Appendices

A Appendix: Proofs

A Proof of Lemma 1

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

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

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

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

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

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

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

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

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

The right-hand side of (15) is weakly smaller than one, because $\mu^h \leq \psi^h$ and $\mu^l = \psi^l = 1$. For the same reason, the left-hand side is weakly larger than one, so it holds that $\frac{\partial}{\partial \beta}(\frac{\partial \pi}{\partial \mu^h}) \geq 0$.

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

B Proof of Theorem 1

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

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

C Proof of Proposition 1

To prove: For $\mu_f^l < \psi_f^l$and $\psi_f^l \neq 1$:

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

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

D Proof of Proposition 2

To prove: For $\mu_f^l = \psi_f^l$ or $\psi_f^l = 1$:

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

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

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

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

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

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

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

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

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

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

E Proof of Proposition 3

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

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

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

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

F Proof of Proposition 4

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

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

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

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

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

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

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

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

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

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

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

Again, this expression is weakly positive, which implies:

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

B Appendix: Data

B.1 Sources

Mine Inspector Reports The main data source is the biennial report of the Bureau of Labor Statistics of Illinois, of which I collected the volumes between 1884 and 1902. Each report contains a list of all mines in each county, and reports the name of the mine owner, the town in which the mine is located, and a selection of variables that varies across the volumes. An overview of all the variables (including unused ones), and the years in which they are observed, is in Tables A7 and A8. Output quantities, the number of miners and other employees, mine-gate coal prices, and information on the usage of cutting machines are reported in every volume. Miner wages and the number of days worked are reported in every volume except 1896. The other variables, which includes information about the mine type, hauling technology, other technical characteristics, and other inputs, are reported in a subset of years.

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

Monthly data The 1888 report contains monthly production data for a selection of 11 mines in Illinois, across 6 counties. I observe the monthly number of days worked and the number of skilled and unskilled workers. I also observe the net earnings for all skilled and unskilled workers per mine per month, and the number of tons mined per worker per month. This allows me to compute the daily earnings of skilled and unskilled workers per month.

B.2 Data cleaning

Employment In every year except 1896, workers are divided into two categories, 'miners' and 'other employees'. In 1896, a different d istinction is made, between 'underground workers' and 'above-ground workers'. This does not correspond to the miner-others categorization because all miners were underground workers, but some underground workers were not miners (e.g. doorboys, mule drivers, etc.). Hence, I do not use the 1896 data. From 1888 to 1896, boys are reported as a separate working category. Given that miners (cutters) were adults, I include these boys in the 'other employee' category. The number of days worked is observed for all years. The average number of other employees per mine throughout the year is observed in every year except 1896; in 1898 it is subdivided into underground other workers and above-ground other workers, which I add up into a single category. The quantity of skilled and unskilled labor is calculated by multiplying the number of days worked with the average number of workers in each category throughout the year. Up to and including 1890, the average number of miners is reported separately for winters and summers. I calculate the average number of workers during the year by taking the simple average of summers and winters. If mines closed down during winters or, more likely, summers, I calculate the annual amount of labor-days by multiplying the average number of workers during the observed season with the total number of days worked during the year.

Wages Only miner wages are consistently reported over time at the mine level. The piece rate for miners is reported. Up to 1894, miner wages per ton of coal are reported separately for summers and winters. I weight these seasonal piece rates wages using the number of workers employed in each season for the years 1884-1890. In 1892 and 1894, seasonal employment is not reported, so I take simple averages of the seasonal wage rates. In 1896, wages are unobserved. From 1898 onwards, wages are no longer reported seasonally, because wages were negotiated biennially from that year onwards. For these years, wages are reported separately for hand and machine miners. In the mines that employed both hand and machine miners, I take the average of these two piece rates, weighted by the amount of coal cut by hand and cutting machines.

Output quantity and price The total amount of coal mined is reported in every year, in short tons (2000 lbs). Up to and including 1890, the total quantity of coal extraction is reported, without distinguishing different sizes of coal pieces. After 1890, coal output is reported separately between 'lump' coal (large pieces) and smaller pieces, which I sum in order to ensure consistency in the output definition. Mine-gate prices are normally given on average for all coal sizes, except in 1894 and 1896, where they are only given for 'lump' coal (the larger chunks of coal). I take the lump price to be the average coal price for all coal sizes in these two years. There does not seem to be any discontinuity in the time series of average or median prices between 1892-1894 or 1896-1898 after doing this, which I see as motivating evidence for this assumption.

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

Deflators I deflate all monetary variables using the consumer price index from the *Handbook of Labor Statistics* of the U.S. Department of Labor, as reported by the Minneapolis Federal Reserve Bank website.[1]

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

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

Mine and firm identifiers The raw dataset reports mine names, which are not necessarily consistent over time. Based on the mine names, it is often possible to infer the firm name as well, in the case of multi-mine firms. For instance, the Illinois Valley Coal Company No. 1 and Illinois Valley Coal Company No. 2 mines clearly belong to the same company. For single-mine firms, the operator is usually mentioned as the mine name, (e.g. 'Floyd Bussard'). For the multi-mine firms, mine names were made consistent over time as much as possible.

Town identifiers and labor market definitions The data set contains town names. I link these names to geographical coordinates using Google Maps. I calculate the shortest distance between every town in the data. For towns that are located less than 3 miles from each other, I merge them and assign them randomly the coordinates of either of the two mines. This reduces the number of towns in the dataset from 448 to 374. The resulting labor markets lie at least 3 miles from the nearest labor market.

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

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

C Appendix: Empirics

C.1 Equilibrium expressions for the empirical model

The equilibrium output of a mine f at time t is denoted Q^*_{ft}. It can be solved for by computing the first order conditions of the profit maximization problem, (11), and using Equations (6a), (7), and (8), which are respectively the production, coal demand, and labor supply functions. The resulting equilibrium output expression is in Equation (17a), which is the empirical analogue of Equation (5a) with Cournot competition upstream and downstream. When assuming that the firm is a monopolist and monopsonist (all market shares become one, and $f = m = n$), and there are no latent differences between coal and labor markets (no $\xi^l_{nt} = \xi^h_{nt} = \zeta_{mt} = 1$), Equation (17a) simplifies to Equation (5a).

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

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

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

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

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

C.2 Alternative production model

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

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

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

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

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

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

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

The markup μ_{ft} can be expressed as the ratio of the output elasticity of miners over the product of its revenue share and markdown:[2]

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

Results The results of this alternative production model are in Table A5. Coal cutting machines are still unskill-biased: the output elasticity of miners is estimated to fall by 0.346 points when adopting a cutting machine, coming from 0.695. In the baseline model, this was a smaller drop of 0.160 points, down from 0.734. In contrast, the usage of mining locomotives is estimated to barely change the output elasticity of skilled workers, in contrast to the main specification where it increased this output elasticity compared to unskilled workers. The scale parameter, ν, is estimated at 0.856, whereas it was assumed to be 0.9 in the main text. Thus, the assumption of decreasing returns to scale is confirmed. The average markup ratio μ is estimated at 1.139, which implies that the coal price is 13.9% above marginal costs. This estimate does not impose any model of competition on the coal market. The homogeneous goods Cournot model in the baseline model delivered a very similar average markup ratio of 1.148, which is close to the markup in the alternative model that does not impose Cournot competition.

C.3 Alternative labor supply model

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

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

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

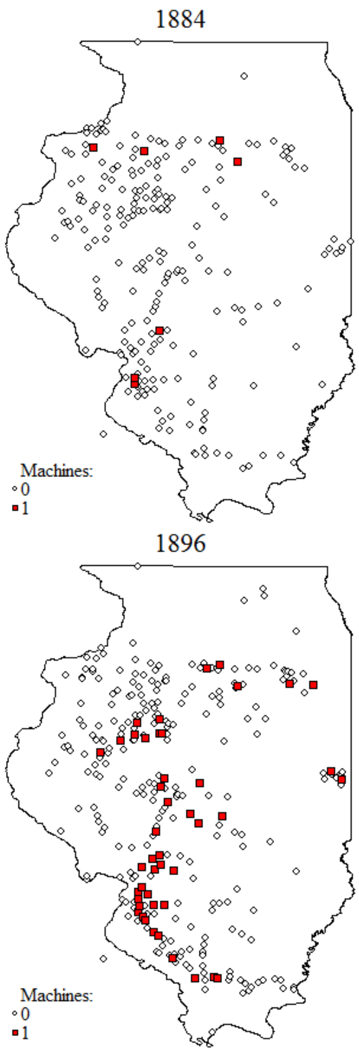
C.4 Cost dynamics

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

C.5 Inverse miner supply elasticity: correlations

I regress the log town-level inverse miner supply elasticity $\ln(\psi^h_{nt})$ on observed town and county characteristics. A higher town-level inverse supply elasticity implies more inelastic miner supply. The results are in Appendix Table A3. First, miner supply is more inelastic if total coal employment as a share of the town population is higher, which implies fewer outside work opportunities. A second regressor is the log of the ratio of the total farmed area in a county divided by the county's surface. Miner supply is more inelastic in areas with less farming (for instance, because of rugged geography), presumably because there are fewer outside work opportunities to switch to. Third, the population share of African Americans in the county does not correlate significantly with the miner supply elasticity. Fourth, the miner supply elasticity does not differ with the share of firms connected to the railroad network, which is in line with historical evidence that railroads were not used to transport workers.[3] Finally, the average wage in manufacturing industries in the same county correlates positively, but not significantly, with the inverse miner supply elasticity, which suggests that the outside option was mainly to work in agriculture, rather than in manufacturing industries, which were in any case scarce in rural Illinois.

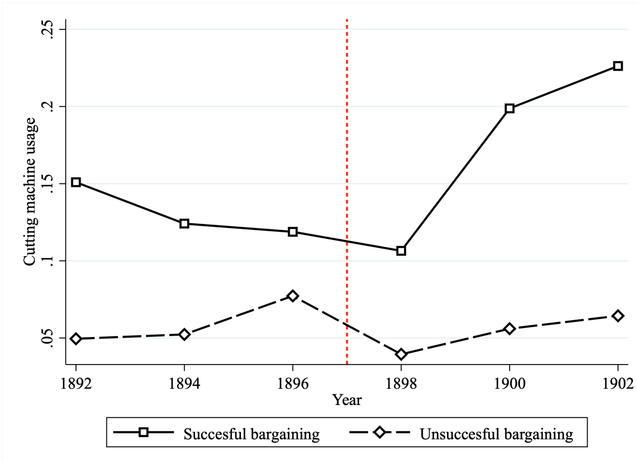
Figure A1: Geographical spread of cutting machines



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

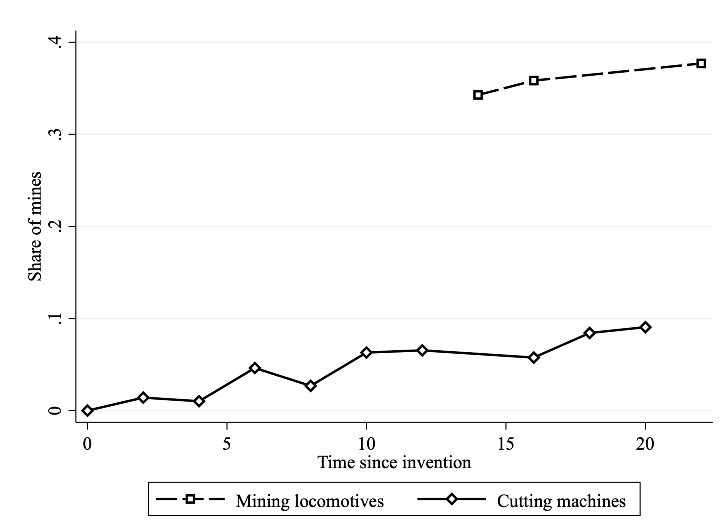
[3]If mining towns would not be isolated due to workers commuting by train, being connected to the railroad network should result in more elastic labor supply.

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



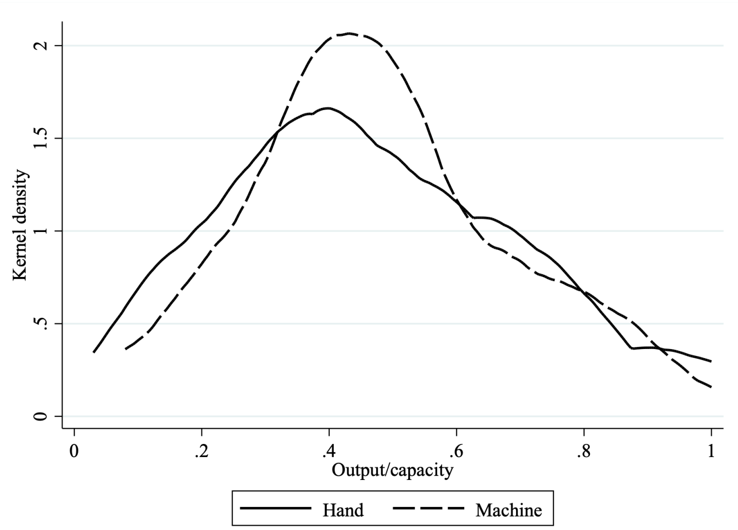
Notes: This figure compares average cutting machine usage between mines at which the introduction of wage bargaining in 1898 resulted in increased miner wages, and mines at which wages remained unchanged. The dotted line indicates the large Illinois coal strikes of 1897, which led to the introduction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

Figure A3: Cutting machine vs. locomotive adoption



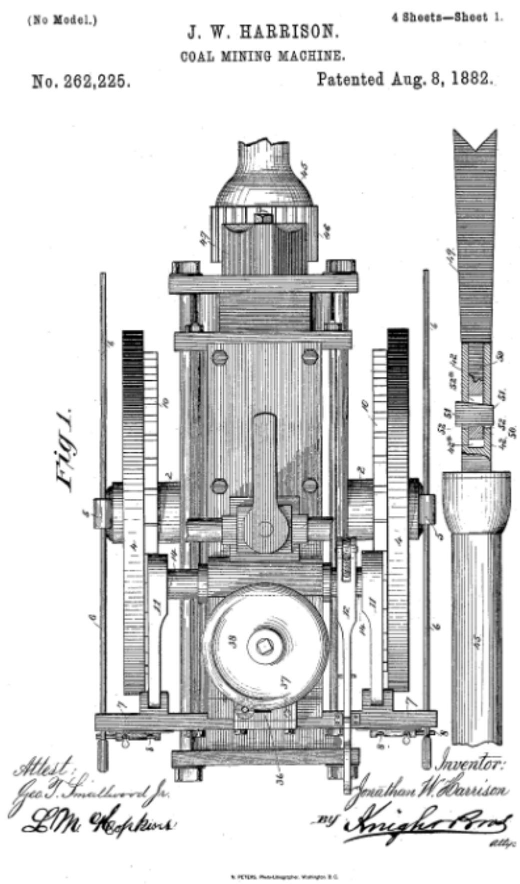
Notes: This figure compares the share of Illinois mines using locomotives and cutting machines against how long these technologies have been invented.

Figure A4: Capacity utilization



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

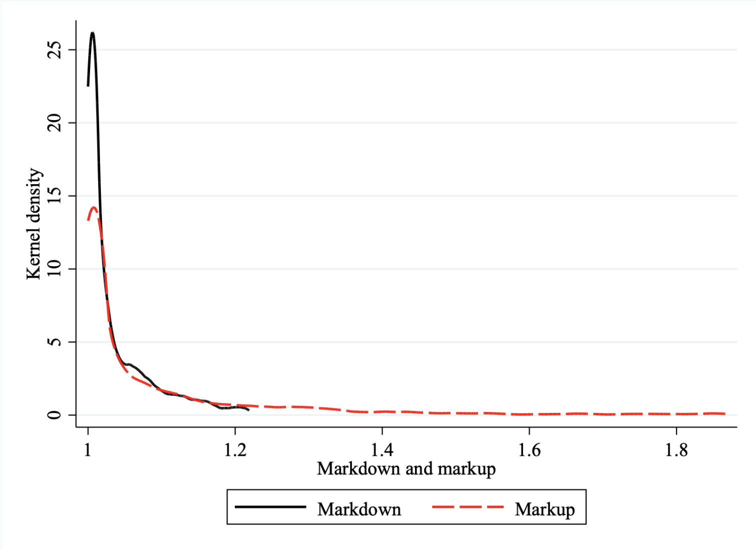
Figure A5: Patent of the Harrison Cutting Machine



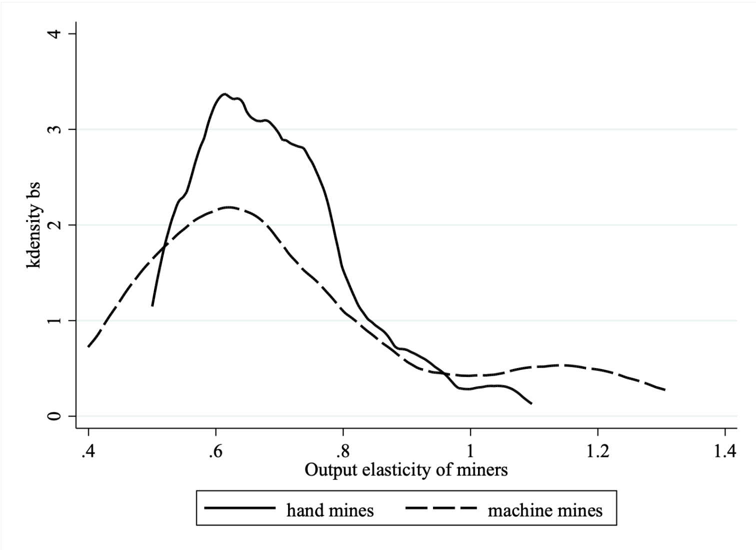
Notes: U.S. patent of the 1882 Improved Harrison Coal Cutting Machine (Whitcomb, 1882). This was the most frequently used coal cutting machine in the data set.

Figure A6: Distributions of latent variables

(a) Markdowns and markups



(b) Skilled labor output elasticity

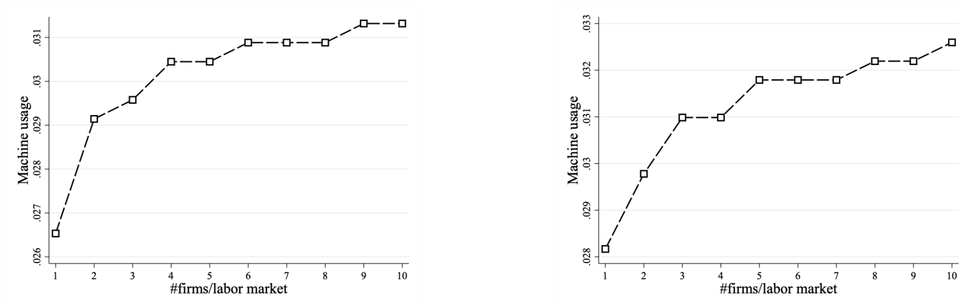


Notes: Panel (a) shows the distribution of the miner wage markdown and coal price markup across mines between 1884-1894. Panel (b) shows the distribution of miner output elasticities for firms using at least one cutting machines ('machine mines' and firms that did not use any cutting machines ('hand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

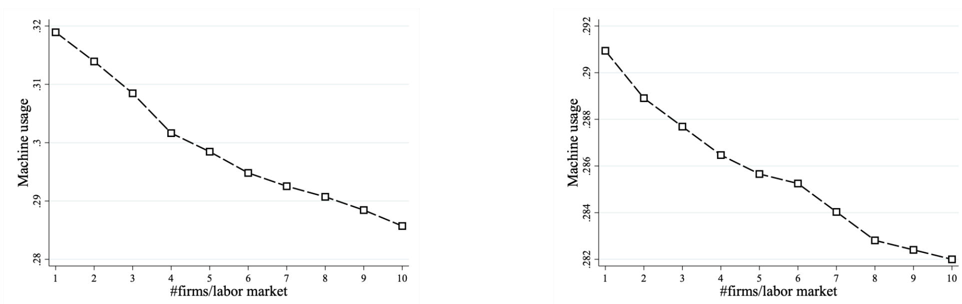
Figure A7: Technology usage and labor market structure - alternative supply model

Main model: **Alternative model:**

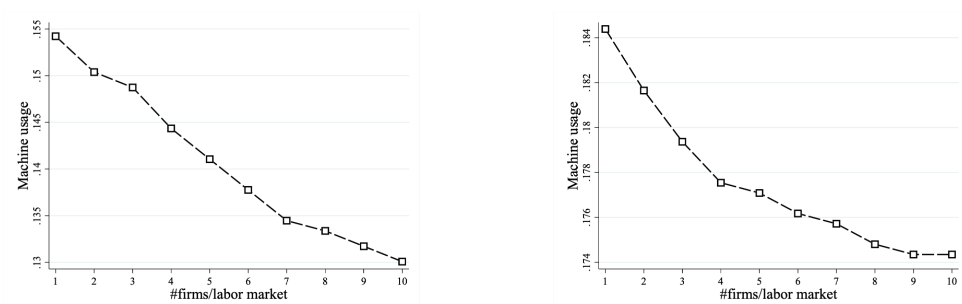
(a) Cutting machines



(b) Locomotives



(c) Skill-biased cutting machines



(d) Hicks-neutral cutting machines

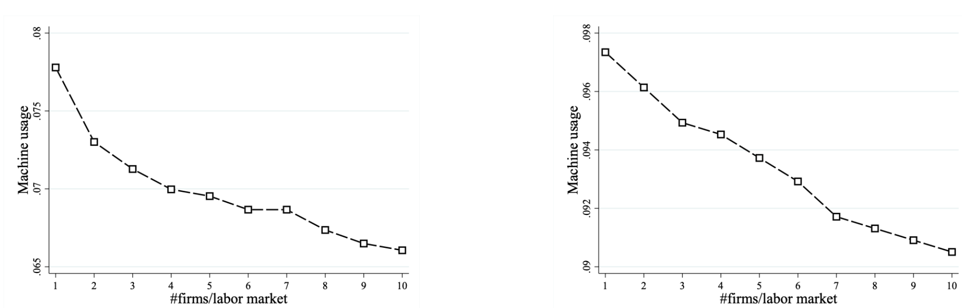
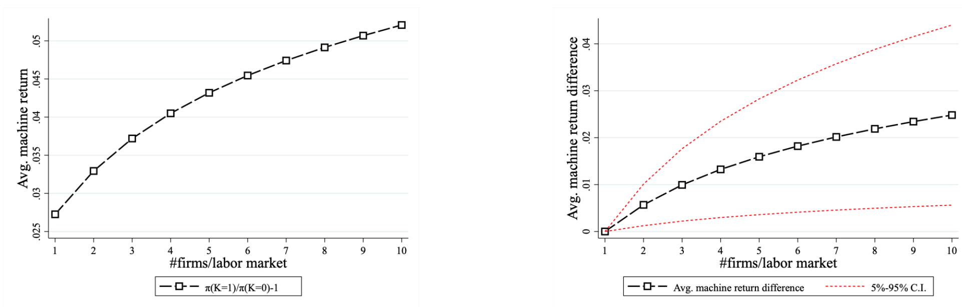
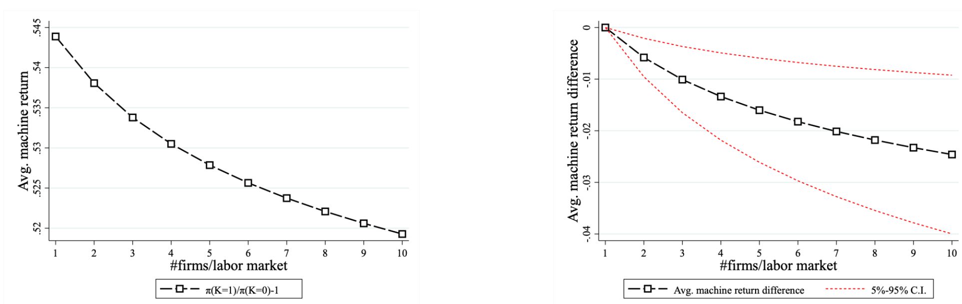


Figure A8: Variable profit returns and labor market structure

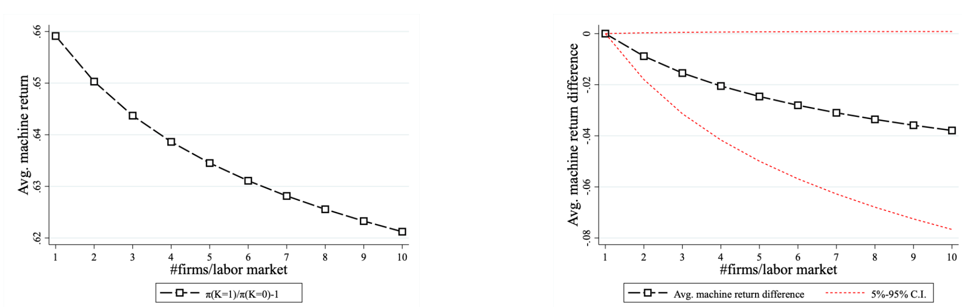
(a) Cutting machines



(b) Locomotives



(c) Skill-biased cutting machines



(d) Hicks-neutral cutting machines

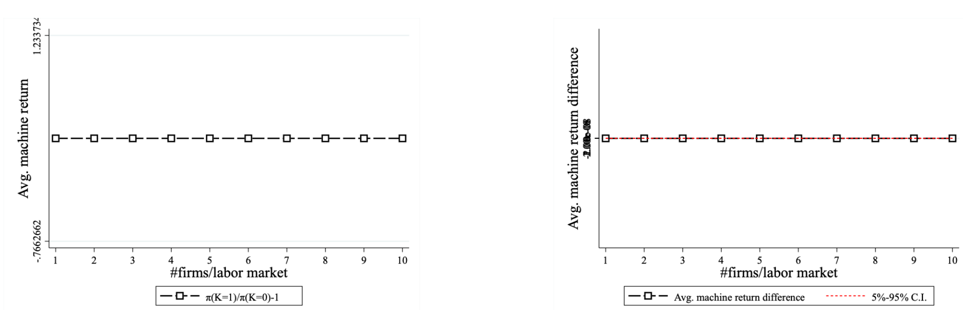


Table A1: Occupations and wages

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

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

Table A2: Cost dynamics

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

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

Table A3: Markdown correlations

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

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

Table A4: Coal demand and production estimates: all coefficients

(a) Coal demand (county-level)	log(Coal price)	
	Est.	S.E.
Coal demand elasticity	-0.465	0.101
1(Railroad connection)	0.353	0.123
1(Railroad crossing)	0.689	0.209
log(Dist. to St. Louis)	-0.069	0.071
log(Dist. to Chicago)	-0.044	0.070
Observations	453	
F-stat 1st stage	70.1	
R-squared	.191	
(b) Output elasticity transition	log(Output elasticity of skilled miners)	
	Est.	S.E.
1(Cutting machine)	-0.160	0.068
1(Locomotive)	0.101	0.029
log(Materials)	0.005	0.017
Year	-0.011	0.013
Constant	21.111	25.417
Observations	1149	
R-squared	.008	
(c) Hicks-neutral productivity transition	log(Hicks-neutral productivity)	
	Est.	S.E.
1(Cutting machine)	0.218	0.157
1(Locomotive)	0.277	0.182
log(Materials)	0.105	0.130
Year	0.028	0.197
Constant	-51.688	.
Observations	1066	
R-squared	.225	

Table A5: Alternative production model

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

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

Table A6: Alternative labor supply model

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

Table A7: All variables per year

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

Table A8: All variables per year (cont.)

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

References

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

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