage.

Figure A3: Geographical Spread of Cutting Machines



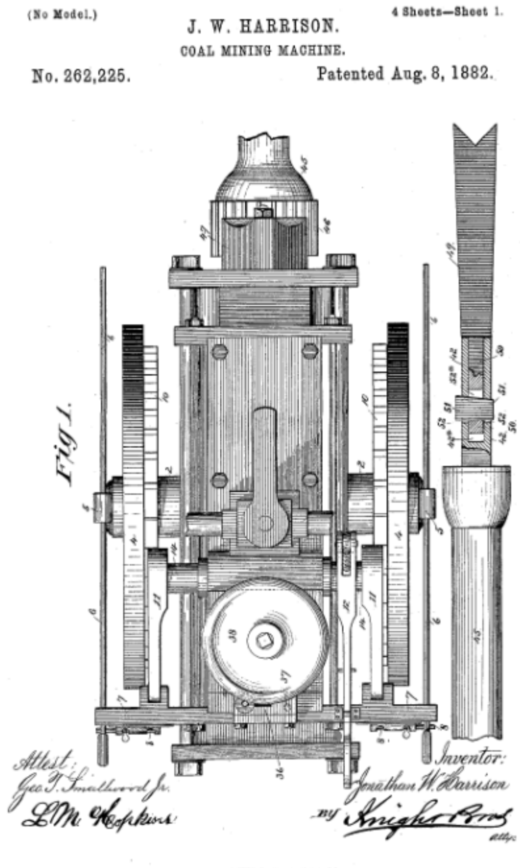
Notes: The dots indicate mining towns, each of which can contain multiple mines. Towns with squares contain at least one machine mine.

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 Illinois in 1898. I distinguish hand mines, which did not use cutting machines, from machine mines, which did.

Figure A5: Patent for the Harrison Coal Mining Machine



Notes: U.S. patent for the 1882 Harrison Coal Mining Machine (Whitcomb, 1882). This was the most frequently used coal-cutting machine in the dataset.

Table A1: Model Fit

	Observed	Model
Cutting machine usage (share of firms)	0.040	0.037
Output (tons)	1170.832	855.928
Coal price (USD/ton)	1.426	2.219
Unskilled labor (days)	172.949	185.388
Skilled labor (days)	524.921	484.025
Skilled wage (USD/day)	1.943	1.065

Notes: This table compare median values for quantities and prices between the observed data and the predicted values in the model. For cutting-machine usage, averages are compared because the median usage is zero. None of the variables are targeted moments in the model estimation, except for cutting-machine usage.

Table A2: 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 from the 1890 Inspector Report (Illinois Bureau of Labor Statistics, 1890). The 20 occupations with the highest employment shares together cover 97% of coal-mining workers in the sample.

Table A3: Cost Dynamics

	Log(Output/(Labor-Days))			
	Est.	S.E.	Est.	S.E.
Log(Cumulative Output)	0.126	0.004	-0.010	0.017
Mine FE	No		Yes	
R-squared	.336		.818	
Observations	3614		3614	

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

Table A4: Inverse Coal Demand: Local vs. Shipping Mines

	Log(Price)		Log(Price)	
	Est	SE	Est	SE
log(Output)	-0.169	0.016	-0.449	0.155
Shipping Mine?	No		Yes	
Observations	2748		379	

Notes: I estimate inverse coal demand on a split sample of mines that do and do not sell locally.

Table A5: Production Function: Extensions

	Nonconstant RTS		Adding Materials		Capital and RTS	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Returns to scale	1.032	0.041	1.051	0.088	0.981	0.118
Labor coefficient	0.009	0.011	0.013	0.029	0.003	1.303
Capital coefficient	0.029	0.162	-0.053	0.191	0.861	1.436
Serial corr. TFP	0.347	0.119	0.515	0.160	0.372	0.260
Materials coefficient		.	0.000	0.042	0.000	0.014
Returns to scale * K		.		.	-0.011	0.022
Observations	668		298		298	

Notes: This table reports the estimates for the various extensions of the production function. Standard errors are block-bootstrapped with 200 iterations.

Table A6: Wage Variation

	R^2	R^2	R^2	R^2
Log(Daily Skilled Miner Wage)	0.099	0.186	0.285	0.734
Year FE	X	X	X	X
County FE		X	X	X
Town FE			X	X
Firm FE				X

Notes: The four columns report the R^2 of regressing log wages on, alternatively, year, county, town, and firm fixed effects.

Table A7: All Variables per Year

Year	1884	'86	'88	'90	'92	'94	'96	'98	1900	'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 dollars)	X									

Table A8: All Variables per Year (cont.)

Year	1884	’86	’88	’90	’92	’94	’96	’98	1900	’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			
Mine Characteristics										
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 of 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 Machines										
Cutting machine dummy					X	X	X	X		
# Cutting machines	X	X	X	X						
# Tons cut by machines									X	X
# Cutting machines, by type			X							