Monopsony power and factor-biased technology adoption

Michael Rubens*

November 30, 2020

Abstract

This paper studies how the degree of monopsony power on input markets affects buyers’ incentives to adopt factor-biased production technologies. I find that monopsony power over the factor on which a technology saves decreases the incentive to adopt that technology, and vice-versa. As an empirical application, I study the mechanization of the Illinois coal mining industry between 1884 and 1902. I construct and estimate a model of monopsonistic labor markets with dynamic capital demand, which I estimate using mine-level production and cost data. I find that if the market for miners would have been perfectly competitive, the usage rate of coal cutting machines, a miner-saving technology, would have increased by 27% per year.

Keywords: Monopsony power, Innovation, Technological change, Productivity

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

*KU Leuven and Research Foundation Flanders, michael.rubens@kuleuven.be

Preliminary, please do not circulate. I thank Jo Van Biesebroeck, Jan De Loecker, Frank Verboven, Chad Syverson, Otto Toivanen, Penny Goldberg, Steve Berry, Suresh Naidu, Daniel Haanwinckel and participants at the KU Leuven and Yale IO prospectus workshops. This is the second chapter of my Ph.D. dissertation at KU Leuven. Financial support from BAEF, the Fulbright program and the Research Foundation Flanders (FWO) is gratefully acknowledged.
1 Introduction

The relationship between competition and innovation has been intensively studied in the economic literature, at least since Schumpeter (1942). While this literature has mainly focused on imperfectly competitive product markets, much less is known about how imperfect factor market competition affects the incentives of buyers to innovate. An increasingly large empirical literature finds evidence for such buyer power to be pervasive across many countries and industries.\(^1\)

This paper therefore examines how such monopsony power affects the incentives of buyers to adopt new production technologies.\(^2\) The focus of the paper lies on factor-biased technologies, as these affect demand on different input markets differently, and on process innovation, which affect the cost side, rather than product demand. The central mechanism of the paper is that a technology that increases the marginal product of an input is more valuable if a firm has high monopsony power over that input. Figure 1 illustrates the intuition behind this mechanism. Suppose a firm produces a homogeneous good that is sold on a competitive market with fixed capacity. It buys two homogeneous inputs \(S\) and \(U\) that enter a constant returns to scale Cobb-Douglas production function and that are sold on monopsonistic markets at prices \(W^S\) and \(W^U\). The inverse supply curve for input \(U\) has the smallest slope, so the firm has more monopsony power over \(S\) than over \(U\). The firm can choose has to choose between two technologies with marginal product curves \(MR_1\) and \(MR_2\) and identical fixed costs. The profits under technologies 1 and 2 correspond to the striped and dotted areas, respectively. Because the firm has higher monopsony power over \(S\), it earns more profits from \(S\) than from \(U\). It hence prefers technology 1, which has a high marginal product of \(S\). If monopsony power over \(U\) would be larger than over \(S\), the firm would prefer to use technology 2.\(^3\)

In an empirical application, I study the mechanization of the Illinois coal mining industry between 1884 and 1902. This is an ideal setting to study the mechanisms proposed by the theoretical model, for three reasons. First, the introduction of the first improved coal cutting machines in the U.S. in 1882, two years before the dataset start provides a major technological shock that was unskill-biased. The mine-level data tracks both the usage of these machines and input and output quantities and prices over a period of 18 years. Secondly, 19th century coal mining towns are a textbook example of monopsonistic labor markets. As miners were paid piece rates, firms could not wage discriminate between workers, as in the classical monopsony model. Thirdly, coal is a nearly homogeneous product, and coal markets were not concentrated. The absence of product market power allows to isolate the effects of monopsony power on innovation.

The central counterfactual question is how different cutting machine usage would have been if the market for coal markets would have been perfectly competitive, rather than monopsonistic. In order

\(^1\)See the literature review by Ashenfelter et al. (2010) and recent papers by, among many others, Naidu et al. (2016); Berger et al. (2019); Rubens (2019); Morlacco (2017); Lamadon et al. (2019); Kroft et al. (2020).
\(^2\)Or oligopsony power, more in general
\(^3\)The model on which figure 1 is based can be found in appendix B.1.
to answer this question, I develop an empirical model that consists of three main elements. First, it features a log-linear skilled labor supply function, of which the slope and intercept vary flexibly across firms and over time. The identification of the labor supply side relies on how miner wages vary with seasonal coal demand shocks during the year. During colder months, coal demand increased, and hence also labor demand. I find that skilled labor wages covaried with these labor demand shocks, while unskilled labor wages did not. My estimates show that skilled miner wages were on average marked down by 12% below their marginal product, while firms did not have monopsony power over unskilled workers.\(^4\) Secondly, demand for both types of workers is derived from a CES production model. In order to identify the labor demand functions, I follow the production function literature by imposing assumptions on the timing of and information set under which input decisions are made by firms. I find that cutting machines were skill-augmenting, and that skilled and unskilled workers were gross complements. Cutting machines were hence unskill-biased: they increased the demand from (unskilled) helpers by more than the demand for (skilled) miners. These estimates are in line with anecdotal historical evidence on the coal industry, and are similar to most other technologies that were invented during the second industrial revolution (Mokyr, 1990; Goldin & Katz, 2009). Thirdly, I estimate a dynamic discrete choice model of coal cutting machine adoption. The model is dynamic

\[^4\]This is logical because the skills of high-skilled workers were only valuable in the coal mining industry. Miner skills, such as building mine roofs, knowing how thick pillars should be in order to avoid collapse, etc., were not easily transferred to other industries. The skills required to perform low-skilled worker tasks, such as tending mules and operating elevators, were similar in other industries, and these workers could therefore switch to similarly paid jobs in other industries.
because the cutting machines were a durable investment. The estimated labor market model delivers an estimate of the equilibrium variable profits when using a cutting machine and when not. This is very helpful to identify the machine adoption model: as the returns to machine adoption are given from the labor market model, only their costs need to be identified, for which I use a rich set of observed technical mine characteristics.\(^5\)

I use the estimated model to compute how cutting machine adoption would have differed when moving from the monopsonistic labor market equilibrium to the competitive equilibrium. I find that the adoption rate of coal cutting machines would have increased by a fourth, from 11.2% to 14.2%, when moving to the competitive equilibrium. As a consequence, average miner productivity would have increased by 1.4% per year. The existence of monopsony power hence led to substantially lower mechanization and productivity growth in the coal industry. The endogeneity of technology adoption also mediates the effects of changes in labor market competition on worker outcomes. If technology adoption would be exogenous, more competitive labor markets would always increase both worker wages and employment. I find, however, that increased labor market competition can lead to lower equilibrium employment: firms react to increased labor market competition by adopting more labor-saving technologies. This also dampens the wage gains to workers when moving to more competitive labor markets. Although the move from monopsonistic to perfectly competitive markets is a theoretical counterfactual, this can inform concrete policies that make labor markets more competitive, such as antitrust policy or merger guidelines that target oligopsonists.

The most frequently used policy measure in the presence of oligopsony power is, however, not competition policy, but minimum wage policies. In a second counterfactual exercise, I therefore simulate the effects of a state-wide minimum wage policy in Illinois on employment outcomes, technology adoption and productivity. I find that minimum wages led to even lower cutting machine adoption rates than in the monopsonistic equilibrium. The reason for this is that one of the benefits of adopting a labor-saving technology is that it reduces equilibrium wages, and hence costs. A minimum wage, however, prevents this drop in equilibrium wages in many cases. Although wages would obviously increase, and employment also depending on the level of the minimum wage, cutting machine usage and productivity would fall in response to a minimum wage.

In reality, product markets are usually not perfectly competitive, and production capacity not fixed. In order to understand the relationship between imperfect competition and innovation, one has to use combine the traditional models on product market competition and innovation with the stylized model of this paper, that isolated the effects of imperfect factor market competition on innovation. The findings of the model are relevant beyond the historical setting of U.S. coal mining. During the second industrial revolution, technologies were mostly unskill-biased, which implies that monopsony power over skilled workers led to slower technology adoption and lower productivity growth. During the last four decades, technological change is believed to have been ‘hollowed out’ the center of

\(^5\)A similar identification approach was used by Peters et al. (2017).
the skill and income distribution (Autor et al., 2006; Goos & Manning, 2007; Goos et al., 2014). Automation incentives therefore fall with the degree of monopsony over these workers at the center of the skill distribution, but rise with monopsony power at the low- and high-end of the skill and income distribution. Knowing both the relative degrees of monopsony power over different types of workers and the direction of technical change is therefore crucial to determine the interaction between technological change and oligopsony power, and for the ex-ante evaluation of minimum wage and competition policies.

This paper makes three main contributions to the literature. First, I build on a large literature that studies the relationship between competition and innovation (Schumpeter, 1942; Aghion, Bloom, Blundell, Griffith, & Howitt, 2005; Collard-Wexler & De Loecker, 2015; Hashmi & Van Biesebroeck, 2016; Igami & Uetake, 2017). In contrast to this literature, I focus on the effect of buyer power rather than seller power, on innovation. Work on this topic is scarce. In a recent theoretical paper, Loertscher and Marx (2020) examine the relationship between countervailing buyer power and investment. Their model relies on a setting with efficient bargaining in which imperfect information is key. In contrast, the mechanism in this paper consists of the combination of classical monopsony power and factor-biased technological change.

Secondly, this paper contributes to the literature on factor-biased technological change. The key difference with the seminal models of directed technical change such as, among others, Autor et al. (2003); Acemoglu (2003) and Antras (2004), is that I allow input prices to be endogenous from the point of view of individual firms. This paper is also related to the ‘induced innovation’ hypothesis of Hicks (1932). Under this hypothesis, higher wages, for instance due to a minimum wage increase, lead to higher adoption of labor-saving technologies, because the cost saving to firms increases with wages. If firms have monopsony power, however, profits do not stem only from costs being low, but also from the wedge between wages and marginal products being high. I show that because of this, increased minimum wages can lead to lower, rather than higher technology adoption, in contrast to Hicks’ hypothesis. This paper is also related to Goolsbee and Syverson (2019), who argue that monopsony power over tenure-track faculty induces universities to substitute these workers for adjunct faculty members. In contrast to their paper, I endogeneize the choice of the production technology, and show that can change the effects of monopsony power on factor ratios to an important

---

6Katz and Margo (2014) argue this also held for technical change during the second industrial revolution.

7Especially the variation of monopsony power along the skill and income distribution is a mostly unanswered empirical question. The labor literature has historically mainly focused on monopsony power over low-skilled workers, such as Card and Krueger (1994). Non-compete clauses are, however, most frequent among high-skilled jobs in the U.S. (Starr et al., 2019), which could imply important monopsony power over these workers.

8In their study of tomato harvesters, Just and Chern (1980) examine how oligopsony power of buyers affects technology adoption of their suppliers, and the same holds for Huang and Sexton (1996); Köhler and Rammer (2012). I focus, in contrast, on technology adoption by the buyers.

9In the existing literature, relative aggregate input prices change due to the general equilibrium effects of factor-biased technical change.

10Dechezleprêtre, Hémous, Olsen, and Zanella (2019) empirically examines this theory using current-day data.
Thirdly, this paper relates to the literature on the welfare effects of market power (De Loecker et al., 2020; Edmond et al., 2018), and of buyer power specifically (Berger et al., 2019; Morlacco, 2017). I contribute by showing that technological change and productivity growth are endogenous to the level of monopsony power. Depending on the direction of technical change and the relative degrees of monopsony power on input markets, monopsony power can be beneficial or detrimental to innovation rates, and hence to aggregate productivity growth. This is an additional channel through which (input) market power shapes aggregate outcomes and, ultimately, welfare. This channel is, moreover, dynamic: current monopsony power affects future productivity growth, and hence also future income and wage growth. A subset of this literature focuses on the productivity consequences of market power through its effects on allocative efficiency (Harberger, 1954; Asker et al., 2019). This paper complements this literature by focusing on the effects of monopsony power on aggregate productivity through endogenous technological change and within-firm productivity changes, rather than through input reallocation.

The remainder of this paper is structured as follows. I start with discussing key facts on the Illinois coal mining industry in section 2. Next, I build a model of labor and capital demand and supply with monopsonistic producers in section 3, from which I derive theoretical results on the relationship between monopsony power and innovation. In section 4, I apply this model to the Illinois coal setting, and discuss its identification and estimation. Section 5.2 uses the estimated model to analyze how technology adoption, productivity growth, and worker outcomes would change in counterfactual worlds with competitive labor markets and minimum wages.

2 Key facts on the Illinois coal mining industry

2.1 Industry background

I study the mining of bituminous coal in Illinois between 1884 and 1902. The dataset covers all mines, of which the number increased from 688 to 895 during the sample period. The number of firms owning these mines increased from 655 to 824. Although only 12% of mines belonged to a multi-mine firm, they produced 43% of total output.

Coal markets

Coal was sold at the mine gate, and there was no vertical integration with post-sales coal treatment, such as coking. Coal markets were integrated across the state of Illinois: 93% of the mines’ coal sales were transported by train from the mine gate to markets. The remaining 7% of output was sold locally.
near the mine, without being transported by train. The main coal destination markets for Illinois mines were St. Louis and, to a lower extent, Chicago.\footnote{11} Railway firms were also major coal consumers. Coal markets were unconcentrated: the average state-wide coal market share was 0.13\%, and 99\% of the firms had a market share below 1.7\%. Historical evidence points to intense competition on coal markets during the last two decades of the 19th century, before the large consolidation wave in the early 1900s Graebner (1974). It is hence unlikely that coal firms could influence coal prices.

**Extraction process**

The coal extraction process consisted of four consecutive steps. First, the coal seam had to be accessed, which usually required either a vertical ‘shaft’, a diagonal ‘slope’ or a horizontal ‘drift’, depending on geography. As large parts of Illinois are flat, 60\% of the mines were ‘shaft’ mines. Less than 2\% of the mines were surface mines that did not require a tunnel. Second, upon reaching the seam, the wall was ‘undercut’. This was traditionally done by hand, but from the 1880s onwards also using mechanical cutting machines. The mechanization of the coal cutting process was the most significant technological change during this time period Fishback (1992), and hence the main topic of this paper. Third, the coal was blasted using explosives and shovelled into a cart. Finally, coal had to be transported back to the surface and sorted from impurities. More than nine out of ten mines used a ‘rooms and pillars’ technique in which miners excavated everything except pillars, which were left to sustain the roof.\footnote{12} Mines used two types of intermediate inputs. First, black powder was used to blast the coal wall. This was purchased by the miners, not by the firm. Secondly, coal itself was used to power steam engines, electricity generators, and air compressors.\footnote{13}

**Types of workers**

I follow the mine inspector reports by classifying workers into two types: ‘miners’ and ‘helpers’. The actual coal cutting was done by the miners, while helpers covered a variety of tasks: hauling coal to the surface by tending mules or by operating locomotives (‘drivers’), clearing the area around the miners of debris (‘miner laborers’), operating doors and elevators (‘doorboys’ or ‘trappers’), and sorting coal from impurities (‘slatepickers’). In line with the anecdotal evidence in the reports, I consider miners to be skilled workers, and helpers to be unskilled workers. Cutting involved a significant amount of skill: in order to determine the thickness of the pillars, miners had to trade off lower output with the

\footnote{11}Chicago also supplied itself with cheaper coal from fields in Ohio, Pennsylvania, and West Virginia, which was transported by railroad and lake steamers Graebner (1974).
\footnote{12}The other mines used so-called ‘longwall’ techniques in which miners temporarily constructed an artificial roof and allowed the room to collapse in a controlled way.
\footnote{13}A fraction of the mine’s coal output was re-used as an energy input. I only observe reused coal inputs in 1902, and the fraction of output that was re-used as an input was on average 5\%, and 0\% for the median mine. As I do not observe this variable in all years, I do not take it into account in the model.
risk of collapse. Miners also built their own roofs at the seam, and operated explosives. Helper tasks required, in contrast, considerably less skills, which were moreover not specific to coal mining, such as tending mules and opening doors.14

**Labor markets**

Miners received a piece rate per ton of coal mined, while helpers were paid a daily wage. Converting the piece rates to daily wages, miners received a net salary15 that was on average 23% higher than helper wages, although they faced similar risks from working underground. Mining areas were sparsely populated: in the average town in the dataset, the number of coal employees constituted on average a third of the town’s population, which was 3090 on average, and 1067 at the median town. Considering that women and children under the age of 12 could not work in mines, this implies that virtually the entire town was employed in coal mining. Of all the villages, 50% had just one coal firm, and another 30% had two or three firms. Two thirds of all employees worked in a village with three or less coal firm. The different villages were connected by railroads, but these were exclusively used for freight: passenger lines only operated between major towns (Fishback, 1992). As most roads were still unpaved, and automobiles not yet introduced, and that miners had to bring their own supplies to the mine, commuting between villages was not an option. Moving to other villages to change employers usually required moving the entire family, which increased the cost of switching employers across villages. The average village was 7.4 miles apart from the closest other village, which was too far to commute on foot on a daily basis.

**Technological change**

Coal cutting in the U.S. was done manually until the early 1880s, using picks. The first mechanical coal cutter in the U.S.A. was invented by J.W. Harrisson in 1877, but it was merely a prototype.16 The Harrisson patent was acquired and adapted by Chicago industrialist George Whitcomb, whose ‘Improved Harrison Cutting Machine’ was released in 1882, of which the patent is pictured in figure A4. Ninety percent of the cutting machines in the dataset are of this type. The spatial diffusion of cutting machines is shown in appendix figure A1. The graph in figure 2(a) shows that the share of mines using a coal cutting machine increased from below 1% to 13% in 1902.17 Mechanized mines were larger: their share of output increased to 40% in 1902. Another technological change was the mechanization of the hauling process, which replaced mules by underground locomotives. The share

---

14Some unskilled workers eventually became skilled, such as boys who started out as ‘slatepickers’ who sorted coal, but became miners as they aged.
15That is, net of material costs and other work-related expenses.
16Simultaneously, prototypes of mechanical coal cutting machines were invented in Northern England in the late 1870s (Reid, 1876; Ackermann, 1902).
17Only 9 mines used a cutting machine in 1884.
of mines using locomotives increased from 33% in 1884 to 42% in 1896, when their measurement ends. The mines that did not use locomotives were, however, tiny: the share of output mined in locomotive mines increased from 80% in 1884 to 95% in 1896, as shown in panel (b) of figure 2.

Figure 2: Technological change

(a) Cutting machines

(b) Locomotives

Unionization

The unionization of the Illinois coal mining industry started around 1860, but were largely unsuccessful in their attempts to raise wages (Boal, 2017). In 1886, 15% of mine workers in Illinois were member of a trade union. The first successful labor union in Illinois was the United Mine Workers of America, founded in 1890. A major strike occurred in 1897-1898, and resulted both in wage raises and in a reduction of working hours to a maximum of eight per day. Various regulations existed to counter unionism, such as the usage of ‘yellow-dog contracts’ which stipulated non-membership of a union as a condition for employment. These contracts were criminalized in Illinois in 1893, with fines of $100 USD\textsuperscript{18} (Fishback, Holmes, & Allen, 2009). There was no minimum wage law and labor markets were largely unregulated (Naidu & Yuchtman, 2017).

2.2 Key facts

Fact 1 Output and output per miner increased, but miner wages fell.

The ratio of industry output over the total number of miner-days and helper-days is in figure 3(a). Daily output per miner increased from 2.5 to 3.75 between 1884 and 1890, and then fell again to

\textsuperscript{18}This was worth on average six monthly miner wages
2.75 ton per worker-day in 1902. Total output in Illinois increased from 10 to 16 Mton between 1884 and 1902. The industry hence grew at a fast rate and became more productive. This growth did not translate into higher wages for miners: figure 3(b) shows that the daily wage per miner fell from $1.8 to 1.4 between 1884 and 1902. Coal prices per ton first dropped from $1.2 to $1.0 until 1890, and sharply increased after 1898. This surge in coal prices was partly due to pass-through of miner wages, which increased after 1898 due to wage bargaining after a series of strikes, and partly due to an increase in aggregate coal demand across the U.S.A.

**Figure 3: Aggregate quantities and prices**

(a) Quantities

(b) Prices

Notes: Panel (a) reports industry output and industry output over the total number of miner-days. Panel (b) reports the aggregate daily miner wage, defined as the total wage bill over total number of miners in the industry. The coal price per ton is the mine-gate price, and is the ratio of total mine revenue over total output.

**Fact 2** Mechanized mines used less miners relatively to helpers.

Figure 4(a) plots the ratio of the total number of miner-days over the total number of helper-days. The blue dotted line reports this ratio for mines that adopted a cutting machine throughout the time period, the red dashed line plots it for the other mines. The aggregate ratio of miners to helpers was 3.6 for both groups of mines in 1884, and decreased to 2.4 in the mechanized mines, and 2.8 in the other mines. Miners were hence replaced by helpers in both types of mines, but more so in the mechanized mines. This trend was the sharpest during the first six years of the panel: the ratio of miners to helpers fell by more than half for mechanized mines until 1890, before increasing again afterwards.

Panel (b) plots daily output per miner for both groups of mines. Machine mines used less workers to mine a ton of coal, but this was already the case in 1884. Miner productivity increased sharply between 1888 and 1890 for the machine mines, but then fell again. The correlations in panels (a)-(b) are not necessarily causal: Hicks-neutral and factor-specific productivity are likely to change the
demand for machines by mines. The evidence does suggest, however, that the main long-run effect of cutting machines was to replace helpers by miners, rather than to increase output per worker. This is consistent with historical evidence: the 1888 report of the Illinois Coal Mine 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” (Illinois Bureau of Labor Statistics, 1888)

Along the same lines, the State Inspector of Mines of Illinois wrote in his 1898 report:

The advantages derived from machinery […] consist not only in the greater execution of the machine, but in the subdivision of labor which it involves. 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

**Fact 3** *Miner wages vary seasonally, helper wages and coal prices do not.*

Coal demand was seasonal: during the colder winter months, energy demand increased compared to the warmer summer months. Figure 5(a) plots monthly mine output in 1890. As there was a time lag between coal extraction and sales in the final market, total coal extraction was high from August to February, at a monthly average of 11 Kton, and low between March and July, at 8.5 Kton. As a result, demand for both miners and helpers was high during the fall and winter, and low during the spring and summer, as shown in figure 5(b). Miner employment was 40% higher in October than in March, helper employment 60% higher. Panel 5(c) shows that both coal prices and helper wages

---

\[19\] This monthly data is based on a sample of firms selected by the Illinois Bureau of Labor Statistics across 5 counties in 1890, which covers 16% of miner employment and 9% of helper employment.
did not co-vary with product and labor demand throughout the year. Miner wages were, in contrast, higher during winter compared to summers. Miners were paid on average $2.2 per day during between August and February, but only $1.8 per day between March and July. Panel 5(d) also shows this by plotting monthly wages for both types of workers against the monthly number of worker-days at each mine in a sample of firms throughout 1890. Miner wages were positively correlated with monthly employment, while this did not hold for helper wages. Moreover, there was a large variation in miner wages across mines and months, but very little variation in helper wages.

**Figure 5: Seasonality**

(a) Coal extraction  
(b) Employment  
(c) Wages and prices  
(d) Wage-employment profile  

Notes: Panel (a) plots average monthly output throughout the year, panel (b) does the same for miner and helper employment. Panel (c) plots average daily wages per month for both types of workers. Panel (d) plots the correlation between monthly employment and wages across mines. All four panels are based on a sample of firms in 1890.
2.3 Data

I observe every bituminous coal mine in Illinois between 1884 and 1902 at two-year intervals, which results in 8307 observations. The data are obtained from the *Biennial Report of the Inspector of Mines of Illinois*. I observe the mine’s owner, yearly coal extraction, average employee counts for both skilled and unskilled workers, days worked, intermediate inputs (black powder) in quantities, dummies for the usage of various technologies (cutting machines, locomotives, ventilators, longwall machines) and technical characteristics such as mine depth, vein thickness and the mine entrance type (shaft, drift, slope, surface). I observe the average piece rate for miners throughout the year and the daily wage for helpers from 1888-1896. For some years I observe additional variables such as mine capacities, the value of the total capital stock and a break-up of coal sales by destination. Wages and employee skill types are not observed in 1896. I deflate all monetary variables using historical CPI estimates from Hoover (1960). The reported monetary values are all in 1884 U.S. dollars, except when indicated otherwise.

Miner wages and employment are separately reported for the summer and winter months between 1884 and 1890. In 1890, the inspection reports contains monthly data on wages and employment for both types of workers, and of production quantities are given for a sample of 11 firms that covers 15% of miners and 9% of helpers. Monthly free-of-board bituminous coal prices in the harbor of New York are collected for the years 1890-1900 from the NBER Macrohistory Database (National Bureau of Economic Research, n.d.).

I supplement the plant-level dataset with town- and county-level information from the 1880 and 1900 population census and the censuses of agriculture and manufacturing. I refer to appendix A for more details regarding the data sources and cleaning procedures.

3 Model: technology adoption and monopsony power

I set up and estimate a model of labor and capital supply and demand. The labor supply and demand models are static, while the capital demand model is dynamic.

3.1 Environment

Production function

Firms $f$ extract $Q_{f,t}$ tons of coal using a constant elasticity of substitution production function, given by equation (1a).\textsuperscript{20} The firm uses two type of workers $\tau \in \{s, u\}$: skilled and unskilled workers, \textsuperscript{20}A more parsimonious model that relies on a Cobb-Douglas production function is in appendix B.1.
which corresponds to miners and helpers in the coal application. The quantity of each type used, measured in working days, is denoted $L_{f,t}^s$ and $L_{f,t}^u$, and they are paid a daily wage $W_{f,t}^s$ and $W_{f,t}^u$. Both labor types are substitutable at an elasticity $\sigma$. Firms differ in terms of their factor-augmenting productivity levels $\omega_{f,t}^s$. Returns to scale are assumed to be constant, which is supported by the fact that labor productivity does not vary with coal demand shocks throughout the year.\footnote{This evidence is presented in appendix C.1.} In the baseline model, I abstract from intermediate inputs.\footnote{Coal firms did use intermediate inputs, such as coal and black powder. In appendix C.3, I extend the model to allow for intermediate inputs that come in fixed proportions of both labor types.}

$$Q_{f,t} = \left( (\omega_{f,t}^s L_{f,t}^s)^{\frac{\sigma-1}{\sigma}} + (\omega_{f,t}^u L_{f,t}^u)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \text{ s.t. } Q_{f,t} \leq Q_{f}^*$$ (1a)

\textbf{Assumption 1} \textit{The production capacity $Q_{f}^*$ is exogenous to each firm.}

Firms face an annual production capacity $Q_{f}^*$, which is specific to each firm. In the baseline model, I assume that firms cannot change this capacity.\footnote{In appendix C.4, I find that coal output did not change in response to cutting machine adoption, while miner employment fell, which is an empirical validation of this assumption. In appendix B.4, I extend the model to allow for endogenous capacity.} The usage of cutting machines is indicated by a dummy $K_{f,t} \in \{0, 1\}$. Cutting machines change the productivity of miners relatively to helpers, as parameterized by equation (1b). Relative miner productivity increases with cutting machines at a rate $\exp(\beta^k)$ with time at a rate $\exp(\beta^t)$. The intercept is $\beta^0$.

$$\frac{\omega_{f,t}^s}{\omega_{f,t}^u} = \exp(\beta^k K_{f,t} + \beta^t t + \beta^0 + \eta_{f,t})$$ (1b)

The variation in the miner to helper productivity ratio that is not explained by cutting machine adoption and time is denoted as the residual $\eta_{f,t}$. I assume that this residual evolves as an AR(1) process, as parametrized by equation (1c).\footnote{This productivity transition rules out a number of dynamic cost mechanisms that are suggested in the natural resource extraction literature. I refer to appendix C.2 for a discussion of potential cost dynamics, and find none of them to be of first-order importance in this industry.}

$$\eta_{f,t} = \rho \eta_{f,t-1} + \varepsilon_{f,t}$$ (1c)

\textbf{Coal markets}

Coal is assumed to be a homogeneous good.\footnote{There is some variation in heat rates between different types of coal, but all mines in the dataset produced bituminous coal, which is therefore assumed to be an undifferentiated good.} Denote the mine-gate coal price as $P_{f,t}$. Variable costs for the coal firm are equal to $W_{f,t}^s L_{f,t}^s + W_{f,t}^u L_{f,t}^u$.\footnote{This evidence is presented in appendix C.1.}
Assumption 2 Coal prices are exogenous to individual coal firms.

In line with the evidence presented in section 2, I assume that coal markets are perfectly competitive: individual firms cannot influence the coal price: Denote the firm’s marginal cost as \( \lambda_{f,t} \equiv \frac{\partial (W^{s}_{f,t} L^{s}_{f,t} + W^{u}_{t} L^{u}_{f,t})}{\partial Q_{f,t}} \), an exogenous variable transport cost as \( \theta_{f,t} \), and the firm’s markup as \( \mu \equiv \frac{P_{t}}{\lambda_{f,t}} \).

Transport costs are not paid for by the coal firms. There is one market-clearing price per year \( P^*_t \), at which aggregate coal supply and demand for coal in Illinois are equal to each other. The mine-gate price at each mine is equal to the price in the destination market, minus an exogenous variable transport cost \( \theta_{f,t} \).

\[
P^*_t = P_{f,t} + \theta_{f,t}
\]

Total coal demand is assumed to be perfectly inelastic on the short run. Ranking all coal firms in Illinois in a given year by marginal and transport cost \( (\lambda_{f,t} + \theta_{f,t}) \), there is one marginal firm with the highest possible cost whose output balances aggregate demand and supply. Denote the marginal and transport cost of this firm, which can change across the years, as \( \lambda^*_t \) and \( \theta^*_t \). The market clearing price \( P^*_t \) is equal to the marginal and transport cost of this marginal producer \( P^*_t = \lambda^*_t + \theta^*_t \). All firms that operate at a lower cost \( \lambda_{f,t} + \theta_{f,t} < \lambda^*_t + \theta^*_t \) operate at a markup \( \mu_{f,t} \) that is larger than one. This also implies that firms operate at full capacity and that both current and future coal output is exogenous to each firm. As total coal demand fluctuates from year to year, the total number of firms does as well. As coal reserves in Illinois were much larger than actual demand, there are sufficient potential entrants to cope with higher aggregate coal demand. A firm exits if its marginal costs are higher than the marginal costs of the marginal producer, meaning that the markup \( \mu_{f,t} < 1 \), and enters if its marginal costs are below those of this marginal producer, meaning that \( \mu_{f,t} > 1 \). I assume there are no sunk entry and exit costs, so market structure adjusts flexibly to demand fluctuations over time.

Labor supply

The inverse supply curve for both types of workers is given by equation (2). Firms cannot wage discriminate between different workers of the same type in their firms. Each firm faces a supply curve with a constant wage elasticity \( (\psi_{f,t}^{\tau} - 1) \) for workers of type \( \tau \). The log-linearity assumption on the labor supply curve is a strong functional form assumption, but the wage elasticity \( \psi_{f,t}^{\tau} =

26 In appendix C.6, I show that transport costs are correlated with the distance between the mine and St. Louis, as expected.  
27 Markups can be estimated using the labor demand and supply model outlined in the next section, but I do not use these in the main analysis. Markup estimation and a discussion of the results is included in appendix C.5.  
28 This assumption is relaxed in appendix B.4.  
29 This is consistent with high entry and exit rates, and with the existence of small drift mines with low fixed costs and high marginal costs.  
30 In some years, the inspector reports contain mine-employee-level wages, which show that miner and helper wage rates were indeed uniform within each mine.
\[ \frac{\partial W^\tau_{f,t}}{\partial L^\tau_{f,t}} W^\tau_{f,t} + 1 \] is allowed to vary fully flexibly across firms and over time. The term \( \alpha^\tau_{f,t} \) captures any non-wage firm characteristics that shifts type-\( \tau \) labor supply, such as working conditions or the firm’s location. If the labor supply curve is flat, meaning that \( \psi^\tau_{f,t} = 1 \), workers receive a wage \( W^*_t \) that is the same for all firms in each year. This can be thought of as the wage rate that workers can receive when working in an industry other than coal mining, and it is the same for both skilled and unskilled workers. Miner skills are hence assumed to be specific to the coal mining industry.

\[ W^\tau_{f,t} = (\alpha^\tau_{f,t} L^\tau_{f,t}) \psi^\tau_{f,t} W^*_t \] (2)

**Assumption 3**  *Labor markets are isolated for both types of labor.*

I assume that firms are monopsonists that operate in isolated local labor markets, in line with the descriptive evidence from section 2. Labor and capital decisions of other firms have no effect on the own equilibrium wages.

**Capital supply**

The adoption of cutting machines at time \( t \) is denoted \( A_{f,t} \in \{0, 1\} \). Once firms adopt a cutting machine, they become a mechanized mine forever, as shown in equation (3).\(^{31}\)

\[ K_{f,t} = \max\{A_{f,t-1}, K_{f,t-1}\} \] (3)

Adopting a cutting machine at time \( t \) requires the firm to pay an upfront sunk cost \( W^K_{f,t} \) at time \( t \). I assume that coal firms take sunk costs as given. This cost has three components. First, a part of sunk costs depends on observable characteristics \( z_{f,t} \), such as mine depth or vein thickness, as these alter installation costs. A second component of sunk costs, \( \chi_{f,t} \) is potentially serially correlated. Thirdly, random sunk cost shocks \( \kappa_{f,t} \) arrive in each period. Variable and fixed capital costs are, finally, set to zero, because only the extensive margin is modeled and because capital adoption is a terminal state. Using \( \gamma \) to parametrize the effects of \( z \), sunk costs are given by:

\[ W^K_{f,t} = \gamma z_{f,t} + \chi_{f,t} + \kappa_{f,t} \]

\(^{31}\)There were 77 instances in which machines were adopted and then scrapped, but only at very small units, which represent merely than 0.17% of industry output. The terminal state assumption does not rule out depreciation of the capital stock: firms are assumed to always modernize/replace their machines when depreciated at zero cost.
3.2 Firm behavior

Profit maximization

The decision timeline is in figure 6. Just before time \( t \), the productivity shock \( \varepsilon_{f,t} \) arrives. After observing this shock, firms choose both labor inputs \( L_{f,t} \) at time \( t \), which are flexibly adjustable. Equally at time \( t \), firms make the capital adoption decision \( A_{f,t} \), which determines the capital stock in the next period, \( K_{f,t+1} \). The capital stock is hence exogenous when firms choose their labor inputs. After these decisions have been made, a new shock \( \varepsilon_{f,t+1} \) arrives, and both the productivity and capital stock update following their transition equations, (1c) and (3). Let variable profits be denoted as \( \pi_{f,t} \equiv P_{f,t} Q_{f,t} - W_{s,t} L_{s,t}^f - W_{u,t} L_{u,t}^f \), and total profits as \( \Pi_{f,t} \equiv \pi_{f,t} - A_{f,t} W_{k,t}^f \).

Figure 6: Information and decisions timeline

I assume that firms are monopsonists: they are the only firm in their local labor market.\(^{32}\) I assume firms choose both labor inputs in order to maximize current variable profits, as stated in equation 4a. Because of assumptions 3 and 4, the labor demand problem is single-agent.

\[
\max_{L_{s,t}^f, L_{u,t}^f} \pi_{f,t} \tag{4a}
\]

Every time period, firms make the capital adoption decision \( A_{f,t} \in \{0,1\} \) that maximizes their expected discounted profit stream, with discount factor \( \delta \).

\[
\max_{A_{f,t}} \mathbb{E}_t \sum_{r=1}^{\infty} (\delta^r \Pi_{f,r}) \tag{4b}
\]

\(^{32}\)In reality, only half the towns have just one firm, and these produce a third of total output. I justify the monopsony assumption in section 4 by making the additional assumption that firms maximize joint profits within the same labor market.
Counterfactual firm behavior

In real life, firms are assumed to maximize profits. In the counterfactual world of interest, variable profits are replaced by a counterfactual value function $\varphi$ with profit weights $\zeta_{f,t}$ on factor prices.

$$\varphi_{f,t} \equiv P_{f,t} - W_{s,f,t}^s \zeta_{s,t}^s L_{s,f,t}^s - W_{u,f,t}^u \zeta_{u,t}^u L_{u,f,t}^u \quad \text{with } \zeta_{f,t}^\tau \in \left[\frac{1}{\psi_{f,t}^\tau}, 1\right]$$

In the counterfactual, firms choose labor inputs to maximize this value function, rather than profits: they hence solve problem (5) rather than problem (4a). The weights $\zeta^\tau$ can be interpreted as the actual amount of monopsony power exerted by the firm over input $\tau$. If firms maximize profits, then $\zeta_{f,t}^\tau = 1$. As will be shown in section 3.3, this corresponds to a ratio of the marginal product of an input over its price of $\psi_{f,t}^\tau$, which is the maximal amount of monopsony power the firm can exert. In the competitive equilibrium, firms do not exert any of their monopsony power, meaning that $\zeta_{f,t}^\tau = \psi_{f,t}^\tau$.

$$\max_{L_{s,f,t},L_{u,f,t}} \varphi_{f,t} \quad (5)$$

The capital market was assumed to be perfectly competitive. I therefore assume that in the competitive markets counterfactual, firms still choose the capital adoption that maximizes the discounted profit flow as in equation (4b). These profits will be different, however, from the monopsonistic case.

3.3 Monopsonistic equilibrium

Labor demand

I start by deriving labor demand if firms maximize profits, which assumed to be the real world. Deriving the first order conditions of the labor demand problem, equation (4a), implies that the supply elasticity $\psi_{f,t}^\tau$ is equal to the ratio of the marginal revenue of an input over its price:

$$P_{f,t} \frac{\partial Q_{f,t}}{\partial L_{f,t}^\tau} = \frac{W_{f,t}^\tau}{\psi_{f,t}^\tau}$$

Marginal revenue Marginal cost

Working this out gives the demand functions for type-$\tau$ in equation (6).

$$L_{f,t}^\tau = Q_{f,t} \left(\omega_{f,t}^{\tau}(K_{f,t})\right)^{\sigma-1}(W_{f,t}^\tau \psi_{f,t}^\tau)^{-\sigma} \quad (6)$$

The supply elasticities $\psi_{f,t}^\tau$ negatively affect input demand: firms internalize the fact that higher employment leads to higher wages. These elasticities can hence be interpreted as a markdown: they are
equal to the wedge between the marginal product of a worker and his wage: \( \frac{\partial (Q^\tau_{f,t}, P_{f,t})}{\partial L^\tau_{f,t}} = \psi^\tau_{f,t} W^\tau_{f,t}. \)

Now consider the effects of an increase in \( \omega^\tau_{f,t} \), i.e. a skill-augmenting productivity shock. If both worker types are gross complements, meaning that \( \sigma < 1 \), then this shock is unskill-biased: it reduces relative demand for skilled compared to unskilled labor. If they are gross substitutes, meaning that \( \sigma > 1 \), the opposite holds (Acemoglu, 2003).

**Labor market equilibrium**

The labor market equilibrium consists of the labor quantities and prices and machine adoption for which supply and demand are equal to each other, which can be found by solving the system of equations (2) and (6). I denote the equilibrium values of any variable \( X \in \{ L^s, L^u, W^s, \pi \} \) as \( \hat{X} \). The reduced-form expressions for the labor quantities and wages are in equation (7a). This corresponds to point \( A \) in figure 7(a). The equilibrium labor quantities and prices depend on machine usage \( K_{f,t} \), because the productivity levels \( \omega^\tau_{f,t} \) depend on machine usage.

\[
\begin{align*}
\hat{W}^\tau_{f,t}(K_{f,t}) &= [Q_{f,t}P^\sigma_{f,t}(\omega^\tau_{f,t}(K_{f,t}))^{\sigma - 1}(\psi^\tau_{f,t})^{-\sigma}]^{ \frac{\psi^\tau_{f,t} - 1}{1 + \sigma(\psi^\tau_{f,t} - 1)} } (\alpha^\tau_{f,t})^{\frac{\psi^\tau_{f,t} - 1}{1 + \sigma(\psi^\tau_{f,t} - 1)}} (W^*_t)^{\frac{1}{1 + \sigma(\psi^\tau_{f,t} - 1)}} \\
\hat{L}^\tau_{f,t}(K_{f,t}) &= [Q_{f,t}P^\sigma_{f,t}(\omega^\tau_{f,t}(K_{f,t}))^{\sigma - 1}(\psi^\tau_{f,t})^{-\sigma}]^{ \frac{1 - \psi^\tau_{f,t}}{1 + \sigma(\psi^\tau_{f,t} - 1)} } (\alpha^\tau_{f,t})^{\frac{(1 - \psi^\tau_{f,t})\sigma}{1 + \sigma(\psi^\tau_{f,t} - 1)}} (W^*_t)^{\frac{1 - \sigma}{1 + \sigma(\psi^\tau_{f,t} - 1)}}
\end{align*}
\]  \( \text{(7a)} \)

Inserting the equilibrium wages and labor quantities into the variable profit function yields equilibrium variable profits \( \hat{\pi}_{f,t} \):

\[
\hat{\pi}_{f,t}(K_{f,t}) = Q_{f,t}P_{f,t} - \sum_{\tau} \left[ [Q_{f,t}P^\sigma_{f,t}(\omega^\tau_{f,t}(K_{f,t}))^{\sigma - 1}(\psi^\tau_{f,t})^{-\sigma}]^{ \frac{\psi^\tau_{f,t} - 1}{1 + \sigma(\psi^\tau_{f,t} - 1)} } (\alpha^\tau_{f,t})^{\frac{\psi^\tau_{f,t} - 1}{1 + \sigma(\psi^\tau_{f,t} - 1)}} (W^*_t)^{\frac{1 - \sigma}{1 + \sigma(\psi^\tau_{f,t} - 1)}} \right]^\tau
\]  \( \text{(7b)} \)

Variable profits depend on the factor-augmenting productivity levels \( \omega^\tau_{f,t} \), which depend on cutting machine usage \( K_{f,t} \in \{0, 1\} \). Denote the conditional variable profits when using cutting machines and when not as \( \hat{\pi}^1_{f,t} \equiv \hat{\pi}_{f,t}(K_{f,t} = 1) \) and \( \hat{\pi}^0_{f,t} \equiv \hat{\pi}_{f,t}(K_{f,t} = 0) \).

**Capital demand**

If mine owners decide on adoption in time \( t \), they get a cutting machine at time \( t + 1 \): capital is fixed at the short run. The variable profit gain from using a cutting machine, \( \hat{\pi}^1_{f,t} - \hat{\pi}^0_{f,t} \), can be computed using equation (7b) and the estimated coefficients from the labor supply and demand model. Machine usage changes \( \tau \)-augmenting productivity following equation (1b), and the level of factor-augmenting productivity affects variable profits. If machines increase the productivity levels of each input, machines lead to lower variable costs, and hence, increased variable profits, so \( \hat{\pi}^1_{f,t} - \hat{\pi}^0_{f,t} > 0 \).
Normalizing per-period profits when not using a cutting machine to zero, the discounted profit stream of using a machine net of the transient sunk cost $\kappa_{f,t}$ is given by $\hat{v}_{f,t}^1$. If mine owners decide on adoption in time $t$, they get a cutting machine at time $(t+1)$: capital is fixed at the short run. Adopting a machine in year $t$ only leads to the variable profits associated with using machines in year $(t+1)$, hence the expected variable profit stream $(\hat{\pi}_{f,t}^1 - \hat{\pi}_{f,t}^0)$ is multiplied with the discount factor $\delta$.

$$\hat{v}_{f,t}^1 = \delta (\hat{\pi}_{f,t}^1 - \hat{\pi}_{f,t}^0) - \gamma z_{f,t} - \chi_{f,t}$$ (8)

The value of not adopting a cutting machine is given by the per-period profit of not using a machine, which was normalized to zero, and the expected value of adopting a cutting machine in the future, $\mathbb{E}_t V_{f,t+1}$:

$$\hat{v}_{f,t}^0 = \delta \mathbb{E}_t \hat{V}_{f,t+1}$$ (9)

Assuming the sunk cost shock $\kappa$ is logistically distributed, the ex-ante value function $\hat{V}_{f,t+1}$ has the usual log-sum form:

$$\hat{V}_{f,t+1} = 0.577 + \ln \left( \exp(\hat{v}_{f,t+1}^1) + \exp(\hat{v}_{f,t+1}^0) \right)$$

Using Scott (2015), I decompose the ex-ante value function $\mathbb{E}_t \hat{V}_{f,t+1}$ into the realized ex-post value function $\hat{V}_{f,t+1}$ and prediction error $e_{f,t} \equiv \hat{V}_{f,t+1} - \mathbb{E}_t \hat{V}_{f,t+1}$

$$\hat{v}_{f,t}^0 = \delta (\hat{V}_{f,t+1} - e_{f,t})$$

The optimization problem in equation (4b) can now be rewritten as:

$$\max_{A_{f,t}} \mathbb{E}_t \{ \hat{v}_{f,t}^1, \hat{v}_{f,t}^0 \}$$

The adoption of the technology $K$ therefore boils down to a trade-off between the variable cost reduction and the sunk costs associated with using the technology. Due to assumption 3, the capital demand problem is single-agent. I assume there are no serially correlated latent sunk costs: $\chi_{f,t} = 0$. I assume the sunk shocks $\kappa_{f,t}^1$ and $\kappa_{f,t}^0$ are type-I extreme value distributed. The probability of adopting the technology in year $t$, $\hat{S}_{f,t}^1$, is then equal to:

$$\hat{S}_{f,t}^1 = \frac{\exp(\hat{v}_{f,t}^1)}{\exp(\hat{v}_{f,t}^0) + \exp(\hat{v}_{f,t}^1)}$$ (10)

Given that sunk costs are exogenous, the capital supply curve is horizontal. The capital market equilibrium is hence given by equation (10). There are a number of assumptions about information and
learning in the capital demand model. First, it is assumed that there is no private information about machine benefits: everyone knows in advance what the effect of cutting machines on productivity is, and everyone draws sunk costs from the same distribution. Secondly, machines have the same effects on productivity at all mines, which is implicit from the fact that $\beta^k$ is homogeneous across firms and time. This also implies that firms are able to perfectly use machines from the start: there is no learning about how to use cutting machines.

**Labor and capital market equilibrium**

The equilibrium capital usage rate $\hat{K}_{f,t}$ can be calculated using the transition rule for capital, equation (3), the equilibrium adoption probabilities $\hat{S}_{f,t}$, and an initial value for the capital stock in the first year. The equilibrium values of any capital-dependent variable $\hat{X}_{f,t}(K_{f,t}) \in \{L^s_{f,t}(K_{f,t}), L^n_{f,t}(K_{f,t}), W^s_{f,t}(K_{f,t}), \pi_{f,t}(K_{f,t})\}$ can then be found by weighting its values with the probabilities of machine usage:

$$\hat{X}_{f,t} = \hat{K}_{f,t} \hat{X}_{f,t}(K_{f,t} = 1) + (1 - \hat{K}_{f,t}) \hat{X}_{f,t}(K_{f,t} = 0)$$

### 3.4 Counterfactual equilibria

**Competitive equilibrium**

The counterfactual labor and demand curves can be derived analogously from the optimization problem in equation (5). The wedge between marginal revenue and input prices is now equal to $\psi_{f,t} \zeta_{f,t}$. The counterfactual of interest corresponds to the competitive equilibrium, in which case $\zeta_{f,t} = \frac{1}{\psi_{f,t}}$. This implies that workers are paid their marginal revenue:

$$\frac{\partial Q_{f,t}}{\partial L_{f,t}} = \frac{W^*_t \psi_{f,t} \zeta_{f,t}}{\psi_{f,t} - 1}$$

Denoting the values of a variable $x$ in the competitive equilibrium as $\tilde{x}$, the competitive equilibrium is given by equation (11). This corresponds to the intersection of the marginal revenue and labor supply curve, point $B$ in figure 7.

$$\begin{align*}
\tilde{W}_{f,t}(K_{f,t}) &= [Q_{f,t}P_{f,t}^c(\omega_{f,t}(K_{f,t}))^{\sigma-1}]^{\psi_{f,t} - 1} (\alpha_{f,t}^{\psi_{f,t} - 1} W^*_t)^{\psi_{f,t} - 1} (W^*_t)^{1 + \sigma(\psi_{f,t} - 1)} \\
\tilde{L}_{f,t}(K_{f,t}) &= [Q_{f,t}P_{f,t}^c(\omega_{f,t}(K_{f,t}))^{\sigma-1}]^{\psi_{f,t} - 1} (\alpha_{f,t}^{\psi_{f,t} - 1} W^*_t)^{1 + \sigma(\psi_{f,t} - 1)} (W^*_t)^{-\frac{\sigma}{1 + \sigma(\psi_{f,t} - 1)}}
\end{align*}$$

The equilibrium variable profits in the competitive counterfactual are given by equation (12). Similarly to before, the conditional variable profits when using cutting machines and when not are denoted
as $\tilde{\pi}^1_{f,t}$ and $\tilde{\pi}^0_{f,t}$.

$$\tilde{\pi}_{f,t}(K_{f,t}) = Q_{f,t}P_{f,t} - \sum_{\tau} \left( [Q_{f,t}P_{f,t}^{\omega_{f,t}^\tau}]^{\sigma - 1} \frac{\psi_{f,t}^\tau}{1 + \sigma(\psi_{f,t}^{\tau-1})} \left( \frac{\psi_{f,t}^{\tau-1}(1-\sigma)}{1 + \sigma(\psi_{f,t}^{\tau-1})} \right) \right) \tag{12}$$

Defining the value functions $\tilde{v}^0_{f,t}$ and $\tilde{v}^1_{f,t}$ similarly to equation (9)-(8), with profits $\tilde{\pi}$ rather than $\hat{\pi}$, the machine adoption probability in the competitive equilibrium is equal to:

$$\tilde{S}_{f,t}^1 = \frac{\exp(\tilde{v}^1_{f,t})}{\exp(\tilde{v}^0_{f,t}) + \exp(\tilde{v}^1_{f,t})} \tag{13}$$

Similarly to before, equilibrium on both the labor and capital markets implies the following equilibrium values for employment, wages, and profits:

$$\tilde{X}_{f,t} = \tilde{K}_{f,t}\tilde{X}_{f,t}(K_{f,t} = 1) + (1 - \tilde{K}_{f,t})\tilde{X}_{f,t}(K_{f,t} = 0)$$

**Figure 7: Counterfactuals: theory**

(a) Equilibria

(b) Minimum wage

**Assumptions**

In order for this counterfactual exercise to be valid, the following variables and coefficients need to be invariant between the monopsonistic and competitive equilibrium: $(\omega^s, 1, \omega^s, 0, \omega^u, 1, \omega^u, 0, W^u, z, \alpha^s, \alpha^u, \beta, \gamma, \delta, \sigma)$. This implies that moving to competitive labor markets does not change the
effects of cutting machines on skilled and unskilled labor productivity, the sunk cost determinants, working conditions, the discount factor, and the elasticity of substitution between skilled and unskilled workers.

**Minimum wage counterfactual**

Minimum wages are the usual policy measure in response to monopsony power on labor markets. In order to assess the effects of a state-wide minimum wage $W_{t}^{\text{min}}$, it is crucial to know whether the minimum wage binds or not, and whether it lies above or below the competitive wage equilibrium. The variable $I_{k,f,t}^{k}$ indicates this relative position of the minimum wage compared to the equilibrium miner wage at firm $f$, with the superscript $k \in \{0, 1\}$ indicating the usage of cutting machines. The minimum wage does not bind for firm $f$ if the monopsonistic equilibrium wage $\hat{W}_{s,k}^{f,t}$ lies above the minimum wage, in which case $I_{k,f,t}^{k} = 1$. If the minimum wage lies above the monopsonistic equilibrium wage, but below the competitive wage, $I_{k,f,t}^{k} = 2$. Finally, $I_{k,f,t}^{k}$ takes the value of three if the minimum wage lies above the competitive equilibrium wage.

$$
\begin{align*}
I_{k,f,t}^{k} = 1 & \iff \hat{W}_{s,k}^{f,t} > W_{t}^{\text{min}} \\
I_{k,f,t}^{k} = 2 & \iff \hat{W}_{s,k}^{f,t} \leq W_{t}^{\text{min}} \leq \tilde{W}_{s,k}^{f,t} \quad \forall k \in \{0, 1\} \\
I_{k,f,t}^{k} = 3 & \iff W_{t}^{\text{min}} > \tilde{W}_{s,k}^{f,t}
\end{align*}
$$

The A-B-C line in figure 7(a) plots the evolution of counterfactual employment under different levels of the minimum wage. As long as the minimum wage is non-binding ($I = 1$), employment is at the monopsonistic equilibrium $A$. As soon as the minimum wage becomes binding, but is below the competitive equilibrium, the equilibrium shifts from points $A$ to $B$, along the labor supply curve, as both the equilibrium wage and employment rise. As soon as the minimum wage exceeds the competitive wage, counterfactual employment falls again, along the labor demand curve, from point $B$ towards $C$. Denote the equilibrium values of any variable $x$ in the minimum wage counterfactual as $\bar{x}$. In equation form, the counterfactual miner employment levels under a minimum wage are given by:

$$
\begin{align*}
\bar{L}_{s,k}^{f,t} & = \hat{L}_{s,k}^{f,t} \quad \iff I_{k,f,t}^{k} = 1 \\
\bar{L}_{s,k}^{f,t} & = \frac{W_{t}^{\text{min}} \psi_{s,f,t}^{f,t} - 1}{\alpha_{s,f,t}} \quad \iff I_{k,f,t}^{k} = 2 \\
\bar{L}_{s,k}^{f,t} & = Q_{f,t}(\omega_{s,k}^{f,t})^{-1}(W_{t}^{\text{min}} \psi_{s,f,t}^{f,t})^{-\sigma} \quad \iff I_{k,f,t}^{k} = 3
\end{align*}
$$

The counterfactual variable profits under the minimum wage, $\bar{\pi}_{1,f,t}$ and $\bar{\pi}_{0,f,t}$, can now be solved for as well. Figure 7(b) shows how a minimum wage alters the variable profit effects of machine adoption.
Suppose cutting machines reduce the marginal product of miners from curve $MR_1$ to $MR_2$, and that the minimum wage is initially not binding, meaning that $T^0 = 1$. When there is no minimum wage, variable profits become the northwest-striped area, as the equilibrium shifts from $A_1$ to $A_2$. With a binding minimum wage, however, the wage level cannot drop below $W_{min}$. The variable profit after machine adoption under the minimum wage counterfactual, $\bar{\pi}_1$, corresponds to the northeast-striped area. This area is smaller than the northwest-striped area: without a minimum wage, wages would drop after machine adoption, which would be a cost saving, but a minimum wage prevents this. More formally, the counterfactual machine usage rate $\bar{S}_{1,t}$ can be computed by plugging the counterfactual variable profits into equation (10), as was done previously in section 5.2:

$$\bar{S}_{1,t} = \frac{\exp(\bar{v}_{1,t}^1)}{\exp(\bar{v}_{0,t}^1) + \exp(\bar{v}_{1,t}^1)}$$

Similarly to the first counterfactual exercise, the equilibrium values for $\bar{x}_{f,t}$ can be calculated by weighting their values with and without capital usage by the capital usage probabilities:

$$\bar{x}_{f,t} = \bar{K}_{f,t}\bar{x}_{1,t}^1 + (1 - \bar{K}_{f,t})\bar{x}_{0,t}^0$$

### 3.5 Monopsony power and the returns to capital

In section 5.2, I will estimate the model in sections 3.1-3.3, and compute the optimal cutting machine adoption rates in reality and in the competitive labor market counterfactual. The direction of this effect can, however, be derived theoretically without knowing the full model parametrization. The relationship of interest is the effect of the ‘conduct’ parameter $\zeta_{f,t}^r \in [\frac{1}{\psi_{f,t}}, 1]$ on the return to mechanization:

$$\frac{\partial}{\partial \zeta_{f,t}^r} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}} \right)$$

The change in variable profits caused by machine usage depends on how machines change factor-augmenting productivity, and on how factor-augmenting productivity affects variable profits. Using the chain rule results in equation (14a).

$$\frac{\partial \pi_{f,t}}{\partial K_{f,t}} = \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^s} \frac{\partial \omega_{f,t}^s}{\partial K_{f,t}} + \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^w} \frac{\partial \omega_{f,t}^w}{\partial K_{f,t}}$$

(14a)

The effects of an increase in $\tau$-augmenting productivity on variable profits is given by equation (14b). An increase in $\tau$-augmenting productivity increases variable profits if inputs are gross complements,
meaning that $\sigma < 1$, because it then reduces the equilibrium amount of input $\tau$ used.

$$
\frac{\partial \pi_{f,t}}{\partial \omega_{f,t}} = \frac{(1 - \sigma)\psi_{f,t}}{1 + \sigma(\psi_{f,t} - 1)} [Q_{f,t} P^\sigma_{f,t} (\bar{\psi}_{f,t})^{\sigma-1} (\omega_{f,t})^{1+\sigma(\psi_{f,t} - 1)^{-1}} - \sigma(\psi_{f,t} - 1)^{-1} (\alpha_{f,t} (\psi_{f,t} - 1)^{-1} W_t^*)]^{1-\sigma}
$$

(14b)

Taking the derivative of the profit effect of productivity shocks with respect to the degree of monopsony power leads to lemma 1:

**Lemma 1** Under the assumptions (1)-(4): If firms exert more monopsony power over type-$\tau$ workers, a change in $\tau$-augmenting productivity changes variable profits by less:

$$
\frac{\partial}{\partial \zeta_{f,t}} (|\frac{\partial \pi_{f,t}}{\partial \omega_{f,t}}|) < 0
$$

**Proof:** cfr. appendix B.2.

The intuition behind lemma 1 follows from the graphs in figure 1. More monopsony power over an input implies a larger wedge between its marginal product and its price. If inputs are gross complements, a $\tau$-augmenting productivity shock reduces demand for input $\tau$. This is less profitable if monopsony power over that input is higher, because the profit extracted from this wedge decreases. If inputs are gross substitutes, a $\tau$-augmenting shock increases demand for that input, which decreases profits (by increasing costs), but less so if firms extract large profits from input $\tau$ because of monopsony power. The effect of monopsony power on the returns to technology adoption can now be found by combining equations (14a), (14b) and (20), which leads to theorem 1.

**Theorem 1** Under the assumptions (1)-(4): The variable profit gain from adopting a technology that decreases (increases) the absolute demand for an input decreases (increases) with the degree of monopsony power over that input:

$$
\frac{\partial}{\partial \zeta_{f,t}} (\frac{\partial \pi_{f,t}}{\partial K_{f,t}}) \begin{cases} < & 0 \iff (\sigma - 1) \frac{\partial \omega_{f,t}}{K_{f,t}} \begin{cases} < & 0 \end{cases} \end{cases}
$$

**Proof:** Cfr. appendix B.3.

A technology can decrease relative demand for input $\tau$ in two cases: either if it is $\tau$-augmenting and inputs are gross complements, or if it augments the other input and they are gross substitutes. Given that output is exogenous, a relative change in demand for $\tau$ is also an absolute change in demand for that input. Monopsony power over input $\tau$ reduces the variable profit gain from adopting a technology if it saves on $\tau$. The intuition is the same as with lemma 1: saving on an input is less profitable if the wedge between that input’s price and marginal product is higher. If the level of monopsony power over each variable input is unrelated to the size of the capital cost, as is assumed, then theorem 1
implies that the adoption of a technology is reduced by monopsony power over the input on which it saves, but increased by monopsony power over the input of which demand increases.

4 Identification and estimation

I now turn to the identification and estimation of the model, which consists of equations (2), (6) and (10): labor supply, labor demand, and capital demand. The dataset comes at the mine-year level, but 42% of output is produced by firms that operate multiple mines. The absence of firm identifiers makes it hard, however, to track firms over time, which is necessary for the identification of both the labor and capital demand models. I therefore use villages as the firm level \( f \). I aggregate all variables up to the village-year level using the procedure that is described in appendix A.3. The median village contains just two firms, and the average village 2.8 firms. I assume that firms collude perfectly on their input market, meaning that firms within the same village maximize their joint profits. The collection of firms in a village is hence modelled to be a single firm. As was motivated in the background section, villages can be thought of as isolated labor markets. The firms in a village are hence assumed to be monopsonists. The words ‘firm’ and ‘village’ hence have the same meaning from this point onwards.

4.1 Labor supply

Identification

I start with the identification of the miner supply function. Taking the logarithm of equation (2) for \( \tau = s \), and denoting logs as lowercases, gives equation (15).

\[
\begin{align*}
w^s_{f,t} - w^*_t = (\psi^s_{f,t} - 1)l^s_{f,t} + (\psi^s_{f,t} - 1) \ln(\alpha^s_{f,t})
\end{align*}
\]

The miner supply elasticity \( \psi^s_{f,t} \) cannot be recovered by simply regressing miner employment on miner wages because of the latent firm characteristics \( \alpha^s_{f,t} \). Firms that are attractive to miners due to a high \( \alpha^s_{f,t} \), for instance because their mine has a good underground air quality, can offer a lower wage to attract the same number of miners. In order to identify the slope of the miner supply curve, a shock to labor demand that is excluded from miner utility is hence necessary. I rely on the seasonal character of coal demand as a source of labor demand variation. As explained in section 2.1, coal demand rises during the fall and winter due to low temperatures. Denote miner employment at firm \( f \) during winter and summer months as \( L^{s,WIN}_{f,t}, L^{s,SUM}_{f,t} \), and the corresponding daily miner wages as \( W^{s,WIN}_{f,t}, W^{s,SUM}_{f,t} \). The supply residuals during winter and summer are \( \alpha^{s,WIN}_{f,t} \) and \( \alpha^{s,SUM}_{f,t} \). I assume that the residual \( \alpha \), which contains unobserved firm characteristics and the outside option of working in another industry than coal mining, does not change between winter and summer: \( \alpha^{s,WIN}_{f,t} = \alpha^{s,SUM}_{f,t} \).
The slope of the skilled labor supply curve can then be calculated using equation (16). This yields a different labor supply curvature for every firm and every time period.

$$\psi_{f,t}^s = \frac{w_{f,t}^{s,WIN} - w_{f,t}^{s,SUM}}{l_{f,t}^{s,WIN} - l_{f,t}^{s,SUM}} + 1$$

(16)

How realistic is the exclusion restriction that non-wage firm characteristics do not vary between summer and winter? The supply residual $\alpha_{f,t}^s$ could differ between seasons if working conditions at coal mines change between these months, or if the outside option of working outside of the coal mining industry changes. Increasing agricultural demand during the summer could, for instance, make working outside of coal mining more attractive during these months.\(^{33}\) Figure 5 shows, however, that the monthly wage profile of unskilled workers did not fluctuate between the different seasons. Unskilled workers, such as mule drivers, could switch between coal mining and other industries, such as agriculture, without a wage loss, as their skills were not specific to coal mining. If outside options of coal mining working conditions would have differed between different months, then unskilled wages should have fluctuated as well, which they did not. Moreover, historical evidence on the Northern Illinois coalfields mentions that most miners were unemployed during the summer months in any case (Joyce, 2009), seasonal changes in labor demand in other industries that could violate the exclusion restrictions are therefore less problematic. Figure 5(d) reveals a very small dispersion in helper wages, which did not change in response to labor demand shocks. I therefore assume that the helper supply function is flat, meaning that $\psi_{f,t}^u = 1$. Helper wages are hence equal to the base salary that can be earned in other industries: $W_{f,t}^u = W_t^*$.\(^{34}\)

**Measurement of wages**

Miners earn a piece rate $W_{f,t}^q$ per ton of coal, which corresponds to a daily wage per miner $W_{f,t}^s \equiv \frac{W_{f,t}^q Q_{f,t}}{L_{f,t}^s}$, while helpers earned a uniform daily wage. Firms had to set a unique piece rate for all miners in a given mine in a given month, and were hence not able to wage discriminate between their workers. I transform these piece rates in the data to daily wages in order to estimate the model.

\(^{33}\)This would, however, be consistent with higher wages during the summer, while lower summer wages are observed.

\(^{34}\)There are other possible explanations for the fact that wages do not react to labor demand shocks, such as behavioral explanations (Kaur, 2019). The key thing to note here is, however, that monthly wage profiles were only flat for unskilled workers, not for skilled workers. Although miner labor contracts differed from helpers in that they received a piece rate rather than a daily wage, both of these contracts were limited to monthly durations or less; it is hence not the case that helper wages did not respond to seasonal demand shocks because they were pre-negotiated for the entire year.
Estimation

I calculate the slope of the labor supply curve for each firm using equation (16). Unskilled worker wages $W^*_t$ are not observed every year, but the years for which they are reveal that they did not meaningfully vary over time. I therefore impute unskilled wages for the other years by assuming that unskilled wages were constant over time. Wages and employment rates are reported separately for winter and summer months between 1884 and 1890. The reported wages are, however, piece rates (wage per ton). This is a problem to the extent that labor productivity differed between winter and summer months. Luckily, output per worker-day is reported together with the monthly wage data, which I denote as $A^{SUM}$ and $A^{WIN}$, and which are assumed to be identical for all mines. I transform the piece rates $W_{f,t}^{WIN}$, $W_{f,t}^{SUM}$ into daily wages by multiplying them by the number of employees per worker: $W_{f,t}^{WIN} = W_{f,t}^{WIN} A^{WIN}$ and $W_{f,t}^{SUM} = W_{f,t}^{SUM} A^{SUM}$. The firm characteristics $\alpha^H_{f,t}$ can be recovered by inverting the labor supply function, using the estimated values for $\psi^s_{f,t}$ and $W^*$:

$$\alpha^s_{f,t} = \left( \frac{W_{f,t}^s}{W^*} \right)^{1/\psi^s_{f,t}}$$

4.2 Labor demand

Identification

I divide the demand for helpers by demand for miners, from equation (6), and take logs. The relative demand for helpers vs. miners in logs is hence given by equation (17).

$$l^u_{f,t} - l^s_{f,t} = \sigma \left( w^u_{f,t} - w^s_t + \ln(\psi^s_{f,t}) \right) + (1-\sigma) \left( \beta^k K_{f,t} + \beta^t t + \beta^0 + \eta_{f,t} \right)$$

The coefficients that need to be identified are (i) the elasticity of substitution between miners and helpers, $\sigma$, and (ii) the effects of cutting machines and time on skill-augmenting productivity, $\beta \equiv (\beta^k, \beta^t, \beta^0)$. There are two reasons why one cannot recover $\sigma$ and $\beta$ by regressing the relative usage of high- vs. low-skilled workers on relative wages, capital usage and time. First, technology adoption is endogenous to the unobserved differences in productivity, which is the classical simultaneity problem when identifying a production function. Second, high-skilled wages are endogenous, and hence also a function of these unobserved productivity differences. In order to identify the labor relative demand function, I combine the timing assumptions on the labor and capital decisions that were made in the previous section, following Olley and Pakes (1996); Ackerberg et al. (2015) with the linearity assumption for the productivity transition rule in equation (1c). I follow Blundell and Bond (2000) by
taking $\rho$-differences of the log labor demand function:

$$l_{t}^{u} - l_{t}^{s} - \rho(l_{t-1}^{u} - l_{t-1}^{s}) = \sigma\left(w_{t}^{u} - w_{t}^{s} + \ln(\psi_{t}^{s}) - \rho(w_{t-1}^{u} - w_{t-1}^{s} + \ln(\psi_{t-1}^{s}))\right) + (1 - \sigma)\left(\beta_{k}(K_{t} - \rho K_{t-1}) + \beta_{t}(t - \rho t + \rho) + \beta_{0}(1 - \rho)\right) + (1 - \sigma)\varepsilon_{f,t}$$

The transient productivity shock $\varepsilon_{f,t}$, from equation (1c), was already assumed to be i.i.d. distributed. The timing assumptions on capital and labor choices can now be used to construct moment conditions. As mentioned before, cutting machines in time $t$ are assumed to be chosen at time $(t - 1)$, before the productivity shock $\varepsilon_{f,t}$ is observed. Capital choices at time $t$ are therefore orthogonal to the shock $\varepsilon_{f,t}$. Secondly, while miner wages $W_{f,t}$ change in response to productivity shocks, the slope of the miner supply curve $\psi_{f,t}$ does not, as the labor supply curve is assumed to be linear. The slope $\psi_{f,t}$ should therefore also be orthogonal to the productivity shock $\varepsilon_{f,t}$. The moment conditions to identify $\sigma$ and $\beta$ are hence given by in equation (18):

$$E[\varepsilon_{f,t}|K_{f,t},\psi_{f,t},t] = 0$$

(18)

The main drawback of the approach followed above is that the productivity transition in equation (1c) has to be linear, which is a strong assumption. An alternative would be to invert out the productivity residual from the input demand conditions, as is common in the productivity literature. The non-scalar nature of productivity rules out the usage of Ackerberg et al. (2015), but alternatives have been developed for factor-augmenting production functions in Doraszelski and Jaumandreu (2017); Demirer (2020). These approaches do not, however, allow for endogenous input prices, which is problematic in the context of this paper. The variation in factor cost shares used to identify factor-augmenting productivity in these papers could just as well be due to variation in monopsony power.

**Estimation**

I estimate the coefficients $\sigma$, $\beta_{k}$, $\beta_{t}$, $\beta_{0}$ and $\rho$ in equation (17) using GMM, with the moment condition (18). The fitted log markdown estimates $\psi_{f,t}^{s}$ from the miner supply function are included in the righthand side, as if they were observed. As I only estimate these markdowns for the first four time periods (8 years), the sample size is reduced to 517. I assume that the labor demand coefficients estimated for this time period hold for the entire sample, until 1902. As the labor supply estimates are used to estimate labor demand, I block bootstrap the entire estimation procedure, with 50 bootstrap iterations. Once the estimates of $\sigma$ and $\beta$ are known, which are denoted $\hat{\sigma}$ and $\hat{\beta}$, miner- and helper-augmenting
productivity can be backed out as follows:

\[
\begin{align*}
\hat{\omega}^u_{f,t} &= Q_{f,t} \left[ \left( \exp(\hat{\beta}^k K_{t,f} + \hat{\beta}^t t + \hat{\beta}^0) L_{f,t}^u \right)^{\frac{\hat{\sigma} - 1}{\hat{\sigma}}} + (L_{f,t}^u)^{\frac{\hat{\sigma} - 1}{\hat{\sigma}}} \right]^{\frac{1}{\hat{\sigma}}} \exp(\hat{\beta}^k K_{t,f} + \hat{\beta}^t t + \hat{\beta}^0) \\
\hat{\omega}^s_{f,t} &= Q_{f,t} \left[ \left( \exp(\hat{\beta}^k K_{t,f} + \hat{\beta}^t t + \hat{\beta}^0) L_{f,t}^u \right)^{\frac{\hat{\sigma} - 1}{\hat{\sigma}}} + (L_{f,t}^u)^{\frac{\hat{\sigma} - 1}{\hat{\sigma}}} \right]^{\frac{1}{\hat{\sigma}}} \exp(\hat{\beta}^k K_{t,f} + \hat{\beta}^t t + \hat{\beta}^0)
\end{align*}
\]

4.3 Capital demand

Identification

Finally, the capital demand function, equation (10), needs to be identified. Using Hotz and Miller (1993), the conditional value of not adopting \(v^0_{f,t}\) can be rewritten as a function of the value of adopting and the choice probability of adopting in the next period.

\[
v^0_{f,t} = \delta(v^1_{f,t+1} - \ln(S_{f,t+1}^1) - e_{f,t})
\]

Substituting this into equation (10) results in the following estimable equation:

\[
s^1_{f,t} - s^0_{f,t} = \frac{\delta(\pi^1_{f,t} - \pi^0_{f,t})}{1 - \delta} - \frac{\delta^2(\pi^1_{f,t+1} - \pi^0_{f,t+1})}{1 - \delta} + \delta s^1_{f,t+1} - \gamma z_{f,t} + \delta \gamma z_{f,t+1} + \delta e_{f,t}
\]

The variable profit gain from cutting machines, \((\pi^1_{f,t} - \pi^0_{f,t})\) can be computed using equation (7b) and the estimated coefficients from the labor demand and supply functions. The discount factor \(\delta\) is assumed to be known, and the empirical choice probabilities \(s_{f,t}\) are observed in the data. The only coefficients that still need to be identified are hence the sunk cost parameters \(\gamma\). I include a series of technical mine characteristics as sunk cost shifters \(z\), which are all assumed exogenous when cutting machines are chosen. First, I control for the average mine depth and vein thickness: it was harder to install machines in deeper mines. Secondly, I control for size: mechanizing larger mines probably requires a larger sunk cost investment, while it also reaps larger returns. Thirdly, I control for the mine type (shaft, slope, drift or surface mine, or a hybrid combination of these). Mines using vertical shafts had to transport cutting machines using elevators, which was hard. I also control for whether hauling was mechanized or not, as electricity generators may have been shared between technologies. Finally, I control for the mine ventilation type, which could be done using a fan, furnace, steamjet, or naturally. I include the the distance to St. Louis and Chicago, where the machines were produced as a sunk cost shifter, and year dummies to capture changes in common sunk costs over time. Finally, I control for the capital intensity in manufacturing firms in the same county, measured by their capital per worker, as some sunk cost drivers may be common between industries. I assume all these sunk cost shifters \(z\) are exogenous at the time of choosing cutting machines: firms do not choose how deep or thick the coal seam is, for instance. The hauling and ventilation technologies are choices by
the firm, but given that these technologies were already largely in place by the time that the cutting process mechanized, they are considered as given throughout the panel.

Estimation

In order to avoid values of zero and one for capital usage on the left hand side, I start by smoothing the choice probabilities, as is usual in the literature. I use a probit model in which I regress machine adoption on output, the coal price, employees of each type, wages, vein thickness and mine depth, all in logs. I also include county and year fixed effects. Next, I use these choice probabilities as $S_{f,t}^1$ and $S_{f,t}^0$ to construct the left-hand side of equation (10). The markdown estimates are only available for the first four time periods (1884-1890). I therefore assume that the level of monopsony power in each town remained constant from 1892 onwards. I do the same for the mine depth, thickness, and shaft type. As these variables are not observed in every year, I average them over the entire time period by village. I set the annual discount factor to 0.95, which implies a biennial discount rate of $\delta = (0.95)^2$. The capital demand model gives an estimated machine adoption probability $S_{f,t}^1$ for each firm. I obtain the estimated machine usage probability, $\hat{K}_{f,t}$ by setting it equal to the actual usage rate in the first year that a village enters the dataset, and by then using the transition equation (3).

5 Discussion of the results

5.1 Model estimates

Labor supply

The model estimates are in table 1. I censor the markdown estimates at the 1st and 99th percentiles of the markdown distribution. The mean miner wage markdown ratio $\psi_{f,t}^s$ is 1.138, which implies that the marginal product of miners lies on average 14% above their wage (i.e., wages are marked down by 12%). The average markdown ratio is significantly higher than one, so mining firms have monopsony power over miners. The distribution of markdowns is plotted in appendix figure A5. The markdown ratio lies mostly between 1 and 1.5. Appendix table A5 provides some descriptive analysis by regressing the markdown estimates on a number of county and firm characteristics. A higher labor supply elasticity, which implies higher monopsony power, is positively correlated with the firms’ employment shares in a county, and with the share of the county’s population that works in coal mining. The supply elasticity does not correlate with the number of manufacturing plants and workers in the county. It is also uncorrelated to both the total farmland area in the county and an

---

This is consistent with the average interest rate on loans in Illinois being around 5.5% in 1902 (Smiley, 1975). I conduct robustness checks with different discount rates in appendix D.1.
indicator of ‘agricultural labor scarcity’ in each village from the Bureau of Labor Statistics’ reports. This suggests that there is a low substitutability between skilled miners in coal mining and in other industries.

Labor demand

The estimates of the high-skilled labor demand function are in panel (b) of table 1. The elasticity of substitution between skilled miners and unskilled helpers is estimated to be 0.077, and is both significantly larger than zero and smaller than one. Miners and helpers were hence gross complements. This low elasticity is consistent with the nature of underground coal mining: in order to extract coal, it needs to be both cut and hauled, which happened by skilled and unskilled workers. As coal cannot be stored underground, one cannot cut more than is hauled, or vice-versa. The tasks carried out by both types of workers were, hence, of a complementary nature. The effect of cutting machines on miner-augmenting productivity, $\beta^K$, is estimated at 0.502, so cutting machines increased miner productivity by 65%. Combining the findings that both worker types were gross complements and that cutting machines were skill-augmenting implies that cutting machines were unskill-biased: they decreased the relative demand for skilled workers. The biennial correlation in high-skilled productivity is 0.451, which corresponds to a yearly serial miner productivity correlation of 0.672.

Capital demand

The estimated sunk cost drivers are, finally, in panel (c) of table 1. The coefficients $\gamma$ are in units of 1,000 US dollars: an increase in the right-hand side variable of one percent hence changes sunk costs by $\gamma * $10 dollars. The average sunk cost of mechanizing a village is estimated at $203K. This was around five times the average annual variable profit, and hence a large investment. The average capital stock of mechanized villages in 1884 was on average $120K larger compared to non-mechanized villages, which is of the same order of magnitude of the sunk cost estimates. Sunk costs increase with size, vein thickness, and mine depth. If the depth of the mine increased by 10%, sunk costs were $2,310 higher, or 1.1% of the average total sunk cost. The other technical characteristics do not significantly alter sunk costs. Panel (a) of figure 8 compares the average cutting machine usage rates per year in reality, the solid line, and the predicted rates using the model, the dashed line. Although the model overestimates usage rates between 1894 and 1900 and underestimates it between 1884 and 1892, it is able to replicate the broad trend and level of average machine usage rates over time. Although the model assumes that machine adoption is permanent, average machine usage rates can fall over time due to entry and exit of villages.

36This is equal to US$5.8M in 2020
Table 1: Model estimates

(a) Miner supply

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miner supply elasticity $\psi^s$</td>
<td>1.138</td>
<td>0.002</td>
</tr>
</tbody>
</table>

(b) Miner demand

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of substitution $\sigma$</td>
<td>0.077</td>
<td>0.016</td>
</tr>
<tr>
<td>Miner-augmenting effect of machines $\beta^k$</td>
<td>0.502</td>
<td>0.022</td>
</tr>
<tr>
<td>Time $\beta^t$</td>
<td>0.084</td>
<td>0.005</td>
</tr>
<tr>
<td>Constant $\beta^0$</td>
<td>-159.261</td>
<td>9.233</td>
</tr>
<tr>
<td>Serial correlation $\rho$</td>
<td>0.451</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Observations: 517  
R-squared: 0.470

(c) Sunk capital costs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>18.872</td>
<td>0.000</td>
</tr>
<tr>
<td>log(Output)</td>
<td>3.029</td>
<td>0.167</td>
</tr>
<tr>
<td>log(Dist. to St. Louis)</td>
<td>-0.405</td>
<td>1.508</td>
</tr>
<tr>
<td>log(Thickness)</td>
<td>62.822</td>
<td>2.142</td>
</tr>
<tr>
<td>log(Depth)</td>
<td>23.095</td>
<td>1.354</td>
</tr>
<tr>
<td>log(Dist. to Chicago)</td>
<td>-5.114</td>
<td>2.445</td>
</tr>
<tr>
<td>log(K/L) in manufacturing</td>
<td>-1.347</td>
<td>0.930</td>
</tr>
<tr>
<td>1(Drift mine)</td>
<td>2.396</td>
<td>22.049</td>
</tr>
<tr>
<td>1(Other type mine)</td>
<td>692.521</td>
<td>415.759</td>
</tr>
<tr>
<td>1(Shaft mine)</td>
<td>1.685</td>
<td>23.036</td>
</tr>
<tr>
<td>1(Slope mine)</td>
<td>-1.443</td>
<td>22.946</td>
</tr>
<tr>
<td>1(Furnace ventilation)</td>
<td>-3.871</td>
<td>4.787</td>
</tr>
<tr>
<td>1(Natural ventilation)</td>
<td>-4.635</td>
<td>3.113</td>
</tr>
<tr>
<td>1(Manual hauling)</td>
<td>-3.776</td>
<td>2.887</td>
</tr>
</tbody>
</table>

Observations: 894  
R-squared: 0.175

Notes: Standard errors are block-bootstrapped with 50 iterations. Estimates in panel (c) are in units of 1000 US dollar: an increase in a right-hand variable of one unit results in a change in sunk costs of $\gamma^*$1000.
Table 2: Counterfactuals

<table>
<thead>
<tr>
<th>(a) Competitive labor markets</th>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.142</td>
<td>0.030</td>
<td>0.003</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.903</td>
<td>0.068</td>
<td>0.005</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8592.386</td>
<td>25.831</td>
<td>43.336</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.528</td>
<td>0.004</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Minimum wage</th>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.071</td>
<td>-0.041</td>
<td>0.026</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.740</td>
<td>-0.094</td>
<td>0.042</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8220.798</td>
<td>-345.757</td>
<td>133.430</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.736</td>
<td>0.213</td>
<td>0.004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Competitive labor markets, K fixed</th>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.112</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.834</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8674.859</td>
<td>108.304</td>
<td>43.336</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.530</td>
<td>0.007</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Minimum wage, K fixed</th>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.112</td>
<td>0.000</td>
<td>0.026</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.834</td>
<td>0.000</td>
<td>0.042</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8176.489</td>
<td>-390.066</td>
<td>133.430</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.733</td>
<td>0.209</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Notes: Standard errors bootstrapped, 50 iterations

5.2 Counterfactuals

With the estimated model at hand, I now examine how cutting machine adoption, productivity, employment and wages would change in the different counterfactual scenarios.

Perfectly competitive labor markets

Panel (a) in table 2 summarizes the average estimated values for cutting machine usage, miner-augmenting productivity, the number of miner-days used, and miner daily wages in the monopsonistic
equilibrium, the first column, and in the competitive equilibrium, the second column. The third and fourth columns tabulate the difference between both equilibria and its standard error. I compare these variables for the period 1886-1902: in 1884, machine usage was by definition the same in the counterfactual and real world, as it takes one time period to install machines. If miner markets would become perfectly competitive, the usage of cutting machines would increase by a fourth, from 11.2% to 14.2%, and this change is statistically significant. As a result, the average miner-augmenting productivity level across all mines increases by 1.4% per year. The number of miners employed increases slightly, but not significantly, while wages increase by 0.3%. These changes in employment and wages are the result of two countervailing forces. On one hand, decreased monopsony power causes firms to shift their input usage towards skilled workers, because their marginal cost falls. On the other hand, it also makes firms use less cutting machines, which are miner-saving. Panel (c) in table 2 shows the counterfactual results if machine usage would, erroneously, be considered exogenous. The increase in the number of miner-days when moving to the competitive equilibrium would be four times higher compared to the endogenous technology model, and the increase in wages twice as high. This is an important finding: without endogenous machine usage, a fall in monopsony power would always benefit miners, as both wages and employment increase. In reality, however, firms respond to this drop in monopsony power by adopting more miner-saving machines, which dampens the increases in employment and wages.

Figure 8: Counterfactual machine usage

(a) Model fit

(b) Cutting machine usage

Notes: Machine usage sometimes falls because of exit and entry of villages out of and into the dataset.
Minimum wage policy

In panel (b) of table 2, I compare the outcome variables of interest between reality and a world in which a minimum wage is set at the 40th percentile of the wage distribution. Machine usage would fall from 11.2% to 7.1%, and this drop is statistically significant. Miner productivity would, as a result, drop by 2%. Wages would increase by 21% on average and employment would drop by 4%, although this drop is not statistically significant. The effects of a minimum wage would, again, be different if machine adoption would be exogenous. As shown in panel (d) of table 2, employment would fall by more and wages increase by less in that case. The decrease in miner-saving machine

Notes: Panels (a)-(d) plot counterfactual average machine usage, employment, wages and variable profits as a fraction of their value in reality, in function of the minimum wage percentile.

37The increases in employment and wages in the model without endogenous machines are still moderate, because of the low elasticity of substitution between miners and helpers. If this substitution elasticity would be higher, then changes in monopsony power would have larger effects on equilibrium labor quantities and wages.
usage dampens the drop in employment and rise in wages after a raise of the minimum wage.

In figure 9, I plot counterfactual machine usage, miner employment, miner daily wages and variable profits as a fraction of their values in reality against the minimum wage level as a percentile of the wage distribution. Machine usage, in panel (a), remains fairly constant up to a minimum wage at the 25th percentile, after which it falls. A minimum wage at the 50th percentile results in a reduction of machine usage by 40%. Miner employment, in panel (b), equally falls with minimum wage levels. In theory, employment could increase with minimum wages, if the minimum wage lies in between the monopsonistic and competitive equilibrium wage, as was shown in figure 7(a). The supply and demand curves differ, however, across firms: a change in the minimum wage is therefore likely to increase employment for some firms, but decrease it for other whose competitive wage is lower. Figure 9(b) shows that even moderate minimum wages lead to decreased employment in this industry. A minimum wage at the 10th percentile reduces employment on average by 5%, a fraction that remains roughly constant up to very high values for the minimum wage. One reason why employment losses are not higher is that decreased machine usage at higher minimum wage levels also implies that more labor-intensive production technologies are used, which require more labor in equilibrium. Miner wages, in figure 9(c), increase with the minimum wage, while variable profits, in (d) fall with minimum wage levels.

Monopsony and skill-biased technologies

Using the estimated model, I find that monopsony power led to slower technological change and productivity growth in the Illinois coal mining industry. Theorem 1 showed, however, that monopsony power can also lead to higher innovation and productivity growth in other settings. The combination of monopsony power over skilled workers and skill-saving technological change is the cause of the negative effect of monopsony power on innovation in the coal example. I now discuss two additional counterfactuals in which the direction of the effect goes the other way. The estimates of the labor demand curve showed that miners and helpers are gross complements, and cutting machines skill-augmenting. Panel (a) of figure 10 conducts the same counterfactual exercise of moving to competitive labor markets, but rather assumes that miners and helpers are gross substitutes, with an elasticity of substitution of 1.20. In line with theorem 1, cutting machine usage now drops from 0.241 to 0.219 when moving to the competitive equilibrium, as cutting machines now move input usage from helpers towards miners, rather than the other way around.

Figure 10(d) keeps the elasticity of substitution at the actual level, but assumes that machines are helper-augmenting, rather than miner-augmenting. I flip the sign on the factor-augmenting productivity coefficient and impose that \( \beta_k = -1 \). Mining locomotives are an example of such a technology: these reduced demand for helpers, as their hauling tasks were mechanized, relatively to miners. I find that the usage rate of such a technology would drop from 0.216 to 0.210 when moving to a per-
fectly competitive miner market. Monopsony power can therefore be conducive to innovation when technologies increase the demand of the inputs over which firms have monopsony power.38

6 Conclusion

In this paper, I investigate how monopsony power affects the adoption of factor-biased technologies. I find that in theory, the adoption of technologies that reduce demand for an input are reduced by the level of monopsony power over that input. I use this model to understand how monopsony power affected the mechanization of the late 19th century Illinois coal mining industry. If miner markets would have become perfectly competitive, usage rates of coal cutting machines would have increased by a fourth, and miner productivity by 1.4%. Minimum wage policies would have countered the labor-saving effects of technological change, but would reduce technology adoption and productivity growth by even more. The main driver for these effects was the combination of the facts that technological change saved on skilled workers, and that firms mainly had monopsony power over these skilled workers. If monopsony power would have been concentrated on unskilled labor markets instead, or if new technologies would have saved on unskilled workers, then monopsony power would lead to higher innovation rates. In order to know how monopsony power affects innovation incentives and productivity growth today, it is therefore crucial to have robust evidence of both the direction of technological change, and of the relative degree of monopsony power across the worker skill and

38Hauling machine usage was implicitly assumed to be exogenous in the counterfactual analysis, which is in contrast to this finding. By 1886, 90% of coal was, however, already hauled using locomotives, and this share was over 95% by 1898. Technology adoption is, in addition, assumed to be irreversible. The drop in locomotive adoption resulting from a move to the competitive equilibrium is hence not an important threat to the validity of the cutting machine counterfactual analysis.
income distribution.
References


Reid, A. (1876). *Transactions of the north of england institute of mining and mechanical engineers*.


Appendices

A  Data

A.1  Data sources

The main data source is the biennial report of the Bureau of Labor Statistics of Illinois between 1884-1902. Every edition contains a list of all mines in each county, the name of the firm or individual operating the mine, and information on mine characteristics, input usage, production and prices. I digitized this data using a data entry firm, and checked the data for consistency by comparing the county totals mentioned elsewhere in the reports with the aggregates of the mine-level data. An overview of all variables (including unused ones) in the production data, and the years in which they are observed, is in tables A6 and A7. I use the 1880 population census to have information on county population sizes, demographic compositions, and areas. I also use the county-level capital stock and employment in manufacturing industries from the 1880 census of manufacturing, and the number of farms and improved farmland area from the 1880 census of agriculture.

A.2  Data cleaning

I adjusted mine, firm, town, and county names in the raw dataset to have consistent names over time. The raw mine and firm names as reported in the data, next to the up firm and mine identifiers which ensure consistency over time. Nevertheless, mine and firm names changed frequently due to ownership changes or other reasons, which will be recorded as false exits and entries. For this reason, the dataset can best be used at the village-level when using panel data methods. In 1898, eight-hour days were enforced by law, which means that the ‘number of days’ measure changes in unit between 1898 and 1900. As the inspector report from 1886 shows that ten-hour days were the standard, I multiply the number of working days after 1898 by 80% in order to ensure consistency in the meaning of a ‘workday’.

A.3  Aggregation to the village-year level

I aggregate all variables from the mine-year-level to the village-year-level, in order to estimate the model. For the number of employee-days, the wage bill, output, and revenue, I sum all variables to the village-year level. For the technical characteristics (depth, thickness, ventilation dummies, shaft type, hauling technology), I take averages weighted by miner usage.
B Theory addendum

B.1 Simple model: Cobb-Douglas production function

In this section, I discuss a simplified model that features a Cobb-Douglas production function and log-linear input supply curves. Consider a monopsonistic firm that produces a homogeneous good $Q$. The firm uses two inputs, $S$ and $U$. Input $S$ has an output elasticity of $\beta$, and there are constant returns to scale. The output elasticity $\beta$ is a function of a technology $K \in \{0, 1\}$.

$$Q = S^{\beta(K)} U^{1-\beta(K)}$$

Assume the goods market is perfectly competitive, meaning that the price of the good has an exogenous price that is normalized to one. Also assume that capacity is fixed: the quantity of $Q$ is hence exogenous from the point of view of the firm.

$S$ and $U$ have prices $W^s$ and $W^u$. The technology $K$ has a fixed cost $W^K$. Variable profits are denoted as $\pi$, total profits as $\Pi$:

$$\pi \equiv Q - W^s S - W^u U$$
$$\Pi \equiv \pi - W^K K$$

**Input supply**  The supply functions for both inputs are upward-sloping and have a constant elasticity $(\psi^s - 1)$ and $(\psi^u - 1)$. The fixed cost $W^K$ is exogenous.

$$\begin{cases} W^s &= S^{\psi^s - 1} \\ W^u &= U^{\psi^u - 1} \end{cases}$$

**Variable input demand**  First, consider technology usage to be exogenous. Assume firms are profit maximizing: $\max_{S,U} \Pi$ Solving the first order conditions, demand for $S$ and $U$ is given by the following system of equations:

$$\begin{cases} S &= \beta(K)Q(W^s)^{\psi^s}^{-1} \\ U &= (1 - \beta(K))Q(W^u)^{\psi^u}^{-1} \end{cases}$$

Plugging these into variable profits gives the following expression. Variable profits are zero if input prices are exogenous (because the goods market is competitive), and positive if the input supply curves
are upward-sloping:

\[ \pi = Q \left( 1 - \frac{\beta(K)}{\psi^s} - \frac{1 - \beta(K)}{\psi^u} \right) \]

**Technology demand**  
Now consider the effects of the technology on variable profits.

\[ \frac{\partial \pi}{\partial K} = Q \left( \frac{\psi^s - \psi^u}{\psi^u \psi^s} \right) \frac{\partial \beta}{\partial K} \]

Suppose the technology is S-biased, meaning that \( \frac{\partial \beta}{\partial K} > 0 \). The profit gain from using the technology then increases with the degree of monopsony power over input \( S \). The reason for this is that using the technology results in higher usage of \( S \), which increases the profit extracted from the wedge between the marginal product and price of \( S \). The opposite holds for input \( U \). Using the technology reduces the marginal product of input \( U \), and hence demand for this input. The firm substitutes away from \( U \), and hence extracts less profits from the wedge between the marginal product and price of \( U \). The larger this wedge is, the smaller the benefits of using the technology.

\[
\begin{align*}
\frac{1}{\psi^s} \left( \frac{\partial \pi}{\partial K} \right) &> 0 \\
\frac{1}{\psi^u} \left( \frac{\partial \pi}{\partial K} \right) &< 0
\end{align*}
\]

Using a S-biased technology increases variable profits as long as the firm has more monopsony power over \( S \) than over \( U \). It increases total profits as long as this variable profit gain is larger than the fixed cost \( W^K \).

**B.2 Proof of lemma 1**

Suppose inputs are gross complements, meaning that \( \sigma < 1 \). From equation 14b, we know that a \( \tau \)-augmenting productivity shock increases profits in that case: \( \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}} > 0 \). It follows that the derivative of this profit change with respect to the degree of monopsony power \( \tilde{\psi}_{f,t} \) is then negative, as shown in equation (20). In the opposite case that inputs are gross substitutes, meaning that \( \sigma > 1 \), a \( \tau \)-augmenting productivity shock decreases profits. The derivative in equation (20) is then positive, so the profit decrease is smaller in its absolute value.

\[
\frac{\partial}{\partial \tilde{\psi}_{f,t}} \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}} = \frac{\sigma(\sigma - 1)(\psi_{f,t}^\tau)^2}{(1 + \sigma(\psi_{f,t}^\tau - 1))^2} \frac{-\sigma \psi_{f,t}^\tau}{\sigma(\psi_{f,t}^\tau - 1)} C_{f,t}^\tau
\]

(20)
with $C_{f,t}^\tau \equiv \left[ Q_{f,t} P_{f,t}^{\sigma} \right]^{\omega_{f,t}^{\tau}(\psi_{f,t}^{\tau}-1)\left(1+\sigma(\psi_{f,t}^{\tau}-1)\right)^{-1}} \left(\omega_{f,t}^{\tau}(\psi_{f,t}^{\tau}-1)W_{t}^{*}\right)^{\frac{1-\sigma}{1+\sigma(\psi_{f,t}^{\tau}-1)}} > 0$

### B.3 Proof of theorem 1

The effect of monopsony power over input $\tau$ on the variable profit gain from $K$ is equal to:

$$\frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}} \right) = \frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\tau}} \right) \frac{\partial \omega_{f,t}^{\tau}}{\partial K_{f,t}}$$

A technology $K$ can decrease the demand for input $\tau$ in two cases: (i) if $\sigma < 1$ and $\frac{\partial \omega_{f,t}^{\tau}}{\partial K_{f,t}} > 0$, and (ii) if $\sigma > 1$ and $\frac{\partial \omega_{f,t}^{\tau}}{\partial K_{f,t}} < 0$. Suppose (i) holds. It can be seen from equation (20) that monopsony power decreases the variable profit gains from the technology:

$$\frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}} \right) = \frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\tau}} \right) \frac{\partial \omega_{f,t}^{\tau}}{\partial K_{f,t}} < 0$$

Suppose, now, that case (ii) holds: inputs are gross substitutes and the technology decreases the productivity of input $\tau$. Again using equation 20), the variable profit gain decreases with monopsony over input $\tau$:

$$\frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}} \right) = \frac{\partial}{\partial \psi_{f,t}^{\tau}} \left( \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\tau}} \right) \frac{\partial \omega_{f,t}^{\tau}}{\partial K_{f,t}} < 0$$

This proves that a technology that decreases the demand for an input increases with the degree of monopsony power over that input. The proof for a technology that increases the demand for an input is analogous.

### B.4 Endogenous output and prices

#### Equilibrium

In the baseline model, output was considered exogenous. In many other settings, the adoption of factor-biased technologies may change equilibrium output. I now extend the model from section B.1 to allow for endogenous output. The equilibrium input and output quantities and input prices are the
solution to the following system of equations:

\[
\begin{align*}
W^s &= S^{\psi^s-1} \\
W^u &= U^{\psi^u-1} \\
S &= \frac{\beta(K)Q}{W^s^{\psi^s\mu}} \\
U &= \frac{(1-\beta(K))Q}{W^u^{\psi^u\mu}} \\
Q &= S^{\beta(K)U^{1-\beta(K)}}
\end{align*}
\]

The solutions for both types of labor quantities are:

\[
\begin{align*}
S &= \left[ \frac{(\beta^{\psi^s})(1-\beta)}{\psi^s\mu} \right]^{\psi^u-\beta} \left[ \psi^s-(1-\beta)(\psi^u-\beta)-\beta(1-\beta) \right] \\
U &= \left[ \frac{(\beta^{\psi^u})(1-\beta)}{\psi^u\mu} \right]^{\psi^s-(1-\beta)} \left[ \psi^s-(1-\beta)(\psi^u-\beta)-\beta(1-\beta) \right]
\end{align*}
\]

**Factor bias and profits**

I simulate the model above with \( \psi^s = 1.2, \psi^u \sim U[1.5, 2.5], \) and \( \beta \in U[0.1, 0.9] \). Figure A7(a) plots how profits change with the output elasticity of input \( S \), over which the firm has the highest monopsony power, in the case that firms do not have any product market power: \( \mu = 1 \). I plot this relationship for the four quartiles of the markdown \( \psi^s \) distribution. The profit gain of moving from a \( U \)-intensive production function, at \( \beta = 0.1 \), to an \( S \)-intensive production function, at \( \beta = 0.9 \), increases with monopsony power over \( S \). This is in line with theorem 1: the profitability of a technology that increases the usage of the input increases with the degree of monopsony power that the firm exerts over that input. Figures A7(b)-(d) plot the same relationship, but for markups ranging in between 1.5 and 2.5. The gradient of the profit curve increases with monopsony power over \( S \) for all these different markup levels.

**C Additional results**

**C.1 Returns to scale**

The model assumes constant returns to scale. I rely on the 1890 monthly data to check this assumption. If returns to scale are constant, output per worker should not change in response to coal demand shocks. If returns to scale are increasing, we would expect labor productivity to be higher during the autumn and winter, when coal demand and output is high. Figure A2 plots the monthly variation in log output per miner and helper. Both do not seem to change by much during the year, which is consistent with constant returns to scale. Regressing log output per worker on a summer/spring
dummy also yields a coefficient that is close to zero and statistically not significant.

C.2 Cost dynamics

There are multiple sources of cost dynamics that would invalidate the productivity transition in equation (1c). If it becomes increasingly costly to operate deeper mines, for instance, productivity would depend on past cumulative output, as Aguirregabiria and Luengo (2017) find for copper mining. Such dependence could also exist due to learning by doing, as in Benkard (2000), but productivity would then increase with cumulative output, rather than fall. I test this by regressing the logarithms of the productivity residuals $\omega^s_{f,t}$ and $\omega^u_{f,t}$ on log cumulative output. The estimated coefficients in table A4 are close to zero and insignificant, and the $R^2$ is near zero as well. As cumulative output does not seem to explain any variation in worker productivity levels, it is unlikely that cost dynamics play an important role in this setting. This may be due to the fact that coal reserves were still plentiful in Illinois: figure A3 shows that there was very little variation in mine depth and vein thickness, two sources of potential cost dynamics, with the time of operation since 1884.

C.3 Intermediate inputs

Firms used some intermediate inputs, such as black powder and coal. These were omitted from the baseline model. I assume that these intermediate inputs are not substitutable with either labor or capital, and enter the production function (1a) in fixed proportions:

$$Q_{f,t} = \min \left\{ \left( \omega^s_{f,t} L^s_{f,t} \right)^{\sigma-1} + \left( \omega^u_{f,t} L^u_{f,t} \right)^{\sigma-1} ; \omega^m_{f,t} M_{f,t} \right\}$$

This way, they do not enter the production function. I also assume that both intermediate input markets were perfectly competitive. For the coal input, this assumption was already made, as the coal product market is perfectly competitive.\footnote{Black powder was not purchased by the coal operators, but by the miners themselves. I do not observe the prices of black powder in every year or by firm. Due to this market being perfectly competitive, it can be abstracted from in the model.}

C.4 Prices and wages: descriptive analysis

Table A1 provides some more descriptive evidence for the assumption that firms did not have any pricing power on the coal market and that quantities were fixed. I regress the change in output, miners, helpers, miner wages, and prices on the change in cutting machines in the previous time period, both
at the town-year level. Cutting machine adoption correlates with a drop in miner quantities and wages, but not in a drop in output, helper quantities, or prices.

C.5 Markups

Estimation  Markups $\mu_{f,t}$ are, by definition, equal to the ratio of coal prices $P_{f,t}$ over marginal costs $\lambda_{f,t}$:

$$\mu_{f,t} \equiv \frac{P_{f,t}}{\lambda_{f,t}}$$

Marginal costs are, by definition, equal to the derivative of variable costs to output:

$$\lambda_{f,t} \equiv \frac{\partial}{\partial Q}(W_{s,t}^s L_{f,t}^s + W_{u,t}^u L_{f,t}^u)$$

Taking the derivative and using the miner piece rate $W_{q,t}^q$, marginal costs can be re-written as the sum of the product of this piece rate and the miner markdown ratio, and the marginal product of helpers. The helper markdown ratio was assumed to be one, and hence does not enter the marginal expression.

$$\lambda_{f,t} = W_{q,t}^q \psi_{f,t} + W_{u,t}^u \frac{\partial L_{f,t}^u}{\partial Q_{f,t}}$$

Using the helper demand function 6, markups can be rewritten as:

$$\mu_{f,t} = \frac{P_{f,t}}{W_{q,t}^q \psi_{f,t} + (W_{u,t}^u)^{1-\sigma} (\omega_{u,t}^u)^{\sigma-1} P_{f,t}^\sigma}$$

Results  The average markup ratio was 1.59 and the median markup 1.36. Markup and markdown ratios are plotted in figure A5, and are both censured at the 1st and 99th percentile of their distribution. Average markups are higher than average markdowns, but this does not imply that market power on coal markets is higher than market power on labor markets. Both fixed costs and capacity constraints can be reasons why inframarginal producers operate at positive markups.

C.6 Transport costs

I assume the lowest observed price in every year is equal to the market-clearing price in St. Louis, $P_t^*$. I then regress log transport costs, $\ln(P_{f,t} - P_t^*)$ on the log distance between each village and both Chicago and St. Louis, in table A2. Transport costs increase with the distance to St. Louis, but not to Chicago. This is consistent with the fact that St. Louis was the most important destination market for Illinois coal mines.
D  Robustness checks

D.1  Discount rates

I re-estimate the model while changing the discount rate from 0.95 to both 0.90 and 0.98. I also estimate a static version of the model, in which the discount factor is set to zero. In this static model, I allow variable profits of machine usage to be immediately earned by the firm when adopting. Equation 8 hence becomes:

\[ v_{f,t}^1 = \frac{\pi_{f,t}^1 - \pi_{f,t}^0}{1 - \delta} - \gamma z_{f,t} - \chi_{f,t} \]

The resulting counterfactual plots are in figure A6. The main conclusions that machine usage rates would be higher in perfectly competitive labor markets and lower when enforcing minimum wages still hold. The effect of moving to perfect competition seems to be lower when discount factors are also lower, while the opposite holds for minimum wages. In the static model, a 40th percentile minimum wage decreases machine usage by half, while the effects are barely noticeable when the discount rate is at 0.98.
Table A1: Prices and wages

<table>
<thead>
<tr>
<th></th>
<th>Δ Cutting machines (t-1)</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ log(Output) (t)</td>
<td></td>
<td>-0.020</td>
<td>-0.020</td>
</tr>
<tr>
<td>Δ log(Miners) (t)</td>
<td></td>
<td>-0.080</td>
<td>0.033</td>
</tr>
<tr>
<td>Δ log(Helpers) (t)</td>
<td></td>
<td>-0.033</td>
<td>0.046</td>
</tr>
<tr>
<td>Δ log(Miner wage) (t)</td>
<td></td>
<td>-0.012</td>
<td>0.007</td>
</tr>
<tr>
<td>Δ log(Price) (t)</td>
<td></td>
<td>-0.006</td>
<td>0.015</td>
</tr>
</tbody>
</table>

**Notes:** I regress the difference of log output, miners, helpers, wages, and prices on the lagged difference in cutting machines at the town-year level.
Table A2: Transport costs

<table>
<thead>
<tr>
<th>log(Transport cost)</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Distance to Chicago)</td>
<td>-0.019</td>
<td>0.040</td>
</tr>
<tr>
<td>log(Distance to St. Louis)</td>
<td>0.301</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Observations: 2,696
R-squared: 0.102

Notes: Standard errors clustered at county level.
### Table A3: Wage cross-elasticity

<table>
<thead>
<tr>
<th></th>
<th>log(Wage)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td># Machines in own village</td>
<td>-0.139</td>
<td>0.069</td>
</tr>
<tr>
<td># Machines in other villages in county</td>
<td>0.012</td>
<td>0.016</td>
</tr>
<tr>
<td>Observations</td>
<td>2,253</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.324</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Unit of analysis is village-year level.
Table A4: Cost dynamics

<table>
<thead>
<tr>
<th></th>
<th>log(Miner productivity)</th>
<th>log(Helper productivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td>log(Cum. output)</td>
<td>-0.0039</td>
<td>0.0028</td>
</tr>
<tr>
<td>Observations</td>
<td>798</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0022</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Unit of analysis is village-year level.
Table A5: Miner supply elasticity: correlations

<table>
<thead>
<tr>
<th>log(Miner employment share)</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Total miners / population)</td>
<td>0.097</td>
<td>0.017</td>
</tr>
<tr>
<td>log(Afro-Americans)</td>
<td>0.006</td>
<td>0.012</td>
</tr>
<tr>
<td>log(Foreign-born)</td>
<td>-0.060</td>
<td>0.041</td>
</tr>
<tr>
<td>log(Population)</td>
<td>0.045</td>
<td>0.073</td>
</tr>
<tr>
<td>log(Area)</td>
<td>0.029</td>
<td>0.158</td>
</tr>
<tr>
<td>log(Mfg. workers)</td>
<td>-0.003</td>
<td>0.026</td>
</tr>
<tr>
<td>log(Mfg. plants)</td>
<td>0.051</td>
<td>0.042</td>
</tr>
<tr>
<td>log(Farmland area)</td>
<td>-0.005</td>
<td>0.105</td>
</tr>
<tr>
<td>1(Agricultural scarcity)</td>
<td>-0.021</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Observations 1,658
R-squared 0.231

Notes: Standard errors clustered at county level.
<table>
<thead>
<tr>
<th>Year</th>
<th>1884</th>
<th>1886</th>
<th>1888</th>
<th>1890</th>
<th>1892</th>
<th>1894</th>
<th>1896</th>
<th>1898</th>
<th>1900</th>
<th>1902</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output quantities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lump</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mine run</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pea</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slack</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shipping or local mine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Shipping quantities</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Input quantities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miners, winter</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miners, summer</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Miners, avg entire year</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Miners, max entire year</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Other employees</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Other employees, underground</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Other employees, above ground</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Other employees winter</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Other employees summer</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Boys employed underground</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mules</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Days worked</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Kegs powder</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Men killed</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Men injured</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Capital (in dollar)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### Table A7: Variables per year, continued

<table>
<thead>
<tr>
<th>Year</th>
<th>1884</th>
<th>1886</th>
<th>1888</th>
<th>1890</th>
<th>1892</th>
<th>1894</th>
<th>1896</th>
<th>1898</th>
<th>1900</th>
<th>1902</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output price</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price per ton at mine</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Price per ton of lump at mine</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Input prices</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer wages per ton</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Winter wages per ton</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Wage hand miners</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Wage machine miners</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Day or piece rate</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Payment frequency</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Net/gross wage</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Oil price</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Technicals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mine type (drift, shaft, slope)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hauling technology</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Depth</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Thickness</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Geological seam type</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Longwall or PR method</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Number egress places</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ventilation type</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>New/old mine</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Number of acres worked</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mine capacity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mined or blasted</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Machine usage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine dummy</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Number of tons cut by machines</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Number of machines, by type</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Figure A1: Mine and machine locations

Notes: The dots represent mining towns, each of which can contain multiple mines. Villages with squares contain at least one machine mine.
Figure A2: Returns to scale

(a) Miners

(b) Helpers
Figure A3: Cost dynamics

(a) Mine depth

(b) Vein thickness
Figure A4: Harrison Cutting Machine
Figure A5: Markups and markdowns
Figure A6: Counterfactuals with various discount rates

(a) $\delta = 0$ (static model)

(b) $\delta = 0.925$

(c) $\delta = 0.95$ (baseline)

(d) $\delta = 0.975$
Figure A7: Technical change and monopsony with endogenous output

(a) Markup = 1

(b) Markup = 1.5

(c) Markup = 2

(d) Markup = 2.5