KU LEUVEN

DEPARTMENT OF ECONOMICS

Managerial knowledge and technology choice: Evidence from U.S. mining schools

Michael Rubens

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS20.05

APRIL 2020

Managerial knowledge and technology choice: Evidence from U.S. mining schools

Michael Rubens*

March 30, 2020

Latest version here

Abstract

How do managers affect firm performance? A key difference between managers and other production inputs is that they choose the production function. I empirically distinguish between the *direct* effects of managers as inputs in the production process, which is the standard way to think about management, and their *indirect* effects as decision-makers of the production technology. I use this model to understand how the introduction of mining engineering degrees in the U.S.A. changed coal mining productivity. I find that conditional on all inputs and technology choices, mines managed by managers with mining degrees were not more productive than other mines. Mining college graduates did, however, tend to select better technologies, which in turn increased productivity by 29% on average. The main mechanism behind these better choices was that mining college graduates had superior ex-ante knowledge about the returns to various new technologies, while other managers had to acquire this information through trial-and-error.

Keywords: Management, Productivity, Technology Adoption, Higher Education
JEL Codes: L25, O14, J24, I23

*Department of Economics, KU Leuven and Research Foundation Flanders, michael.rubens@kuleuven.be
I thank Jo Van Biesebroeck, Jan De Loecker, Frank Verboven, Otto Toivanen, Chad Syverson, Penny Goldberg, Naomi Lamoreaux and participants at the KU Leuven IO lunch, Yale IO prospectus workshop, Business History Conference and RES Conference. Financial support from the Research Foundation Flanders, Belgian American Educational Foundation and Fulbright Commission is gratefully acknowledged. Any errors are my own.

1 Introduction

> *"With no means of educating miners to their work, the conduct of mines in this country is a lamentable story of mismanagement, energy wrongly directed, and consequent great losses."* John A. Church, mine superintendent, 1871

Recent evidence points to the importance of managers as drivers of firm performance. 'Good management' is usually viewed either as a well-defined set of best practices (Bloom & Van Reenen, 2007; Bloom, Sadun, & Van Reenen, 2017), or as a time-invariant individual trait (Hoffman & Tadelis, 2018; Bertrand & Schoar, 2003; Lazear, Shaw, & Stanton, 2015).[1] The exact mechanism through which managers affect firm performance is, however, less clear.

Prior research tended to view managers as inputs in the production function or as shifters of the production isoquant.[2] Managers could, however, also choose the production function itself, by adopting different technologies. In this paper, I empirically distinguish the *direct* effects of managers as inputs in the production process from their *indirect* effects as decision-makers of the production technology. This distinction is crucial for the empirical evaluation of the productivity effects of managers. Suppose that managers with, for instance, high human capital levels systematically choose production functions with a higher production possibilities frontier, but are not inputs to production themselves. Conditional on all input choices, the presence of such managers will not correlate with total factor productivity (TFP), even if they do enhance output by choosing better technologies.

If managers differ in terms of their technology choices, there must be an underlying mechanism driving these differences. I propose three different mechanisms: managerial characteristics can alter the returns to technologies, their costs, or information about either returns of costs of new technologies. I provide an empirical framework to distinguish between these three effects.

The empirical application of this paper examines how the introduction of college-level mining engineering degrees in the U.S. affected productivity and technology choices in the coal mining industry. I track the very first mining engineering graduates as they entered managerial positions in Pennsylvania coal mines between 1900 and 1914. I specify and estimate an empirical model of coal extraction in which mine managers either held a mining engineering degree, another college degree, or no degree at all. The educational background of the manager can both enter the production function as an input,

[1] An example of such traits is the possession of good interpersonal skills. These two views of management have very different implications for how to 'manage managers'. In the first case, selection and allocation of managers is crucial (Benson, Li, & Shue, 2019; Terviö, 2009), and their education and training merely serves as a signaling mechanism. In the second case, managers be can actively improved, for instance by attending a business school or by hiring external consultants (Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013).

[2] With a Cobb-Douglas production function, both amount to the same. Managers could, however, also have factor-augmenting effects (Caroli & Van Reenen, 2001; Van Biesebroeck, 2003; Bender, Bloom, Card, Van Reenen, & Wolter, 2018).

shifting the isoquant, or enter demand for different production technologies. I focus on how to haul coal to the surface, which was a crucial determinant of mine productivity. There were four alternative technologies (i) mules, the traditional technique, (ii) steam locomotives, (iii) electrical locomotives, and (iv) compressed air locomotives. I find that mining school graduates did not shift the production frontier significantly, conditional on all other inputs. In other words, they were not an input in the production function themselves. They were, however, more than twice as likely to choose electrical locomotives compared to other managers. These types of engines had a higher marginal product compared to other engines, while their costs were similar.

Next, I examine the underlying mechanism of why mining engineering graduates differed from other managers in terms of their technology choices. I find no evidence for different technology returns or operating costs between mining engineers and other managers. The ex-ante information about each technology was, however, very different between managers. Mining engineers knew beforehand that electrical engines were superior to alternative innovative technologies. As time progressed, the differences in technology choices between different managers disappeared.

The historical setting of early 20th century coal mining is unique and ideally suited to examine the relationship between management and technology choices, for three reasons. First, the spread and rise of mining engineering colleges throughout the United States during this period provides a large shock to managerial human capital. Second, the first decade of the 20th century saw rapid technological change, as electrification fundamentally changed many industries, including coal mining. I observe detailed technology choices down to the mine level, together with standard production and cost variables. Finally, physical productivity was the main driver of firm performance in this industry, as coal is a homogeneous product that was sold on competitive markets. This offers a stylized setting to think about firm performance.

The underlying questions are relevant, however, beyond this historical setting. Technological innovations, such as artificial intelligence, and educational innovations, such as 'Tech MBAs',[3] usually happen simultaneously. When evaluating the effects of such educational innovations on firm performance, a narrow view of managers as production inputs will potentially underestimate their added value compared to a model where these managers also choose production technologies. Secondly, distinguishing the source of such differences in technology choices between managers is crucial to know whether these are transitory or permanent. If new educational programs mainly affect the information set of managers concerning new technologies, as I find in the mining engineering example, then these benefits are entirely conditional on information about these novel technologies being imperfect. As soon as information on new technologies becomes common knowledge, the benefits of better educated managers would vanish. If educational innovations would, however, also change the costs or returns of operating new technologies, then this would imply much more persistent returns to

[3]https://www.economist.com/whichmba/tech-mbas-catching-up

education.

This paper makes three main contributions to the literature. First, I contribute to the literature on management and productivity, such as Bloom and Van Reenen (2007); Bloom et al. (2013, 2017). This literature tends to view managers as inputs in the production function.[4] I allow, in contrast, the production function itself to be a choice made by managers. As already mentioned above, this distinction changes the evaluation of the effects of managers on productivity and firm performance.

Secondly, I contribute to the literature on the productivity effects of higher education. The usual approach in this literature is to correlate TFP measures with managerial education (Bertrand & Schoar, 2003; Braguinsky, Ohyama, Okazaki, & Syverson, 2015). This again views managers as production inputs. I find, in contrast, that the main effect of mining engineering degrees was that it changed technology choices.[5] Interestingly, these effects held uniquely for mining college degrees. Managers with other, mostly liberal arts, college degrees were not different from uneducated managers in terms of their technology choices, even if they had attended elite institutions. My findings hence imply that the diversification of the U.S. higher education landscape during the late 19th century was instrumental in driving adoption of innovative technologies.

Finally, I contribute to the literature on technology adoption with imperfect information on technology benefits, such as Levin et al. (1987); Besley and Case (1993); Jovanovic and Nyarko (1996); Foster and Rosenzweig (1995); Munshi (2004); Conley and Udry (2010). I contribute by showing that managers with more specialized educational backgrounds had ex-ante information on new technologies, which allowed them to adopt better technologies faster than other managers. This aligns with the hypothesis in Rosenzweig (1995) that education could be substitutable with 'learning by doing' about new technologies.[6] The implication of this is that the benefits from technically educated managers were both transitory and conditional on technological change: as information spread throughout the economy, all managers ultimately found out which technology to use.

The remainder of this paper is structured as follows. In section 2, I discuss the industry background and data sources. Section 3 contains a model of production and technology choices, which is estimated in section 4. Section 5 concludes.

[4] An alternative approach is to deduct the gains from good managers from variation in managerial compensation, using a revealed-preferences argument (Terviö, 2009). I do not use such an approach as I do not observe data on managerial compensation.

[5] Other papers have found higher education increases technology adoption (Wozniak, 1984, 1987; Skinner & Staiger, 2005; Lleras-Muney & Lichtenberg, 2005), but did not examine whether the type of education matters, and did not contrast this view to the 'manager as input' channel. Toivanen and Väänänen (2016); Bianchi and Giorcelli (2019) finds that technical higher education led to more technology invention, but does not examine the effects on technology adoption.

[6] Rosenzweig did not find empirical evidence for this, however, probably because his setting involved agents that enjoyed primary schooling at most, which is unlikely to provide concrete knowledge of new technologies, in contrast to mining engineering degrees.

2 Industry background

2.1 Pennsylvania anthracite mining, 1900-1914

This paper studies anthracite extraction in Pennsylvania between 1900 and 1914. Anthracite is the coal type with the highest energy content, and was the main coal product in Pennsylvania at the time. The extraction process consisted of three main steps. First, a shaft or tunnel had to be dug to reach the underground coal seam. Next, coal was excavated using either picks and black powder or mechanized cutting machines. Finally, it had to be hauled back to the surface, using mules or underground mining locomotives. Anthracite is a nearly homogeneous product, with limited differentiation in caloric content. It was sold as a fuel after extraction without any further processing, unlike other coal types.[7] Pennsylvania mines transported on average 85% of their output over the railroad network towards urban markets, with the remainder being either sold locally or re-used as an input. Coal markets were hence likely to be perfectly competitive. There are 287 unique anthracite firms and 615 unique mines in the dataset.

Coal mine management

Coal firms were headed by a general manager, often also the owner, who was based in a nearby city. Daily management at the mine level was delegated to 'superintendents' (henceforth 'managers'), who are the main object of interest in this paper. The lowest level of operational management was carried out by 'foremen', of which two thirds worked below the surface. Managers had a wide range of responsibilities, including technical procurement, human resources management, production line design, financial analysis and cost reporting (Ochs, 1992). As coal markets were competitive, low marginal costs were the main driver of firm performance.

Technological change: underground mine locomotives

Transporting coal to the surface was a crucial part of the coal extraction process. In a mining journal article, Hodges (1905) asserts that *"the problem of getting coal from the working face to the surface in the most economical way is one of the most serious which the mine manager has to solve."* Mules were traditionally used for hauling, but were gradually replaced by underground mining locomotives from the 1880s onwards. Three main locomotive types existed: steam-powered, electrical and compressed air locomotives. Images of all three types are shown in figure 1.

[Figure 1 here]

[7] Bituminous coal is often degassed into 'cokes' before usage in order to improve its properties.

Steam locomotives were invented first, and were already relatively common by 1900. They were more efficient compared to mules, but also more dangerous: underground air quality deteriorated and there was a risk of explosions due to mine gas (Randolph, 1905). These concerns led to the development of electrical and compressed air locomotives during the 1890s. Compressed air locomotives were a relatively safe technology, but had the disadvantage of having to be refilled frequently, resulting in a limited range. Electrical locomotives, finally, required the installation of overhead lines, and could lead to electrocution in humid or flooded mines (Gairns, 1904). Their massive adoption indicates their usefulness. Electrification was a new technology, and electricity experts were rare, as the *Transactions of the Institute of Mining Engineers* noted:

> *"The machinery used with compressed air so closely resembles that used with steam, that mechanics familiar with the one have little to learn in managing the other. [...] men competent to manage pneumatic plants are easily obtained, while experts in electricity are scarce."* (Randolph, 1905)

Figure 2 depicts the total number of locomotives used of each type, and the share of mines using them. As mines could use several locomotive types concurrently, the shares add up to more than one. In 1900, around 2000 steam locomotives were already being used, while barely any of the other two types were in use. Up to 1904, the number of electrical and air locomotives grew at similar rates, after which electricity took over as the standard technology. The share of electrified mines increased from 10% to 60% between 1900 and 1908. Compressed air engines were never used in more than 40% of the mines. Weighting these shares by mine output delivers somewhat higher shares as mines using locomotives were larger, but the evolution per type is very similar to the the unweighted version in figure 2.

[Figure 2 here]

Educational change: mining engineering programs

Up to the 1880s, the U.S.A. lagged far behind Europe in terms of technical higher education. While different continental European nations already had specialized engineering colleges from the early 18th century onwards, American universities such as MIT and Columbia only started offering engineering degrees during the 1860s (Lundgreen, 1990). Rapid technological change during the second industrial revolution increased demand for engineers. Lehigh University was home to an important mining college in Pennsylvania and phrased its 1872 mission statement as follows:

> *"To introduce branches which have been heretofore more or less neglected in what purports to be a liberal education [...] especially those industrial pursuits which tend to develop the resources of the country."*

Mining engineering was such a 'neglected' branch. The annual number of U.S. graduates with an *Engineer of Mines (E.M.)* title[8] quadrupled from 80 to 320 between 1898 and 1914, as shown in figure 3. As the supply of mining engineers grew, these graduates started to enter the coal mining industry. The solid red line in figure 3 shows that the share of mines managed by a college-level mining engineer increased from none in 1898 to 6% in 1914. The fraction of mines managed by a graduate with other college degrees grew from 2 to 6% as well. They slowly replaced an older generation of non-educated managers who had usually entered the mines around the age of twelve.[9] Managers with a mining degree were on average responsible for 6 different mines, twice as many as other managers. Their mines also employed nearly twice as many workers.

[Figure 3 here]

What did mining colleges teach? Appendix table A10 contains summary statistics on the curricula of 7 important mining colleges. They were heavily dominated by the natural sciences and various engineering branches, which made up for around 95% of credits. The remaining credits were used for non-science subjects such as language courses and administrative topics such as accounting and law. The rise of electricity was anticipated by mining colleges: by the year 1900, most mining engineering programs had compulsory courses in electrical engineering and applied electricity in their junior or senior years.[10]

Mining engineering graduates in Pennsylvania mainly graduated from local schools such as Lafayette College and Lehigh University. Managers with college degrees in other fields hailed, in contrast, predominantly from elite universities: two thirds of these managers graduated from Ivy League colleges such as Yale, Princeton, Cornell and Penn. Managers with college degrees were on average in their early thirties, or 15 years younger than the average non-educated manager. Summary statistics on managers of different backgrounds are in appendix table A9.

Figure 4 compares output per worker between mines managed by mining college graduates and other mines. I calculate labor productivity by dividing annual output by the number of worker-days.[11] Labor productivity grew on average by 3% per year between 1900 and 1914 in mines managed by a mining college graduate, compared to merely 1.7% for the other mines. The productivity gap between both groups was largest around 1910 and closed again by 1914.

[Figure 4 here]

[8]To fix terminology, when I mention 'college-level mining programs', I mean four-year undergraduate degrees with a specific focus on mining engineering, delivering an *Engineer of Mines (E.M.)* degree. This definition also excludes non-college-level technical schools and military academies.

[9]This becomes clear from reading biographies of important managers and state inspectors in the *Reports*: most of these men were born in England or Wales, had entered the mines at an early age, and had then migrated to the U.S.A.

[10]I observe this, for instance, in the catalogs of the University of Utah and New Mexico School of Mines, where electricity was introduced in the mining curriculum during the late 1890s.

[11]I include mines that are not managed by a mining engineer, but were in the past three years in the 'mining engineering' category: the effects of mining engineers outlasts their tenure at the mine.

The visual correlation between managerial education and productivity could reflect that more productive firms chose better managers, or vice-versa. Besides, it is more meaningful to compare total factor productivity between these mines, rather than just output per worker. A more comprehensive production model is hence necessary.

2.2 Data sources

Production

Mine output and input data are obtained from the *Report of the Bureau of Mines* by the Department of Internal Affairs of Pennsylvania. It includes 615 Pennsylvania coal mines between 1898-1914. I observe annual coal extraction in tons and the share of output that is shipped, sold locally or re-used as inputs. Labor is measured in employee counts multiplied by the number of days worked. Intermediate inputs include black powder and dynamite, which are measured in quantities (kegs and pounds, respectively). I also observe the number of mules used at each mine.

Managers

The given, middle and surnames of all firm and mine managers and of their deputies are observed. I only focus on mine superintendents because firm managers and foremen were almost never college-educated. I match full manager names with population census records and with college alumni records using *Ancestry.com*. I also cross-checked all managers with alumni records from U.S. mining schools between 1870 and 1914, in order to ensure all mining college graduates are flagged as such.

Technology

A complication is that the number of locomotives of each type are observed at the county-firm-year level, while all other variables are observed at the mine-year level. The average firm operated 2.6 mines, but 80% of firms operated just one mine. The average county in the dataset contained 28 mines, the median county just four. Both attributing locomotives to the mine-level and mine managers to the firm-level requires ad-hoc weights. I choose to bring the entire dataset to the mine-level and assign locomotive usage evenly to all mines in a given county-firm-year pair. As such, it is assumed that upon adopting a locomotive, firms install them in all mines in a given county. A map with mining village locations and mining engineers is in appendix figure A1, and further details concerning the data sources and cleaning are in appendix A.

3 Managers, productivity and technology choice

3.1 Production and costs

Mines i extracted Q_{it} tons of coal in year t using variable inputs $\mathbf{V}_{it}$ and fixed assets $\mathbf{K}_{it}$. Variable inputs included labor, materials (black powder), and mules. Fixed assets included three types of mining locomotives: steam, electrical and/or compressed air locomotives. These types are denoted as $\tau \in \{\text{st, el, ca}\}$. The number of locomotives used of each type is denoted K_{it}^{τ}. The vector $\mathbf{K}_{it}$ contains these three locomotive types: $\mathbf{K}_{it} = (K_{it}^{st}, K_{it}^{el}, K_{it}^{ca})$. It was possible to operate mines without any locomotive, in which case $\mathbf{K}_{it} = (0,0,0)$. This was the case for a third of all mines. Multiple locomotive types could also be operated simultaneously. Some capital was, finally, required to dig the tunnels or shafts to reach the coal seam, but these costs are considered sunk and are not part of the production model.

Mines were managed by superintendents, with a dummy $X_{it} \in \{0,1\}$ indicating whether they obtained a mining college degree. Managers often managed multiple mines simultaneously. I will, however, consider input decisions to be independently made at each mine as markets are assumed to be perfectly competitive, and costs to evolve independently across mines. Let the production function be given by equation (1), with parametrization $\boldsymbol{\beta}$. In line with most of the literature, I rule out unobserved heterogeneity in the production parameters $\boldsymbol{\beta}$ across mines or over time. The educational background of a manager enters the production function as an input:

$$Q_{it} = F\Big(\mathbf{V_{it}}, \mathbf{K_{it}}, X_{it}; \boldsymbol{\beta}\Big)\exp(\omega_{it}) \tag{1}$$

As is usual in the productivity literature, the residual ω is assumed to be a scalar. As the main specification, I implement a Cobb-Douglas production function for $F(.)$., except that I allow for interaction effects between managerial education and each locomotive type. I hence allow for the presence of a manager with an engineering degree to change the output elasticity of each locomotive type, which is captured by the interaction effect β_{kx}. Denoting logarithms of variables in lowercases, the estimable equation is (2):

$$q_{it} = \beta_v \mathbf{v}_{it} + \boldsymbol{\beta}_k \mathbf{k}_{it} + \boldsymbol{\beta}_x \mathbf{X}_{it} + \boldsymbol{\beta}_{kx} \mathbf{k}_{it} \circ \mathbf{X}_{it} + \omega_{it} \tag{2}$$

In appendix B.1, I relax the functional form of $F(.)$ by allowing for factor-augmenting effects of mining locomotives, but this does not change the output elasticity of mining college graduates, which is the main coefficient of interest, by much.

Variable vs. fixed inputs

Labor, materials, mules and managers are assumed to be flexibly adjustable every year. Black powder and mules were sold on spot markets, and U.S. labor markets were very flexible and unregulated at the time (Naidu & Yuchtman, 2017). Locomotives $\mathbf{K}_{it}$ are assumed to be fixed assets. In the baseline model, I assume they were statically chosen every time period. The reason for this is that linking mines over time is bound to a lot of measurement error, which is problematic when estimating a dynamic technology choice model.[12] In an extension in appendix B.4, I allow for dynamic capital accumulation.

Variable input demand

I assume that mine managers chose variable inputs annually in order to minimize per-period variable costs. Let both product and input markets be perfectly competitive. I hence assume all mines made input decisions independently from each other. Prices of inputs $\mathbf{V}$ are denoted as $\mathbf{W}^{\mathbf{V}}$. Variable inputs were chosen by minimizing static costs in each period, taking output Q^* as given:

$$\mathbf{V_{it}} = \arg\min \left[\mathbf{W^V_{it}} \mathbf{V_{it}} - \lambda_{it}\Big(Q^*_{it} - F(.)\Big)\right] \tag{3}$$

I assume that there was perfect information about prices and output elasticities of all variable inputs. As managers are assumed to be variable inputs as well, demand for managers follows the same minimization problem as above, with the only difference that X is a binary variable.

Cost dynamics and capacity constraints

I abstract from cost dynamics and capacity constraints. Cost dynamics would be important if cumulative past productivity affected current productivity. This could have been the case for coal mines. Coal that can be reached at the lowest cost is usually mined first. Marginal costs are hence likely to increase as more coal is extracted (Aguirregabiria & Luengo, 2015; Asker, Collard-Wexler, & De Loecker, 2019). On the other hand, there could have been some ‘learning by doing’, as in Benkard (2000). I test for cost dynamics in appendix B.5, and find no strong evidence for them to be of first order importance. I also assume there were no capacity constraints. Data from Illinois coal mines from the same period show that 95% of mines operated under 90% of their maximum extraction capacity, which suggests that capacity constraints were mostly non-binding.

[12] As discussion in appendix A, I do not observe unique mine identifiers, but rather mine names which changed frequently.

3.2 Technology choice

Next, I examine how technologies were chosen. I distinguish three ways in which a mining degree could have affected the technology choices of a manager: by changing the returns to a technology, the costs of using a technology, or information about technology returns. The extent to which any of these three channels played a role is an empirical question.

Technology returns

As was shown in equation (2), I allow the human capital level of managers to affect the output elasticity of each locomotive type. As mechanical and electrical engineering courses were an important part of mining engineering curricula, it is possible that these graduates knew better how to operate each locomotive technology, thereby increasing their returns.

Technology costs

Using a locomotive of type τ required a fixed cost Φ_{it}^{τ} and a variable cost μ_{it}^{τ}. These costs could depend on the manager's educational background, for the reasons mentioned in the previous paragraph. Some fixed costs, such as laying railroad tracks, could have been shared between different locomotive types. I therefore let fixed costs depend on the usage of other locomotive types, which are denoted as $\tilde{\mathbf{K}}_{it}^{\tau}$. Fixed costs also depend on the location of each mine, due to the costs of transporting the technology to the mine, so I include dummies δ^{ℓ} for each town ℓ as a fixed cost shifter. Finally, there is a mine-year specific residual fixed cost ν_{it}^{τ}. I assume that managers observe all these cost components when deciding on which technology to use.

$$\Phi_{it}^{\tau} = \Phi^{\tau}(\mathbf{X_{it}}, \tilde{\mathbf{K}}_{it}^{\tau}, \delta_{\ell}) + \nu_{it}^{\tau} \tag{4}$$

Information and learning

As different locomotive types were newly invented, information about their returns was more likely to be imperfect compared to the other inputs, such as horses or materials.[13] I therefore assume that the manager of mine i had perfect information about all production function and cost coefficients, except for locomotive type coefficients $\boldsymbol{\beta_k}$. Together with the managerial education dummy and the interaction effect β_{kx}, these coefficients determine the output elasticity of each locomotive type θ_{it}^{τ}. At time $t = 0$, when each technology is invented, a manager of mine i has a private prior expectation

[13] One could also allow for imperfect information about costs, this could not be identified from imperfect information about returns in the current framework.

$\hat{\theta}^{\tau}_{i0}$ about the output elasticity of each locomotive type τ. The distribution $\mathcal{D}$ of this prior $\hat{\theta}^{\tau}_{i0}$ has the true output elasticity θ^{τ}_{it} as its mean, and a standard deviation σ^{τ}: $\hat{\theta}^{\tau}_{i0}(X_{it}) \sim \mathcal{D}\Big(\theta^{\tau}_{it}, \sigma^{\tau}(X_{i0})\Big)$

The standard deviation $\sigma^{\tau}(X_{i0})$ is a function of whether the manager has a mining degree, because these degrees contained information about new technologies. The hypothesis is that the standard deviation was lower for mines managed by a manager with a mining degree, as their prior was better informed. Whether this was indeed the case is an empirical question that will be addressed in the next section.

Each year, a signal u^{τ}_{it} arrived about the output elasticity of every locomotive type. The signal u^{τ}_{it} has an unbiased distribution $\mathcal{E}$, with a standard deviation ς^{τ} that is the same for all mines.

$$u^{\tau}_{it} \sim \mathcal{E}\Big(\theta^{\tau}_{it}, \varsigma^{\tau}\Big)$$

Using Bayesian updating, the updated prior in period t, $\hat{\theta}^{\tau}_{it}$, is a weighted average of the previous period's prior and of the signal u^{τ}_{it}, as long as the mine does not operate any locomotives of type τ. As soon as a locomotive of type τ is installed, managers immediately observe its output elasticity θ^{τ}_{it}. Denoting the indicator function as $\mathbb{1}(.)$, the updated prior is given by:

$$\hat{\theta}^{\tau}_{it} = \Big[\Big(1 - \alpha^{\tau}_{it}\Big)\hat{\theta}^{\tau}_{it-1} + \alpha^{\tau}_{it}u^{\tau}_{it}\Big]\left(\mathbb{1}(K = 0) + \theta^{\tau}_{it}\mathbb{1}(K \geq 0)\right)$$

The Bayesian weights α^{τ}_{it} depend on the relative standard deviations of the original prior, σ and of the information shocks, ς. In contrast with the prior literature in which managers have common priors, such as in Munshi (2004), I allow the weights α^{τ} do differ across managers: a mining engineer knows his priors are more precise because of his educational background.

$$\alpha^{\tau}_{it} = \alpha^{\tau}(X_{it})$$

If managers have a perfect signal about each technology, meaning that $\sigma^{\tau}(1, .) = 0$, then they do not place any weight on the signal u: $\alpha^{\tau}_{it} = 0$. Eventually, as time goes by, everyone becomes informed about all locomotive benefits. The signal u^{τ} then becomes a very precise estimate of the true coefficient β^{τ}, so managers only base their prior on this signal: α_{it} approaches unity.

Decision problem

I assume managers are risk neutral.[14] Each year t, they decide whether to use locomotives of each type. Normalizing (exogenous) coal prices to one, the net return to technology τ is denoted $\Delta^\tau\Pi$; this is the additional output generated by the technology τ minus the costs of using it:

$$\begin{aligned}\Delta^\tau \Pi_{it} &\equiv \Pi(., K_{it}^\tau = 1) - \Pi_{it}(., K_{it}^\tau = 0)\\ &= Q(., K_{it}^\tau = 1) - Q(., K_{it}^\tau = 0) - \mu_{it}^\tau K_{it}^\tau - \Phi_{it}^\tau\\ &= \theta_{it}^\tau Q_{it} - \mu_{it}^\tau K_{it}^\tau - \Phi_{it}^\tau\end{aligned}$$

As returns are uncertain, managers base their choice on their expectation about these returns, which is given by equation (5). As shown in the theoretical model above, expectations of the output elasticity $\hat{\theta}_{it}^\tau$ are a function of managerial education and of time. All other unobserved variables, such as fixed and variable costs, are observed to the manager before choosing the technology.

$$\mathbb{E}_{it}[\Delta^\tau \Pi_{it}] = \hat{\theta}_{it}^\tau(\theta_{it}^\tau, X_{it}, t) Q_{it} - \mu_{it}^\tau K_{it}^\tau - \Phi_{it}^\tau \tag{5}$$

The manager of mine i uses a locomotive τ if the expected additional profit from doing so is positive:

$$K_{it}^\tau \geq 0 \Leftrightarrow \mathbb{E}_{it}[\Delta^\tau \Pi_{it}] \geq 0$$

This extensive margin decision depends on both fixed and variable costs of the locomotive type, and on expected returns of that type. Conditional on using at least one locomotive of type τ, a manager chooses the number of locomotives that minimizes costs, similarly to the variable input decisions in equation 3. Fixed locomotive costs do not affect this intensive margin decision, only the output elasticity and variable costs do.

Static vs. dynamic problem

I assume managers chose locomotives by solving a static problem every year. In reality, this choice problem was likely to have a dynamic dimension, because locomotives were an investment that carried through for a number of years. I do not estimate a dynamic problem, however, for the same reason mentioned before: there is considerable measurement error involved when linking mines over time. Given that I do not aim to separately identify fixed from sunk costs, a reduced-form static approach suffices to compare technology usage across managers.

[14] Most existing papers on technology adoption under uncertainty assume risk aversion. This would, however, make the decision model dynamic: managers without a mining degree would value waiting longer before adopting, because their risk of adopting the wrong locomotive would become smaller. In the current model, managers only care about net returns from locomotives, which allows for a static decision model.

4 Empirical analysis

4.1 Production function

Distinguishing direct and indirect productivity effects of managers

The production function in equation (2) needs to be estimated in order to know the output elasticities of all technologies and of managerial education. I estimate two versions of equation (2). First, I omit the locomotive vector $\mathbf{k}_{it}$ from the production function.

$$q_{it} = \beta_v \mathbf{v}_{it} + \boldsymbol{\beta}_x \mathbf{X}_{it} + \omega_{it} \tag{2a}$$

Second, I include all locomotives as inputs to the production function, and interact them with the other inputs, but do not interact the locomotives variable with the mining engineering dummy, in order to get the average output elasticity of mining engineers controlling for all technology choices. If mining school graduates enter the production function directly as inputs, then their output elasticity should be higher than zero, even when controlling for all technologies. If these graduates do not enter the production function directly, but do choose better technologies, then their output elasticity should be positive when not controlling for all technology choices, but zero when doing so.

$$q_{it} = \beta_v \mathbf{v}_{it} + \boldsymbol{\beta}_k \mathbf{k}_{it} + \boldsymbol{\beta}_x \mathbf{X}_{it} + \omega_{it} \tag{2b}$$

Finally, I estimate the full version of equation (2), by also interacting each locomotive type with the managerial education dummy. This allows me to know whether mining school graduates changed the returns to each locomotive type:

$$q_{it} = \beta_v \mathbf{v}_{it} + \boldsymbol{\beta}_k \mathbf{k}_{it} + \boldsymbol{\beta}_x \mathbf{X}_{it} + \boldsymbol{\beta}_{kx} \mathbf{X}_{it} \circ \mathbf{k}_{it} + \omega_{it} \tag{2c}$$

The main coefficient of interest is the mining school graduates coefficient β_x, and how it changes between versions (2a)-(2c).

Identification

As more productive mines had higher demand for better educated managers, the usual endogeneity problem applies when estimating β_x. The usual approach to solve this endogeneity problem is to make assumptions about when the mines selected their managers, and under which information set, such as in Ackerberg, Caves, and Frazer (2015).[15] There are, however, two problems with taking this approach in the context of this paper. First, on the practical side, timing series variation within mines becomes crucial, but is subject to serious measurement error, as was already mentioned. Secondly, these 'control function approaches' take into account the decision process of the mines when choosing inputs, but it is *prima facie* not clear whether they also allow for a more complex matching process in which managers also choose mines, and in which there is assortative matching.

I use an instrumental variables approach instead. Suppose that there are mine characteristics Z_{it} that affected the likelihood of getting a mining college graduate manager, but did not enter the production function. These characteristics can be used as an instrument for managerial education X_{it}. In appendix B, I outline an entry model of mining college graduates into mine locations that yields two instruments $\mathbf{Z}_{it}$. These are (i) the average age of the other managers in town ℓ at time t, and (ii) the average number of mining college graduates in competing mines in town ℓ at time t. The reasoning of why these instruments are relevant is as follows. Given that mining engineering graduates were scarce, they had some wage-setting power, and hence preferred to enter labor markets without any prior mining engineering graduates. This is in a similar vein to oligopoly entry models such as Bresnahan and Reiss (1991). Secondly, assume that location tastes were correlated among managers of the same age cohort. Given that mining school graduates were much younger than the average manager, they probably preferred to live in locations where the average age of the managerial workforce was younger as well.

I add a third instrument, which relies on mining college locations. I construct a dummy for whether the mine lies in a county that is adjacent to both counties with mining colleges.[16] Mines that are located near these mining colleges were more likely to attract their graduates. The locations of these universities were decided many years before these colleges offered mining engineering programs, and therefore were not based on the characteristics of the anthracite mines in the panel.

Discussion of the exclusion restriction

The exclusion restriction for these instruments $\mathbf{Z}_{it}$ is that they affect productivity of mine i only through their effect on managerial education X_{it}, and not by directly changing total factor productivity ω_{it}:

$$\mathbb{E}(\mathbf{Z}_{it}'\omega_{it}) = 0$$

The main underlying assumption is that there were no productivity spillovers between managers of

[15] When mentioning 'mines' as a decision-maker, I implicitly mean the mine owners.
[16] These were Lehigh county and Northampton county.

different mines in the same town. Hence, managerial characteristics of other mines do not affect a mine's productivity directly, but do affect its likelihood to attract a mining school graduate as manager. Given that managers were competing against each other, they did not have incentives to augment productivity levels of each other's mines .

I support this exclusion restriction with two placebo tests. First, I regress the productivity residual on the average age of other managers in the town both before and after the first introduction of mining college graduates (in 1900). If managerial characteristics of other mines would affect productivity in other ways than through the effect on hiring a mining college graduate, then the estimated coefficient should be positive both before and after the introduction of mining college graduates. Panel (a) of appendix table A1 shows this was not the case. Before 1900, the average age of other managers did not correlate significantly with productivity, but after 1900 it became negatively correlated, which is consistent with the managerial entry model.

Second, I regress the productivity residual on the share of other managers with a mining degree for both mines which were family-operated, and mines which were not.[17] Family-operated mines did not hire external managers, and none of these mines therefore hired mining college graduates. If the exclusion restriction is met, then the presence of mining college graduates in other mines of the same town should not be correlated with productivity for these family-operated businesses. Panel (b) in appendix table A1 shows this was again the case.

Finally, I test for overidentifying restrictions using the mining locations instrument in panel (c) of table A1. The Sargan ξ^2 statistic is 4.23. If the exclusion restriction is met for the adjacent mining college instrument, it is therefore also met for the other two instruments.

Estimation

I estimate the production functions (2a)-(2c) by instrumenting for managerial education. The coefficients on the other variables are therefore still biased, but I am mainly interested in the mining school coefficients.[18] As shown in equation (2), I measure both output and all inputs in logs, except for the mining college graduate dummy. For each locomotive type, I calculate the log number of locomotives of that type, but add one, in order to capture their effects both at the intensive and extensive margin. I control for a time trend and for other managerial characteristics, such as the manager's age and whether he obtained a college degree in a field other than mining engineering.

I construct the instruments as follows. For each mine, I calculate the average age of all other managers of other mines in the same town. I also calculate the share of other mines managed by different

[17]Details on how I measure family ownership are in appendix B.2.

[18]In appendix B.4, I combine the instruments for managerial education with a control function to identify the other production function coefficients.

managers who have a mining college graduate as manager. Both these averages are calculated at the town level. In model (2c), I interact each locomotive type with the mining school graduate dummy. I therefore also interact all three instrumental variables with each locomotive type.

Results

The estimated production function coefficients are in panel (a) of table 1. I only report the coefficients of interest. When not controlling for locomotive usage, in column (2a), the mining engineer coefficient is 0.254. This implies that a mine's output is 29% higher when managed by a mining college graduate, keeping all inputs, except locomotives, fixed. This large effect is consistent with the visual evidence from figure 4. Once I control for the usage of all locomotive types, in column (2b), the mining college graduate coefficients drops, however, to 0.069 and is no longer significantly different from zero. This implies that managers do not enter the production function directly as inputs, but do indirectly by choosing different technologies. Column (2c) reports, finally, the interaction terms between managerial education and locomotive types. The coefficient on the mining degree β_x is now no longer an output elasticity. None of the interaction effects with locomotive types are significant. Managers with mining degrees therefore did not increase locomotive returns.

[Table 1 here]

The coefficients on the locomotive types are included as well, but they are not instrumented for, so caution is required when interpreting them. It is, nevertheless, remarkable that the output elasticity of electrical and steam locomotives is significant and positive, while it is negative for compressed air locomotives. This is consistent with historical evidence that compressed air engines were plagued by many problems, as discussed previously in section 2.

The first stage estimates are in panel (b). Mining engineers are around 25% more likely to join mines in the county adjacent to a mining college. They are also more likely to join mines where the average age of other managers is lower, and where there are less managers with mining college degrees. The F-statistic lies between 40 and 70 across specifications, and the R-squared between 0.35 and 0.40, so the instruments are sufficiently strong.

4.2 Technology choice

Estimable equation

The choice model in equation (5) implies that the usage of locomotives τ was a function $g^{\tau}(.)$ of expected technology returns $\hat{\theta}^{\tau}_{it}$, fixed technology costs Φ^{τ}_{it}, variable technology costs μ^{τ}_{it}, and output:

$$\mathbb{1}(K^{\tau} \geq 0) = g^{\tau}(\hat{\theta}^{\tau}_{it}, \mu^{\tau}_{it}, \Phi^{\tau}_{it}, Q_{it})$$

From equation (4), we know that fixed costs depended on the mine's location δ^{ℓ}, managerial education X_{it}, other locomotive types $\tilde{K}^{\tau}_{it}$, and residual fixed cost variation ν^{τ}_{it}. Expected technology returns depended on time t, because of Bayesian updating, managerial education X_{it}, and on variable input usage $\mathbf{v}_{it}$ (because of potential factor-augmenting technology effects). Substituting these expressions into the choice function, the estimable equation therefore becomes:

$$\mathbb{1}(K^{\tau} \geq 0) = \tilde{g}^{\tau}(t, X_{it}, Q_{it}, V_{it}, \tilde{K}^{\tau}_{it}, \delta_{\ell}, \nu^{\tau}_{it}, \mu^{\tau}_{it}, \theta^{\tau}) \tag{6}$$

Estimation

I estimate equation (6) by regressing locomotive dummies for each type on a linear time trend, a mining school dummy, town fixed effects, log output and log variable inputs, and on usage of other locomotive types. The residual contains both fixed and variable costs. I use a linear probability model as the main specification, but estimate a discrete choice model as a comparison in appendix B.3.

I relay on the same identification strategy to account for the endogeneity of managerial characteristics as when estimating the production function: I use the average age and educational background of managers in competing mines as instruments for the own mine's managerial characteristics. Besides ruling out productivity spillovers across managers, I now also rule out such spillovers in terms of fixed costs and information. In an environment of imperfect information, managers have all the incentives to keep technology returns private. I no longer use the adjacent county dummy as an instrument, as I now control for town fixed effects. I cluster standard errors at the town-year level.

Results

The estimates of the technology choice model from equation (6) are in panel (a) of table 2. I report the average effects at the mean and do not report the estimates for all control variables. The first stage regressions are the same as in the production model. Managers with mining degrees are 62.3

percentage points more likely to use electrical locomotives than other managers,[19] which is a very large effect: initially, almost all electrical locomotives were adopted by mining college graduates. These managers were, however, less likely to adopt both other locomotive types. Managers with other college degrees did, in contrast, adopt less locomotives of all type, and were more likely to keep using mules as a hauling technology.

[Table 2 here]

Managers with mining degrees were less likely to use steam locomotives, even if their output elasticity was estimated to be positive. Steam locomotives were, however, already very widely used at the start of the sample period (at 60% of mines), and their usage did not grow much over time. The mines that did not use steam locomotives usually did not because of the presence of mine gas, which made subterranean combustion risky. The real choice was therefore between electrical and compressed air engines.

4.3 Why did mining engineers adopt different technologies?

The previous section documented that mining college graduates were much more likely to use electrical locomotives, compared to other managers. As outlined before, the reason for this can be that these managers in terms of (i) technology returns, (ii) fixed technology costs, (iii) variable technology costs, and (iv) information about technology benefits.

Variable vs. fixed cost differences

Mechanism (i), a different return, was already ruled out due to the production function estimates. Mechanism (iii), different variable costs, can be tested by comparing intensive to extensive margins: the dependent variable in equation (6) was $\log(K_{it}^{\tau}+1)$, and hence contained information on both the intensive and the extensive margin. If the educational mechanism related to fixed costs or information, only the decision to adopt electrical locomotives should differ between managers. If variable costs were different, then mining college graduates should also use more electrical locomotives.

The estimates in panel (b) of table 2 show that this was not the case. Conditional on using at least one electrical locomotive, mining engineers did not adopt significantly more electrical locomotives than other managers. If there was any difference in technology operating costs between managers, it should hence be a difference in fixed costs, not in variable costs.[20] Managers with college degrees in

[19] In the discrete choice model in appendix B.3, usage probabilities are bound between zero and one, and the effect of mining college graduates on electrical locomotive usage falls to 30.3 pp.

[20] Mining engineers did use more steam locomotives conditional on having at least one steam locomotive (which they were less likely to have), but steam locomotives were mainly adopted before the sample period started anyway.

other fields used less locomotives of all types.

Fixed costs vs. information

The two remaining mechanisms behind the different locomotive choices are differences in fixed costs and in information. I rely on two empirical implications from the information acquisition model in section 3 to test these mechanisms. First, if the information channel is important, then the differences in electrical locomotive usage between mining college graduates and other managers should narrow over time, and end up being zero, as everyone eventually finds out which locomotives work best. If fixed cost differences were the main reason for adoption differences, then these differences should persist over time. Second, the adoption rate of compressed air locomotives should fall over time for managers without a mining degree, as they gradually find out that these engines were not very useful. If mining college graduates were, in contrast, perfectly informed from the start, then this pattern should not hold for them.

I test both predictions in table 3. In panel (a), I estimate equation (6) for three different time blocks of five years. The difference in electricity usage between managers were the largest in the first five years of the sample, narrowed to 45.8 percentage points in the next five years, and dropped to being not significantly different from zero in the last 4 years of the sample. This is consistent with the information mechanism, but not with the fixed costs mechanism.

[Table 3 here]

In panel (b), I look how adoption rates of compressed air locomotives, being defined as going from no air locomotives to at least one air locomotive, evolved over time for both types of managers. I use the same specification as in equation (6), but with adoption rather than usage on the left-hand side. I report the coefficient on the time trend, which is measured by the number of decades elapsed since 1900. I find that compressed air adoption did not change significantly over time for mining engineers, but fell for other managers. Over a period of 10 years, compressed air adoption dropped by 3.1 percentage points on average for these other managers, on an average adoption rate of 3.9 pp. In other words, practically all the adoption of compressed air locomotives by non-educated managers took place in the beginning of the time period. This is consistent with the hypothesis that non-educated managers were not informed about compressed air locomotives initially, but learned about them over time.

Why not all firms hired mining engineers

Given that their technology choices were much better, it seems surprising that so few mine owners hired mining college graduates. The supply of mining college graduates was, however, still very low during the first decade of the 20th century. In 1909, after the large entry wave of mining colleges, only 4000 people had ever graduated from a U.S. mining college program, and some of these could have already died by then. In comparison, there were over 6000 coal mines in the U.S.A. and many more in other extractive industries such as copper or gold mining. College yearbooks show that many mining engineers migrated to work in Mexican, Canadian or Australian mines, and many also entered different industries, became civil servants or academic researchers (Ochs, 1992). Coal mining was also more competitive, and hence less profitable, than some other extractive industries. It was therefore a less attractive industry for highly solicited mining college graduates. Finally, the imperfect information about technology returns also imply that there was imperfect information about mining college graduate returns, at least initially. As the only difference between mining college graduates and other managers lay in their technology choices, owners had to know which technologies worked best in order to know which managers were making the best technology choices.

5 Conclusion

In this paper, I examine the effects of a shock to managerial human capital on both total factor productivity and on technology choices. I show that managers can have a positive effect on total factor productivity, even if their output elasticity is zero, if they make better technology choices. I illustrate this using the historical case of the entry of mining college graduates in managerial positions in the early 20th century Pennsylvania coal mining industry. I find that these managers did not increase mining productivity directly, as inputs in the production function, but did increase productivity by large amounts *indirectly*, by choosing better mining locomotive technologies. Finally, I find that access to better information about new technologies was the main driver of these better decisions, rather than differences in the costs or returns of new technologies. The findings in this paper imply that the differentiation of U.S. higher education from the late 19th century onwards had important consequences for innovation and productivity growth of firms. These gains were, however, conditional on the arrival of new technologies, and on the imperfect nature of information about these technologies.

References

Ackerberg, D., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*(6), 2411-2451.

Aguirregabiria, V., & Luengo, A. (2015). *A microeconometric dynamic structural model of copper mining decisions.* (Mimeo)

Asker, J., Collard-Wexler, A., & De Loecker, J. (2019). (Mis) Allocation, market power, and global oil extraction. *American Economic Review*, *109*(4), 1568-1615.

Bender, S., Bloom, N., Card, D., Van Reenen, J., & Wolter, S. (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics*, *36*(S1), S371–S409.

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

Benson, A., Li, D., & Shue, K. (2019). Promotions and the Peter principle. *The Quarterly Journal of Economics*, *134*(4), 2085–2134.

Bertrand, M., & Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, *118*(4), 1169-1208.

Besley, T., & Case, A. (1993). Modeling technology adoption in developing countries. *The American economic review, Papers and Proceedings*, *83*(2), 396–402.

Bianchi, N., & Giorcelli, M. (2019). Scientific education and innovation: from technical diplomas to university stem degrees. *Journal of the European Economic Association*.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? evidence from India. *The Quarterly Journal of Economics*, *128*(1), 1-51.

Bloom, N., Sadun, R., & Van Reenen, J. (2017). *Management as technology?* (NBER Working Paper No. 22327). National Bureau of Economic Research.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, *122*(4), 1351-1408.

Braguinsky, S., Ohyama, A., Okazaki, T., & Syverson, C. (2015). Acquisitions, productivity and profitability: evidence from the Japanese cotton spinning industry. *The American Economic Review*, *105*(7), 2086-2119.

Bresnahan, T., & Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, *99*(5), 977-1009.

Caroli, E., & Van Reenen, J. (2001). Skill-biased organizational change? evidence from a panel of British and French establishments. *Quarterly Journal of Economics*, *116*(4), 1449-1492.

Church, J. A. (1871). Mining schools in the United States. *The North American Review*, *112*(230), 62-81.

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American*

Economic Review, *100*(1), 35-69.

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing from others: Human capital and technical change in agriculture. *Journal of Political Economy*, *103*(6), 1176-1209.

Gairns, J. (1904). Industrial locomotives for mining, factory and allied uses. In (Vol. 26, p. 291-309). Cassier's Magazine.

Hodges, C. (1905). Compressed air locomotives for mine haulage. *Mining Magazine*, *11*(3), 305-310.

Hoffman, M., & Tadelis, S. (2018). *People management skills, employee attrition, and manager rewards: An empirical analysis* (Tech. Rep.). National Bureau of Economic Research.

Jovanovic, B., & Nyarko, Y. (1996). Learning by doing and the choice of technology. *Econometrica*, *64*(4), 1299-1310.

Lazear, E. P., Shaw, K. L., & Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, *33*(4), 823–861.

Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G., Gilbert, R., & Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brookings papers on economic activity*, *1987*(3), 783–831.

Lleras-Muney, A., & Lichtenberg, F. R. (2005). Are the more educated more likely to use new drugs? *Annales d'Économie et de Statistique*, 671–696.

Lundgreen, P. (1990). Engineering education in Europe and the U.S.A., 1750-1930: The rise to dominance of school culture and the engineering professions. *Annals of Science*, *47*(1), 33-75.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of development Economics*, *73*(1), 185–213.

Naidu, S., & Yuchtman, N. (2017). Labor market institutions in the gilded age of American history. *The Oxford Handbook of American Economic History*, *Forthcoming*.

Ochs, K. H. (1992). The rise of American mining engineers: A case study of the Colorado School of Mines. *Technology and Culture*, *33*(2), 278-301.

Randolph, B. S. (1905). Comparison of electric and compressed-air locomotives in American mines. In M. W. Brown (Ed.), *Transactions of the institution of mining engineers* (Vol. 27). Institution of Mining Engineers.

Rosenzweig, M. R. (1995). Why are there returns to schooling? *The American Economic Review*, *85*(2), 153–158.

Skinner, J., & Staiger, D. (2005). *Technology adoption from hybrid corn to beta blockers* (Tech. Rep.). National Bureau of Economic Research.

Terviö, M. (2009). Superstars and mediocrities: Market failure in the discovery of talent. *The Review of Economic Studies*, *76*(2), 829–850.

Toivanen, O., & Väänänen, L. (2016). Education and invention. *Review of Economics and Statistics*, *98*(2), 382–396.

Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *The Review of Economic Studies*, *70*(1), 167-198.

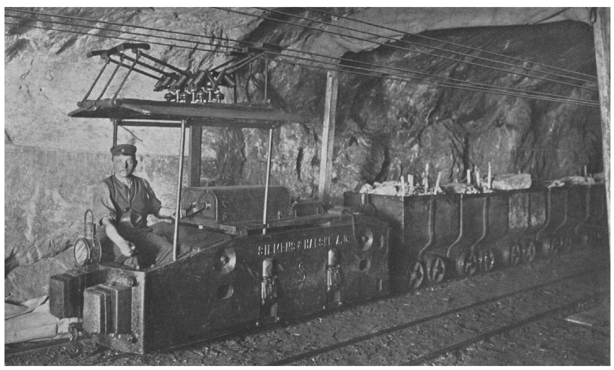
Wozniak, G. D. (1984). The adoption of interrelated innovations: A human capital approach. *The Review of Economics and Statistics*, 70–79.

Wozniak, G. D. (1987). Human capital, information, and the early adoption of new technology. *Journal of Human Resources*, 101–112.

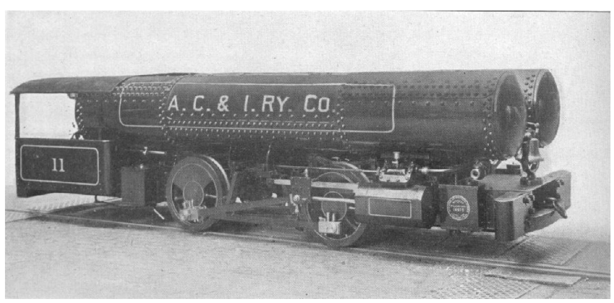
Figure 1: Mining locomotive types

(a) Steam

(b) Electrical



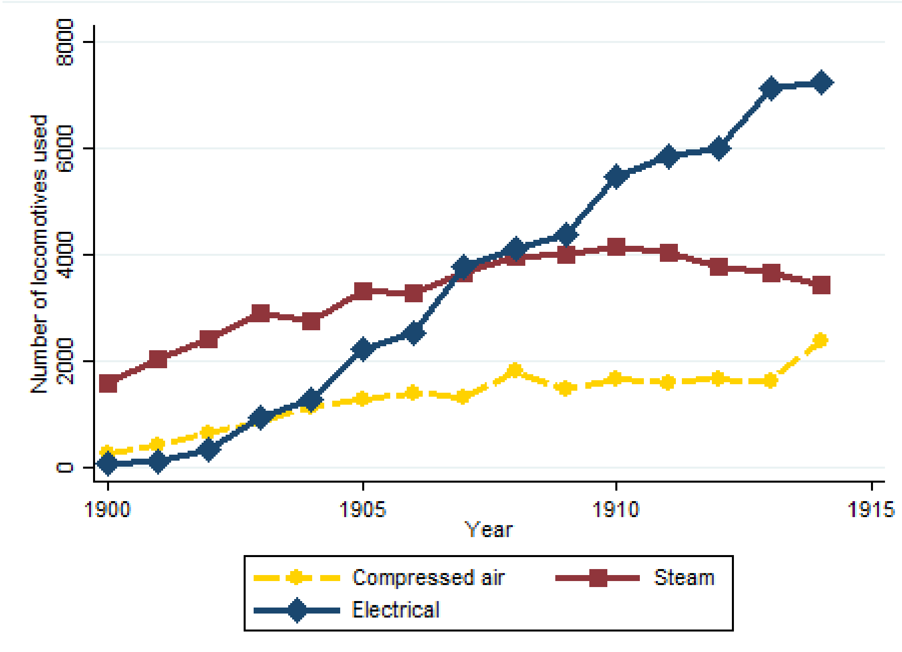
(c) Compressed air



Source: Gairns (1904)

Figure 2: Usage of mining locomotives

(a) Total number in use



(b) Share of mines

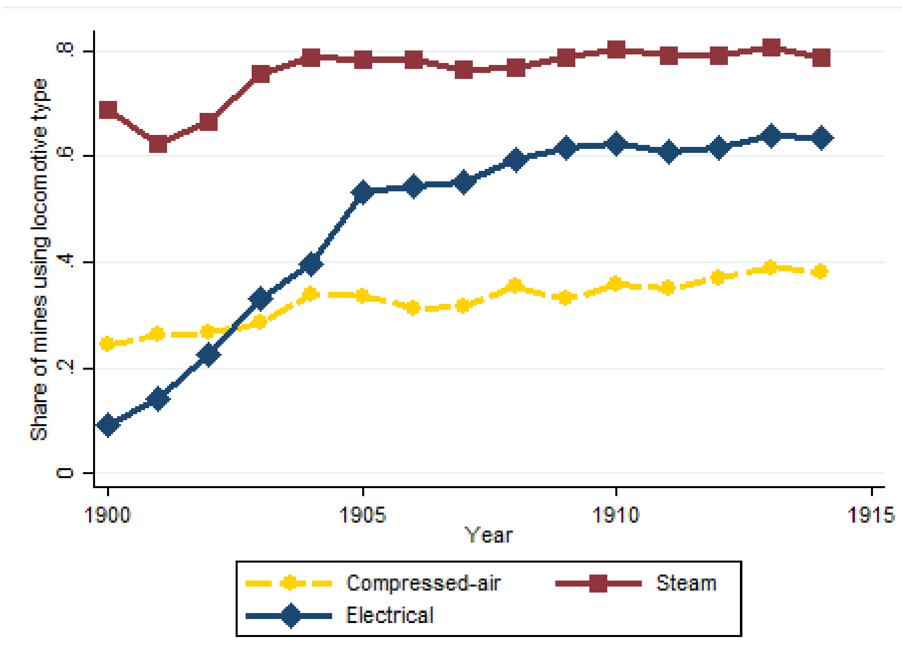
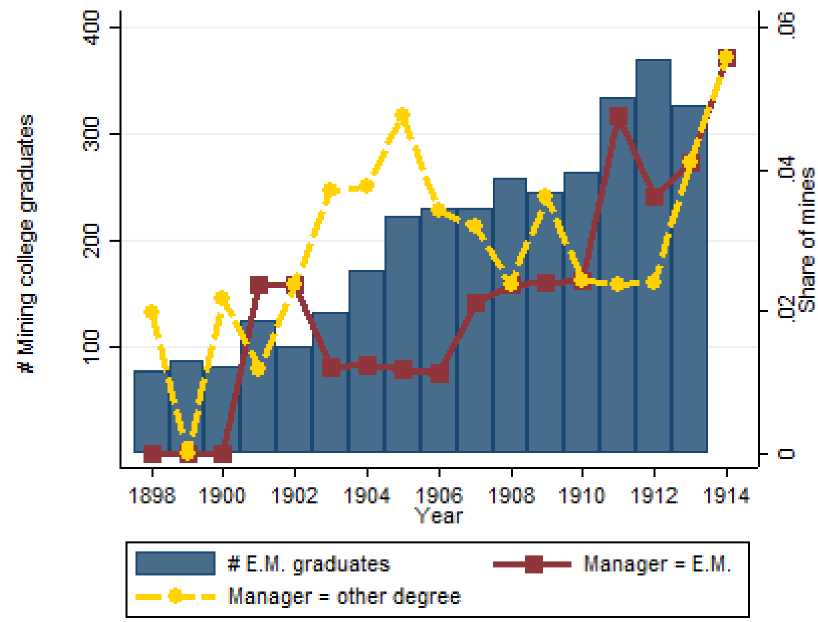
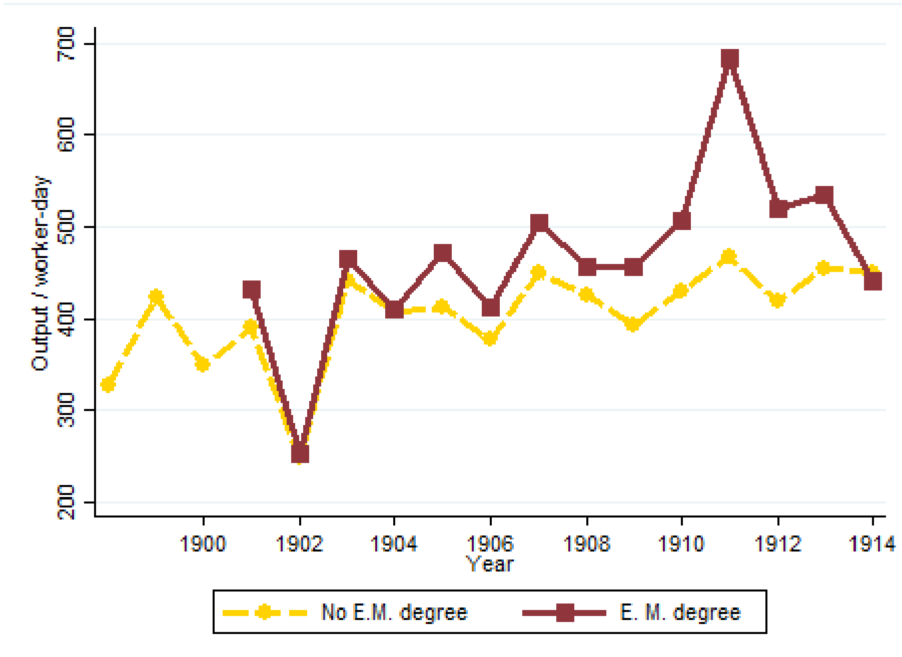


Figure 3: The rise of U.S. mining colleges



Notes: The blue bars show the annual number of graduates (left axis) with an Engineer of Mines (E.M.) degree from all U.S. mining colleges. The solid red line plots the share of Pennsylvania anthracite mines with a manager with such an E.M. degree, the dotted yellow line does the same for other college degrees.

Figure 4: Labor productivity and managerial education



Notes: The solid red line shows plots the evolution of output per worker per day in mines managed by a mining college graduate. The dotted yellow line does the same for mines without such a manager.

Table 1: Production function

Panel (a): Production function	Model (2a)		Model (2b)		Model (2c)	
	Estimate:	SE:	Estimate:	SE	Estimate:	SE
β_x (mining degree)	0.254	(0.120)	0.069	(0.126)	-0.018	(0.301)
β_k^{el} (electrical loc.)			0.034	(0.008)	0.044	(0.009)
β_k^{st} (steam loc.)			0.020	(0.010)	0.009	(0.014)
β_k^{ca} (compr. air loc.)			-0.018	(0.008)	-0.010	(0.011)
β_{xk}^{el} (md x elec. loc.)					-0.203	(0.117)
β_{xk}^{st} (md x steam loc.)					0.161	(0.137)
β_{xk}^{ca} (md x compr. air loc.)					-0.109	(0.277)
Observations	3,349		3,349		3,349	
R-squared	0.871		0.874		0.870	

Panel (b): First stage estimates	Model (2a)		Model (2b)		Model (2c)	
	Estimate:	SE:	Estimate:	SE	Estimate:	SE
Avg. age * 10	-0.058	(0.007)	-0.055	(0.007)	-0.041	(0.010)
Avg. mining degrees	-0.734	(0.101)	-0.803	(0.116)	-0.073	(0.157)
Carbon county	0.248	(0.101)	0.230	(0.055)	0.177	(0.102)
F-stat	66.00		52.68		36.61	
R-squared	0.353		0.363		0.407	

Notes: Panel (a) contains the second-stage production function coefficients. Dependent variable is log output. All variables in logs, except for the mining college dummy. Model (2a) does not control for the different locomotive types used, model (2b) does control for all locomotive types, and model (2c) interacts the locomotive types with the managerial education dummy. Panel (b) contains the first-stage estimates. Dependent variable is the managerial education dummy.

Table 2: Technology choice

Panel (a): Extensive margin	$\mathbb{1}$(Elec. loc.)		$\mathbb{1}$(Steam loc.)		$\mathbb{1}$(Air loc.)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
1(Mining degree)	0.623	(0.136)	-0.352	(0.148)	-0.064	(0.163)
1(Other degree)	-0.197	(0.084)	-0.120	(0.085)	-0.034	(0.066)
Observations	3,357		3,357		3,357	
Mean usage:	0.426		0.676		0.289	
R-squared	0.467		0.439		0.429	

Panel (b): Intensive margin	log(Elec. loc.)		log(Steam loc.)		log(Air loc.)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
1(Mining degree)	0.519	(0.504)	-0.740	(0.435)	0.506	(0.215)
1(Other degree)	-1.742	(0.193)	-1.772	(0.230)	-0.511	(0.111)
Observations	1,548		1,101		2,370	
R-squared:	0.536		0.449		0.522	

Notes: Dependent variables are dummies for usage of each locomotive type in panel (a) and log numbers of each locomotive in panel (b). Panel (a) uses a linear probability model with same instrumental variables as in production model. Controls: size, variable input usage, other locomotive types usage, time trend, town dummies.

Table 3: Mechanism: fixed costs or information?

Panel (a): Electricity usage over time	Additional usage by E.M. managers	
	Estimate:	SE:
1900-1905	0.694	(0.129)
1906-1910	0.458	(0.100)
1911-1914	-0.291	(0.628)
Panel (b): Compressed air adoption over time	Time trend coefficient	
	Estimate:	SE:
Mining engineers	0.077	(1.289)
Other managers	-0.031	(0.008)

Notes: Panel (a) reports the effect of mining school graduates on electrical locomotive usage compared to other managers for three different time period. Panel (b) reports the estimated coefficient on the time trend on additional adoption of compressed air locomotives for both mining engineers and other managers.

Appendices

A Data sources

A.1 Production and cost data

Raw data

Data on output, inputs, managers, technical characteristics, ownership and locations of mines were obtained from the *Report of the Bureau of Mines* by the Department of Internal Affairs of Pennsylvania. Geographical coordinates were obtained from Google Maps. A full list of all variables used, and their characteristics, is in table A7. The data structure is unchanged between 1900 and 1914 and is composed of four tables per county. A first table lists all mines, their owners, the managers, a post office location and the railroad to which it is connected. This is shown in table A2. A second table provides production and cost data at the mine level, a sample is in figure A3. Thirdly, technology choices are reported in a third table, at the firm-county-year level, see A4. Fourthly, the occupational breakdown of labor is given in a fourth table, again at the firm-county-year level. Yearly prices of Pennsylvania anthracite coal are obtained from the *Statistical Abstract of the United States*.

Data cleaning

I make unique mine identifiers by tracking mine name changes over time. It happens that mines have multiple sub-units which merge or split over time. I collect these sub-units to the mine-level in order to have a unit which does not change over time. I sum all inputs and outputs of the sub-units to this mine-level. For the number of days worked, I calculate the means across sub-units. Locomotives are given at the county-firm-year level, rather than the mine level. I assign locomotives evenly to all mines belonging to the same firm-county-year pairs.

Table A8 shows some summary statistics on the mines. Annual extraction was on average 0.21 Mton. The average mine had 428 employees who worked for 195 days per year.

A.2 Management data

As explained in the main text, I matched the managers in the production data to their full names, years and residences with Federal Census records using *Ancestry.com.* I checked whether the listed occupations were correct (e.g. 'Coal operator' or 'Mine superintendent'). Next, I retrieved birth years and matched them to the full names and birth years with the alumni records, both in my own mining college alumni lists and through *Ancestry.com*.

In principle, full manager names are given in the data. Sometimes, however, only the first letter of the second name is given, or a shorter version of the given name (Joe vs. Joseph). I encode unique manager identifiers by looking the managers up in the U.S. census through *Ancestry.com* and comparing their location in the data to the location in the census, and the observed years in the data with their age in the census.

A list of mining schools on which I have curriculum information is in table A11. Six out of fifteen institutions are specialized in mining engineering, offering no other fields of study. Columbia University was the only private elite university offering a mining degree, all other mining schools and universities in our sample are public and generally younger. Only two schools trained mining engineers before 1885. The mean annual cohort does not exceed 30 students, and is on average 17 students. These small class sizes were customary in engineering education these days, and were considered as being beneficial for educational quality by contemporary professors (Church, 1871).

B Managerial entry model

A mining school graduate m has to choose a mine i in year t, which is located in location ℓ. The utility of manager m joining mine i in year t is denoted U_{imt} and depends on the managerial wage W_{it}^{m}, and on the mine location dummy δ_i^{ℓ}, for which each manager has a valuation ζ_m^{ℓ}:

$$U_{imt} = U(W_{it}^{m}, \delta_{it}^{\ell}, \zeta_m^{\ell})$$

Wages increase a manager's utility: $\frac{\partial U}{\partial W^m} \geq 0$ and $\frac{\partial U}{\partial \delta} \geq 0$. Suppose locations ℓ are also the local labor market, and let $N_{\ell t}$ be the number of mining college graduates in location ℓ. I assume mining college graduates have some market power when setting their salary. This is likely to be true, because there were few mining college graduates at the time. Similar to oligopoly entry models, such as Bresnahan and Reiss (1991), managerial wages will be lower in locations with more mining college graduates: $\frac{\partial W_{it}^{m}}{\partial N_{\ell t}} < 0$. Managers will hence be more likely to enter locations with less mining college graduates.

Suppose now that the taste parameter for location characteristics, β_m^{ℓ} is correlated between managers of the same age cohort. A manager will then be more likely to enter a location where the age composition is more similar to his own age. Given that mining college graduates were much younger than the average manager, they were more likely to enter locations where other managers were, on average, younger.

B.1 Factor-biased effects

The main specification of the production function was Hicks-neutral in all inputs. I now extend this model to allow for interaction effects between the locomotive technologies and the other inputs. In the first two columns of table A2, I estimate production function (2b), but allow for interaction effects between each locomotive type and both labor and mules. Both electrical and steam locomotives interact negatively with mules, as expected: locomotives replace mules. The interaction effect between compressed air engines and mules is, however, positive. yet the key thing to note is that the coefficient on mining college graduates is very similar as in the main specification of equation (2b).

In the last two columns of table A2, I interact between locomotive types. I find no evidence for complementarities between locomotive types, except for a negative interaction effect between steam and compressed air locomotives. The estimated coefficient on mining college graduates is again very similar to the baseline model.

B.2 Additional controls

I add various additional controls to the production function which may drive both managerial change, technology choice and productivity. First, I include a dummy for ownership changes, as this potentially affects both management and firm performance (Braguinsky et al., 2015). I define an ownership dummy to be one if the firm owning the mine changed.

Next, I include a dummy which indicates whether a mine ships at least some of its output over the railroad network to other towns. This distinguishes purely local producers from 'exporters' (even if they did not necessarily export to foreign countries). Connection to the railroad network is also relevant because it may have affected the transport cost to acquire a mining locomotive.

Thirdly, I control for whether the mine was a family business by checking whether the managers and foremen had the same surname (but a different given name). This is relevant because the choice of the manager and the ownership structure of the mine may both be correlated with firm performance and innovation decisions.

Both the production function and technology choice estimates with these additional controls are in panels (a) and (b) of table A3. They are very similar to those in the main specification. More productive mines are more likely to be connected to the railroad network, as expected, and more likely to change ownership. Being a family business was not significantly correlated with either productivity or technology choices.

B.3 Discrete choice model

The main specification used a linear probability model. In table A5, I estimate a probit model instead, while still instrumenting for managerial education using the same instruments as before. I report the marginal effects at the mean. Mining college graduates are now 30 percentage points more likely to use electrical locomotives, which is a smaller effect compared to the linear probability model. They are also 25pp less likely to use steam locomotives, and as likely to use compressed air locomotives as the other managers. The number of observations varies now depending on the locomotive type, because towns where none of the mines (or all mines) use a certain locomotive type are dropped.

B.4 Alternative production function identification strategy

In the main text, I estimated the output elasticity of mining college graduates using instrumental variables. In this appendix, I use the approach of Ackerberg et al. (2015) as a comparison. In addition to the model in the main text, an equation of motion is now imposed on both the capital stock and total factor productivity. Capital depreciates at rate $(1-\rho)$ and grows through investment $\mathcal{I}$, as shown in equation (7):

$$K_{it}^{\tau} = K_{it-1}^{\tau}\rho + \mathcal{I}_{it-1}^{\tau} \tag{7}$$

Productivity follows an AR(1) process, with unexpected shocks ξ_{it}:

$$\omega_{it} = h(\omega_{it-1}) + \xi_{it}$$

Identification

As in the main text, I assume that labor, mules, materials and managers are variable static inputs, using the classification in Ackerberg et al. (2015). Capital is still fixed, but is now chosen using a dynamic model through capital investment $\mathcal{I}$. The moment conditions follow Ackerberg et al. (2015): it is assumed that variable inputs and managers are chosen after the productivity shock ξ is observed by the manager, while capital investment is chosen before this shock materializes. I use the same instruments for mining college graduates, $\mathbf{z}$, as before.

$$\mathbb{E}\left\{\xi_{it}(\boldsymbol{\beta}_v, \boldsymbol{\beta}_k, \boldsymbol{\beta}_{vk}, \boldsymbol{\beta}_x)\begin{pmatrix}\mathbf{v_{it-1}} \\ \mathbf{k_{it}} \\ \mathbf{z_{it}} \circ \mathbf{k_{it}} \\ \mathbf{z_{it}}\end{pmatrix}\right\} = 0 \tag{8}$$

Estimation

I follow the estimation procedure from $Ackerberg\ et al.(2015)$. In the first stage, I estimate a second-order polynomial in all inputs, and include the instruments for mining college graduates in the input demand conditions. The second stage includes the same functional forms as used in the main text, models (2a)-(2c).

Results

The main estimated coefficients of interest of equation (2) are in table A6. The estimated coefficient for mining engineers is now much higher compared to the instrumental variables specification, at 0.608, but it is much less precisely estimated and not significantly different from zero. This is not abnormal when using ACF on smaller-scale datasets. The same pattern still holds, however, that the coefficient estimate shrinks to close to zero once technology choice is controlled for.

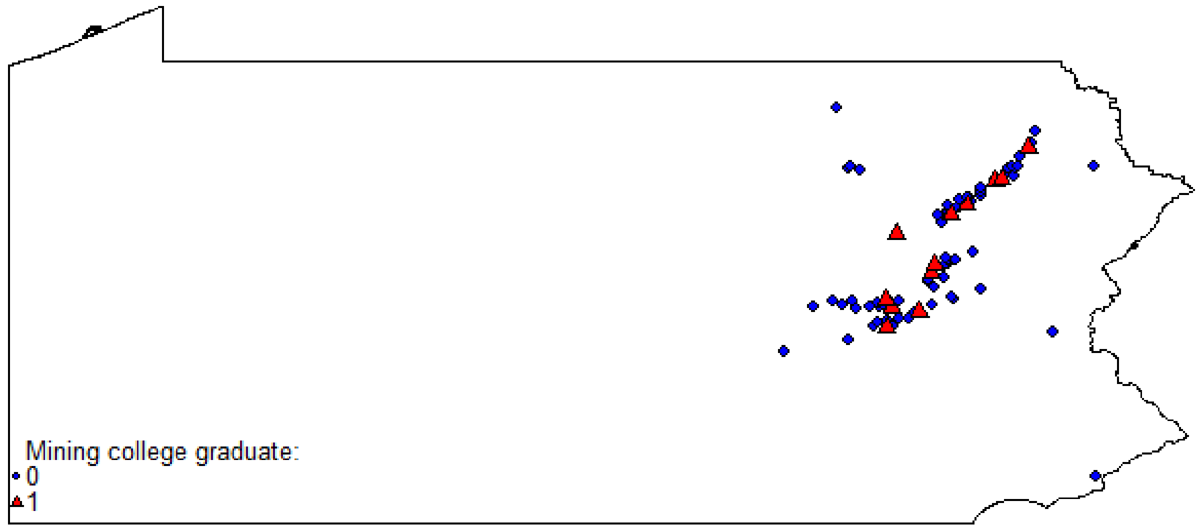
B.5 Cost dynamics

Suppose there would be important cost dynamics in coal mining. The productivity transition then depends on past cumulative output:

$$\omega_{it} = \tilde{h}(\omega_{it-1}, C_{it-1}) + \xi_{it} \quad \text{with} \quad C_{it} = \sum_{k=1}^{t} Q_{ik}$$

I test whether such cost dynamics are important by regressing the productivity residual from the static production model, (2b), on both current output (to allow for scale returns) and on lagged cumulative output. The results are in appendix table A4. In the first two columns, I do not include mine fixed effects, and therefore look at both cross-sectional and time series variation in productivity. I find evidence for scale economies, but not for cost dynamics: the coefficient on cumulative past output is close to zero and precisely estimated. Including mine fixed effects, in the last two columns, does not change this conclusion.

Figure A1: Map with mining towns



Note: Red triangles are towns in which there was at least one mining college graduate managing an anthracite mine between 1900 and 1914. Blue circles are towns where this was not the case.

Figure A2: Data example: ownership, management and location

TABLE I—Showing Names of Operators, Railroads, etc., etc., and Location of Collieries in the Third Anthracite District for the year 1901.

Names of Operators and Collieries.	County.	Name of General Superintendent.	P. O. Address.	Name of Superintendent.	P. O. Address.	Railroad to Mine.
Pennsylvania Coal Company.		Sidney Williams,	Dunmore,			Erie and Wyoming.
Barnum No. 1 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John Popling and	Pittston,	Erie and Wyoming.
Barnum No. 2 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid	Moosic,	Erie and Wyoming.
Barnum No. 3 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
Laws shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 13 shaft,	Lackawanna,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 9 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 10 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 10 Jr. shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 1 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 8 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 7 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 4 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
Hoyte shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 6 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 5 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 11 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 14 shaft,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 14 tunnel,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 6 washery,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
No. 8 washery,	Luzerne,	Sidney Williams,	Dunmore,	John W. Reid,	Moosic,	Erie and Wyoming.
Lehigh Valley Coal Company.						
Prospect shaft,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Oakwood shaft,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Midvale slope,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Wyoming Hillman slope,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Wyoming shaft,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Henry shaft,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.
Maltby shaft,	Luzerne,	W. A. Lathrop,	Wilkes-Barre,	Eli P. Conner,	Wilkes-Barre,	Lehigh Valley Railroad.

Figure A3: Data example: production, sales and inputs

TABLE II—Gives the total number of tons of coal mined in each colliery, number of days worked, number of employes, number of persons killed and injured, number of kegs of powder, etc., used in the Third Anthracite District for the year ending December 31, 1900.

Names of Operators and Collieries.	County.	Shipments of coal in tons by rail or otherwise.	Number of tons used for steam and heat at colliery.	Sold to local trade and used by employes—tons.	Total production of coal in tons.	Number days worked.	Number persons employed.	Number fatal accidents.	Number non-fatal accidents.	Number kegs powder used.	Number pounds of dynamite used.	Number horses and mules.
Pennsylvania Coal Company.												
Barnum No. 1, 2 and 3 shafts,	Luzerne,	252,138.16	7,509.19	....	259,648.15	159.50	769	3	4	9,682	511	55
Laws and No. 13 shafts,	Luzerne,	183,273.15	4,863.04	....	188,136.19	162.75	552	1	2	4,646	1,181	49
Shafts No. 9, 10 and 10 Jr.,	Luzerne,	185,192.04	11,488.14	....	196,680.18	161	800	......	2	5,825	823	71
Shafts No. 1 and 8,	Luzerne,	128,717.18	2,848.00	....	131,565.18	160.50	385	1	......	3,773	819	43
Shafts No. 4, 7 and Hoyte,	Luzerne,	215,247.19	11,805.07	....	227,053.06	141.50	885	2	3	8,037	1,285	72
Shafts No. 5, 6 and 11,	Luzerne,	239,683.16	10,962.01	....	250,645.17	141.50	944	2	10	9,412	6,340	67
No. 14 shaft and tunnels,	Luzerne,	197,369.04	8,509.15	....	205,878.19	154.50	628	3	2	5,537	1,414	57
No. 6 washery,	Luzerne,	55,781.13	2,867.16	....	58,649.09	155	36	......	......	....	....	2
No. 8 washery,	Luzerne,	76,488.02	2,978.07	....	79,466.09	162	60	......	......	....	....	....
Total,	....	1,533,893.07	*63,833.03	....	1,597,726.10	155.33	5,059	12	23	46,912	12,373	416
Lehigh Valley Coal Company.												
Prospect and Oakwood shafts,	Luzerne,											
Wyoming and Midvale slopes,	Luzerne,	242,619.18	23,775.00	6,273.02	272,668.00	146.75	816	2	5	4,839	37,707	90

Figure A4: Data example: technology usage

TABLE II—Continued.

Name of Operators.	County.	Number of Boilers.				Total horse power.	Locomotives.			Number steam engines of all classes.	Total horse power.	Number pumps delivering water to surface.	Capacity in gallons per minute.	Quantity delivered to surface per minute—gallons.	Number electric dynamos.	Number air compressors.
		Cylindrical.	Horse power.	Tubular.	Horse power.		Steam.	Air.	Electric.							
Pennsylvania Coal Company,	Luzerne,	35	1,400	51	7,605	9,005	10	3	......	133	15,041	30	23,792	11,290	......	6
Lehigh Valley Coal Company,	Luzerne,	27	837	34	6,592	7,429	7	......	......	92	13,661	26	19,033	13,080	3	4
Butler Mine Company, Limited,	Luzerne,	24	280	6	440	720	2	......	......	16	350	14	3,000	800	......	......
Delaware, Lacka. & West. R. R. Co.	Luzerne,	41	1,000	7	905	1,905	......	......	......	31	1,674	16	5,900	2,950	2	......
Temple Iron Company,	Luzerne,	30	800	20	4,675	5,475	4	......	3	72	3,600	10	11,450	5,450	4	6
Miscellaneous Coal Companies.																
Seneca Coal Company,	Luzerne,	27	1,135	1	100	1,235	4	......	......	16	716	11	3,500	1,500	......	1
Old Forge Coal Company,	Luzerne,	6	200	4	450	650	......	......	......	7	386	......	......	800	......	......
Delaware and Hudson Canal Co.,	Luzerne,	15	450	5	750	1,200	......	......	......	19	1,350	3	1,760	1,250	......	1
Raub Coal Company,	Luzerne,	13	440	5	610	1,060	1	......	......	14	793	1	500	300	......	1
John C. Haddock,	Luzerne,	15	288	10	1,070	1,358	1	......	......	36	1,902	3	1,580	1,208	......	1
Clear Spring Coal Company,	Luzerne,	8	275	5	750	1,025	......	......	......	8	800	2	1,200	600	......	......
Florence Coal Company, Limited,	Luzerne,	14	350	3	150	500	1	......	......	12	366	6	2,600	1,800	......	......
W. G. Payne and Company,	Luzerne,	24	444	2	200	644	......	......	......	12	1,453	1	2,000	1,200	......	1
Traders' Coal Company,	Luzerne,	8	160	2	190	350	......	......	......	10	315	2	600	400	......	......

Table A1: Placebo tests

Panel (a): Age of other managers	log(TFP)		log(TFP)	
	Year $<$ 1900		Year $\geq$ 1900	
	Estimate:	SE:	Estimate:	SE:
log(Avg. age other managers)	-0.024	(0.167)	-0.216	(0.073)
Observations	382		2,967	
R-squared	0.025		0.006	

Panel (b): Education of other managers	log(TFP)		log(TFP)	
	Family business		Non-family business	
	Estimate:	SE:	Estimate:	SE:
Share of other managers with mining degree	-0.499	(1.907)	-0.275	(0.120)
Observations	71		2,624	
R-squared	0.004		0.008	

Panel (c): Overidentification test	Sargan test	
	ξ^2:	p-value:
	4.23	(0.121)

Notes: In panel (a), I regress log productivity residuals on the average age of other managers in the same town, both before and after the introduction of mining engineers in 1900. The log age of the manager itself is controlled for, and a linear time trend is included. In panel (b), I regress log productivity residuals on the share of other managers in the same town with a mining college degree, both for family businesses and other firms. I again control for the log age of the manager and for a linear time trend. In panel (c) I report the Sargan test for overidentifying restrictions with and without including the mining college adjacent county dummy as an instrument.

Table A2: Factor-biased technology effects

	log(Output)		log(Output)	
	Estimate:	SE:	Estimate:	SE:
1(Mining degree)	0.055	(0.131)	0.049	(0.130)
log(Mules)*log(Elec. loc)	-0.082	(0.013)		
log(Mules)*log(Steam loc)	-0.044	(0.022)		
log(Mules)*log(Air loc)	0.072	(0.022)		
log(Labor)*log(Elec.loc)	0.034	(0.013)		
log(Labor)*log(Steam loc)	-0.072	(0.023)		
log(Labor)*log(Air loc)	0.018	(0.021)		
log(Steam loc.)*log(Elec. loc)			-0.004	(0.007)
log(Elec. loc)*log(Air loc)			-0.004	(0.006)
log(Steam. loc)*log(Air loc)			-0.044	(0.011)
Observations	3,349		3,349	
R-squared	0.892		0.875	

Notes: The first column allows for interaction effects between each locomotive type and both labor and mules. The second column allows for interaction effects between all locomotive types. The same instruments as before are being used.

Table A3: Additional controls

Panel (a): Production function	Model (2a)		Model (2b)		Model (2c)	
	Estimate:	SE:	Estimate:	SE	Estimate:	SE
β_x (mining degree)	0.160	(0.092)	0.009	(0.096)	-0.343	(0.764)
β_k^{el} (electrical loc.)			0.031	(0.008)	0.036	(0.010)
β_k^{st} (steam loc.)			-0.003	(0.009)	-0.024	(0.033)
β_k^{ca} (compr. air loc.)			-0.002	(0.007)	0.014	(0.012)
β_{xk}^{el} (md x elec. loc.)					-0.044	(0.155)
β_{xk}^{st} (md x steam loc.)					0.235	(0.328)
β_{xk}^{ca} (md x compr. air loc.)					-0.271	(0.277)
Railroad dummy	1.203	(0.373)	1.257	(0.374)	1.270	(0.371)
Family business	-0.063	(0.054)	-0.026	(0.053)	-0.029	(0.057)
Ownership change	-0.124	(0.036)	-0.114	(0.036)	-0.103	(0.042)
Observations	2,526		2,526		2,526	
R-squared	0.821		0.824		0.816	
Panel (b): Technology choice	1(Elec. loc.)		1(Steam loc.)		1(Air loc.)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
1(mining degree)	0.599	(0.152)	-0.322	(0.159)	0.007	(0.174)
Railroad dummy	-0.291	(0.117)	0.276	(0.164)	-0.465	(0.124)
Family business	-0.221	(0.071)	0.071	(0.092)	-0.265	(0.075)
Ownership change	0.011	(0.051)	0.010	(0.044)	-0.056	(0.054)
Observations	2,635		2,635		2,635	
R-squared	0.463		0.362		0.447	

Notes: Panel (a) contains the second-stage production function coefficients. Dependent variable is log output. All variables in logs, except for the mining college dummy. Model (2a) does not control for the different locomotive types used, model (2b) does control for all locomotive types, and model (2c) interacts the locomotive types with the managerial education dummy. Panel (b) contains the technology choice estimates. Dependent variables are dummies for usage of each locomotive type.

Table A4: Cost dynamics

	log(TFP)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:
log(Cum. past output)	-0.035	(0.022)	0.039	(0.028)
log(Current output)	0.139	(0.025)	0.186	(0.033)
Mine FE	No		Yes	
R-squared	0.092		0.601	
Observations	3,008		3,008	

Notes: Productivity residuals from production model (2b). Linear time trend included.

Table A5: Technology choice: probit model

	1(Elec. loc.)		1(Steam loc.)		1(Air loc.)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
1(Mining degree)	0.303	(0.084)	-0.246	(0.159)	0.070	(0.157)
Average usage	0.426		0.676		0.290	
Observations	3,109		3,342		2,863	

Notes: Probit model with same instruments as before. Report marginal effects at the mean.

Table A6: Production function estimates using Ackerberg, Caves and Frazer (2015)

	Model (2a)		Model (2b)		Model (2c)	
	Estimate:	SE:	Estimate:	SE	Estimate:	SE
β_x (mining degree)	0.608	(0.450)	0.019	(0.310)	-0.043	(0.214)
β_k^{el} (electrical loc.)			0.140	(0.064)	0.128	(0.055)
β_k^{st} (steam loc.)			0.044	(0.037)	0.036	(0.056)
β_k^{ca} (compr. air loc.)			-0.055	(0.147)	-0.049	0.040
β_{xk}^{ca} (md x compr. air loc.)					0.145	(0.212)
β_{xk}^{el} (md x elec. loc.)					0.043	(0.143)
β_{xk}^{st} (md x steam loc.)					-0.001	(0.283)
Observations	2,765		2,765		2,765	
R-squared	0.841		0.881		0.885	

Table A7: List of variables

	Unit of measurement	Level	Frequency	Flow/stock
(a) Output and sales				
Coal extracted	short tons (2000 lbs)	Mine	Annual	Flow
Coal shipped over railroad	short tons	Mine	Annual	Flow
Coal sold locally	short tons	Mine	Annual	Flow
Coal reused as input	short tons	Mine	Annual	Flow
(b) Inputs				
Employees	average counts	Mine	Annual	Stock
Days worked	counts	Mine	Annual	Flow
Powder	kegs of 25lbs	Mine	Annual	Flow
Mules	counts	Mine	Annual	Stock
(c) Management				
Superintendent name	string	Mine	Annual	N/A
Foreman name	string	Mine	Annual	N/A
Mining degree	binary	Manager	Annual	N/A
College degree	binary	Manager	Annual	N/A
Superintendent Age	string	Manager	Annual	N/A
(d) Technology				
Locomotives	counts	Firm-county	Annual	Stock
Compressed-air locomotives	counts	Firm-county	Annual	Stock
Electrical locomotives	counts	Firm-county	Annual	Stock
Steam locomotives	counts	Firm-county	Annual	Stock
Location	coordinates	Village	Time-invariant	N/A

Table A8: Mine summary statistics

	Mean	Std. Dev.	Min.	Max.	Observations
(a) Output:					
Coal extracted, Mtons	0.21	0.19	0	3.52	4730
Output shipped, share	0.87	0.15	0	1	4572
(b) Variable inputs:					
Employees	512.71	427.8	0	6595	4730
Powder, 1000 kegs	52.42	138.24	0	1748.43	4730
Coal inputs, Ktons	20.92	21.85	0	494.48	4730
Dynamite, 1000 pounds	26.96	47.83	0	654.89	4730
(c) Capital inputs:					
Mules	51.03	43.01	0	276	4730
Mining locomotives	25.37	36.15	0	172	4730
Machinery horsepower x 1000	29.67	44.18	0	212.92	4730
(d) Managerial inputs:					
Manager has college mining degree	0.07	0.25	0	1	4730
Manager has other college degree	0.01	0.1	0	1	4730

Table A9: Manager summary statistics

	Mean	Std. Dev.	Min.	Max.	Observations
(a) Age	46.02	11.34	19.5	77.40	309
if mining degree	33.53	6.98	25.87	43.5	7
if other degree	31.84	5.76	20	43	11
if no degree	46.85	11.06	19.5	77.40	291
(b) Years in firm	1.56	1.97	0	8.54	308
if mining degree	1.33	1.56	0	4.48	7
if other degree	1.11	1.67	0	4.47	11
if no degree	1.58	1.99	0	8.54	290
(c) # Mines managed	2.71	5.17	1	37.32	309
if mining degree	5.72	9.62	1	26.85	7
if other degree	1.2	0.43	1	2.33	11
if no degree	2.7	5.12	1	37.32	291
(d) Output, Mtons	0.12	0.12	0	0.66	310
if mining degree	0.22	0.18	0.06	0.63	7
if other degree	0.13	0.22	0	0.66	11
if no degree	0.12	0.11	0	0.63	292

Table A10: Mining school curriculums

Subject	Course examples	% Credits	Usual phase
Science	Mathematics, Chemistry, ...	33.7	Freshman / sophomore
Mining engineering	Drilling, Mine construction, Geology, ...	34.3	Junior / senior
Other engineering	Electricity, Mechanics, ...	24.3	Sophomore / junior
Languages	Foreign languages, writing, retorics	4.7	Freshman
Thesis	Master project	2.0	Senior
Management	Mining economics, mining law, contracts, ...	1.0	Senior

Table A11: School data sources

School name	Document name	Type	Years
Arizona School of Mines	Alumnal Record of the University of Arizona	Alumni record	1916
Colorado School of Mines	Quarterly of the Colorado School of Mines	Bulletin	1908, 1912-1914
Columbia College of Mines	Catalogue of Columbia University	Catalogue	1867-1914
Michigan College of Mines	Graduates of the Michigan College of Mines	Alumni record	1910
	Year Book of the Michigan College of Mines	Catalogue	1910-1914
Missouri School of Mines	School of Mines and Metallurgy Bulletin	Catalogue	1914
Montana School of Mines	Annual Catalogue	Catalogue	1908-1914
Nevada Mackay School of Mines	Register of the University of Nevada	Catalogue	1908-1914
New Mexico School of Mines	Register of the New Mexico School of Mines	Catalogue	1909-1914
Penn State School of Mines	Alumni Directory	Alumni record	1913
South Dakota School of Mines	Annual Catalogue	Catalogue	1912
University of Minnesota	School of Mines Announcement	Catalogue	1897-1914
West Virginia University	Register of Faculty, Alumni and Students	Alumni record	1920