Monopsony power and factor-biased technology adoption

Michael Rubens*

November 19, 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 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\(^2\) affects the incentives of buyers to adopt new production technologies. The focus of the paper lies on factor-biased technologies, as these affect demand on different input markets differently. If a firm has monopsony power on an input market, it earns profits from the difference between the marginal product and price of that input. If adopting a technology lowers demand for that input, the profit extracted from this wedge falls, which decreases the incentive to adopt such a technology. The opposite holds if technology adoption increases demand for the input. This is illustrated in figure 1, for an input \(L\) with price \(W_L\), and a technology that shifts input demand down from \(D_1\) to \(D_2\). If the input market is competitive, in panel (a), variable profits increase by the striped area, which corresponds to the cost saving. If the firm is a monopsonist, it earns profits from the wedge between the marginal cost curve \(MC\) and the price \(W\). An increase in productivity reduces costs by the striped area, but also reduces the total wedge between the marginal input product and input price by the dotted area. Although the total profit, which is the difference between the striped and dotted area, still increases, this increase is clearly smaller compared to the competitive setting.

I formalize this idea using a model of input supply and demand with a monopsonistic producer that decides on the adoption of a factor-biased production technology. The model is stylized in the sense that both goods and inputs are homogeneous, that returns to scale are constant, and that both product quantities and prices are exogenous. The input demand side builds on the seminal models of Acemoglu (2003) and Antras (2004), but with the key distinction that input prices are now endogenous from the perspective of individual firms. I find that monopsony power over the factor on which a technology saves decreases the incentive to adopt the technology, and that the opposite holds if absolute demand for the input increases due to the technology.

I use this model to 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 replaced manual labor in the cutting process. The mine-level data tracks both the usage of these machines and input and

\(^{1}\)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
output quantities and prices over a period of 18 years. Secondly, 19th century coal mining towns are a textbook example of oligopsonic 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 empirical model 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. Secondly, I derive the labor demand functions for both types of workers using 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 because the cutting machines were a durable investment that carried over through time. 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.4

I use the estimated model to conduct two counterfactual exercises. First, I 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 13.9%, when moving to the competitive equilibrium. As a consequence, average miner productivity would have increased by 1.2% 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.

These findings 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

4A similar identification approach was used by Peters et al. (2017).
four decades, technological change is believed to have mainly ‘hollowed out’ the center of 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.\footnote{Katz and Margo (2014) argue this also held for technical change during the second industrial revolution.}

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 Biesenbroeck, 2016; Igami & Uetake, 2017). In contrast to this literature, I focus on the effect of buyer power rather than seller power, on innovation.\footnote{Especially 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.} 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.\footnote{In their study of tomato harvesters, Just and Chern (1980) examine how oligopsony power of buyers affects technology adoption of their suppliers, and the same holds for Huang and Sexton (1996); Kühler and Rammer (2012). I focus, in contrast, on technology adoption by the buyers.} This paper is also related to the ‘induced innovation’ hypothesis of Hicks (1932).\footnote{In the existing literature, relative aggregate input prices change due to the general equilibrium effects of factor-biased technical change.} 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 relates, finally, to work on the identification of production functions with non-scalar productivity residuals, as in Doraszelski and Jaumandreu (2017); Demirer (2020). In contrast to these papers, I allow for endogenous input prices.\footnote{Dechezleprêtre, Héamous, 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 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.\(^{10}\) Railway firms were

\(^{10}\)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).
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. Appendix C.1 provides some more empirical evidence for this.

**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{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.} 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{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.}

**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 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.\footnote{Some unskilled workers eventually became skilled, such as boys who started out as ‘slatepickers’ who sorted coal, but became miners as they aged.}
**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 salary\(^{14}\) 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.\(^{15}\) 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 A2. 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.\(^{16}\) 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 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.

**Unionization** The unionization of the Illinois coal mining industry started around the 1860, and the Knights of Labor were a first attempt to form a union. These initiatives were largely unsuccessful in their attempts to raise wages (Boal, 2017). In 1886, 15% of mine workers were member of a trade union. The first succesful labor union in Illinois was the United Mine Workers of America, founded

---

\(^{14}\)That is, net of material costs and other work-related expenses.

\(^{15}\)Simultaneously, prototypes of mechanical coal cutting machines were invented in Northern England in the late 1870s (Reid, 1876; Ackermann, 1902).

\(^{16}\)Only 9 mines used a cutting machine in 1884.
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. Due to their level of violence, these strikes became known as the ‘Illinois coal wars’. 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\(^1\) (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 per miner increased, output per helper did not

The evolution of the ratio of total coal output over total miner-days and helper-days is in panel(a) of figure 3. Daily output per miner increased from 2.5 to 3.3 ton between 1884 and 1902, and increase of 31%. Daily output per helper fell, however, from 9.11 to 8.42 ton over that same time period. Miners were, hence, being replaced by helpers. This increase in miner productivity translated into lower miner piece rates and coal prices. Panel (b) of figure 3 plots the evolution of total revenue and the total miner wage bill divided by total coal output. Miner wages per ton fell by 40% from 1884 to 1898, and then increased again between 1898 and 1902. Wages per ton were 30% lower in 1902 compared to 1884. The aggregate coal price per ton fell from $1.25 to $0.88 between 1884 and 1896, and then increased to $1.97 by 1902. This increase was partly driven by increased aggregate demand for coal, as coal prices also rose outside of Illinois.

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

\(^1\)This was the equivalent of six average monthly miner wages
Panel (a) of figure 4 plots the total number of miner-days as a share of total employment over time. The blue line with squares reports this ratio for mines that adopted a cutting machine throughout the time period, the red line connecting diamonds plots it for the other mines. The aggregate ratio of miners to helpers was 3.59 for both groups of mines in 1884, and decreased to 2.42 in the mechanized mines, and 2.77 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.

Panel (b) plots the ratio of total output over all workers for both groups of mines. Machine mines used less workers to mine a ton of coal, but this was already the case in 1884. Output per worker rose in both mechanized and non-mechanized mines, and even more so non-mechanized mines. These correlations cannot simply be interpreted as picking up the causal effect of cutting machines on productivity, as both Hicks-neutral and factor-specific productivity are likely to change the demand for machines by mines. The evidence does suggest, however, that the main 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 highly seasonal: during cold winters, residential coal demand increased compared to warm summers. Panel (a) of figure 5 As there was a time lag between coal extraction and sales in the final market, total coal extraction in Illinois was high from August to February, at a monthly average of 11 Kton, and low between March and July, at a monthly average 8.5 Kton. This is plotted in panel (a) of figure 5. As a result, demand for both miners and helpers was also high during fall and winter, and low during spring and summer, as shown in panel 5(b).\(^{18}\) Panel 5(c) shows that both coal prices and helper wages 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.

### 2.3 Data

**Biennial 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, employee counts for both skilled and unskilled workers, days worked, intermediate inputs (black powder) in quantities,  

\(^{18}\)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.
Figure 5: Seasonality

(a) Coal extraction

(b) Employment

(c) Wages and prices

(d) Wage-employment profile

Notes: Panel (a) reports average employment per month for both worker types, panel (b) plots daily wages per month, and panel (c) the correlation between monthly employment and wages. All three panels are based on a sample of firms in 1890.

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). 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.

Seasonal and monthly data 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.).

Additional data 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 A model of technology adoption by labor monopsonists

I set up and estimate a model of labor and capital supply and demand. The model features static labor supply and static labor demand by monopsonistic firms, and a dynamic adoption model of a factor-biased technology. I rely on a Constant Elasticity of Substitution production function; a more parsimonious model that uses a Cobb-Douglas production function is in appendix B.1.

3.1 Environment

Production function Firms $f$ extract a homogeneous product $Q_{f,t}$, coal. The production function is given by equations (1a)-(1c).\(^\text{19}\) The firm uses two type of workers $\tau \in \{s, u\}$: skilled and unskilled workers, which correspond to miners and helpers in the coal application. The number of worker-days used of each type is denoted $L_{\tau f,t}$, and they are paid daily wages $W_{\tau f,t}$. Both labor types are substitutable at an elasticity $\sigma$, and all intermediate inputs in $M_{f,t}$ are assumed to be a perfect complements to both labor types. Firms differ in terms of their factor-augmenting productivity terms $\omega_{\tau f,t}$. I assume there are constant returns to scale. Mines also differ in their intermediate input requirements $\omega_{m f,t}$.

$$Q_{f,t} = \min \left\{ \left( (\omega_{s f,t}^s L_{s f,t})^{\frac{\sigma - 1}{\sigma}} + (\omega_{u f,t}^u L_{u f,t})^{\frac{\sigma - 1}{\sigma}} \right)^{\frac{\sigma}{\sigma - 1}} ; \omega_{m f,t}^m M_{f,t} \right\} \quad (1a)$$

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 $\beta_0$. The variation in log relative miner productivity that is not explained by cutting machine

\(^{19}\)I discuss the relationship between monopsony power and technological change using the special case of a Cobb-Douglas production function in B.1.
adoption or time is denoted as the residual $\eta_{f,t}$.

$$\frac{\omega^g_{f,t}}{\omega^u_{f,t}} = \exp(\beta^k K_{f,t} + \beta^t t + \beta^0 + \eta_{f,t})$$  \hspace{1cm} (1b)

I assume that the residual variation in the logarithm of skill-augmenting productivity evolves following the AR(1) process in equation (1c):

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

The productivity evolution assumed by equation (1c) rules out a number of dynamic cost mechanisms that are suggested in the natural resource extraction literature. Productivity could, for instance, depend on past cumulative output, if it becomes increasingly costly to operate deeper mines, as Aguirregabiria and Luengo (2017) find for copper mining. Another reason why this may hold is learning by doing as in Benkard (2000). In all these cases, equation (1c) would be violated. I discuss cost dynamics in appendix D.2, and find them not to be of first-order importance in this industry.

**Variable profits** The mine-gate price of coal at mine $f$ in year $t$ is $P_{f,t}$. Materials, such as black powder, were paid for by miners, so they do not enter the firm’s profit function. Variable profits are hence equal to $\pi_{f,t} \equiv P_{f,t} Q_{f,t} - W^s_{f,t} L^s_{f,t} - W^u_{t} L^u_{f,t}$. I abstract from entry and exit of firms into and out of the industry.

**Coal markets** In line with the evidence presented in section 2, I assume that coal markets are perfectly competitive: individual firms cannot influence coal prices. There is one market-clearing price per year $P^*_t$, at which supply and demand for coal in St. Louis and Chicago are equal to each other. The mine-gate price at each mine is equal to the price in the destination market, minus a transport cost $\theta_{f,t}$. In appendix C.4, I show that transport costs are correlated with the distance between the mine and both St. Louis and Chicago, as expected.

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

The market clearing price $P^*_t$ is therefore determined by the coal operator with the highest marginal cost $\lambda_{f,t}$, all other coal mines have a positive markup. I assume that capacity is fixed on the short term: firms can therefore sell up to $Q_{f,t}$ tons of coal at price $P_{f,t}$, but not more. 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. They are therefore included in appendix C.3.
3.2 Labor market

**Labor supply** The inverse supply curve for both types of workers is given by equation (2). Each firm faces a supply curve with a constant wage elasticity $\psi_{f,t}^{\tau}$ for workers of type $\tau$. The term $\alpha_{f,t}^{\tau}$ captures any firm characteristic that shifts labor supply, other than wages. Examples of such characteristics can be working conditions, the firm’s location, etc. If the labor supply curve is flat, meaning that $\psi_{f,t}^{\tau} = 1$, 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 a different industry than coal mining, and it is the same for both skilled and unskilled workers. In other words, miner skills are only valuable in the coal mining industry.

$$W_{f,t}^{\tau} = \left(\alpha_{f,t}^{\tau} L_{f,t}^{\tau}\right)^{\psi_{f,t}^{\tau}-1} W^*_t$$

Equation (2) implies that the wage elasticity of the supply curve for worker type $\tau$ is a firm-year-specific constant $\psi_{f,t}^{\tau} = \frac{\partial W_{f,t}^{\tau}}{\partial L_{f,t}^{\tau}} L_{f,t}^{\tau} W_{f,t}^{\tau} + 1$. 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. The classical monopsony model hence applies: in order to hire an additional worker, a mine therefore did not merely need to raise the wage of this additional worker, but of all existing workers as well. Miner supply also increases non-wage mine characteristics $\alpha_{f,t}^{s}$, which includes unobservables such as working conditions. These unobservables can be choice variables of the mines.

I assume that firms colluded in their own village, but that these local labor markets are isolated. This means that both the labor and capital demand models are single-agent optimization problems: the labor and capital decisions of other firms outside the same village have no effect on the own equilibrium wages.\(^{20}\)

**Labor demand** Let both labor inputs $L_{f,t}^{\tau}$ be flexibly adjustable and chosen at time $t$. As will be discussed later, the capital stock at time $t$ is assumed to be chosen at time $(t-1)$, and hence exogenous when firms choose their labor input mix. I assume firms choose both labor inputs to maximize current variable profits: $\max_{L_{f,t}^{s},L_{f,t}^{u}} \pi_{f,t}$. Solving the first order conditions, the demand for labor type $\tau$ is given by equation (3).

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

\(^{20}\)I provide evidence for this in appendix C.2.
The supply elasticities $\psi^\tau$ negatively affect input demand: firms internalize the fact that higher employment leads to higher wages. They be interpreted as the wedge between the marginal product of a worker and his/her wage: $\frac{\partial Q^\tau_{f,t}}{\partial L^\tau_{f,t}} = \psi^\tau_f W^\tau_f$. Now consider the effects of an increase in $\omega^s$, 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 $L^s$ compared to $L^u$. If they are gross substitutes, meaning that $\sigma > 1$, the opposite holds (Acemoglu, 2003).

**Monopsonistic equilibrium**  The equilibrium labor quantities and wages can be found by solving the system of equations (3) and (2). The reduced-form expressions for the labor quantities and wages are in equation (4):

$$
\begin{align*}
W^\tau_{f,t} &= \left[ Q_{f,t} P^\sigma_{f,t} (\omega^\tau_{f,t})^{\sigma-1} (\psi^\tau_{f,t})^{-\sigma}]^{\psi^\tau_{f,t,-1}} (\alpha_{f,t}) \right]^{(1+\sigma)(\psi^\tau_{f,t,-1})^{-1}} (W^*_f) \left( \frac{1}{1+\sigma(\psi^\tau_{f,t,-1})} \right) \\
L^\tau_{f,t} &= \left[ Q_{f,t} P^\sigma_{f,t} (\omega^\tau_{f,t})^{\sigma-1} (\psi^\tau_{f,t})^{-\sigma}]^{\psi^\tau_{f,t,-1}} (\alpha_{f,t}) \right]^{(1+\sigma)(\psi^\tau_{f,t,-1})^{-1}} (W^*_f) \left( \frac{1}{1+\sigma(\psi^\tau_{f,t,-1})} \right)
\end{align*}
$$

(4)

**Competitive equilibrium** So far, it was assumed that firms maximize profits, which implies a wedge of $\psi^\tau_f$ between the wage and marginal product of input $\tau$. Now suppose that firms do not maximize profits, but choose the amount of inputs that drives a wedge $\tilde{\psi}^\tau \in [1, \psi]$ between marginal products and wages. The monopsonistic equilibrium corresponds to $\tilde{\psi} = \psi$, while the competitive equilibrium corresponds to $\tilde{\psi} = 1$.

$$
\frac{\partial Q^\tau_{f,t}}{\partial L^\tau_{f,t}} = \tilde{\psi}^\tau_f W^\tau_{f,t}
$$

The counterfactual equilibrium that corresponds to a wedge $\tilde{\psi}$ is denoted $(\tilde{L}^\tau_{f,t}, \tilde{W}^\tau_{f,t})$:

$$
\begin{align*}
\tilde{W}^\tau_{f,t} &= \left[ Q_{f,t} P^\sigma_{f,t} (\omega^\tau_{f,t})^{\sigma-1} (\tilde{\psi}^\tau_{f,t})^{-\sigma}]^{\psi^\tau_{f,t,-1}} (\alpha_{f,t}) \right]^{(1+\sigma)(\psi^\tau_{f,t,-1})^{-1}} (W^*_f) \left( \frac{1}{1+\sigma(\psi^\tau_{f,t,-1})} \right) \\
\tilde{L}^\tau_{f,t} &= \left[ Q_{f,t} P^\sigma_{f,t} (\omega^\tau_{f,t})^{\sigma-1} (\tilde{\psi}^\tau_{f,t})^{-\sigma}]^{\psi^\tau_{f,t,-1}} (\alpha_{f,t}) \right]^{(1+\sigma)(\psi^\tau_{f,t,-1})^{-1}} (W^*_f) \left( \frac{1}{1+\sigma(\psi^\tau_{f,t,-1})} \right)
\end{align*}
$$

(5)

**Equilibrium variable profits** Inserting the equilibrium wages and labor quantities into the variable profit function yields the following expression for variable profits $\pi_{f,t}$:

$$
\pi_{f,t} = Q_{f,t} P_{f,t} - \sum_{\tau} \left[ Q_{f,t} P^\sigma_{f,t} (\omega^\tau_{f,t})^{\sigma-1} (\tilde{\psi}^\tau_{f,t})^{-\sigma}]^{\psi^\tau_{f,t,-1}} (\alpha_{f,t}) \right]^{(1+\sigma)(\psi^\tau_{f,t,-1})^{-1}} (W^*_f) \left( \frac{1}{1+\sigma(\psi^\tau_{f,t,-1})} \right)
$$

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 $\pi^1_{f,t}$ and $\pi^0_{f,t}$.
3.3 Capital market

The choice to adopt cutting machines at time $t$ is denoted $A_{f,t} \in \{0, 1\}$. Once firms choose to adopt a cutting machine, they become a mechanized mine forever, as shown in equation (6).\textsuperscript{21} If mine owners decide on adoption in time $t$, they can install a cutting machine in that same year: there is no adjustment time lag.

$$K_{f,t} = \max\{A_{f,t}, K_{f,t-1}\}$$  \hspace{1cm} (6)

**Capital supply**  Adopting a cutting machine requires the firm to pay an upfront sunk cost $W^K_{f,t}$. I assume that coal firms take sunk costs $W^K$ as given. This cost has three components. First, a part of sunk costs depend on observable characteristics $z_{f,t}$, such as mine depth or vein thickness, as these alter installation costs. A second component of fixed costs, $\nu_{f,t}$ is potentially serially correlated. Thirdly, random sunk cost shocks $\varphi_{f,t}$ 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:

$$W^K_{f,t} = \gamma z_{f,t} + \nu_{f,t} + \varphi_{f,t}$$

**Capital demand**  Total profits are, by definition, equal to variable profits minus the sunk cost of adopting a cutting machine:

$$\Pi_{f,t} \equiv \pi_{f,t} - W^K_{f,t} A_{f,t}$$

The variable profit gain from using a cutting machine, $(\pi^1_{f,t} - \pi^0_{f,t})$, can be computed using the estimated model of labor supply and demand. 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 costs, and hence, increased variable profits, so $(\pi^1_{f,t} - \pi^0_{f,t}) > 0$. Normalizing per-period profits when not using a cutting machine to zero, and denoting the discount factor as $\delta$, the value of using a machine net of $\varphi$ is given by $v^1_{f,t}$. This is the discounted variable profit gain from using a cutting machine, minus the sunk cost of doing so. 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 $(\pi^1_{f,t} - \pi^0_{f,t}) \frac{(1 - \delta)}{1 - \delta}$ is multiplied with

\textsuperscript{21}In reality, 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.
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, $\bar{E}_t V_{f,t+1}$:

\begin{equation}
v^0_{f,t} = \delta \bar{E}_t V_{f,t+1}
\end{equation}

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

\begin{equation}
\bar{V}_{f,t+1} = 0.577 + \ln \left( \exp(v^1_{f,t}) + \exp(v^0_{f,t}) \right)
\end{equation}

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

\begin{equation}
v^0_{f,t} = \delta (\bar{V}_{f,t+1} - e_t)
\end{equation}

I assume that every period, firms decide on machine adoption in order to maximize their discounted expected profit stream, as long as they do not yet have any cutting machines:

\begin{equation}
\max A_{f,t} \{v^1_{f,t+1}, v^0_{f,t+1}\}
\end{equation}

The adoption of the technology $K$ therefore boils down to a trade-off between the variable cost reduction due to decreased high-skill labor usage and the fixed costs associated with using the technology. For now, I assume there are no serially correlated latent fixed costs: $\nu_{f,t}$ is i.i.d. over time. I assume the shocks $\varphi^1$ and $\varphi^0$ are type-I extreme value distributed. The probability of using the technology in year $t$, $S^1_{f,t}$, is then equal to:

\begin{equation}
S^1_{f,t} = \frac{\exp(v^1_{f,t})}{\exp(v^0_{f,t}) + \exp(v^1_{f,t})}
\end{equation}

**Capital demand: caveats** 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 effects 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.

Intermediate inputs Intermediate inputs consist of black powder. Powder is purchased by the miners at a price $W_{t}^{M}$, which is exogenous to the miners. As intermediate inputs do not enter the firm’s cost function and enter the production function in fixed proportions, they do not play a role in the estimation of the model.

### 3.4 Comparative statics

Monopsony power and the returns to capital In section 5, I will use the estimated model to compute technology adoption in both the observed monopsonistic equilibrium, and a counterfactual competitive equilibrium. Such an exercise is necessary to know the amount by which changes in monopsony power change technology adoption. Some key results about the relationship between monopsony power and technology adoption can, however, be derived without estimating the full model. The relationship of interest is the effect of the level of monopsony power exerted by the firm, $\psi_{f,t} \in [1, \psi_{t}^{\tau}]$, which is:

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

The effect of capital adoption on variable profits depends on its effects on the factor-augmenting productivity levels, and on the effects of changes in productivity on variable profits, as expressed in equation (10a).

$$\frac{\partial \pi_{f,t}}{\partial K_{f,t}} = \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\sigma}} \frac{\partial \omega_{f,t}^{\sigma}}{\partial K_{f,t}} + \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\eta}} \frac{\partial \omega_{f,t}^{\eta}}{\partial K_{f,t}}$$

(10a)

The effects of an increase in $\tau$-augmenting productivity on variable profits is given by equation (10b). 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.

$$\frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^{\sigma}} = \frac{(1 - \sigma)\psi_{f,t}}{1 + \sigma(\psi_{f,t} - 1)} \left[ Q_{f,t} P_{f,t}^{\sigma}(\psi_{f,t})^{-\sigma} \left( (\psi_{f,t}^{\tau})^{(\sigma - 1)} \right)^{\frac{\psi_{f,t}}{\psi_{f,t}^{(\sigma - 1)}}} \left( \frac{\psi_{f,t}^{\tau}}{\psi_{f,t}^{(\sigma - 1)}} \right)^{\frac{\psi_{f,t}^{(\sigma - 1)}}{\psi_{f,t}^{(\sigma - 1)}}} \left( (\alpha_{f,t})^{(\psi_{f,t}^{\tau} - 1)} W_{t}^{\sigma} \right) \right]$$

(10b)

Taking the derivative of the profit effect of productivity shocks with respect to the degree of monopsony power leads to lemma 1, a proof of which is in appendix B.2.
Lemma 1 If firms exert more monopsony power over type-τ workers, then the effects of a change in τ-augmenting productivity on variable profits are smaller:

\[ \frac{\partial}{\partial \psi_{f,t}^\tau} \left( |\frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^\tau}| \right) < 0 \]

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 τ-augmenting productivity shock reduces demand for input τ. 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 τ-augmenting shock increases demand for that input, which decreases profits (by increasing costs), but less so if firms extract large profits from input τ because of monopsony power.

The effect of monopsony power on the returns to technology adoption can now be found by combining equations (10a), (10b), and (16), which leads to theorem 1. A proof is in appendix B.3.

Theorem 1 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:

\[ \frac{\partial}{\partial \psi_{f,t}^\tau} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}^\tau} \right) \begin{cases} < 0 \iff (\sigma - 1) \frac{\partial \omega_{f,t}^\tau}{K_{f,t}^\tau} \begin{cases} > 0 \end{cases} \end{cases} \]

A technology can decrease demand for input τ in two cases: either if it is τ-augmenting and inputs are gross complements, or if it augments the other input and they are gross substitutes. Monopsony power over input τ reduces the variable profit gain from adopting a technology if it saves on τ. 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), (3), and (9): labor supply, labor demand, and capital demand.

Unit of observation 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.\(^{22}\) The words ‘firm’ and ‘village’ hence mean that same 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 \( \tau = s \) gives equation (11).

\[
w^s_{f,t} - w^s_t = \left( \psi^s_{f,t} - 1 \right) l^s_{f,t} + \left( \psi^s_{f,t} - 1 \right) \ln(\alpha^s_{f,t})
\]

(11)

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, will be able to attract the same number of miners as other mines at a lower wage rate. In order to identify the slope of the miner supply curve, a shock to labor demand that is excluded from miner utility is 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 \( L^s_{f,WIN}, L^s_{f,SUM} \), and the corresponding daily miner wages as \( W^s_{f,WIN}, W^s_{f,SUM} \). The supply residuals during winter and summer are \( \alpha^s_{f,WIN} \) and \( \alpha^s_{f,SUM} \). 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: \( \alpha^s_{f,WIN} = \alpha^s_{f,SUM} \). Denoting logs as lowercases, the slope of the skilled labor supply curve can then be calculated using equation (12). This yields a different labor supply curvature for every firm and every time period.

\[
\psi^s_{f,t} = \frac{w^s_{f,WIN} - w^s_{f,SUM}}{l^s_{f,WIN} - l^s_{f,SUM}} + 1
\]

(12)

How realistic is the exclusion restriction \( \alpha^s_{f,WIN} = \alpha^s_{f,SUM} \)? The supply residual \( \alpha^s_{f,t} \) could differ

\(^{22}\)Because of this, the capital adoption model is single-agent: if labor markets would be oligopsonistic, the adoption decision at one firm would affect wages, and hence also machine usage decisions, at other firms in the same labor market.
across 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. The main evidence that supports the exclusion restriction comes, however, from the monthly wage profile of unskilled workers in figure 5. The wages of these workers did not change significantly between different months. 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) showed a very small dispersion in helper wages, which did not change in response to helper 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: $W_{f,t}^u = W_t^*$. 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 wages are only rigid for helpers, not miners, which is hard to reconcile without invoking differences in competition between these two markets. 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 (12). I do not observe the unskilled worker wage $W_t^*$ every year, but do so in 1890. I 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 $\tilde{W}_{f,t}^{WIN}$, $\tilde{W}_{f,t}^{SUM}$ into daily wages by multiplying them by the number of employees per worker: $W_{f,t}^{WIN} = \tilde{W}_{f,t}^{WIN} A^{WIN}$ and $W_{f,t}^{SUM} = \tilde{W}_{f,t}^{SUM} A^{SUM}$.

The firm characteristics $\alpha_{f,t}^H$ can be recovered by inverting the labor supply function, using the

---

23This would, however, be consistent with higher wages during the summer, while lower summer wages are observed.
estimated values for $\psi_{f,t}$ and $W^*$:

$$\alpha_{f,t}^s = \frac{(W_{f,t}^s)^{\frac{1}{1-\sigma}}}{L_{f,t}^s}$$

### 4.2 Labor demand

**Identification**  I divide the demand for miners by demand for helpers, from equation (3), and take logs. The relative demand for helpers vs. miners in logs is hence given by equation (13).

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

(13)

The coefficients that need to be identified are (i) the elasticity of substitution between miners and helpers, $\sigma$, and (ii) the determinants of miner vs. helper productivity, $\beta$. 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 relative labor demand function, I make combine the assumptions made about the timing of the decisions on labor and capital, as in Olley and Pakes (1996); Ackerberg et al. (2015), with the linear equation of motion for relative worker productivity in equation (1c). I follow Blundell and Bond (2000) by taking $\rho$-differences of the log labor demand function:

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

I denote the right-hand side of the equation above as $\zeta_{f,t} - \rho\zeta_{f,t-1}$. 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 transient shock $\varepsilon_{f,t}$ is observed. Capital choices at time $t$ are therefore orthogonal to the shock $\varepsilon$. Secondly, while miner wages $W_{f,t}^s$ change in response to productivity shocks, the slope of the miner supply curve $\psi_{f,t}^s$ does not, as the labor supply curve is assumed to be linear. The slope $\psi_{f,t}^s$ should therefore also be orthogonal to the productivity shock $\varepsilon$. The moment conditions to identify $\sigma$ and $\beta$ are hence given by in equation (14):

$$E \left[ \zeta_{f,t} - \rho\zeta_{f,t-1} | K_{f,t}, \psi_{f,t} \right] = 0$$

(14)
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 (3) using GMM, with the moment conditions mentioned earlier. The fitted log markdown estimates $\psi_{s,t}$ from the labor supply function are added to log wages on the right-hand side. As I only observe markups for the first four time periods (8 years), the sample size is reduced to 405. 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 we know the estimates of $\sigma$ and $\beta$, $\hat{\sigma}$ and $\hat{\beta}$, we can impute miner- and helper-augmenting productivity as follows:

$$
\hat{\omega}^u_{f,t} = Q_{f,t} \left[ \exp \left( \hat{\beta}^k K_{f,t} + \hat{\beta}^t t + \hat{\beta}^0 \right) \right] \frac{\sigma - 1}{\sigma} \left( L^u_{s,f,t} \right)^{\frac{\sigma - 1}{\sigma}} + \left( L^u_{s,f,t} \right)^{\frac{\sigma - 1}{\sigma}} \frac{1}{\sigma - 1} - \sigma
$$

$$
\hat{\omega}^s_{f,t} = Q_{f,t} \left[ \exp \left( \hat{\beta}^k K_{f,t} + \hat{\beta}^t t + \hat{\beta}^0 \right) \right] \frac{\sigma - 1}{\sigma} \left( L^u_{s,f,t} \right)^{\frac{\sigma - 1}{\sigma}} + \left( L^u_{s,f,t} \right)^{\frac{\sigma - 1}{\sigma}} \frac{1}{\sigma - 1} \exp \left( \hat{\beta}^k K_{f,t} + \hat{\beta}^t t + \hat{\beta}^0 \right)
$$

4.3 Capital demand

**Identification** Finally, we need to identify the fixed cost determinants in the capital demand model, equation (9). 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.

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

Substituting this into equation (9) results in the following estimable equation:

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

Given that the variable profit gain from cutting machines, $\pi^1_{f,t} - \pi^0_{f,t}$, the discount factor $\delta$, and the empirical choice probabilities $s_{f,t}$ are known, the only coefficients that need to be identified are the sunk cost parameters $\gamma$. I include a series of technical mine characteristics as fixed cost shifters.
x, 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, and cutting machines were more useful when coal veins were thicker. Secondly, I control for mine size: it may again be harder to install machinery in mines that operate at a larger scale, as the coal seam is further away in that case (even when controlling from depth). Thirdly, I control for the mine type (shaft, slope, drift, or surface mine). Mines using vertical shafts had to transport cutting machines using elevators, which was hard. I control for the hauling and ventilation technologies used; if these were mechanized, some costs such as electricity generators may have been shared between technologies. I include the county’s surface to proxy for the extent to which mines were isolated, and a linear time trend to capture changes in common fixed costs over time. I assume all these fixed cost shifters z are exogenous at the time of choosing cutting machines: firms do not choose how deep or thick the coal seam is. 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 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 $S^1$ and $S^0$ to construct the left-hand side of equation (9). 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 was at the same level during the time period 1892-1902 as it was on average during the time period 1884-1890. 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 assume that I set the annual discount factor to 0.95.24 The biennial interest rate is hence equal to $\delta = (0.95)^2$.

The capital demand model gives an estimated machine adoption probability $S^1_{f,t}$ 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 using the transition equation (6).

### 4.4 Results

**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 markdown ratio $\psi^H_{f,t}$ is 1.138, so the marginal product of skilled miners lies on average 14% above their wage (i.e., that wages are marked down by 12%). The average markdown ratio is significantly higher than one, so mining firms have

---

24 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.
monopsony power over miners. The distribution of markdowns is plotted in appendix figure A3. Markdowns lie mostly between 1 and 1.5. Appendix table A5 provides some descriptive analysis by regressing the labor supply elasticity 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 indicator of ‘agricultural labor scarcity’ in each village, as reported in the Bureau of Labor Statistics’ reports. This suggests that there is a low substitutability between coal mining and both manufacturing and agriculture for skilled miners.

**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 are 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 62%.\(^{25}\) Combining the findings that both worker types are gross complements and that cutting machines are skill-augmenting implies that cutting machines are unskill-biased: they decrease 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 unit changes sunk costs by $\gamma$ dollars. Sunk costs increase with size, vein thickness, and mine depth. The other technical characteristics do not significantly alter sunk costs. The average sunk cost of mechanizing a village is estimated at $203K.\(^{26}\) 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. Panel (a) of figure 7 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

\(^{25}\) $= \exp(0.502) - 1$

\(^{26}\) This is equal to US$5.8M in 2020
## Table 1: Model estimates

(a) **Miner supply**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miner supply elasticity $\psi_s$</td>
<td>1.138</td>
<td>0.007</td>
</tr>
</tbody>
</table>

(b) **Miner demand**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of substitution $\sigma$</td>
<td>0.077</td>
<td>0.026</td>
</tr>
<tr>
<td>Miner-augmenting effect of machines $\beta^k$</td>
<td>0.502</td>
<td>0.158</td>
</tr>
<tr>
<td>Time $\beta^t$</td>
<td>0.084</td>
<td>0.020</td>
</tr>
<tr>
<td>Constant $\beta^0$</td>
<td>-159.261</td>
<td>37.427</td>
</tr>
<tr>
<td>Serial correlation $\rho$</td>
<td>0.451</td>
<td>0.046</td>
</tr>
</tbody>
</table>

(c) **Sunk capital costs**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>22.018</td>
<td>0.000</td>
</tr>
<tr>
<td>log(Output)</td>
<td>3.029</td>
<td>1.238</td>
</tr>
<tr>
<td>log(Dist. to St. Louis)</td>
<td>-0.405</td>
<td>1.154</td>
</tr>
<tr>
<td>log(Thickness)</td>
<td>62.822</td>
<td>21.130</td>
</tr>
<tr>
<td>log(Depth)</td>
<td>23.095</td>
<td>5.716</td>
</tr>
<tr>
<td>log(Dist. to Chicago)</td>
<td>-5.114</td>
<td>3.024</td>
</tr>
<tr>
<td>log(K/L) in manufacturing</td>
<td>-1.347</td>
<td>2.192</td>
</tr>
<tr>
<td>1(01Drift mine)</td>
<td>2.396</td>
<td>38.256</td>
</tr>
<tr>
<td>1(Other type mine)</td>
<td>692.521</td>
<td>4637.029</td>
</tr>
<tr>
<td>1(01Shaft mine)</td>
<td>1.685</td>
<td>38.604</td>
</tr>
<tr>
<td>1(01Slope mine)</td>
<td>-1.443</td>
<td>39.013</td>
</tr>
<tr>
<td>1(Furnace ventilation)</td>
<td>-3.871</td>
<td>4.266</td>
</tr>
<tr>
<td>1(Natural ventilation)</td>
<td>-4.635</td>
<td>3.982</td>
</tr>
<tr>
<td>1(Manual hauling)</td>
<td>-3.776</td>
<td>34.659</td>
</tr>
</tbody>
</table>

**Notes:** Standard errors are block-bootstrapped with 50 iterations. Estimates in panel (c) are in units of $1000$ US$: an increase in a right-hand variable of one unit results in a change in sunk costs of $\gamma^{*}$1000. The average sunk cost is estimated to be $515,249.$

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.

5 Counterfactuals

With the estimated model at hand, I now examine how cutting machine adoption, productivity, employment and wages would change in two counterfactual scenarios. In section 5.1, I consider the effects of moving from monopsonistic to perfectly competitive miner markets. In section 5.2, I study the effects of the introduction of a state-wide minimum wage in Illinois.

5.1 Perfectly competitive labor markets

Monopsonistic vs. competitive equilibrium The estimated model in the previous section describes the real world, which was assumed to be the monopsonistic equilibrium. I collect the outcome variables of interest; labor quantities and prices, productivity, and variable profits, into a vector \( x_{1f,t} = (l_{s,1f,t}, l_{u,1f,t}, w_{s,1f,t}, \pi_{1f,t}, \omega_{s,1f,t}, \omega_{u,1f,t}) \) if a firm uses cutting machines, and \( x_{0f,t} = (l_{s,0f,t}, l_{u,0f,t}, w_{s,0f,t}, \pi_{0f,t}, \omega_{s,0f,t}, \omega_{u,0f,t}) \) if it does not. The fitted values of these variables using the model, i.e. in the monopsonistic equilibrium, are denoted as \( \hat{x}_{1f,t} \) and \( \hat{x}_{0f,t} \). The estimated values for all the variables in \( x_{f,t} \) is hence their expected value, weighted by the probability of using cutting machines \( \hat{K}_{f,t} \):

\[
\hat{x}_{f,t} = \hat{K}_{f,t} \hat{x}_{1f,t} + (1 - \hat{K}_{f,t}) \hat{x}_{0f,t}
\]

The competitive equilibrium can be computed by setting the parameter \( \tilde{\psi}_{s,f,t} \) in the miner demand function (5) to one. This corresponds to the intersection between supply and demand, rather than between the marginal cost and demand curves. Graphically, this implies moving from point A to B in figure 6a. I denote the values for \( (x_{0f,t}, x_{1f,t}) \) in the competitive equilibrium as \( (\tilde{x}_{0f,t}, \tilde{x}_{1f,t}) \).

Assumptions In order for this counterfactual exercise to be valid, a number of variables and coefficients are assumed to be invariant between the monopsonistic and competitive equilibrium: \( \omega^{s,1}, \omega^{u,1}, \omega^{s,0}, \omega^{u,0}, W^u, z, \alpha^s, \alpha^u, \beta, \gamma, \delta, \sigma \). 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.

Machine usage in competitive equilibrium The long-run variable profits derived from using cutting machines and not doing so in the competitive equilibrium, \( \tilde{v}_{1f,t} \) and \( \tilde{v}_{0f,t} \), can be calculated using the counterfactual variables in \( (x_{0f,t}, x_{1f,t}) \) and using the value functions (7) and (8). Plugging these
Figure 6: Counterfactuals: theory

(a) Equilibria

\[ \psi = 1 \quad B \]
\[ \psi = \psi_A \]
\[ C \]
\[ I = 1 \]
\[ I = 2 \]
\[ I = 3 \]

(b) Minimum wage

\[ W_{min} \]

\[ \hat{W}_s \]
\[ W^\prime_2 \]
\[ D_1^s \]
\[ D_2^s \]
\[ S^s \]
\[ MC^s \]
\[ L^s \]
\[ W^s \]

\[ S^s \]
\[ D^s \]
\[ W^s \]
\[ L^s \]

into the estimated capital demand function, (15), gives the estimated machine adoption rates in the competitive equilibrium:

\[ \tilde{S}_{j,t}^1 = \frac{\exp(\tilde{v}_{j,t}^1)}{\exp(\tilde{v}_{j,t}^0) + \exp(\tilde{v}_{j,t}^1)} \]

Using these adoption rates and the capital transition equation (6), the machine usage probability in the competitive equilibrium, \( \tilde{K}_{f,t} \), can be calculated. The counterfactual labor quantities, prices, productivities and profits \( \tilde{x}_{f,t} \) can now be found using the counterfactual machine adoption probabilities as weights:

\[ \tilde{x}_{f,t} = \tilde{K}_{f,t} \tilde{x}_{j,t}^1 + (1 - \tilde{K}_{f,t}) \tilde{x}_{j,t}^0 \]

Results 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. Table 3 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.\textsuperscript{27} 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.

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>0.007</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.903</td>
<td>0.068</td>
<td>0.018</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8592.386</td>
<td>25.831</td>
<td>73.387</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 at p40</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>0.022</td>
</tr>
<tr>
<td>Miner productivity</td>
<td>4.834</td>
<td>4.740</td>
<td>-0.094</td>
<td>0.048</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8220.798</td>
<td>-345.757</td>
<td>422.701</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.736</td>
<td>0.213</td>
<td>0.021</td>
</tr>
</tbody>
</table>

5.2 Minimum wage policy

Minimum wages are the usual policy measure in response to monopsony power on labor markets. In this counterfactual, I examine how the introduction of a state-wide minimum wage in Illinois would have affected equilibrium employment, miner wages, machine adoption, and miner productivity.

\textsuperscript{27}The 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.
Table 3: Counterfactuals with exogenous machine usage

(a) Competitive labor markets

<table>
<thead>
<tr>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.112</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8674.859</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.530</td>
</tr>
</tbody>
</table>

(b) Minimum wage at p40

<table>
<thead>
<tr>
<th>Reality</th>
<th>Counterfactual</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutting machine usage</td>
<td>0.112</td>
<td>0.112</td>
</tr>
<tr>
<td>Miner-days</td>
<td>8566.555</td>
<td>8176.489</td>
</tr>
<tr>
<td>Miner daily wage</td>
<td>1.523</td>
<td>1.733</td>
</tr>
</tbody>
</table>

Notes: Standard errors bootstrapped, 50 iterations

Figure 7: 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.

Counterfactual employment and wages In order to assess the effects of a minimum wage $W_{min}^t$, 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 $I_{k,t}^f$ indicates this relative position of the minimum wage compared to the equilibrium miner wage at firm $f$, with the index $k \in 0, 1$ indicating the usage of cutting machines. The minimum wage does not bind for firm $f$ if the monopsonistic equilibrium wage $W_{k,t}^f$ lies above the minimum wage, in which case $I_{k,t}^f = 1$. If the minimum wage lies above the monopsonistic equilibrium wage, but below the competitive wage, $I_{k,t}^f = 2$. Finally, $I_{k,t}^f$ takes the...
value of three if the minimum wage lies above the competitive equilibrium wage.

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

The red line in figure 6(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, and 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 $\bar{x}$. Translating this into equations, the counterfactual miner employment levels under a minimum wage are given by:

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

**Counterfactual machine usage** The counterfactual variable profits under the minimum wage, $\bar{\pi}_{f,t}^1$ and $\bar{\pi}_{f,t}^0$, can now be solved for as well. Figure 6(b) shows how a minimum wage alters the variable profit effects of machine adoption. Suppose cutting machines reduce miner demand from curve $D_{1}^s$ to $D_{2}^s$, and that the minimum wage is initially not binding, meaning that $I^0 = 1$. When there is no minimum wage, variable profits become the red area, as the wage drops to the level $\hat{W}_{2}^s$. With a binding minimum wage, however, the wage level is fixed at $W_{t}^{min}$. The variable profit after machine adoption under the minimum wage counterfactual, $\bar{\pi}^1$, corresponds to the blue area. The blue area is smaller than the red 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 $\bar{S}_{f,t}^1$ can be computed by plugging the counterfactual variable profits into equation (9), as was done previously in section 5.1:

$$\bar{S}_{f,t}^1 = \frac{\exp(\bar{v}_{f,t}^1)}{\exp(\bar{v}_{f,t}^0) + \exp(\bar{v}_{f,t}^1)}$$
Similarly to the first counterfactual exercise, the equilibrium capital usage probabilities, labor quantities and prices, and profits, can be calculated as follows:

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

**Figure 8: Minimum wage levels**

(a) Machine usage  
(b) Employment and wages

**Results**  In panel (b) of table 2, I compare the outcome variables of interest between reality and a world under a minimum wage that 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, but 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 were exogenous. As shown in panel (b) of table 3, employment would fall by more and wages increase by less in that case. The decrease in miner-saving machine usage dampens the drop in employment and rise in wages after a raise of the minimum wage.

In figure 8, I plot counterfactual miner employment and machine usage in function of the level of the minimum wage as a percentile of the wage distribution. Machine usage, the solid blue line in panel (a), drops with the minimum wage level, except at very low levels of minimum wages: as explained earlier, the variable profit gains from machine adoption fall with a higher minimum wage, as adoption no longer leads to wage cuts. The equilibrium employment quantity, in panel (b), could in theory increase if the minimum wage would lie in between the monopsonistic and competitive equilibrium, as shown in figure 6(a). These wage equilibria are, however, different for every firm: a state-wide minimum wage will hence affect different firms differently. As shown in figure 8, minimum wages do not increase employment in this industry despite that there is monopsony power, because the employment losses from firms being pushed above their equilibrium wage outsizes the employment gains from
firms for which the minimum wage lies between the competitive and monopsonistic equilibrium. A minimum wage seems not to harm employment too much up to the 50th percentile, after which employment falls, as there is considerable monopsony power in the industry. The cost of such a policy would, however, be a drop in labor productivity due to decreased mechanization.

5.3 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.

Gross substitutes In reality, miners and helpers were found to be gross complements, and cutting machines to be skill-augmenting. Panel (a) of figure 9 conducts the same counterfactual exercise of moving to competitive labor markets, but assuming that miners and helpers are gross substitutes, with an elasticity of substitution of 1.10. In line with theorem 1, cutting machine usage now drops from 0.444 to 0.437 when moving to the competitive equilibrium, as cutting machines now move input usage from helpers towards miners, rather than the other way around.

![Figure 9: Skill-biased technologies](image)

| (a) Gross substitutes ($\sigma = 1.10$) | (b) Unskill-augmenting ($\beta^k = -1$) |

Unskill-augmenting machines Figure 9(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: $\beta^k = -1$. 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 0.294 to 0.268 when moving to a perfectly 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.\textsuperscript{28}

6 Conclusion

In this paper, I investigate how monopsony power affects the adoption of factor-biased technologies. I find that in theory, technologies that reduce demand for an input are reduced by the level of monopsony power over that input. Applying this model to late 19th century Illinois coal mining, I find that if miner markets would have become perfectly competitive, usage rates of coal cutting machines would have increased by a fourth, and hence also miner productivity. 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 (i) technological change saved on skilled workers, and (ii) firms mainly had monopsony power over these same skilled workers. If monopsony power would have been concentrated rather on unskilled labor markets, or if new technologies would have saved on unskilled workers instead, then monopsony power could increase, rather than decrease, 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 how monopsony power varies across the skill distribution.

\textsuperscript{28}Hauling 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.
References


coal-mining-in-northern-illinois)


History, 35(3), 591-620.

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 Special case: Cobb-Douglas production function

Consider a Cobb-Douglas production function, instead of the CES specification in equation (??). The output elasticity \( \omega_f \) now captures the employment share of skilled workers. A skill-saving shock corresponds to a drop in \( \omega_f \).

\[
Q_f = (L_f^s)^{\omega_f}(L_f^u)^{1-\omega_f}
\]

**Labor demand**  First, consider \( \omega_f \) to be exogenous. Firms choose the mix of both labor types that maximizes variable profits \( \pi_f \), taking output \( Q_f \) as given, as in the main model. Labor demand is hence given by the following system of equations:

\[
\begin{aligned}
L_f^s &= \omega_f Q_f (W_f^s \psi_f^s)^{-1} \\
L_f^u &= (1 - \omega_f) Q_f (W_f^u \psi_f^u)^{-1}
\end{aligned}
\]

**Labor supply**  Let the inverse labor supply functions be given by the same constant elasticity specification as before:

\[
\begin{aligned}
W_f^s &= (L_f^s)^{\psi_f^s} \\
W_f^u &= (L_f^u)^{\psi_f^u}
\end{aligned}
\]

**Equilibrium variable profits**  Variable profits are increasing in the input supply elasticities, and are zero if input prices are exogenous:

\[
\pi_f = Q_f \left( 1 - \frac{\omega_f}{\psi_f^s} - \frac{1 - \omega_f}{\psi_f^u} \right)
\]

**Capital demand**  Now let the coefficients \( \omega_f \) be dependent on the usage of a technology \( K_f \). The effect of the technology on variable profits is equal to:

\[
\frac{\partial \pi_f}{\partial K_f} = Q_f \left( \frac{\psi_f^u - \psi_f^s}{\psi_f^s \psi_f^u} \right) \frac{\partial \omega_f}{\partial K_f} \begin{cases} < 0 \end{cases}
\]

The technology increase profits as long as firms have more monopsony over \( L^u \) than over \( L^s \), meaning that \( \psi_f^u > \psi_f^s \). The reason for this is that machines change input usage from skilled to unskilled workers: this is only profitable if firms extract more profits out of unskilled than out of skilled workers.
The profit gain from technology adoption increases in the degree of monopsony power over the input to which the technology redirects input usage, and decreases in the degree of monopsony power over the other input.

\[
\frac{\partial}{\partial \psi_f} \left( \frac{\partial \Pi_f}{\partial K_f} \right) = \frac{\partial \omega_f}{\partial K_f} \frac{1}{(\psi_f^*)^2} < 0 \iff \frac{\partial \omega_f}{\partial K_f} < 0
\]

B.2 Proof of lemma 1

Suppose inputs are gross complements, meaning that \( \sigma < 1 \). From equation 10b, we know that a \( \tau \)-augmenting productivity shock increases profits in that case: \( \frac{\partial \pi_{f,t}}{\partial \omega_{f,t}^\tau} > 0 \). It follows that the derivative of this profit change with respect to the degree of monopsony power \( \psi_{f,t}^\tau \) is then negative, as shown in equation (16). In the opposite case that inputs are gross substitutes, meaning that \( \sigma > 1 \), a \( \tau \)-augmenting productivity shock decreases profits. The derivative in equation (16) is then positive, so the profit decrease is smaller in its absolute value.

\[
\frac{\partial}{\partial \psi_{f,t}^\tau} \left( \frac{\partial \pi_{f,t}}{\partial K_{f,t}} \right) = \frac{\sigma(\sigma - 1)(\psi_{f,t}^\tau)^2}{(1 + \sigma(\psi_{f,t}^\tau - 1))^2 (\psi_{f,t}^\tau)^{\frac{1 - \sigma}{\sigma(\psi_{f,t}^\tau - 1)}} - 1} \kappa_{f,t}^\tau
\]

with \( \kappa_{f,t}^\tau \equiv \left[ Q_{f,t} P_{f,t}^\sigma \right]^{\psi_{f,t}^\tau (\sigma - 1)} \left( \omega_{f,t}^\tau \right)^{\psi_{f,t}^\tau (\sigma - 1)} \left( (\alpha_{f,t}^\tau)^{(\psi_{f,t}^\tau - 1)} W_t^* \right)^{\frac{1 - \sigma}{\sigma(\psi_{f,t}^\tau - 1)}} > 0 \)

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 (16) 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 (16), the variable profit gain decreases with monopsony
over input $\tau$:

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

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.

C Additional results

C.1 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. I regress both coal prices and miner wages on cutting machine usage, the average labor market share, and the share of the total population in county that works in the coal mining industry. The regression is at the town-year level, and I add town fixed effects. The coefficients therefore measure how these variables co-move over time. Coal prices do not seem to change when firms install cutting machines, and do not co-move with labor market structure. Miner wages, in contrast, fall when cutting machines are installed, and are lower when labor markets are more concentrated. Although these coefficients have to be interpreted with care, as both market shares and machine usage rates are endogenous outcomes, the fact that prices are uncorrelated and wages correlated to machine adoption and labor market structure is consistent with imperfectly competitive labor markets and perfectly competitive coal markets.

C.2 Isolated markets

Villages are assumed to be isolated labor markets, which allows to treat firms as monopsonists, rather than oligopsonists. This means that the cross-wage elasticity across different villages is zero. I test this by regressing the wage rate in a village on the usage of cutting machines in the other villages of the same county, and take village fixed effects. I hence look at how wages change as firms in other villages in the same county adopt machines. The estimates are in table A3. Wages drop if machines are adopted in the own village, in line with the theory. Wages do not change, however, with machine adoption in other villages inside the same county, which supports the isolated markets assumption.
C.3 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 derivate of variable costs to output:

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

Taking the derivative and using the miner piece rate $W_{q,f,t}$, 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.

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

Using the helper demand function 3, markups can be rewritten as:

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

**Results** The average markup ratio was 1.59 and the median markup 1.36. Markup and markdown ratios are plotted in figure A3, 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.4 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 7 hence becomes:

\[ v_{f,t}^1 = \frac{\left(\pi^1_{f,t} - \pi^0_{f,t}\right)}{1 - \delta} - \gamma z_{f,t} - \nu_{f,t} \]

The resulting counterfactual plots are in figure A4. 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.

D.2 Cost dynamics

If cost dynamics are important, the productivity level of both worker types should depend on cumulative past production, as in Benkard (2000). I test this by regressing the logarithms of the productivity residuals \( \omega^g_{f,t} \) 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.
Table A1: Prices and wages

<table>
<thead>
<tr>
<th></th>
<th>log(Price)</th>
<th></th>
<th>log(Wage)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td>b_ldchi</td>
<td>0.040</td>
<td>0.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b_ldsl</td>
<td>0.025</td>
<td>0.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutting machine</td>
<td>0.019</td>
<td>0.029</td>
<td>-0.138</td>
<td>0.062</td>
</tr>
<tr>
<td>log(Miner share)</td>
<td>-0.009</td>
<td>0.011</td>
<td>-0.064</td>
<td>0.021</td>
</tr>
<tr>
<td>log(Coal mining employment share)</td>
<td>0.028</td>
<td>0.015</td>
<td>-0.086</td>
<td>0.032</td>
</tr>
<tr>
<td>FE</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2,392</td>
<td></td>
<td>2,234</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.627</td>
<td></td>
<td>0.358</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered at county level.
Table A2: Transport costs

<table>
<thead>
<tr>
<th></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>log(Cum. output)</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>-0.0039</td>
<td>0.0028</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Observations: 798

R-squared: 0.0022 0.0025

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

| Log(Mineral employment share) | 0.068 | 0.008 |
| Log(Total miners / population) | 0.097 | 0.017 |
| Log(Afro-Americans) | 0.006 | 0.012 |
| Log(Foreign-born) | -0.060 | 0.041 |
| Log(Population) | 0.045 | 0.073 |
| Log(Area) | 0.029 | 0.158 |
| Log(Mfg. workers) | -0.003 | 0.026 |
| Log(Mfg. plants) | 0.051 | 0.042 |
| Log(Farmland area) | -0.005 | 0.105 |
| 1(Agricultural scarcity) | -0.021 | 0.027 |

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></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mine run</td>
<td>X</td>
<td></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>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slack</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>Shipping or local mine</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>Shipping quantities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miners, avg entire year</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>Miners, max entire year</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>Other employees</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>Other employees, underground</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other employees, above ground</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other employees winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other employees summer</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys employed underground</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital (in dollar)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>

52
**Figure A1: Mine and machine locations**

Notes: The blue dots represent mining towns, each of which can contain multiple mines. Villages with red squares contain at least one machine mine.
Figure A2: Harrison Cutting Machine
Figure A3: Markups and markdwons
Figure A4: Counterfactuals with various discount rates

(a) $\delta = 0$ (static model) 
(b) $\delta = 0.90$

(c) $\delta = 0.95$ (baseline) 
(d) $\delta = 0.98$