Technology Adoption and Productivity Growth: Evidence from Industrialization in France*

Réka Juhász
(UBC, NBER, and CEPR)

Mara P. Squicciarini
(Bocconi and CEPR)

Nico Voigtländer
(UCLA, NBER, and CEPR)

First version: December 2019
This version: October 2023

Abstract

New technologies tend to be adopted slowly and – even after being adopted – take time to be reflected in higher aggregate productivity. One prominent explanation for these patterns is the need to reorganize production, which often goes hand-in-hand with major technological breakthroughs. We study a unique setting that allows us to examine the empirical relevance of this explanation: the adoption of mechanized cotton spinning during the First Industrial Revolution in France. The new technology required reorganizing production by moving workers from their homes to the newly-formed factories. Using a novel hand-collected *plant-level* dataset from French archival sources, we show that productivity growth in mechanized cotton spinning was driven by the disappearance of plants in the lower tail – in contrast to other sectors that did not need to reorganize when new technologies were introduced. We provide evidence that this was driven by organizational challenges such as developing optimal plant layout. A process of 'trial and error' led to initially low and widely dispersed productivity, and – in the subsequent decades – to high productivity growth as knowledge diffused through the economy and new entrants adopted improved methods of organizing production.

JEL: F63, O14

Keywords: Industrialization, Technology Adoption, Productivity

*We thank the editor, Chad Syverson, and three anonymous referees for their constructive and insightful feedback. We also thank Sandy Black, Serguey Braguinsky, Bruno Caprettini, Jan De Loecker, Michela Giorcelli, Tarek Hassan, Walker Hanlon, Michael Huberman, Amit Khandelwal, Thomas Le Barbanchon, Nils Lehr, Bob Margo, Joel Mokyr, Dávid Krisztián Nagy, Gianmarco Ottaviano, Tommaso Porzio, Todd Schoellman, Eric Verhoogen as well as numerous seminar and conference audiences for helpful comments and suggestions. We thank Gaia Dossi, Johanna Gautier, Felix Iglhaut, Angelo Krüger, Lorine Labrue, Simon Margolin, Marie Michaille, Alberto Mola, Sebastian Ottinger, Verónica C. Peréz and Sarah Wappel for excellent assistance throughout the construction of the dataset, Guillaume Daudin and Walker Hanlon for kindly sharing data, and Eugene White for sharing his expertise of hyperinflation during the French Revolution. Financial support from the National Science Foundation (Grant # 1628866) is gratefully acknowledged.

> [T]here were strong pairwise complementary relations between factory organization and machinery [...] employers needed to simultaneously determine the choice of technique, the level of worker effort, and the way incentives were set up and communications and decisions flowed through the firm hierarchy. [...] Factories were the repositories of useful knowledge ... but they were also the places in which experimentation took place. – Mokyr (2010, pp. 345-46)

1 Introduction

The diffusion of innovation is at the core of aggregate productivity growth in the long run. Yet, many technologies that ended up being widely adopted were slow to diffuse across firms (Griliches, 1957; Mansfield, 1961; Rosenberg, 1976; Hall, 2004; Comin and Hobijn, 2010). This slow adoption is particularly puzzling given that new technology can provide a substantial boost to firm productivity (Syverson, 2011; Bloom, Eifert, Mahajan, McKenzie, and Roberts, 2013; Giorcelli, 2019). There is also a second, well-documented puzzle: When major innovations such as information technology (IT) or electricity spread across firms, the widely expected boost in aggregate productivity has proved hard to document in the data. This prompted Robert Solow to remark in 1987 that "You can see the computer age everywhere but in the productivity statistics" (*The New York Times, July 12, 1987*).

One prominent explanation for both puzzles is the need to modify and reorganize the production process when adopting major breakthrough technologies (David, 1990; Brynjolfsson, 1993; Brynjolfsson and Hitt, 2000; Hall and Khan, 2003; Brynjolfsson, Rock, and Syverson, 2021). Initially, many firms operate the new technology inefficiently – often because complementary organizational innovations are missing. If these challenges are indeed important during the early phase of technology adoption, we expect them to be reflected in a highly dispersed productivity distribution – due to a lack of standardized organizational knowledge that adopting firms can draw from. However, empirical evidence on this mechanism is scarce, as measuring the productivity distribution specific to new adopters is challenging for numerous reasons. First, standard data sources rarely make it possible to observe the use of specific technologies. Second, it is difficult to observe whether an adopting plant has also reorganized production. Third, productivity under the old and new technology are typically correlated.

This paper shows how the need to reorganize production affects the productivity distribution of adopting plants during the diffusion of a major new technology. We bypass the typical challenges by studying a unique historical setting – the adoption of mechanized cotton spinning in France during the 19th century. Importantly, the macro-inventions that mechanized cotton spinning (the famous spinning jenny, the water-frame, and the mule) went hand-in-hand with the need to reorganize production on a revolutionary scale. Prior to mechanization, workers produced in their homes in a cottage-industry setting. Adopting the new technology required setting up factories from scratch, and moving workers there from their homes because of the reliance on inanimate power sources and the need to monitor workers more closely (Williamson, 1980; Szostak, 1989). This led to one of the most dramatic shifts in the organization of production in economic history (Mokyr, 2011). While the key elements of the new spinning technology itself were well-known across France (Horn, 2006), and multiple domestic producers supplied firms with the machinery (Chassagne, 1991), its adoption occurred in the absence of complementary knowledge on how the new cotton spinning plants should be organized (Pollard, 1965).

A number of features of this setting allow us to isolate the productivity distribution of adopters and study its evolution over a long time horizon (three decades), making headway on the typical empirical challenges faced by the literature. In particular, the sharp break in the *location* of production due to its organization in plants makes it possible to distinguish adopters of the new technology from old producers. This addresses the first two empirical challenges mentioned above. Moreover, the first generation of mechanized cotton spinners did not typically have a background in the old technology, suggesting that productivity under the two technologies were not systematically related, which addresses the third challenge.

Our empirical analysis is based on a novel, hand-collected plant-level dataset from historical surveys covering three sectors (mechanized cotton spinning, metallurgy, and paper milling) at two points in time, around 1800 and in the 1840s. To help distinguish the effect of reorganization from broader trends such as general productivity growth, political and institutional change, or enhanced regional integration, we compare the evolution of the plant productivity distribution in mechanized cotton spinning to two comparison sectors (metallurgy and paper milling). This is similar in spirit to a difference-in-differences strategy. Crucially, in both comparison sectors, production was already organized in plants for centuries before the Industrial Revolution because of their reliance on water power and high-fixed-cost machinery. As a consequence, these sectors possessed fairly standardized knowledge about how to organize plant-based production. Moreover, during our sample period, all three sectors witnessed the arrival of new technologies that could be introduced fairly seamlessly into the *existing* organization of production.[1] Thus, while all three sectors were adopting new technologies, only mechanized cotton spinning had to adapt to a radically reorganized production process.

We document four main findings for mechanized cotton spinning plants: 1) we observe a highly dispersed productivity distribution in the initial period (1806) relative to 1840; 2) we estimate that the industry underwent a substantial (82%) increase in plant productivity between 1806 and 1840 *after* mechanization had already been adopted; 3) this aggregate productivity growth was largely

[1] In mechanized cotton spinning, there were improvements to the existing vintages of machinery and in preparatory processes (Allen, 2009). In paper milling, one part of the production process was mechanized (André, 1996) and in metallurgy, charcoal was replaced with coal as the source of fuel (Pounds and Parker, 1957). As we discuss in more detail below, the three industries shared important similarities in adopting these new technologies during our sample period. This supports our implicit assumption that technological additions to the *existing* production setups did not lead to differential productivity trends.

driven by the disappearance of plants in the lower tail of the distribution (which we refer to as 'lower-tail bias' of productivity growth); 4) the disappearance of the lower tail took place almost exclusively through plant exit and entry. Inefficient producers were replaced by more efficient entrants. In the comparison sectors, we also find a sizeable increase in plant productivity during the sample period (68% in metallurgy and 33% in paper milling). However, the lower-tail bias of productivity growth is unique to mechanized cotton spinning. In contrast, in the comparison sectors, the entire productivity distribution shifted right. Taken together, we interpret these results as suggestive of a link between the lower-tail bias of productivity growth and the feature unique to the mechanized cotton spinning industry – the need to reorganize production.

The second part of the paper examines why the need to reorganize production would lead to a lower-tail bias in productivity growth. Central to our argument is the fact that, at early stages of technology adoption, plants need to learn about optimal organizational forms. To fix ideas, we provide a simple framework in which plants endogenously learn about the optimal organization of *multiple* inputs or tasks in the spirit of Perla and Tonetti (2014). These tasks reflect organizational challenges such as optimal plant layout, and they exhibit complementarities in the production function. We show that these features initially, when plants have little knowledge about the optimal ways to perform these tasks, lead to a fat lower tail in the plant productivity distribution. Over time, as plants learn about the efficient organization of inputs, the lower tail disappears. We present both historical and empirical evidence consistent with this.

According to the historical literature, there were two broad classes of challenges that early cotton spinning mills faced. First, they needed to contend with a range of issues related to mill layout and design. As Allen (2009, p.184) writes: "The cotton mill, in other words, had to be invented as well as the spinning machinery *per se*." Second, a set of labor management innovations were required for setting up and operating spinning mills at a scale not seen elsewhere in the economy (Pollard, 1965). We examine the data for evidence consistent with the spread of organizational practices. We provide evidence for the spatial diffusion of knowledge during the early phase of industrialization, by showing that cotton plants located closer to high-productivity peers were themselves more productive. Strikingly, this spatial productivity pattern is not present in the comparison sectors, where plant-based production methods were more mature, nor in cotton spinning in the long-run, once organizational knowledge had diffused. We show, using a rich set of controls and placebo tests, that these results are unlikely to be driven either by selection into productive locations or by omitted variables.

While these results are suggestive of a role for the spatial diffusion of knowledge, they do not distinguish between learning about the new technology itself (i.e., how to operate and maintain the new machines) and learning about efficient organizational forms. We provide direct evidence for the latter using detailed metrics on cotton spinning plant design from our sample period. We find that during the early phase of adoption, two key design features (number of floors and squareness of its layout) were chosen almost at random – consistent with plants experimenting with how to organize the factory floor optimally. Later, as best-practice knowledge spread, these metrics converged towards much narrower distributions around the optimum. We provide additional complementary evidence that helps us to distinguish the two forms of learning. If plants were learning mostly how to operate the newly adopted technologies, we would expect incumbents to have an advantage relative to newer entrants. On the other hand, if organizational knowledge on plant design diffused over time, we would expect later adopters to have an advantage in setting up their factories. Two features in the data point to the latter: First, cotton spinning plants that entered the market *later* had higher productivity during the initial phase of adoption. This holds even after controlling for newer capital vintages, and it does not hold for the comparison sectors or for cotton spinning in later periods. Second, the exit rate of plants in mechanized cotton spinning was substantially higher than in other sectors between 1800 and 1840, and buildings were also abandoned for use by the industry at higher rates. These findings suggest that knowledge of how to set up and organize cotton plants spread over time (and space), giving an edge to new entrants.

We examine alternative explanations that could account for our results. While our DID-style evidence based on the comparison sectors addresses many potential concerns, it is possible that some alternative channels affected mechanized spinning *differentially*. We control for a large set of these directly, showing that the lower-tail bias of productivity growth remains robust. For example, our results hold with region fixed effects, suggesting that the sorting of cotton plants into areas with better location fundamentals, better market access, or better input markets are unlikely to drive our results. Region fixed effects also make it unlikely that the Napoleonic Blockade, which had a regionally differential effect across France in mechanized cotton spinning (Juhász, 2018), drives our findings. We also account for possible more localized confounders, such as battles during the Napoleonic Wars or location-specific market access. Finally, plant-specific features such as size, output quality, capital deepening, and plant life-cycle characteristics do not confound our results, and they are also robust to excluding all plants that could be smaller 'spinning workshops' (which shared some, but not all characteristics of mature factory-based production). For a systematic overview of alternative mechanisms and the corresponding robustness, we refer the reader to the summary table in Appendix F.

Related Literature and Contribution. Our paper is closely related to a literature on innovation and technology adoption in manufacturing – particularly the strand that has studied the productivity effects of the adoption of IT.[2] Interestingly, some of the patterns we document for mechanized cotton spinning have been found in other settings. For example, Syverson (2011) discusses that the

[2] See Hall and Khan (2003) and Hall (2004) for an overview of the literature on technology diffusion. Brynjolfsson and Hitt (2000) and Syverson (2011) discuss the literature on the productivity effects of IT.

adoption of IT capital is associated with increased within-industry productivity dispersion. Foster, Grim, Haltiwanger, and Wolf (2018) provide empirical support for the argument by Gort and Klepper (1982) that periods of rapid innovation are associated with a surge in firm entry, followed by a period where experience with the new technology is accumulated, eventually leading to a shakeout where unsuccessful firms (or plants) exit. Brynjolfsson and Hitt (2000) conjecture that the surge in aggregate productivity in the 1990s was explained in part as a return on the large, intangible complementary organizational innovations that firms had undertaken in prior decades to make efficient use of IT. Our paper contributes to this literature in two ways. First, we provide evidence that these patterns generalize to other settings in which major new technologies are adopted. Second, our unique setting allows us to more closely tie these patterns to the need to reorganize production that often accompanies major technological change. In particular, initial information disparities about the optimal organization of production can help to explain why breakthrough technologies tend to be adopted slowly and – even after being adopted – take time to be reflected in higher aggregate productivity. Along this dimension, our paper relates to recent work that documents how a variety of organizational barriers can impede technology adoption (Atkin, Chaudhry, Chaudry, Khandelwal, and Verhoogen, 2017; Feigenbaum and Gross, 2021).

In addition, our paper brings the insights of the firm productivity literature to the most important structural break in economic history – the First Industrial Revolution, which saw unprecedented growth in manufacturing productivity (Crafts, 1985; Crafts and Harley, 1992; Galor, 2011). So far, productivity growth during this period has been studied mostly at the country level, or – in some cases – at the aggregate sectoral level.[3] Our paper is the first to study the contribution of plant dynamics to manufacturing productivity improvements during the Industrial Revolution. Our focus on the overall plant productivity distribution allows us to shed new light on how productivity growth evolved during this important period.[4] In particular, we isolate and track the productivity distribution of newly created adopting plants in cotton spinning. This goes beyond previous work (including with modern data), where major new technologies are typically introduced by existing producers, so that the changes in the productivity distribution reflect *both* the productivity differential of the new technology and subsequent gains due to organizational improvements. We are also the first to show that the extensive margin of plant entry and exit contributed decisively to productivity growth during the Industrial Revolution. Finally, we complement a rich historical literature by providing the first systematic empirical evidence for the importance of organizational innovations in driving productivity growth during the Industrial Revolution (Pollard, 1965; Sokoloff,

[3]Closely related, Clark (1987) studies cross-country differences in the productivity of mechanized cotton spinning, but only at the sectoral level. Braguinsky, Ohyama, Okazaki, and Syverson (2015) study the Japanese cotton spinning industry in the late 19th century and early 20th century. Rather than technology adoption, their paper focuses on the effects of acquisitions on acquired plants.

[4]In related work, Braguinsky, Ohyama, Okazaki, and Syverson (2021) study how cotton firms grow by innovating vertically and horizontally in Japan's Meiji period in the late 19th and early 20th century.

1984, 1986).

The paper is structured as follows. The next section discusses the historical context. Section 3 describes the data, while Section 4 presents and discusses our empirical results. Section 5 sheds light on the underlying mechanism, and Section 6 concludes.

2 Historical Background

In this section we describe the key features of the historical setting. First, we discuss why mechanized cotton spinning and its adoption in France presents a well-suited setting to examine the productivity effects of major technological breakthroughs. Second, we introduce the three sectors that we analyze in the paper. We first describe the process of mechanization in cotton spinning and discuss the reorganization of production that it entailed. Finally, we introduce our two comparison sectors. We summarize the most important aspects in the main text. Appendix A contains a more extensive discussion.

2.1 The Industrial Revolution in Britain and its Spread to France

Early nineteenth century France presents an opportune setting for our study of technology adoption. First, the development of mechanized cotton spinning during this time period is widely seen as a macro-invention (or general purpose technology), whose effects on the economy were similar to the development of the steam engine, electrification, or the information technology (IT) revolution.[5] Mechanized cotton spinning belongs in this category partly because its path-breaking innovations enabled the subsequent mechanization in other branches of the textile industry such as wool, linen, and silk (c.f. Jenkins, 2003). Freeman and Louçã (2001) include innovations in the cotton textile industry among the first major wave of industrialization, with steam power and railways fueling the second wave, and electrification, the third. As Freeman and Louçã (2001, p. 156) put it: "many of the organizational as well as technical innovations in cotton were followed later by other branches of the textile industry and by manufacturing more generally." Similarly, Pollard (1965); Mokyr (2016) emphasize that mechanization in cotton spinning was the birth of the modern factory system, which fundamentally altered the organization of production – first in this industry itself, and later in manufacturing more broadly.

Second, studying mechanized cotton spinning in France – a follower country – allows us to focus on technology *adoption*, as opposed to *innovation*. As is well-known, the flagship inventions of the Industrial Revolution – most notably the spinning jenny and the coke-fired blast furnace – were developed in Britain. While England was the first country to industrialize, France was an early follower. The latest estimates for France find an acceleration in economic growth around 1800,

[5]Macro-inventions, or general purpose technologies, are typically defined as generic products, processes or organizational forms (Lipsey, Carlaw, and Bekar, 2005) that generate spillovers across sectors and set in motion a whole stream of advances that result in large productivity improvements (Bresnahan and Trajtenberg, 1995; Lipsey et al., 2005; Allen, 2009). See Dudley (2010) for a discussion.

well after growth in England took off (Ridolfi and Nuvolari, 2021). During the initial phase of French industrialization, new technologies arrived from Britain (Mokyr, 2021). The processes of innovation and adoption differed in the two countries. While Britain may have had a comparative advantage in *developing* commercially viable machines and mills due to its highly skilled workers (artisans, engineers, millwrights), France was arguably the world leader in science (Gillispie, 2004), which may have facilitated technology adoption. Indeed, Nuvolari, Tortorici, and Vasta (2023) use patent data to show that France was able to effectively absorb key technologies from Britain in this period. However, the mere adoption of these technologies was only one step. Their integration in factory settings and efficient operation presented major challenges (Mokyr, 2021). The tacit, non-codified aspect of British industrial know-how is important in this context, as it explains why the French needed to undertake costly experimentation to efficiently operate the technology. Appendix A.1 contains further details about the Industrial Revolution in France.

2.2 Cotton Spinning: Mechanization and Reorganization of Production

This section discusses the main historical features and challenges in the transition to mechanized cotton spinning. Appendix A.2 and A.3 provide further detail.

Development of mechanized spinning. Cotton textiles was the flagship industry of the First Industrial Revolution, contributing one-quarter of TFP growth in Britain during the period 1780-1860 (Crafts, 1985). Cotton spinning is the process by which raw cotton fiber is twisted into yarn. Traditionally, this task was performed mostly by women in their homes, using a simple spinning wheel (see Figure A.1 in the appendix). With this old technology, each spinner was able to spin only one thread of yarn at a time. The industry was rurally organized and generally centered around a local merchant-manufacturer who would supply spinners with the raw cotton, collect their output, take care of the marketing, and who often also owned the spinning wheels (Huberman, 1996).

The breakthrough 'macro-inventions' in spinning were forged in Britain in the 1760s and 1770s, when three new vintages of machinery (the spinning jenny, the water-frame and the mule) were developed in quick succession. The left panel in Figure A.2 depicts the mule; Allen (2009) provides an in-depth discussion of each vintage. These new machines made it possible to spin multiple threads simultaneously, as twist was imparted to the fibre not by using the workers' hands, but rather by using spindles. These innovations required production to move from workers' homes to the factory floor for two reasons. First, the machines almost always used inanimate power sources (typically water power), which led to the concentration of production in one location. Second, mechanized production increased the need for monitoring workers because of the complementarity of their tasks, but also because the machinery with which they worked was both more complex and more expensive (Williamson, 1980; Geraghty, 2007; Mokyr, 2010).

The productivity effect of these innovations was enormous. Allen (2009) estimates that the first vintage of the spinning jenny alone led to a threefold improvement in labor productivity.

Correspondingly, the price of yarn declined rapidly in the late 18th century (see Appendix Figure A.3), especially for the highest-quality yarn, where prices declined from 1,091 pence per pound to 76 pence per pound in real terms between 1785 and 1800 (Harley, 1998).

During the early decades of the 19th century, the mechanized cotton-spinning industry was characterized by a steady stream of micro-inventions (Allen, 2009, p. 206). Importantly, the next major innovation, the self-acting mule (a completely new vintage of spinning machinery), did not spread widely until the 1840s, i.e., until after our sample period (Huberman, 1996). Thus, there was no major technology switching during our period of study, but rather a steady stream of inventions that improved existing vintages.

Adoption of mechanized spinning in France. Mechanized spinning was adopted with some lag in France. Efforts to adopt the technology had begun with state support during the *Ancien Régime*. By the time of our first cotton spinning survey in 1806, the large-scale expansion of the industry documented in Juhász (2018) had just begun. The existence of the technology was known throughout the country (Horn, 2006), and a number of domestic spinning machine-makers had been established (Chassagne, 1991). The spinning machinery itself was produced locally (using British blueprints) because of a ban on exporting machinery (and the emigration of engineers and skilled workers) from Britain until 1843 (Saxonhouse and Wright, 2004). All three vintages (the spinning-jenny, the water-frame, and the mule) were used in France. Importantly, when widespread adoption in France began, the technology – and in particular its optimal organization – was far from being mature in Britain. Thus, while France was a follower country, it could not copy a 'mature' technology, as we discuss in more detail below.

The challenging transition to factory-based production in cotton spinning. The transition of workers from their homes to the factory floor has been characterized as "one of the most dramatic sea changes in economic history" (Mokyr, 2010, p. 339). It fundamentally altered people's lives and, importantly for our setting, posed a host of challenges for the first generation of large-scale factories.

In the case of cotton spinning, adopting the first generation of mechanized spinning machinery went hand-in-hand with the need to organize production in plants rather than homes. While cotton spinning was not the first sector to organize production in plants, the industry faced challenges for which a standard set of solutions did not exist at the time. Partly, this was because the knowledge required was largely technical and hence industry-specific (Pollard, 1965, p. 158). In addition, mechanized cotton spinning mills pioneered *flow production* – that is, the production of standardized goods in huge quantities at low unit costs by "arranging machines and equipment in line sequence to process goods continuously through a sequence of specialized operations" (Chapman, 1974, p. 470). This led to a finer division of labor and larger-scale plants than what had been seen before in other sectors, raising novel challenges (Chapman, 1974).

Flow production meant that machines and equipment had to be spatially organized and coordinated such that the "continuous (twenty-four hours a day) synchronisation of a sequence of highly specialised machines" (Chapman, 1974, p. 472) could be achieved. This organizational innovation, pioneered in cotton spinning, distinguishes 'proto-factories' from the 'factory proper' (Chapman, 1974; Markus, 2013). Under the former, older, system, "several batch processes are centralised into a large unit with a systematic grouping of machines, with or without water power, without linked semi-automatic processes" (Markus, 2013, p. 262). In the factory proper, "power and automatic machinery are organised for flow production in 24-hour operation. Water or steam power is a necessary but not sufficient defining feature" (Markus, 2013, p. 262).

Mechanized cotton spinning firms needed to resolve a range of issues related to mill layout and design. Allen (2009, p. 202) discusses some of the key challenges in developing the first mills in Britain: "...design issues emerged regarding the spatial location of the various machines, the flow of materials from one to the next, and the provision of power throughout a multi-story building." It is useful to highlight three complementary aspects of mill design challenges that spinners faced. First, as we have seen, flow production meant that the production line needed to be synchronized so that the layout (i.e., the floorplan) of the building became important. For example, factories needed to be sufficiently wide for a mule operative to work two mules alternately, which were placed back to back in pairs (see Markus, 2013, p. 265, also illustrated on the right-hand side of Figure A.2). Second, flow production relied on the mechanization of each step of the production process (Chapman, 1974), which meant that power needed to be continuously distributed throughout the building. In Britain, Chapman (1970, p. 239) claims that there were only a handful of millwrights, *qualified from experience*, capable of undertaking the construction of the gearing for the new cotton mills. Third, a mechanized production line at this scale introduced a host of structural challenges. For example, building structures needed to withstand the stress they faced from the vibrations of machines (Chassagne, 1991, p. 435). Iron rods with plates held beams to the masonry walls to prevent the vibrations of machines from shaking the walls apart (Langenbach, 2013). Similarly, buildings had to be well-lit, which created design challenges to let as much daylight as possible reach the spindles, and it implied massive fire hazards when gas lights were used, due to the highly inflammable cotton dust (Markus, 2013).

These were novel, complementary challenges for which a standard set of solutions did not exist at the time. The industry experimented through a process of trial and error. Successful mill designs in England were observed and copied (Chapman, 1970, p. 239). It took time for design defects to be improved; for example, contemporaries were aware of ventilation problems in the Arkwright-style mills, but continued to use the same layout regardless (Fitton and Wadsworth, 1958, p. 98). Appendix A.3 provides further detail on the process of trial and error in developing solutions to building design challenges.

In addition to design-related organizational challenges, a large set of management innovations were required to run spinning mills efficiently. In the 'Genesis of Modern Management,' Pollard (1965, p. 160) described the development of efficient labor management practices as the primary management challenge facing early factories. There were three salient aspects of this for cotton spinning mills; i) how to get workers, who were used to the independence of the domestic system, to adapt to the rhythm and hierarchy of factory work, ii) how to coordinate and implement a fine division of labor, and iii) how to solve monitoring problems. Appendix A.3 discusses these challenges in more detail. Here, we give one illustrative example.

Renumeration in the pre-industrial home spinning system was characterized by piece rates. This made sense, given that the merchant had no way to monitor worker effort. The move to continuous flow production made piece rates unworkable. The speed of production was controlled by the entire production line, not individual workers. This is illustrated by Karl Marx, who quoted a large cotton manufacturer, Henry Ashworth: "When a laborer lays down his spade, he renders useless for that period, a capital worth eighteen pence. When one of our people leaves the mill he renders useless a capital that cost *£100,000*" (Clark, 1994, p. 129). Piece rates became impractical because the teamwork inherent to flow production made it difficult to determine the contribution of individual workers (Mokyr, 2016, p. 344). The alternative, a time wage, raised incentive issues because monitoring individual worker effort was difficult in cotton mills (Huberman, 1996). The solution was not obvious, and it was context-specific. In cotton mills, employers experimented with a variety of techniques: some paid the entire team (which in turn created issues with the internal distribution within teams); other early factories solved the team production issue by hiring entire families as a work unit and paying them a piece rate (Mokyr, 2016). Huberman (1996) estimates that it took two generations for efficient labor management practices to be developed in cotton spinning. Finally, around 1830, the industry in Britain settled on efficiency wages (Huberman, 1996).

The first generation of mechanized cotton spinners faced these complementary design and management challenges all at once. Not only did best-practice solutions emerge slowly, but it also took time for this new body of knowledge to coalesce. According to Pollard (1965), the process was more or less complete around 1830 in Britain: "a cotton mill was so closely circumscribed by its standard machinery, and there was so much less scope for individual design, skill or new solutions to new problems, by 1830, at least, ... that little originality in internal layout was required from any but a handful of leaders" (Pollard, 1965, p. 90).

Knowledge diffusion in mechanized cotton spinning. How did the first generation of mechanized cotton spinners learn the solution to these organizational challenges? Given our discussion above, it is unlikely that there were important spillovers from home spinning to mechanized plants. The sharp break in organizational form under the new technology rendered experience with the old technology effectively useless. Chassagne (1991, p.274) presents suggestive evidence to support this assertion: According to data on owners of 148 mechanized cotton spinning establishments between 1785-1815 in France, "traders, bankers and commercial employees" accounted for the vast majority (62.5%) of entrepreneurs. While these figures have to be interpreted with caution (they probably oversample larger, better-known plants), they point to the importance of commercial knowledge as opposed to previous experience with handspinning in setting up cotton spinning factories. This suggests that productivity under the old and new technology were not systematically related.

Knowledge spillovers *across* sectors are also unlikely to have played an important role. As we discussed above, mechanized cotton spinners had to contend with new challenges that had not been encountered in other sectors, even those that were organized as 'proto-factories' in 1800. For example, André (1996) notes that there were no standardized mill designs in paper milling (one of our comparison sectors) – arguably because it did not feature flow production, making it less crucial that all production steps were efficiently integrated. Moreover, much knowledge was highly industry-specific (Pollard, 1965). Again, Chassagne (1991) presents suggestive evidence to support this pattern. Only 10% of early cotton spinners in his sample had a background in another industry – of these, most came from cotton printing, a relatively close industry, and almost none came from other sectors whose production was organized in plants (Chassagne, 1991, p. 274).

For French cotton spinners, the most important source of knowledge about plant design and organization was Britain. However, spillovers from Britain were limited for a number of reasons: It wasn't until the 1830s that the British began to codify best practice in manuals (Pollard, 1965), and even that was limited because a lot of the industry's knowledge was tacit (Mokyr, 2001, 2010). In addition, there were British bans on knowledge transfer (Horn, 2006; Chassagne, 1991). Finally, the best practice that the British industry eventually converged to was not necessarily applicable to plants in France, as local conditions were different. For example, the abundance of water power meant that French spinners mostly relied on hydropower throughout our sample period, in contrast to the British reliance on steam power (Cameron, 1985). Chassagne (1991) notes that the French solution eventually developed for mitigating fire hazards in cotton mills was to build factories with fewer floors. This may have been possible in France as water power allowed the industry to remain more rural than was the case in Britain, where the industry moved into dense, urban environments.

When knowledge about plant design diffused, this often occurred in spatial proximity. For both Britain and France, the historical record is full of examples of knowledge diffusion embodied in engineers and other skilled workers, who were hired by local entrepreneurs to design and build plants (Chapman, 1970; Chassagne, 1991). For example, English engineers were recruited in Toulouse (circumventing legal bans) to build a cotton spinning plant. Once they had finished, a number of them were hired by other local entrepreneurs, and some set up their own mills nearby

(Chassagne, 1991, pp. 243-244). Additionally, the French (local and central) government often incentivized the diffusion of knowledge. Horn (2006, pp. 83-84) describes how the Bureau of Encouragement at Amiens (in Picardy) provided capital for an English machine-builder to install machines for one French plant. In exchange, the firm had to commit to sharing "their techniques and technical know-how" with other firms. Other entrepreneurs in Picardy came to the plant in Amiens to study their machines and processes. The technology transfer that took place was not passive. Local workers tinkered with the machines installed by the English and improved on their designs in various ways, adapting them to local needs (Horn, 2006).

2.3 Comparison Sectors: Metallurgy and Paper Milling

We have highlighted the challenges in reorganizing production in mechanized cotton spinning. To distinguish these from other, broader, trends at the time, we examine two sectors that did not need to reorganize production during this period – our 'comparison sectors.' We summarize the most important characteristics of these industries for our purposes below. Appendix A.4 contains a more detailed discussion.

Metallurgy, the sector that supplied iron and steel to the rest of the economy, was a flagship industry of the Industrial Revolution. Paper milling – while not particularly important for other sectors – also underwent mechanization, which renders it a useful comparison sector. Despite the obvious differences in the production processes, metallurgy and paper milling share an important characteristic that differentiates them from mechanized cotton spinning.

Difference with cotton spinning: plant production before 1800. Both metallurgy and paper milling had already organized production in plants since well before the Industrial Revolution. In metallurgy, plant production was mostly due to a reliance on high fixed-cost machinery such as the furnaces used both in smelting and refining. In paper milling, production was organized in plants because of a reliance on water power. The early start to plant-based production meant that these sectors had already accumulated industry-specific expertise in building design and labor management. Thus, both comparison sectors fit well the characterization of 'proto-factories' described above. Appendix A.5 uses *Encyclopédie* plates to show that best-practice methods and codified knowledge already existed for the two comparison sectors in the late 18th century. These plates illustrate crafts, processes, and inventions, thus representing a unique source of information to study the existence of codified manufacturing knowledge at the time. As André (1996, p. 21) writes about the entries referring to paper milling, "The classic reference for technical descriptions of paper milling is Diderot and D'Alambert's famous 18th century *Grande Encyclopédie* ... [Here], one will find the different practices, with the terminology used in different provinces..." For metallurgy and paper milling, there were many *Encyclopédie* plates illustrating information on plant organization and production technology, while there are no such plates for mechanized cotton spinning (Figure A.9). Note that the absence of plates on mechanized cotton spinning is not surprising, since this technology had just been invented; the most important take-away from these data is that codified knowledge was indeed available for our two comparison sectors in the late 18th century.

Similarities between the comparison sectors and mechanized cotton spinning. Once cotton spinning had mechanized and shifted to plant production, the subsequent development shares important similarities with the comparison sectors. All three experienced a steady flow of productivity-enhancing 'micro'-inventions over our sample period, and in all cases, these were integrated into *existing* plants. In each sector, workers needed to be retrained to work with the new equipment (Gille, 1968; Horn, 2006; André, 1996). However, there was no need to re-organize production, as the plant setup was not affected.

The source country of the technology in each case was Britain. Thus, similar barriers to diffusion applied to all sectors, and machines were typically built in France because of the ban on exports from Britain. Appendix A.6 presents patenting data from Britain, showing that all three sectors witnessed the consistent arrival of patents during our sample period. Spinning was the third-most patent intensive out of the 146 categories; metallurgy ranked ninth, and paper milling twenty-first. It should be noted that spinning patents include those for all textile fibers, so that patent intensity in *cotton* spinning was actually closer to those in the comparison sectors.

In what follows, we briefly discuss the most important innovations in the comparison sectors, showing that they were integrated in the existing plant settings. In paper milling, the main technological innovation was the mechanization of forming paper with the Fourdrinier machine (one step in the production process). This invention still forms the basis of paper making today. Changes in the factory layout were not required to introduce the Fourdrinier machine (see Figures A.6 and A.7). It was thus uncommon to establish new plants for the sole purpose of mechanization. Modifications and enlargements of existing plants were often undertaken without having to substantially modify other parts of the production process, and – in contrast to flow production in cotton milling – different parts of the paper milling process could be hosted in different buildings. For example, when plants adopted the Fourdrinier machine, they typically merely reconstructed the two sections hosting the cylinders and the machine, while re-using the buildings previously devoted to the other operations (André, 1996, p. 178). Similarly, while workers had to be trained to operate the new technologies, there is little surviving record of substantial labor conflicts or management challenges (André, 1996, p. 246).

Metallurgy in France also witnessed technology adoption in *existing* plants (Gille, 1968), suggesting that major reorganization of production was not necessary. The most prominent innovation was the switch from charcoal to coal. Given that the new equipment itself was similar to the older one (Pollard, 1965), its introduction merely required modification or replacement of existing machines and ovens (see the illustration in Figures A.4 and A.5). The main reason for setting up new plants was not technological, but to locate closer to coal sources.

In summary, the three industries share important similarities in adopting new technologies during our sample period. However, the need to reorganize production, and the introduction of flow production, was unique to cotton spinning.

3 Data

Our analysis is based on a novel, plant-level dataset for the initial phase of industrialization in France. The data have a panel-like structure covering three industries: mechanized cotton spinning, metallurgy, and paper milling. We observe plants in these sectors at two points in time: around 1800 and around 1840. We construct the dataset from four main sources that we describe below. We also discuss the construction of the main variables in our analysis: plant-level labor productivity, plant location, and plant survival for all sectors. Appendix B contains further information on data sources and processing.

3.1 Plant-Level Industrial Surveys

Detailed, plant-level industrial surveys form the basis of our dataset. We build on large-scale data collections by the French state over the period 1789-1815. France's innovations in data gathering during this period are frequently praised as the basis of modern statistical data collection (Perrot and Woolf, 1984). We use data from three industry-specific surveys conducted around 1800 and link these to the first manufacturing census in France, 1839-47.

The survey for paper milling was implemented in 1794 during the French Revolution; it contains data on 520 plants. The most important survey for our analysis – mechanized cotton spinning – was conducted by the Napoleonic regime in 1806, covering 340 plants. The survey in metallurgy in 1811 covers 470 plants. Finally, the first manufacturing census in France was initiated in 1839, and the results were published in 1847. While this census covers all manufacturing establishments, we only use data for cotton spinning (528 plants), metallurgy (896 plants), and paper milling (347 plants). For simplicity, we use "1800" throughout the paper to refer to the period of the three early surveys, and "1840" to refer to the later manufacturing census.

The quality of our data sources is high. Each has been scrutinized by economic historians (Bonin and Langlois, 1987; Chanut, Heffer, Mairesse, and Postel-Vinay, 2000; Chassagne, 1976; Woronoff, 1984), allowing us to understand their strengths, as well as their limitations. The three industrial surveys and the first industrial census were conducted in a similar way. The central government sent detailed, standardized questionnaires to local government officials (usually prefects at the département level). The officials and their subordinates (subprefects and mayors) were tasked with identifying and enumerating the relevant plants in their jurisdiction. Plants themselves typically submitted the requested information to local officials, who were also tasked with validating that the submitted records were correct (Ministère de l'Agriculture et du Commerce, 1847, p. xviii). In all cases, the motivation for collecting the data was to gather statistical information as opposed to tax collection.

The fact that the surveys were carried out by local officials contributed to the high data quality, as they were able to cross-check the responses against other sources and use expert guidance (Perrot and Woolf, 1984, p. 161).[6] Furthermore, France had experience in conducting industrial surveys dating back to the *ancien régime*. Thus, in many cases officials merely needed to update existing knowledge as opposed to starting from scratch (Chassagne, 1976; Ministère de l'Agriculture et du Commerce, 1847). For all surveys, the geographic coverage is close to complete, with only a handful of *départements* failing to submit returns. These surveys have been characterized as a true administrative feat of the French government (Chassagne, 1976, p. 350).[7] Figure A.14 shows the spatial distribution of plants around 1800 and in 1840 for the three industries. Although plants are more concentrated in some regions than others, we have broad coverage across all of modern-day France. For the three industry-specific surveys from the 1800s, the data that have survived are the handwritten returns submitted by the *département* and located in the National Archives in Paris. These are at the plant level in all cases, but were not cleaned or harmonized by the central authorities in any way at the time. We use the data assembled by Juhász (2018) for the mechanized cotton spinning industry. For the other two sectors, we collected, digitized, cleaned, and harmonized the handwritten surveys. For the manufacturing census from 1840, the data were cleaned, harmonized, and organized centrally. The results were published by the Ministry of Agriculture and Commerce in four volumes in 1847 *"(Statistique de La France: Industrie)"*. These volumes served as the raw data for the digitization undertaken by Chanut et al. (2000). We use the latter source for our analysis.

Appendix B describes the specific context for each individual survey and assesses their quality along a number of dimensions. A limitation of our overall setup arises from the different years in which the initial surveys were carried out across the three sectors (1794, 1806 and 1811). This may present a challenge if the economic environment changed over this period. We confront this issue by examining the robustness of our results to some of the most important shocks in this period (in particular, the Napoleonic blockade and the Napoleonic Wars).

3.2 Main Data Construction Steps

In what follows, we discuss the most important steps in the construction of our dataset.

[6] For example, Perrot (1975, p. 161) writes; "Woronoff has demonstrated the high level of reliability of the statistics on mines and forges, where the visitations by engineers corresponded to the sophisticated elaboration of questionnaires calculated to trap forge owners into revealing accurate figures."

[7] In particular, Chassagne (1976, pp. 350-351) describes the mechanized cotton survey as follows: "the survey obviously testifies first of all to the efficiency of the prefectural system. Of the 109 prefects questioned, 107 responded [...]. This represents a real administrative performance, since none of the prefects had the information immediately available to respond to requests from the central administration. The realization of this work certainly made it possible, at all levels of the administrative hierarchy, to appreciate the competence and to stimulate the zeal of the civil servants, who were always candidates for a promotion."

3.2.1 Estimating labor productivity

Our main variable of interest is plant-level labor productivity, defined as log revenues per worker. We use this measure in our baseline estimates because it can be constructed for all sectors and in both time periods. For mechanized cotton spinning, we can also construct plant-level TFP. We face two challenges in constructing consistent productivity measures across plants and time. First, while the surveys for the three sectors around 1800 report output *quantity* (and some information on product-specific prices and quality), the census in 1840 reports plant-specific *revenues* (but not output quantities). To render productivity measures comparable over time, we have to construct revenues for 1800. Second, worker categories are not consistently reported across all plants in 1800 in metallurgy and paper milling. We describe how we deal with each of these issues below.

3.2.2 Estimating plant revenues in 1800

In cotton spinning, the 1806 survey reports the quantity of yarn spun as well as the minimum and maximum count of yarn spun – where the count of yarn is the standardized measure of quality in the sector.[8] We construct plant-level revenue by multiplying the quantity of plant-level output by the price of the average quality of yarn produced by the plant. We use a schedule of prices for different counts of yarn reported by the French government (*Archives Nationales, F12/533*). In practice, the adjustment in price for the different qualities produced is not crucial for our results (which we confirm with a battery of robustness checks). The reason for this is that the quality produced by the majority of plants is fairly similar. The interquartile range (25th to 75th percentile) for the average quality of yarn produced by the plants in our sample is 20 – 47.5. That is, most plants produced relatively low quality cotton yarn in this period (high count yarns typically start around 100). This is consistent with the British experience (Harley, 1998).

In metallurgy, the 1811 survey asked for the quantity of output produced (by product, e.g., pig iron, natural steel, etc.), as well as the price charged by the plant, by product type. While product-specific output quantity is reported by all plants, the product-specific price is only reported by a subset of plants. We compute the average price for each product using the subset of plants where the price information is available. We obtain plant revenues by multiplying product-specific plant output by the average price for the respective product, and summing across products of a given plant.

In paper milling, the 1794 survey reports the total quantity of paper products produced, but it does not provide plant-specific output prices. To construct revenues, we multiply plants' output quantity with the average price of paper products as reported in the *Tableaux du Maximum* – a data source compiled in 1794 during the French Revolution that provides detailed data on goods prices and

[8]We use the (unweighted) average of the minimum and maximum count of the yarn produced by the plant as a proxy for its average output quality. The maximum and minimum count is the only information that plants provided on the quality of yarn that they produced.

trade links across French regions.

Finally, to compare revenues in the earlier periods and in 1840, for all three sectors, we deflate revenue data using the wholesale price index for the respective survey years from Mitchell (2003). We note in passing that potential errors in the deflators would affect our estimates for *average* growth rates in the three sectors between 1800 and 1840, but they would not change the growth patterns over the plant distribution (e.g., the lower-tail bias in cotton spinning).

3.2.3 Constructing consistent labor variables

Next, we describe how we harmonize labor input in our two comparison sectors. In cotton spinning, we observe the total number of workers, so that no adjustment needs to be made.

In metallurgy, about 40% of the plants reported either 'internal' labor only, or both 'internal' and 'external' labor, separately. The remainder of plants reported only total labor, with no indication of whether this includes external labor.[9] To construct a consistent measure of labor input, we estimate the size of the internal labor force for the 60% of plants that reported only total labor. We use a matching procedure based on plant characteristics that is described in Appendix B.2. We also show below that our empirical results for metallurgy hold when we restrict the 1811 data to the 40% of plants with direct information on internal labor.

In paper milling, many plants only reported male labor in 1794, while the 1840 survey reports both male and total labor. In order to compare output per worker consistently, we need to impute total employment in 1794. We scale male labor in 1794 by the average proportion of total employees to male employees in 1840.[10] We show that our results are robust to using only male employees in both periods.

3.2.4 Distinguishing plants from home production in cotton spinning

Mechanized spinning was operated in centralized locations (plants), while the old technology relied on home production. We can thus identify the users of the new technology (i.e., all *plants* observed in our data). Consequently, we are able to isolate the productivity distribution for plants that used the new, mechanized technology under the new organizational form.

3.2.5 Plant linking and plant survival

We link plants across the two sample periods within communes based on two metrics: i) the plant had the same (or very similar) owner name in both periods; ii) there was only one plant in the

[9] Woronoff (1984, p. 138) describes external labor as only having very loose ties to the plant. These workers did not typically work at the location of the plant, their work was not supervised by the manager, and their identity was often not even formally known to the manager. They performed tasks such as driving, collecting charcoal for the plant, or performing other jobs without belonging to the hierarchy or reporting to superiors in the chain of command. Thus, external workers were unlikely to be considered formal salaried employees of the plant in the 1840 census.

[10] The validity of this method hinges on the assumption that the ratio of total employment to male employment was constant over time. We are able to check this using the 18 plants in 1794 that reported both types of labor. We find a ratio of 2.11 in 1794, which is very similar to the ratio of is 2.29 in 1840 (among all plants).

respective sector in the commune in 1800 and *at least* one plant in the same sector in 1840. In what follows, we describe each step of this process and the assumptions it entails.

All three surveys from around 1800, as well as the 1840 census, report the name of the owner and the location of the plant up to the commune, which is the lowest administrative unit in France. In bigger cities such as Paris, the *arrondissement* is also reported. We construct a consistent measure of plant location across surveys by assigning each plant to its modern-day commune, *département*, and region (as described in Appendix B.1 - B.4 for each respective survey).

Linking plants. We use two pieces of information to link plants over time. First, we match plants by their owner names in a given commune in the respective industry (see Appendix B.5 for detail on the implementation). Since the name of the owner may change even if the physical structure of the plant was the same, we also match by location in a second step. We match locations that had only *one* plant in the respective sector in 1800, and at least one plant active in the same sector in 1840. This turns out to be fairly common in the data. An obvious concern is whether this 'local matching' indeed identifies the same plant. This is likely, given a fortuitous feature across all three of our industries: their reliance on water power. Only a small number of locations in a typical commune were suitable for setting up a water-powered mill, as rapid stream flow was needed to yield sufficient power. Moreover, the backwater created by one mill meant that another mill could not be located in close proximity. Consistent with this, Crafts and Wolf (2014) argue that agglomeration in the cotton textile industry was not observed until steam became the common source of power in Britain. Consequently, our 'local matching' arguably identifies plants that had the same location within communes. Whether these were owned by the same entrepreneur (or their descendants), or whether they had passed on to a different owner is not crucial for our analysis.

Plant survival. Our main measure of plant survival is based on the combination of matching by owner name and 'local matching' that we described above. We define the survival rate as the percentage of plants from the initial period that survive into the later period. The numerator counts all plants that fulfill at least one of the following two conditions: i) the plant had the same owner in both periods; ii) there was only one plant in the respective sector in the location in the initial period and *at least* one plant in the same sector in 1840. The denominator is the sum of *all* plants in the given sector in the initial period. We provide a verification of this methodology in Appendix B.5.

Note that our baseline measure of plant survival does not adjust for the fact that the number of plants located in communes that had only one plant varies across the three sectors in our sample.[11] Thus, we may mechanically find higher survival rates in a sector where single-plant communes were relatively more frequent. To address this issue, we also construct the 'restricted sample'

[11] Among the 520 plants in paper milling in 1794, 211 (40.6%) were the only plants active in their commune in this sector. For cotton spinning in 1806, the proportion is 25.6% (87 out of the 340 plants), and in metallurgy in 1811, 69% (324 out of 470 plants).

survival rate as a robustness check. This measure is based solely on single-plant locations. The numerator of the 'restricted sample' survival rate counts the number of communes that had only one plant in the respective sector in the initial period, and at least one plant in 1840. The denominator is the set of all single-plant communes in 1800.

3.3 Descriptive Statistics

Tables A.2-A.4 contain descriptive statistics for all variables used in the analysis. Several important features of the data stand out. First, the scale of plants (measured by the number of employees) is striking for cotton spinning plants (see Table A.2). The average spinning plant in 1806 had 64 employees. This is larger than metallurgy and paper milling, where plants had on average 23 and 13 employees, respectively. Second, between 1806 and 1840, the mechanized cotton spinning industry expanded substantially, and this expansion was accompanied by an increase in the number of plants active in the market (from 340 in 1806, to 528 in 1840). That is, for every plant that exited the market, more than one new plant entered. As such, the results we present below constitute more than a 'shake-out' of unsuccessful plants.

4 The Pattern of Productivity Growth

In this section, we study the evolution of the plant productivity distribution in mechanized cotton spinning after the new technology had been adopted. Similar in spirit to a difference-in-differences strategy, we contrast the observed patterns with those in the two comparison sectors – metallurgy and paper milling. This allows us to distinguish the unique feature in mechanized cotton spinning – the need to reorganize production – from other common factors that affected productivity growth in all three sectors over our sample period.

4.1 Average and Quantile Productivity Growth

We begin by examining average annual labor productivity growth. Column 1 in Table 1 shows that all three sectors experienced a significant increase in labor productivity between 1800 and 1840. This is consistent with the historical evidence that gradual innovations were incorporated in all three sectors during this time period. The largest productivity gains were achieved in cotton spinning (2.42% per year), followed by metallurgy (2.33%) and paper milling (0.72%).[12] It is noteworthy that the large productivity increase in spinning reflects improvements *within* the

[12]Given that we discount revenues using price indices, all our productivity calculations reflect price-adjusted revenue-based productivity. To obtain the average annual growth rates between the two time periods, we first regress log output per worker $lnProd$ on a dummy for 1840 (separately for each sector, including the data from both time periods). This coefficient measures the percentage growth in output per worker over the entire time period between the respective survey years. We then annualize these values (and corresponding standard errors) by dividing by the number of years between the surveys in each sector. Note that this method delivers average annual growth figures, not accounting for compound growth. In cotton spinning, the overall growth over the period 1806-40 amounts to 82% (2.42% per year x 34 years). In metallurgy, it is 68% (2.33% per year x 29 years), and in paper milling it is 33% (0.72% per year x 46 years).

mechanized technology.

In which part of the productivity distribution were these gains concentrated? Figure 1 plots the distribution of labor productivity in the three sectors at the beginning and at the end of our sample period, illustrating our main results. In cotton spinning, two features stand out. First, the initial dispersion in labor productivity was large in 1800 relative to that in 1840.[13] Second, the productivity gains are almost exclusively concentrated in the lower tail: The lower tail disappeared over our sample period, while increases in productivity at the upper tail were modest. In other words, productivity growth occurred largely due to the distribution shifting towards the productivity frontier. The contrast between cotton spinning and our two comparison sectors is striking. In metallurgy and paper milling, the entire productivity distribution shifted to the right between 1800 and 1840. Quantile regressions confirm this pattern. Columns 2-6 in Table 1 report these results for the three sectors, estimating regressions for productivity growth at different quantiles of the productivity distribution. Figure 2 displays the corresponding coefficients. In cotton spinning, the bias towards productivity growth in the lower tail is marked. Productivity growth at the 25th percentile was twice as large as that at the 75th percentile (3.3% per year relative to 1.65%), and the difference is fourfold between the 10th and the 90th percentile (3.9% and 1.0%, respectively). In the comparison sectors, productivity growth occurred relatively evenly across the distribution; if anything, growth was concentrated in the upper tail in metallurgy. These differences between cotton spinning and the comparison sectors constitute suggestive evidence that the aspect unique to mechanized spinning – the need to reorganize production – was associated with the lower tail bias in productivity growth.

4.2 Robustness of the Lower-Tail Bias to Construction of Productivity

Before we examine mechanisms behind the upper-tail bias in productivity growth in cotton spinning, we document that it is robust to using alternative measures of productivity.

Output quality. Could the different productivity-growth pattern in cotton spinning be driven by differential trends in output quality? Recall that our data for cotton spinning in 1800 enable us to use quality-adjusted prices to compute revenues from output quantities. Panel B in Table A.5 in the appendix presents quantile regressions without quality-adjustments in 1800, i.e., using the same sector-level price across all plants in cotton spinning. The magnitude of the lower-tail bias is slightly smaller, but it still holds. Quality adjustment does not substantially alter the pattern of productivity growth because most plants produced yarn of a fairly similar (low) quality: The interquantile range for the quality of yarn produced in our data is 20 – 47.5; high-count yarns

[13]The 90-10 percentile productivity range (the difference in log output per worker between plants in the 90th and 10th percentile) decreased from 2.17 to 1.17. Thus, in 1800, 90th percentile cotton spinning plants were 8.7 times more productive than 10th-percentile plants, and this ratio fell to 3.2 in 1840. The latter is comparable to the average 90-10 percentile productivity ratio within 4-digit US manufacturing sectors in 1977, where this factor was about 4 (Syverson, 2004).

typically start around count 100 (Harley, 1998). The low quality output during the initial phase is consistent with the British experience (Harley, 1998).

Mark-up heterogeneity. In cotton spinning, our plant productivity measure in 1800 is computed from physical output (adjusted by quality-specific sector-level prices), while the 1840 values are based on *revenues*. The latter also reflect differences in markups across plants (Garcia-Marin and Voigtländer, 2019). If more productive plants also charged higher markups (De Loecker and Warzynski, 2012), the heterogeneity in markups would lead to a *more* dispersed productivity distribution in 1840. Thus, this data limitation – if quantitatively important – would work against our finding of a *tightening* productivity distribution.

Robustness to measuring productivity as TFP. Our baseline productivity measure is log output per worker. For cotton spinning, we can also compute TFP, using data on physical capital (number of spindles) – see Appendix E.1 for detail. Panel C in Table A.5 confirms the lower-tail bias of productivity growth in cotton spinning using TFP.

Robustness to imputed variables in metallurgy and paper milling. Appendix E.2 shows that our quantile regression results are robust to i) using only those plants in 1800 in metallurgy with direct information on internal labor (Table A.6, Panel B); ii) using the plant-product-specific prices for those metallurgy plants that reported them to construct productivity, while dropping the remaining metallurgy plants in 1800 (Table A.6, Panel C); iii) using only male labor in both periods in paper milling (Table A.7).

5 Mechanism: Learning About Best Practice in Factory-Based Production

We have documented that the lower-tail bias of productivity growth was present only in cotton spinning. The historical evidence in Section 2 suggests that the need to reorganize production was also unique to this sector. In what follows, we link these two features: the need for reorganization can explain the lower-tail bias. We first discuss a stylized framework that captures the key elements of the historical evidence. Then, we turn to examining the data for evidence compatible with the learning effects that reorganizing production entailed.

5.1 A Stylized Framework

We summarize the key features of our stylized model here; Appendix D provides detail and shows simulation results. The model features heterogeneous plants, and we distinguish three phases: Initial market entry, exit of particularly unproductive plants, and endogenous search for better technology among the surviving plants. To model the observed thick lower tail in the initial productivity distribution, we use a production function with multiple *complementary* inputs (tasks). We think of these as the organizational challenges discussed in Section 2.2 such as plant design, power supply, or management of workers. For each input, a plant receives a random efficiency draw, reflecting the plant's organizational knowledge.[14] Due to the strong complementarity across the inputs, even having one bad draw (say, inappropriate factory layout) substantially reduces overall efficiency. This gives rise to the thick lower tail in the initial productivity distribution.

Our model then features two mechanisms that lead to the disappearance of the lower tail over time: First, exit of the least productive firms eliminates the lowest part. This standard mechanism (c.f. Hopenhayn, 1992) reflects the significant exit rates of cotton spinning firms in the early mechanization period. Second, motivated by the historical evidence on the diffusion of organizational practices, we model a process à la Perla and Tonetti (2014) whereby relatively unproductive firms can search and copy the organizational knowledge of more productive firms: Firms with productivity below an endogenously determined cutoff halt production and search for better organizational knowledge; they are matched to a randomly-drawn firm among the higher-productivity firms that continue production. Searching firms thus forego profits, but they expect higher profits in the future due to the improved organization of production. This process leads to the disappearance of the lower tail, while the productivity frontier remains unaffected.

Figure 3 illustrates the simulated productivity distributions over the three phases. The dashed line reflects the initial productivity distribution. In the first phase, firms with very low productivity draws exit. In the second phase, among the surviving firms, the relatively unproductive ones halt production and search for better organizational knowledge among the more productive firms, who continue to produce. This gives rise to the solid blue productivity distribution, where the initially fat lower tail has disappeared, while the productivity frontier (plants that already operate with high organizational efficiency) remains unchanged. This pattern of productivity growth mirrors the one for cotton spinning in the data (see Figure 1). Note also that the final productivity distribution resembles the one observed for our comparison sectors, where factory production had been adopted earlier, so that the process of exit and organizational learning had already occurred by 1800.[15]

While our stylized theoretical framework is not the only one that gives rise to lower-tail bias in productivity growth, it is a simple setup that represents the key features of the historical context. The empirical results below provide evidence for these mechanisms, suggesting that the diffusion of organizational practices across plants was an important dimension in the observed productivity dynamics.

[14] Of course, these draws can also be interpreted as technology. Our model is agnostic about the distinction between technological and organizational knowledge. However, our historical accounts and empirical evidence suggest an important role for the latter.

[15] Note that in order to replicate the right-shift of the whole productivity distribution over time in the comparison sectors, we would have to introduce productivity growth for *all* plants. In contrast, we do not need this additional feature to rationalize the dynamics in cotton spinning, i.e., learning from more productive plants is sufficient to deliver the observed lower-tail bias in productivity growth.

5.2 Learning About Re-Organizing Production

In what follows, we shed light on *what* plants needed to learn to re-organize production efficiently, followed by *how* plants went about learning. Through the lens of our stylized framework, plants initially had a wide array of efficiency draws, reflecting the fact that they largely experimented with different solutions, rather than having standardized answers to the organizational challenges of factory production. Over time, less productive plants either exited the market or they learned better organizational practices from their peers. We present evidence consistent with this initial experimentation and subsequent learning, using one central aspect in the move to factory-based production: optimal mill design.

Building design. The historical literature highlights building design as the foremost challenge faced by French mechanized cotton spinners (Bonin and Langlois, 1987). We are able to examine two aspects of mill design for a subset of 59 historical mechanized cotton spinning plants, using data on their floor plans. The data are from Chassagne (1991).[16] For each plant, we observe the length and width of the plant building, as well as the number of floors. The historical evidence suggests that both dimensions were important: A building with many floors presented a fire hazard, which is why over time, the French converged to buildings with fewer stories (Chassagne, 1991). In addition, a more rectangular (less 'square') building shape was more suitable for the optimal layout of machines (Markus, 2013, pp. 265-267).

In light of this, we examine the number of floors a building had, as well as its 'squareness' (defined as $S \equiv \frac{length \times width}{max\{length, width\}^2}$). This measure of squareness is size-invariant. That is, if buildings simply got bigger over time without changing their shape, our squareness measure would remain unchanged. In light of our model, we expect there to be an initially wide distribution of these dimensions, reflecting the dispersed initial draws. Over time, as plants learned about what worked, we expect convergence to best-practice designs. Importantly, the data contain information on the year in which a plant was set up. We thus split the sample into an 'early' experimental period (defined as plants set up before 1820) and a 'mature' period (defined as plants set up after 1820). Of the 59 plants for which we have building dimensions, 58 have a date of foundation ranging between 1789-1845. Exactly one-half of these plants are assigned to each period.

Figure 4 plots the estimated kernel densities for the number of floors (Panel A), and for building squareness (Panel B). Consistent with our model and the historical evidence, there is wide variation along both dimensions in the early period. Buildings had anywhere from 0-7 stories and all sorts of shapes ranging from a squareness measure of $S = 0.1$ (indicating very asymmetric width and length) to perfect squares ($S = 1$). Moreover, as the historical literature claims, over time, best

[16] Appendix E.3 contains a more detailed description of this source. It is important to note that the data are unlikely to be a representative sample as records of larger, more important plants were more likely to survive. However, note that this bias would likely lead us to *underestimate* the extent of initial experimentation, as we arguably undersample less successful plants with less efficient organizational draws.

practice converged to buildings with a moderate number of floors (around 3) and a more rectangular shape ($S \approx 0.5$).

Was better building layout correlated with productivity? Appendix E.3 explores this question. Since the Chassagne (1991) data do not include the necessary information to compute plant productivity, we examine a proxy: plant survival until the 1840 census (which collected plant data in 1839-47). Table A.8 regresses a dummy for plant survival on each of the two dimensions of plant layout, allowing for a non-linear relationship. We find a hump-shaped pattern. The estimated coefficients imply that the optimal number of floors was about 3.7, and the predicted optimal squareness was $S = 0.5$, which is close to what the industry converged to after 1820, according to Figure 4. These results on building design support the historical account as well as the mechanism outlined in our stylized framework: Cotton spinning plants initially experimented with a wide range of organizational practices, and as the industry matured, they converged to best-practice designs.

Strikes. Another organizational challenge that the historical literature highlights is that firms needed to develop new labor management practices (see Section 2.2). While we do not observe historical management practices, we follow recent work that points to a proxy: Bianchi and Giorcelli (2022) show that firms that improved their labor management as a result of a training intervention saw a significant decline in worker complaints and strikes. Historical evidence from our sample period in France also links key labor management challenges in factory-based cotton spinning to strike activity. Through the 1820s and 1830s, there were several strike episodes in France where workers demanded payment as a time wage as opposed to a piece-rate, as well as a decrease in the length of working hours (Chassagne, 1991). The former in particular is a key labor management challenge that we described in detail in Section 2.2. In stark contrast, André (1996, p. 246) notes that paper milling firms' records do not mention similar conflicts with labor during the mechanization process.

Based on this historical evidence, we examine whether strikes were more frequent in textiles relative to our comparison sectors that did not experience similar labor management challenges. Appendix E.4 provides suggestive evidence, using data on strike activity at the sector-*département* level. Table A.9 shows that the textile sector was subject to about 30% more frequent strike activity, relative to the comparison sectors – metallurgy and paper milling. This holds conditional on controlling for employment at the sector-*département* level, total manufacturing employment (across all sectors), and *département* fixed effects. Of course, factors other than management challenges may also be responsible for the differential strike activity in cotton spinning. We thus interpret the evidence from strikes with caution, and merely view it as complimentary to the historical record in pointing to labor management challenges in mechanized cotton spinning.

5.3 Spatial Diffusion of Knowledge

In what follows, we shed light on how learning about organizational practices took place. The historical background discussed in Section 2 showed that plants copied successful designs and setups of the production process from each other. To examine this channel, we estimate whether a plant's own productivity was higher in the proximity of other high-productivity plants. We use the following specification:

$$lnProd_{ij} = \beta_0 + \beta_1 lnDist_{ij}^{p90} + FE_j + \epsilon_{ij} \ ,$$

where $lnProd_{ij}$ is labor productivity (log output per worker) for plant i located in *département* j; $lnDist_{ij}^{p90}$ is log distance to the nearest plant (in the same sector) with productivity in the 90th percentile (in the distribution of *all* plants in the sector across France). Plants that are themselves in the top productivity decile are excluded from the sample to avoid introducing a mechanical relationship. All specifications include *département* fixed effects (FE_j) to absorb unobserved location characteristics that may make all plants in a given area more productive, irrespective of local spillovers. Thus, the coefficient of interest, β_1, reflects the extent to which plants in the same *département* benefit from being located closer to a high-productivity plant (which may be located in the same or in another *département*). We do not interpret these correlations as causal effects, but as evidence that is compatible with spatial spillovers of knowledge. We estimate the specification separately for the three sectors, and in both time periods. Standard errors are clustered at the *département* level to account for spatial correlation.

Before presenting the results, we first examine the spatial distribution of high-productivity plants across our sectors and time periods. Figure A.17 plots the spatial distribution of cotton spinning, metallurgy, and paper milling plants, distinguishing those in the 90th percentile of the productivity distribution. Unsurprisingly, some regions have a larger concentration of high-productivity plants than others. Due to our use of *département* fixed effects, these regional differences do not affect our results.

Figure 5 visualizes our baseline results, and Table A.10 in the appendix reports the corresponding regressions. To allow for direct comparability, we report standardized beta coefficients of $lnDist^{p90}$ for all three sectors in the two periods. The estimated coefficient for cotton spinning in 1800 is negative, statistically significant and large in magnitude. A one-standard-deviation (std) increase in distance to a high productivity plant is associated with a 0.84 std decline in labor productivity. The pattern is much weaker in the two comparison sectors in 1800 – the coefficients are less than one-third in magnitude as compared to cotton spinning. In addition, in 1840, all three sectors show at best a muted relationship: The distance-coefficient for mechanized cotton spinning is reduced to less than one-fifth of its initial size, and it is only marginally statistically significant.

In the comparison sectors, the coefficients of interest are also further reduced slightly, and they are no longer statistically distinguishable from zero. Thus, proximity to high-productivity plants mattered the most in cotton spinning in 1800, i.e., in the period before knowledge about the optimal organization of production had spread widely. Distance mattered much less when organizational knowledge had diffused: In cotton spinning in 1840, and in the comparison sectors in both time periods. These findings are consistent with spatial learning during the early phase of mechanized cotton spinning.

Alternative explanations for the distance results. In Appendix E.5, we examine possible alternative explanations for the relationship between $lnDist^{p90}$ and productivity. While FE_j capture unobserved differences that vary at the *département* level, they cannot account for unobserved differences at a finer spatial level. To address this possibility, we implement several additional checks. First, we control directly for prominent location fundamentals at the commune level such as the availability of fast-flowing streams (as a source of water power), proximity to coal (for steam power), and the share of forest cover (for access to charcoal – a major input in metallurgy). Table A.11 shows that our results are highly robust to these controls. Second, we check whether our results are affected by more general agglomeration externalities, as opposed to learning. In Table A.12, we control for the density of production at the commune level (measured as the log of total output in the sector, excluding a plant's own output). Our results are essentially unaffected by adding this control. Third, Table A.13 conducts a placebo exercise, showing that, in cotton spinning, plant productivity in 1800 was *not* related to the distance to high-productivity plants in 1840. This suggests that the large estimated coefficient in our baseline specification is not driven by persistent location fundamentals within *départements*. Fourth, we examine whether *ex-ante* high-productivity plants may have selected into 'productive locations' (i.e., chose to locate near existing high-productivity plants). Since we observe plant age in cotton spinning in 1806, we can examine selection patterns. Table A.14 shows that our result holds in a subsample of plants that entered *before* the nearest high-productivity plant, so that the timing of entry rules out the type of selection described above. Finally, we examine whether there is evidence for learning *across* sectors. Table A.15 shows that there is no consistent pattern in the data: Mechanized cotton spinning plants located closer to high-productivity metallurgy or paper milling plants in 1800 were not more productive.

In summary, the consistently larger distance coefficient estimated in cotton spinning in 1806, in combination with a series of robustness checks, points to the spatial diffusion of knowledge as one mechanism through which learning across plants took place. The historical accounts of spatial learning in Section 2.2 corroborate this interpretation.

5.4 Plant Layout vs. Technology: Evidence from Plant Survival and Age Profiles

Which type of knowledge diffused across space? The results presented above cannot distinguish between the spatial diffusion of knowledge about the technology itself (for example, how to operate and maintain the machines efficiently) and organizational knowledge (i.e., how to design mills and organize workers within the plant). We now present two patterns in the data that are more consistent with organizational learning.

Productivity handicap of exiting plants. If an important component of learning in mechanized cotton spinning occurred along the organizational dimension of factory design, we would expect initially low plant survival rates relative to the other two sectors. Building design and layout are either sunk at the time of building or costly to change, implying that plants who got this wrong were likely to exit the market. On the other hand, inefficient operation of the new technology itself could be adjusted within an existing factory so that, if anything, we would expect incumbents to have an edge over new entrants. The data speak in favor of the former. We find substantially larger exit rates in cotton spinning relative to the other two sectors. Table 2 reports plant survival rates over our sample period, using the two measures defined in Section 3.2 in each of the three sectors. Based on our baseline measure, survival rates in spinning (5%) were lower than in paper milling (10.8%) and much lower than in metallurgy (37.7%). Note that the actual difference in plant survival between spinning and paper milling was likely larger, because the survey for the latter was conducted in 1794, more than 10 years earlier than the cotton spinning survey (1806).

The 'restricted sample' survival rates in Table 2 allow a more direct look at the role of building layout and design in plant survival. By using only single-plant locations in the initial period, we are in effect testing whether a building used for production in a particular industry in 1800 continued to be used in the later period in the same industry (irrespective of who the owner was). The differences across the three sectors are even starker. By this measure, the survival rate in spinning (13%) is much lower than in the comparison sectors: 52.5% in metallurgy and 24.6% in paper milling (with the latter being an under-estimate, as discussed above). The low survival rate observed in cotton spinning means that many locations lost their (only) cotton mill – owners that had invested in a mill with poor layout had to exit the market, and the structure of the mill was not subsequently used by other plants in cotton spinning.

Appendix E.6 provides further results on plant survival. We first explain that the differential survival rates are not driven by the shift from water to steam power (which occurred more slowly in France than in Britain). Next, Table A.18 shows that exiting plants in mechanized cotton spinning were much less productive than those that survived. In the comparison sectors, the productivity handicap of exiting plants is also present, but less pronounced. In other words, early cotton plants that 'got it wrong' were particularly unproductive, which can explain the fat productivity lower tail in this sector. These plants eventually exited the market, and for many of them, the same building was not used by another plant in the industry. This pattern is consistent with large organizational challenges and low initial guidance in switching to factory-based production in cotton spinning.

Age profile of plant productivity. To further distinguish the role of best-practice organizational methods as opposed to learning about technology, we now examine the age profile of plant productivity. For now, assume that spinning technology itself did not change over time (we relax this in the next subsection). Then, if learning how to use the (already installed) technology was the dominant dimension, we would expect older plants to have accumulated more experience and hence have a productivity advantage. On the other hand, if efficient organizational design of the plant was more important, younger plants had a larger pool of knowledge to draw from, as they set up their design later. This would render younger plants more productive than older plants that were locked into less efficient designs (see also our evidence on plant design in Section 5.2).

We exploit the richness of our data to test this in both 1800, when best-practice mill design was still evolving, and in 1840, when according to Pollard (1965), the industry had reached maturity – at least in Britain. The 1806 survey in cotton spinning contains the year of foundation of plants. This allows us to compute a dummy for 'young' plants, defined as below-median age (with the median age in 1806 being three years). Column 1 in Table 3 shows that 'young' plants were 58% more productive in 1806. This could be driven by mechanisms other than the one discusses above. For example, new entrants may have used the most recent vintage of capital, leading to higher physical productivity (Foster, Haltiwanger, and Syverson, 2008). To address this issue, we control for several important plant characteristics in columns 2-6 of Table 3. These include the capital intensity of the plant (measured as log spindles per worker), the number of workers in the plant, and the vintage of machinery (binary variables for the three main vintages of machinery, from oldest to newest).[17] The productivity advantage of 'young' plants remains quantitatively very similar and statistically highly significant when we add these controls.

Appendix E.6 explores the 'young' plant productivity differential further. Table A.19 shows that in 1840, when the mechanized cotton spinning technology had reached maturity, younger plants did not have a productivity advantage anymore. For comparison, we also examine the age-productivity pattern for metallurgy plants, where best-practice knowledge had already been established by 1800. Correspondingly, we find that younger metallurgy plants did not have a productivity advantage in either of the two periods (Tables A.20 and A.21). In paper milling, data limitations prevent us from examining these patterns.

In summary, the evidence on productivity-age profiles is in line with best-practice organizational methods in cotton spinning spreading slowly over time and space, so that newly constructed plants were more productive around 1800. This advantage eroded when organizational knowledge

[17]The three different vintages of machinery are the spinning jenny (oldest), the water-frame (throstle), and the mule-jenny (newest). These are not mutually exclusive categories, as some plants used multiple vintages. Young plants tended to be more capital intensive, employ fewer workers, and use the newest vintage of spinning machinery.

diffused more broadly by 1840.

5.5 Robustness to Alternative Explanations

In the final part of this section, we consider additional alternative mechanisms that could also explain our results. We examine these along a number of different dimensions.

Gradual innovation and capital mix: Common patterns across all three sectors. We observe the lower-tail bias in productivity growth only in cotton spinning, and not in the other two sectors. For this reason, it is unlikely that factors that also affected the comparison sectors in similar ways can explain our findings. For example, the fact that mechanized spinning experienced innovations during our sample period seems unlikely to explain the lower-tail bias, as both comparison sectors also witnessed the introduction of new technologies. In a similar vein, improvements to power sources (notably water power, which remained the dominant source of power in cotton spinning) affected the other two sectors similarly (see also Appendix E.6). Finally, all three sectors used a mix of different vintages of their core capital equipment, suggesting that the choice of capital itself is not a confounder. In fact, below, we show that the lower-tail bias of productivity growth remains intact if we remove all plants that used the earliest vintage of capital: the spinning jenny.[18]

Regional differences in productivity distributions. It is important to highlight the robustness of our results to using only within-region variation (see Appendix E.7). Table A.22 shows that the lower-tail bias of productivity growth in mechanized cotton spinning remains intact if we add fixed effects for 22 French regions.[19] The lower-tail bias of productivity growth is somewhat muted, but still striking. Productivity growth in cotton spinning is almost twice as high at the lowest relative to the highest decile. The inclusion of region fixed effects absorbs the effect of different regional fundamentals, input markets, market access, or institutions. In what follows, we examine possible alternative mechanisms at a finer geographic level, i.e., potentially even within regions.

Market integration. Could increased market integration in cotton spinning explain our results? As the French economy became more integrated over time, it is possible that lower-productivity plants faced tougher competition and had to exit the market.[20] We address this concern in Appendix E.8.

[18]In this regard, the metallurgy sector is a helpful comparison as we observe the use of different vintages of capital in 1811. Some plants used the so-called 'direct technology' widely known in France as the Catalan forge, while others used the 'indirect technology' which separates the production process into smelting and refining. The direct technology was gradually replaced by the indirect technology during our sample period (Pounds and Parker, 1957). See Appendix A.4 for further discussion.

[19]Regions are larger than *départements*: There are 22 regions in France and 86 *départements*. Our data in the three sectors do not have sufficiently many observations to include *département* fixed effects, i.e., to estimate meaningful productivity distributions within *départements*. In contrast, our results on spatial diffusion (Figure 5) do include *départements* fixed effects, because they examine the effect of proximity to high-productivity plants on the *average* productivity of plants, rather than on the full distribution.

[20]Market integration arguably increased during our sample period both for policy reasons such as the abolition of internal barriers to trade during the French Revolution (Daudin, 2010), and because of infrastructure improvements that reduced transport costs such as the introduction of railways in the late 1820s.

We first use data on trade flows in 1794 from Daudin (2010) to show that market integration was initially *higher* in cotton yarn than in our comparison sectors – cotton yarn (and textiles more generally) are high value-to-weight products, which made them more easily tradable over long distances than iron and steel, or paper (Figure A.18). Given its higher starting point, if anything, we expect further market integration after 1800 to have been *less* pronounced in cotton spinning than the comparison sectors. This renders it historically unlikely that differential (i.e., higher) growth of market integration in cotton yarn drives our results. We complement this argument by controlling for measures of market access (both within France and within Europe), as well as for access to overseas trade (Table A.23). The coefficients of interest change only marginally, and the lower-tail bias of productivity growth in cotton spinning remains strong.

Napoleonic Blockade and Napoleonic Wars. Juhász (2018) shows that temporarily higher trade protection from British competition shifted the location of the mechanized cotton spinning industry within France. Since our results hold within regions, where the pattern of protection was very similar, it is unlikely that they are affected by the Napoleonic Blockade (1806-14). Figure A.19 presents further evidence that varying trade protection does not drive our results, by splitting the sample into plants in northern and southern regions in France (corresponding to the main dimension along which protection varied). The productivity distributions in the north and south are remarkably similar, and in both regions, productivity growth until 1840 was due to a disappearing lower tail. Appendix E.9 provides further evidence and shows that i) the Napoleonic Blockade did not drive the differential plant survival in cotton spinning; ii) conscription of soldiers as well as battles on French soil during the Napoleonic Wars (1803-15) are not associated with cotton spinning plant productivity.

Early spinning workshops. Another potential concern is that our results may be driven by the disappearance of small cotton spinning plants (or 'workshops'). Indeed our data may include relatively small plants that operated early vintages of mechanized spinning jennies and did not necessarily need inanimate sources of power. This small-scale setup may have been inherently different from the larger-scale factories powered by inanimate power sources. While the move to factory-based production was swift, systematic differences of smaller mechanized cotton workshops could account for the lower-tail bias of productivity growth in this sector. Our data does not differentiate between these two types of plants, as we observe the capital vintage, but not the power source in the 1806 survey. However, we can examine the extent to which our results may be driven by these forces.

In Appendix E.10, we adopt a stricter definition of 'factory-production' and omit plants with fewer than 10 employees from all sector-year pairs. This should exclude the majority of the smaller workshops that may have been organized as factory-based production along some but not all dimensions. It also addresses the concern that that smaller plants were under-sampled in the 1840 manufacturing census (see Appendix B.4). Table A.27 shows that the lower-tail bias in productivity growth is robust to using only larger plants, and that it remains unique to mechanized cotton spinning. In other words, small plants are not responsible for the fat lower tail in 1806. In addition, we implement an even more conservative definition of factory production. We drop the 76 plants from the 1806 survey that used the earliest vintage of machinery – the spinning jenny. These were the types of machines that could, in principle, have been operated also in small workshops without inanimate power sources. Table A.28 shows that the lower-tail bias of productivity growth remains striking. Thus, early spinning workshops that shared some, but not all features of factory-based production do not confound our results.

Plant scale. Appendix E.11 shows that it is unlikely that increasing plant size drives our results, as all sectors witnessed an increase in scale. In addition, as the results in Table A.29 demonstrate, controlling for the number of workers (at the plant level, in all sectors and in both periods) does not alter our findings.

Capital deepening. Over time, spinning machines were equipped with more spindles, and hence less labor was needed to produce a unit of output. We address this in Appendix E.12, where Table A.30 shows that the lower-tail bias of productivity growth remains robust and similar in magnitude when we control for the capital-labor ratio at the plant level (measured as log spindles per employee).

Machine quality. Machine production, and even maintenance, was typically in the hands of external regional suppliers (see Appendix A.2 under 'Historical evidence about machinery producers'). Given that we find the lower-tail bias of productivity growth also within regions, it is unlikely that heterogeneity in machine quality was an important driver.

Output quality. In Section 4.2, we already examined whether cotton output quality could account for our core result by affecting the computation of productivity. This is not the case – the lower-tail bias of productivity growth continues to hold even when we use output prices in 1806 that do not account for quality differences across plants (see Panel B of Table A.5). In Appendix E.13, we provide one additional check. Output quality could still drive our results indirectly if it led to higher sales and thus larger plant size over time, leading to scale economies. To examine this possibility, we estimate the quantile regressions *not* adjusting for quality differences in prices across plants in 1806, and controlling for the number of workers. Table A.31 shows that the lower-tail bias of productivity growth continues to hold.

Age profile of plants. The median age of plants in mechanized spinning in 1806 was strikingly small (3 years). This is consistent with historical accounts that the period witnessed the birth of mechanized spinning in France. Could it be that young plants are always more dispersed in terms of productivity, and that this drives the fat lower tail? Our earlier results in Table 3 render this interpretation unlikely. We have shown that younger plants were substantially *more* productive than older ones. Thus, if anything, a predominance of younger plants would tend to lead to a thicker *upper* tail.

Altogether, the findings of this section show multiple pieces of evidence that point to an important role for reorganizing production in explaining the unique lower-tail bias of productivity growth in mechanized cotton spinning. While this is not the only possible explanation, we have shown that numerous prominent alternative mechanisms are incompatible with the data. Together with the corroborating historical evidence, this leaves organizational challenges in technology adoption as the most promising candidate to explain the observed patterns in the data.

6 Conclusion

The unique setting examined in this paper allows us to shed new light on important open questions in the technology adoption literature. First, our findings speak directly to why the *aggregate* productivity effect of major technological breakthroughs, such as IT and electricity, may be hard to pin down in the data. As pointed out by David (1990), the full effects of a new technology may take significant time to materialize, as plants still need to learn how to organize production efficiently. In our context, adopting mechanized cotton spinning required producers to reorganize production from households to factories. Our results suggest that initially, many plants operated the new technology in combination with inefficient complementary organizational practices. This led to a widely dispersed productivity distribution and relatively low average productivity. Observers estimating the productivity effect of switching from handspinning to mechanized spinning would significantly underestimate the long-run aggregate productivity gain if they only looked at the initial data around 1800.

Second, our results also shed light on the slow adoption of major new technologies. When there is uncertainty about how to operate a new technology efficiently, and the organizational knowledge – once acquired – is observable to competitors, plants face a strategic incentive to delay adoption. The high exit rates observed in cotton spinning relative to other sectors, alongside the higher productivity observed for younger plants in 1806, suggest that plants that entered later were at an advantage. If plants understood the significant uncertainty they faced when setting up a spinning mill at early stages of adoption, they had an incentive to delay the switch to the new technology in order to take advantage of the learning externalities generated by other early adopters.

In summary, our unique setting allows us to speak to a dimension of productivity growth that is usually hidden. Productivity differences across plants reflect both the underlying technology and the complementary organizational practices with which the respective technology is used. Both features play important roles in the decision to adopt new technologies: What are the potential productivity gains of a new technology (i.e., its frontier), and is the organizational knowledge needed to achieve these gains (i.e., operate at the frontier) readily available? Separating these features empirically is difficult because of data limitations. Our results suggest that the need to reorganize production is an important dimension of technology adoption. Approaching the frontier of a new technology via organizational improvements can take a long time, and it can explain some of the salient features in the adoption of major innovations.

Finally, our paper provides a first look at how the unprecedented growth in manufacturing productivity during the First Industrial Revolution played out at the plant level. We show that in mechanized cotton spinning – the flagship industry of the period – a substantial proportion of productivity growth materialized along the extensive margin of plant exit and entry. Our results suggest that throughout this process, organizational innovations (alongside the traditionally emphasized technological ones) were an important driver of productivity growth. Future research – building on the increasing availability of historical data – should examine whether these findings constitute a common feature of the structural transformation from agriculture to modern, factory-based manufacturing. Our paper lays the groundwork by using comparative historical analysis to deepen our understanding of why the diffusion of innovation is often a complex and slow process.

References

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*. Cambridge University Press.

André, L. (1996). *Machines à Papier: Innovation et Transformations de l'Industrie Papetière en France : 1798-1860*. Editions de l'Ecole des Hautes Études en Sciences Sociales.

Atkin, D., A. Chaudhry, S. Chaudry, A. K. Khandelwal, and E. Verhoogen (2017). Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan. *Quarterly Journal of Economics 132*(3), 1101–1164.

Bianchi, N. and M. Giorcelli (2022). The Dynamics and Spillovers of Management Interventions: Evidence from the TWI Program. *Journal of Political Economy 130*, 1630–75.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *Quarterly Journal of Economics 128*(1), 1–51.

Bonin, S. and C. Langlois (1987). *Atlas de la Révolution Française: Economie*, Volume 10. Editions de l'École des Hautes Études en Sciences Sociales, Paris.

Braguinsky, S., A. Ohyama, T. Okazaki, and C. Syverson (2015). Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry. *American Economic Review 105*(7), 2086–2119.

Braguinsky, S., A. Ohyama, T. Okazaki, and C. Syverson (2021). Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization. *American Economic Review 111*(12), 3795–3826.

Bresnahan, T. F. and M. Trajtenberg (1995). General Purpose Technologies: 'Engines of Growth'? *Journal* *of Econometrics 65*, 83 – 108.

Brynjolfsson, E. (1993). The Productivity Paradox of Information Technology. *Communications of the ACM 36*(12), 66–77.

Brynjolfsson, E. and L. M. Hitt (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives 14*(4), 23–48.

Brynjolfsson, E., D. Rock, and C. Syverson (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics 13*(1), 333–72.

Cameron, R. (1985). A New View of European Industrialization. *Economic History Review 38*(1), 1–23.

Chanut, J., J. Heffer, J. Mairesse, and G. Postel-Vinay (2000). *L'Industrie Française au Milieu du 19e Siècle: Les Enquêtes de la Statistique Générale de la France*. Editions de l'École des Hautes Études en Sciences Sociales, Paris.

Chapman, S. (1970). Fixed Capital Formation in the British Cotton Industry, 1770-1815. *Economic History Review 23*(2), 235–253.

Chapman, S. (1974). The Textile Factory Before Arkwright: A Typology of Factory Development. *Business History Review 48*(4), 451–478.

Chassagne, S. (1976). L'Enquête, dite de Champagny, sur la Situation de l'Industrie Cotonnière Française au Début de l'Empire (1805-1806). *Revue d'Histoire Économique et Sociale 54*(3), 336–370.

Chassagne, S. (1991). *Le Coton et ses Patrons: France 1760-1840*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Clark, G. (1987). Why Isn't the Whole World Developed? Lessons from the Cotton mills. *The Journal of Economic History 47*(1), 141 – 173.

Clark, G. (1994). Factory Discipline. *The Journal of Economic History 54*(1), 128–163.

Comin, D. and B. Hobijn (2010). An Exploration of Technology Diffusion. *American Economic Review 100*(5), 2031–59.

Crafts, N. (1985). *British Economic Growth During the Industrial Revolution*. Oxford: Oxford University Press.

Crafts, N. and K. Harley (1992). Output Growth and the British Industrial Revolution: A Restatement of the Crafts-Harley View. *Economic History Review 45*(4), 703–730.

Crafts, N. and N. Wolf (2014). The Location of the UK Cotton Textiles Industry in 1838: A Quantitative Analysis. *Journal of Economic History 74*(4), 1103–1139.

Daudin, G. (2010). Domestic Trade and Market Size in Late-Eighteenth-Century France. *Journal of Economic History 70*(3), 716–743.

David, P. (1990). The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *American Economic Review 80*(2), 355–361.

De Loecker, J. and F. Warzynski (2012). Markups and Firm-Level Export Status. *American Economic Review 102*(6), 2437–2471.

Dudley, L. (2010). General Purpose Technologies and the Industrial Revolution. Papers on Economics and Evolution No. 1011.

Feigenbaum, J. J. and D. P. Gross (2021). Organizational Frictions and Increasing Returns to Automation: Lessons from AT&T in the Twentieth Century. *National Bureau of Economic Research w29580*.

Fitton, R. and A. Wadsworth (1958). *The Strutts and the Arkwrights, 1758-1830: A Study of the Early*

Factory System. Manchester University Press.

Foster, L., C. Grim, J. Haltiwanger, and Z. Wolf (2018). Innovation, Productivity Dispersion, and Productivity Growth. NBER Working Paper 24420.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review 98*(1), 394–425.

Freeman, C. and F. Louçã (2001). *As Time Goes By: From the Industrial Revolutions to the Information Revolution*. Oxford University Press.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press.

Garcia-Marin, A. and N. Voigtländer (2019). Exporting and Plant-Level Efficiency Gains: It's in the Measure. *Journal of Political Economy 127*(4), 1777–1825.

Geraghty, T. (2007). The Factory System in the British Industrial Revolution: A Complementarity Thesis. *European Economic Review 51*(6), 1329–1350.

Gille, B. (1968). *La Sidérurgie Française au XIXe Siècle: Recherches Histioriques*. Librairie Droz.

Gillispie, C. C. (2004). *Science and polity in France: the revolutionary and Napoleonic years*. Princeton University Press.

Giorcelli, M. (2019). The Long-Term Effects of Management and Technology Transfers. *American Economic Review 109*(1), 1–33.

Gort, M. and S. Klepper (1982). Time Paths in the Diffusion of Product Innovations. *Economic Journal* (367), 630–653.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica 25*(4), pp. 501–522.

Hall, B. (2004). Innovation and Diffusion. NBER Working Paper 10212.

Hall, B. and B. Khan (2003). Adoption of New Technology. NBER Working Paper 9730.

Harley, C. (1998). Cotton Textile Prices and the Industrial Revolution. *Economic History Review 51*(1), 49–83.

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150.

Horn, J. (2006). *The Path Not Taken: French Industrialization in the Age of Revolution, 1750-1830*. Cambridge MA: MIT Press.

Huberman, M. (1996). *Escape from the Market*. Cambridge University Press.

Jenkins, D. T. (Ed.) (2003). *The Cambridge History of Western Textiles*. Cambridge University Press.

Juhász, R. (2018). Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade. *American Economic Review 108*(11), 3339–76.

Langenbach, R. (2013). Better than Steel? Part 2: Tall Wooden Factories and the Invention of "Slow-burning" Heavy Timber Construction. Article URL accessed June 20, 2020.

Lipsey, R. G., K. I. Carlaw, and C. T. Bekar (2005). *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth*. Oxford University Press.

Mansfield, E. (1961). Technical Change and the Rate of Imitation. *Econometrica 29*(4), 741–766.

Markus, T. A. (2013). *Buildings and Power: Freedom and Control in the Origin of Modern Building Types*. Routledge.

Ministère de l'Agriculture et du Commerce (1847). *Statistique de la France: Industrie*, Volume 1. Imprimerie Royale.

Mitchell, B. (2003). *International Historical Statistics. Europe, 1750-2000*. Palgrave Macmillan.

Mokyr, J. (2001). The Rise and Fall of the Factory System: Technology, Firms, and Households since the Industrial Revolution. *Carnegie-Rochester Conference Series on Public Policy 55*, 1–45.

Mokyr, J. (2010). *The Enlightened Economy: An Economic History of Britain 1700-1850*. Yale University Press.

Mokyr, J. (2011). The Economics of Being Jewish. *Critical Review 23*(1-2), 195–206.

Mokyr, J. (2016). *Culture of Growth: Origins of the Modern Economy*. Princeton University Press.

Mokyr, J. (2021). "The holy land of industrialism": Rethinking the Industrial Revolution. *Journal of the British Academy 9*, 223–47.

Nuvolari, A., G. Tortorici, and M. Vasta (2023). British-French Technology Transfer from the Revolution to Louis Philippe (1791–1844): Evidence from Patent Data. *The Journal of Economic History*, 1–41.

Perla, J. and C. Tonetti (2014). Equilibrium Imitation and Growth. *Journal of Political Economy 122*(1), 52–76.

Perrot, J. (1975). *L'ăge d'Or de la Statistique Régionale Française (An IV-1804)*. Paris: Société des Etudes Robespierristes.

Perrot, J. and S. Woolf (1984). *State and Statistics in France, 1789-1815*. Taylor & Francis.

Pollard, S. (1965). *The Genesis of Modern Management: A Study of the Industrial Revolution in Great Britain*. Harvard University Press.

Pounds, N. and W. Parker (1957). *Coal and Steel in Western Europe: The Influence of Resources and Techniques on Production*. Bloomington: Indiana University Press.

Ridolfi, L. and A. Nuvolari (2021). L'histoire immobile? A reappraisal of French economic growth using the demand-side approach, 1280–1850. *European Review of Economic History 25*(3), 405–428.

Rosenberg, N. (1976). Factors Affecting the Diffusion of Technology. *Explorations in Economic History 10*(1), 3–33.

Saxonhouse, G. and G. Wright (2004). Technological Evolution in Cotton Spinning. In D. Farnie and D. Jeremy (Eds.), *The Fibre that Changed the World: The Cotton Industry in International Perspective, 1600-1990s*, pp. 129–152. Oxford University Press.

Sokoloff, K. (1984). Was the Transition from the Artisanal Shop to the Nonmechanized Factory Associated with Gains in Efficiency? Evidence from the US Manufacturing Censuses of 1820 and 1850. *Explorations in Economic History 21*(4), 351–382.

Sokoloff, K. (1986). Productivity Growth in Manufacturing During Early Industrialization: Evidence from the American Northeast, 1820-1860. In *Long-Term Factors in American Economic Growth*, pp. 679–736. University of Chicago Press.

Syverson, C. (2004). Product Substitutability and Productivity Dispersion. *Review of Economics and Statistics* (2), 534–50.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature 49*(2), 326–365.

Szostak, R. (1989). The Organization of Work: The Emergence of the Factory Revisited. *Journal of Economic Behavior and Organization 11*(3), 343–358.

Williamson, O. (1980). The Organization of Work: A Comparative Institutional Assessment. *Journal of*

Economic Behavior and Organization 1(1), 5–38.

Woronoff, D. (1984). *L'Industrie Sidérurgique en France Pendant la Révolution et l'Empire*. Paris: École des Hautes Études en Sciences Sociales.

FIGURES

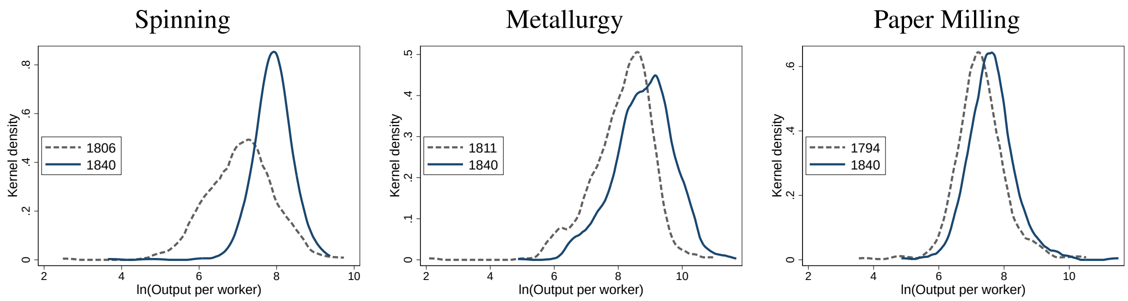


Figure 1: Changes in the Productivity Distributions in the Three Sectors

Notes: The figure shows the distribution of log(output per worker) for the three sectors at the beginning of our sample period (around 1800) and in 1840. Productivity growth in spinning was mainly due to a disappearing lower tail. In contrast, in metallurgy and paper milling, the whole distribution shifted to the right.

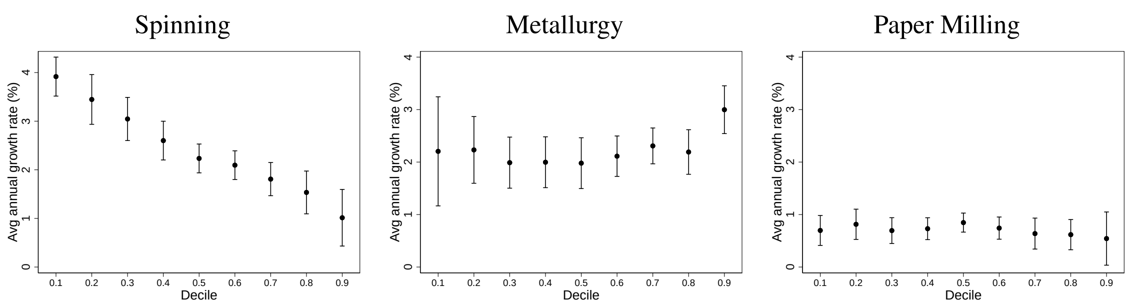


Figure 2: Productivity Growth at Different Quantiles of the Distribution

Notes: The figure visualizes the results of quantile regressions for growth in log(output per worker) for the three sectors, estimated at each decile. Productivity growth in spinning was concentrated in the lower tail of plant productivity. In contrast, in metallurgy and paper milling, productivity growth occurred relatively evenly across the distribution.

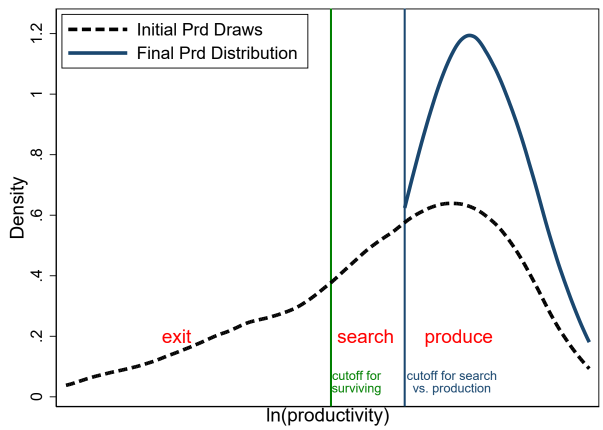


Figure 3: Productivity Dynamics in the Stylized Model

Notes: The figure illustrates the productivity dynamics in our stylized model: In the initial period, the least productive plants exit. Subsequently, in the innovation period, surviving plants decide whether to search for better organizational knowledge, foregoing production. The searching plants are randomly matched to continuing producers as in Perla and Tonetti (2014). In the final period, searching plants adopt the improved organizational knowledge from more productive producing plants. Thus, the lower tail of the productivity distribution disappears, and the mass shifts towards higher productivity draws. At the same time, the productivity frontier remains unchanged.

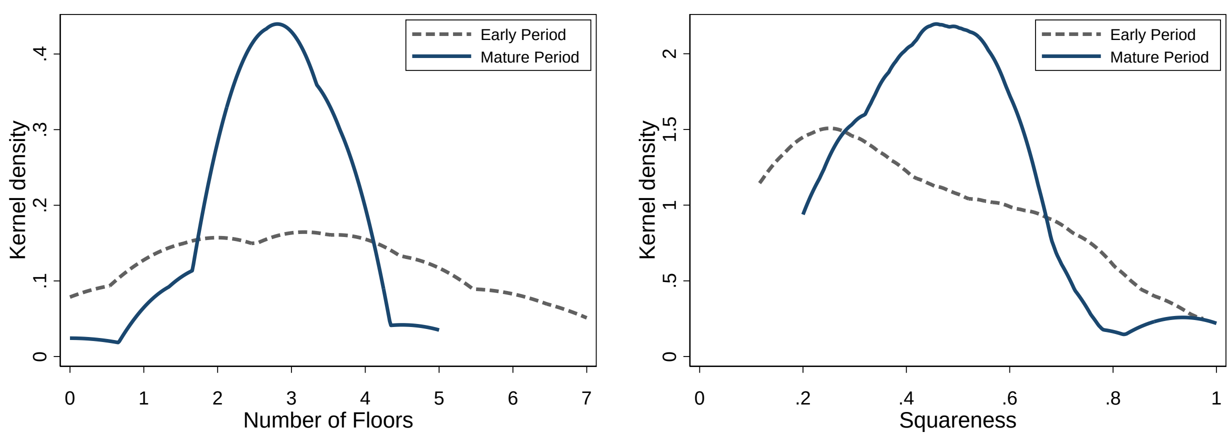


Figure 4: Experimentation and Convergence in Mill-Design for Cotton Spinning Plants

Notes: The figure shows the distributions of two important features of cotton mill design, in an 'early' experimental period (built before 1820) and a later 'mature' period (built after 1820). Panel A plots the distribution of the number of floors of cotton mills. Panel B plots the distribution of 'squareness,' as defined in Section 5.2. Data on the length, width, and number of floors of overall 59 cotton plants are from Chassagne (1991).

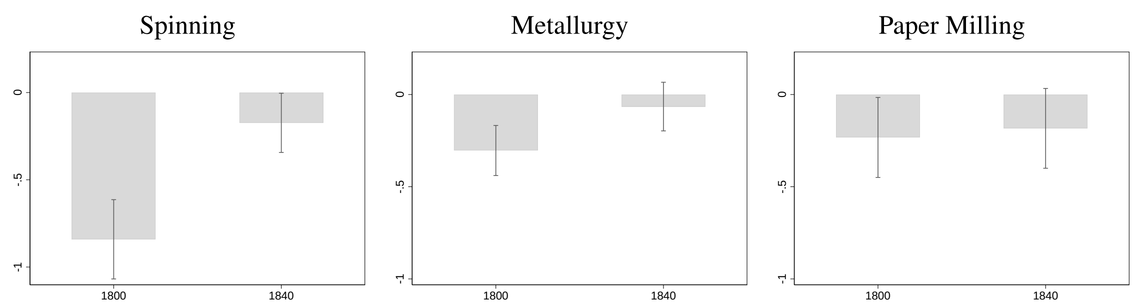


Figure 5: Proximity to High-Productivity Plants

Notes: The figure shows that proximity to high-productivity plants mattered the most in mechanized cotton spinning at the beginning of our sample period (around 1800), when the technology had just been introduced in France. The figures plot the standardized beta coefficients of $lnDist^{p90}$, which measures the log distance to the closest plant with productivity in the 90th percentile (in the same sector and in the same period – 1800 and in 1840, respectively). The dependent variable is log(output per worker). All regressions include *département* fixed effects (see Table A.10). Whiskers indicate 90% confidence intervals.

TABLES

Table 1: Annual Productivity Growth (in %) at Different Quantiles of the Distribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	2.420***	3.917***	3.293***	2.234***	1.651***	1.014***	868
	(0.154)	(0.204)	(0.229)	(0.151)	(0.167)	(0.297)	
Metallurgy (1811-1840)	2.328***	2.205***	2.068***	1.979***	2.285***	2.998***	1366
	(0.183)	(0.530)	(0.317)	(0.247)	(0.193)	(0.232)	
Paper milling (1794-1840)	0.719***	0.697***	0.717***	0.846***	0.691***	0.542**	867
	(0.111)	(0.145)	(0.139)	(0.092)	(0.130)	(0.258)	

Notes: The table reports the average annual productivity growth (in %) between the initial sample period (around 1800) and 1840 for the three sectors (column 1), and annual productivity growth estimated at different quantiles (columns 2-6). Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2: Survival Rates Across Sectors

	Spinning	Metallurgy	Paper milling
Period	1806-1840	1811-1840	1794-1840
Survival rate	5.0%	37.7%	10.8%
Number of plants	340	470	520
Restricted sample survival rate	12.6%	52.5%	24.6%
Number of plants	87	324	211

Notes: The "survival rate" is defined as the percentage of plants from the initial period that survived to the later period based on matching either by name or location (see Section 3.2). The "restricted sample survival rate" uses only the subset of plants located in communes that had only one plant in the initial period.

Table 3: Cotton Spinning in 1806: Productivity and Plants' Age Profile

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)	(5)	(6)
Young plant	0.575***	0.534***	0.608***	0.543***	0.575***	0.493***
	(0.088)	(0.085)	(0.086)	(0.083)	(0.089)	(0.085)
log(Spindles per worker)		0.336***				
		(0.070)				
log(Workers)			0.107***			
			(0.025)			
Spinning jenny				-0.626***		
				(0.087)		
Throstle					-0.003	
					(0.092)	
Mule jenny						0.481***
						(0.086)
R^2	0.11	0.17	0.14	0.20	0.11	0.18
N	340	340	340	340	340	340

Notes: The table shows that mechanized cotton spinning plants that had just entered the market by 1806 had significantly higher productivity. 'Young' is a dummy variable equal to one for cotton spinning plants with below-median age (with the median age in 1806 being three years). The number of spindles is a standard measure of a spinning machine's production capacity, irrespective of vintage. Spinning-jenny, throstle and mule-jenny are binary indicators equal to one for plants using the earliest (spinning-jenny), intermediate (throstle), and latest (mule-jenny) vintage of spinning machinery, respectively. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix

Technology Adoption and Productivity Growth: Evidence from Industrialization in France

Réka Juhász   Mara P. Squicciarini   Nico Voigtländer

A Historical Background: Additional Detail

This section provides additional information about industrialization in France in the first half of the 19th century, followed by further detail on the three sectors that we examine: mechanized cotton spinning as well as metallurgy and paper milling.

A.1 The Industrial Revolution in France

This appendix provides further detail on the Industrial Revolution in France, complementing Section 2.1 in the paper. An earlier literature "derogated the economic development of France as a story of retardation or relative backwardness" (O'Brien and Keyder, 1978, p. 194). This view has been largely revised in recent decades, leading to a new consensus that economic growth accelerated in France as early as in the mid-18th century (Rostow, 1975). The "retardation view" of French industrialization was defended by traditionalists who criticized the emerging cliometric approach and its quantitative methods and data. Numerous studies following the work of French historian Jan Marczewski gave credence to the cliometric approach (most prominently Maddison, 2001). These studies weakened – and eventually eliminated – the idea that French economic growth had stagnated in the 19th century. The consensus view that emerged holds that French growth had in fact been substantial (Crouzet, 2003). Recent estimates suggest that economic growth took off around 1800 (Ridolfi and Nuvolari, 2021). Illustrative of the similarities across the two countries, Horn (2006, p.10) writes that "[i]n an astonishing number of sectors, French entrepreneurs of the 1780s competed successfully with their English counterparts."

As we note in the main text, during the early stages of industrialization, France largely depended on the adoption of major British technological breakthroughs. Industrial espionage became widespread and, despite the attempts of the British government to block it, detailed reports and descriptions of English technology were sent across the Channel (Harris, 1998; Bradley, 2010). Additionally, "industrially minded" people in Britain and France entertained an intense correspondence on scientific and technological advances (Mokyr, 2005). Correspondingly, upper-tail human capital played an important role in industrialization (Squicciarini and Voigtländer, 2015). The fact that France *adopted* the major new technologies from Britain in this period renders the setting well-suited to examine technology adoption.

The adoption of new technologies from Britain was the primary source of innovation in the early 19th century. However, as French industrialization proceeded, "technological progress became indigenous, built in to the economy, so that ... France became at mid-[19th]century a centre of invention and diffusion for modern technologies" (Crouzet, 2003, p.234). We note, however, that the dynamism of innovation in early 19th century France is still a somewhat open question in the literature. While Horn (2006) argues the state facilitated growth with its innovation policies, Khan (2020) suggests that state activism was detrimental and stifled innovation. Consistent with the former view, Nuvolari, Tortorici, and Vasta (2023) finds that the French patent system was successful in absorbing British technologies.

A.2 Mechanized Cotton Spinning

In Section 2.2, we discussed the development of mechanized cotton spinning in Britain as well as its adoption in France. Figure A.1 provides an illustration of how cotton spinning was traditionally performed, mostly by women in their homes, using a simple spinning wheel. With this technology, each spinner was able to spin only one thread of yarn at a time. The invention of the spinning jenny by James Hargreaves in 1765 made it possible to spin multiple threads simultaneously, as twist was imparted to the fibre by using spindles rather than by the workers' hands. Throughout the 1760s – 70s, Richard Arkwright and Samuel Crompton developed two subsequent vintages: the water frame and the water-powered spinning mule, respectively (see Figure A.2, left panel). The mechanization of preparatory processes was also well-underway prior to the 19th century. These new technologies entailed a move from home-based to factory-based production (right panel in Figure A.2). This was partly due to the machines' reliance on inanimate power sources, and partly to an increased need to monitor workers more closely (Williamson, 1980; Szostak, 1989).[1]

These innovations had enormous productivity effects. The first vintage of the spinning jenny alone led to a threefold improvement in labor productivity (Allen, 2009). As a consequence, the price of yarn declined in the late 18th century, especially for the highest-quality yarn. This can be seen in Figure A.3, which shows price data for three different qualities of yarn: 18, 40 and 100 count yarn.[2] While all counts saw striking price declines, this trend was most pronounced for the finest, highest-quality varieties, where prices dropped from 1,091 pence per pound to 76 pence per pound in real terms between 1785 and 1800. Machine spinning had the largest impact on the fine high-quality yarn, which British hand-spinners had not been able to effectively produce and to which the mule-jenny (a subsequent vintage of the machine introduced in the late 18th century) was well-suited (Riello, 2013). Note that our data on French cotton spinning include information on the type of yarn produced, allowing us to account for quality differences across plants.

[1] The spinning jenny was typically hand-powered.

[2] Harley (1998) collected price data for three different qualities of yarn from British sources: 18, 40 and 100 count yarn. The count is an industry-wide standard that refers to the length per unit mass, implying that higher counts are finer. Finer count yarns are used to produce higher-quality cloth, while lower counts are used to produce heavier, cheaper cloth.

The Spinning Wheel | Home Spinning

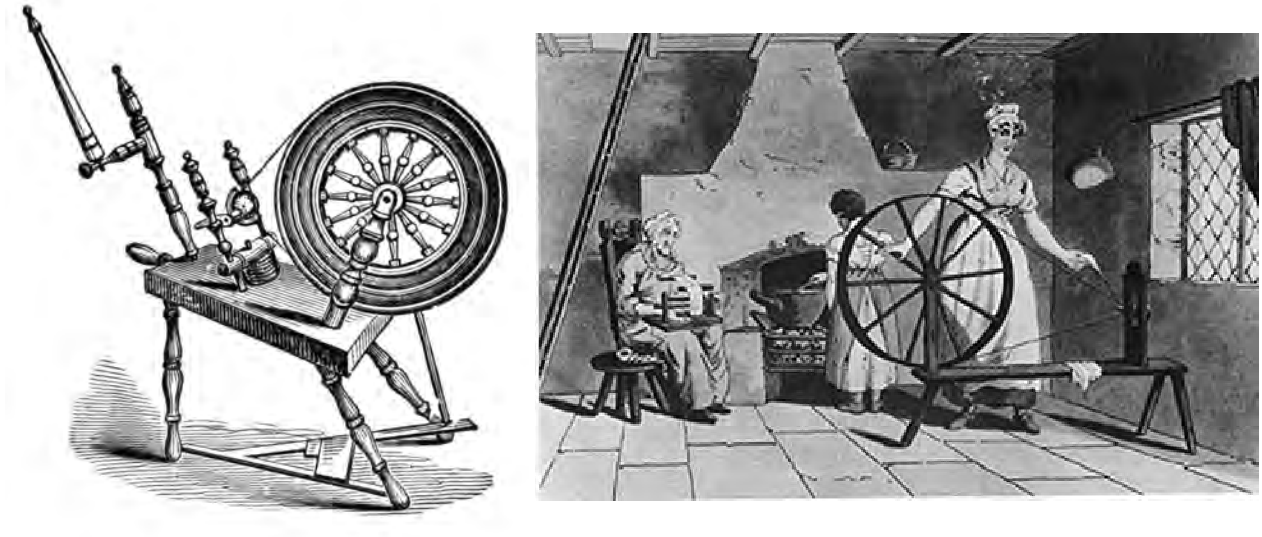


Figure A.1: Old Handspinning Technology

Source: https://etc.usf.edu/clipart/7700/7797/wheel_7797.htm (left panel) and https://digitalcollections.nypl.org/items/510d47dc-dcb3-a3d9-e040-e00a18064a99 (right panel).

Water-Powered Spinning Mule | Spinning Mule Operated in a Mill

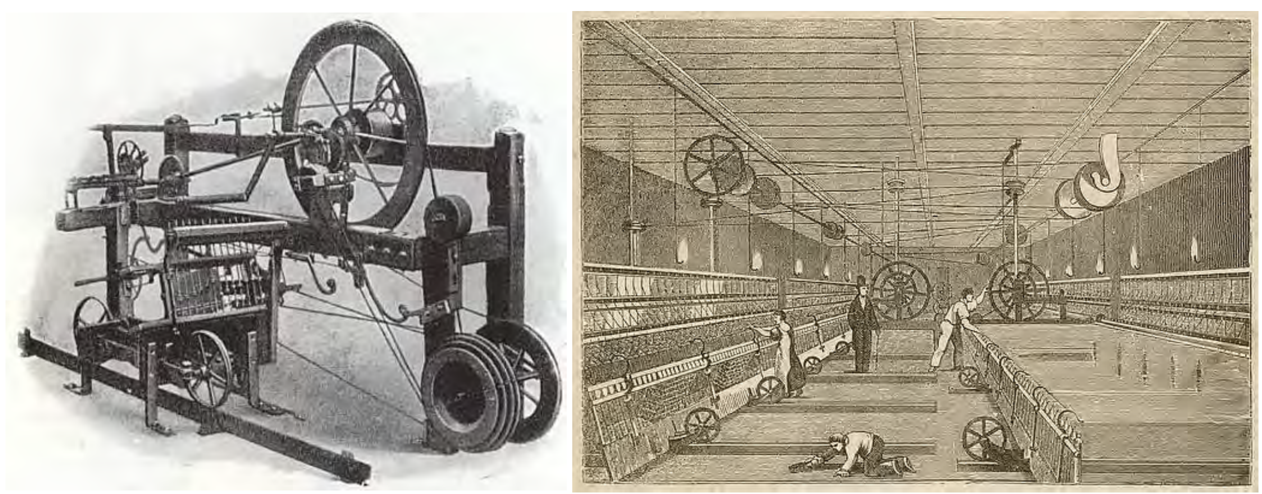


Figure A.2: New Mechanized Technology in Cotton Spinning

Source: https://powerloom.weebly.com/uploads/3/3/4/7/3347452/1722116.jpg?270 (left panel) and https://commons.wikimedia.org/wiki/File:Cotton_Mule_Spinning,_1835.jpg (right panel).

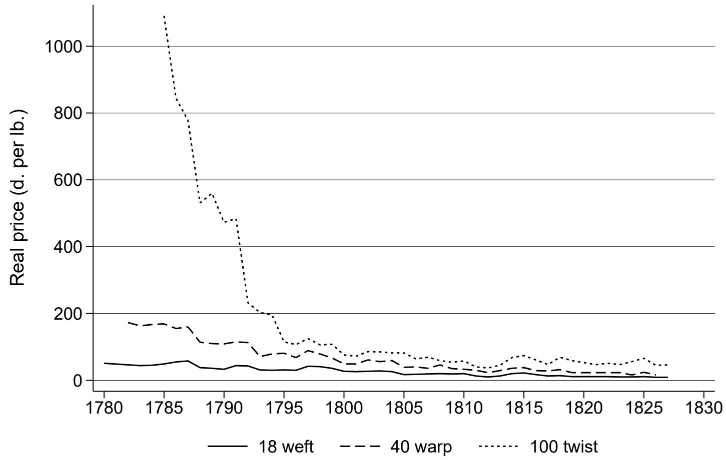


Figure A.3: Price of Different Counts of Yarn in Britain

Notes: Data are from Harley (1998), who collected prices for three different qualities of yarn: 18, 40 and 100 count yarn. The count is an industry-wide standard that refers to the length per unit mass (implying that higher counts are finer). Machine spinning had the largest impact on the fine high-quality yarn, which British hand-spinners had not been able to effectively produce.

Historical evidence about machinery producers. As we discussed in Section 2, machines were mainly produced domestically in France. Importantly, master mechanics and builders were typically not employees of the firm – they were paid by the factories to install, maintain and repair equipment (Cookson, 1997). Technologically complex tasks were 'outsourced' to engineers (Mokyr, 2010). This suggests that plants had access to broadly similar markets for the capital equipment within the same regions.

No large-scale switch to steam power. In contrast to Britain, mechanized cotton spinners in France did not switch from water to steam power to a large extent, owing to the fact that France was not particularly well-endowed with coal (Cameron, 1985).[3] Thus, we can think of the power source used as remaining mostly constant over the time period. Moreover, improvements to the technology used to operate water wheels should have a similar effect on productivity growth in paper milling, one of our comparison sectors, as this sector was also reliant on water power.

A.3 The Challenging Transition to Factory-Based Production in Mechanized Cotton Spinning: Additional Details

This appendix complements Section 2.2 in the paper, where we discussed some of the key challenges regarding the move to factory-based production in mechanized cotton spinning. Here, we provide additional evidence and examples.

[3]This is confirmed in our data for 1840, showing that the majority of cotton spinning plants were still using water power (see Table A.2).

Building design challenges. We illustrate the trial and error process of overcoming building design challenges using the example of constructing buildings better able to withstand fires. Cotton textile mills introduced the so-called "fire-proof building" in Britain in the late 18th century, which entailed leaving no timber surfaces exposed by using cast-iron columns instead of wood (Johnson and Skempton, 1955). However, it quickly became apparent that fireproof mills were not actually fireproof, because "steel or wrought iron, when heated, will fail by buckling or bending very much sooner than the equivalent beam of post or wood" (Boston Mutual Fire, 1908, p. 3). US textile mills developed what became known as "slow-burning mills" in the 1820s, recognizing that fires could not be prevented, but their effects could be curtailed by better mill design. Partly, this entailed moving back to using wood: "Timber posts offer more resistance to fire than either wrought-iron, steel, or cast iron pillars, and in mill construction are preferable in many respects (Boston Mutual Fire, 1908, p. 3). Chassagne (1991, p. 340) posits that early 19th century French mills consisted of multiple buildings and covered vast spaces (as opposed to building vertically), partly in order to minimize the fire hazard.

Labor management challenges. Cotton spinning plants needed to develop organizational and management practices for running spinning mills at a scale not seen elsewhere in the economy. Here, we provide additional evidence regarding labor management challenges.[4] There were three salient aspects of this for cotton spinning mills as described in the main text; i) how to get workers to adapt to the rhythm of factory work, ii) how to coordinate labor in a factory setting and, iii) how to solve monitoring problems.

First, from the workers' side, the move to factory-based production fundamentally altered both the location and the nature of work (Clark, 1994). Under the factory system, the employer "dictated when workers worked, their conduct on the job and that they steadily attend to their assigned tasks." (Clark, 1994, p. 128). Following instructions, showing up to work on time, or getting along with other employees was a challenge for the first generation of factory workers, who had been used to the high degree of independence in the domestic spinning system (Pollard, 1965). As Pollard (1965, p. 181) writes, "What was needed was regularity and steady intensity in place of irregular spurts of work; accuracy and standardization in place of individual design; and care of equipment and material in place of pride in one's tools." Simply put, an industrial labor force needed to be created where none had existed before (Mokyr, 2010).

Second, as we noted in the main text, coordination of labor was crucial, as flow production meant that one worker could hold up the entire production line. This is illustrated by the Karl Marx quote in Section 2.1.

Third, monitoring worker effort, much of which was hard to observe (Huberman, 1996), was

[4] Pollard (1965) and Mokyr (2010) provide more general discussions of other management challenges facing firms at the time.

another novel aspect of factory work. Huberman (1996, p. 11) describes the need for monitoring in mechanized cotton spinning: "If there were multiple breakages of yarn on the larger machines, the mule had to come to a complete stop to piece the broken threads. There was also doffing, when the reels were full of spun cotton, the mule had to be stopped and the reels removed. Finally, there was cleaning. At all times, the spinners could expend effort as they were motivated to, and without proper supervision or incentives they could disguise how hard they could in fact work." This created a strong need for monitoring, so that even early hand-powered machines (a particular vintage of spinning machinery) were housed in the "garrets of cottages and later in sheds" (Huberman, 1996, p. 11) in order to enable a direct supervision of workers.

As we discuss in the main text, overcoming these challenges proceeded via a slow process of trial and error. The industry eventually settled on efficiency wages in the 1830s in Britain (Huberman, 1996).

A.4 Comparison Sectors: Metallurgy and Paper Milling

This appendix complements Section 2.3; it discusses our two comparison sectors during the First Industrial Revolution in France in more detail. We give an overview of the production processes in each, and discuss new technologies developed during our sample period and how they were adopted into the existing organization of production.

Metallurgy

Iron was a flagship product of the Industrial Revolution. It was used for railways, steamships, and for machines. The fundamental process of producing iron has remained the same over centuries. Despite large productivity improvements achieved during the Industrial Revolution (Allen, 2009), the actual changes to methods of production were modest.

Historical production process. Iron is extracted from iron ore in a process called 'smelting' – freeing the iron by combining carbon with the oxygen of the ore under heat. The difficulty comes from the fact that the iron also needs to be separated from other metallic substances in iron ore. This is achieved by controlling the heat of the furnace so that most of the foreign matter separates out with the lowest-possible expenditure of fuel. In the Medieval period, the production process of iron used 'direct' technology. Smelting with this technology produced malleable iron directly from iron ore in a bloomery (a type of furnace) where the temperature was low enough for the iron not to melt. The product of this technology is wrought iron. This process is referred to as 'direct' because iron was produced in a near-finished condition in a single process. One vintage of this technology, the Catalan forge, survived into our study period and beyond in certain parts of France (Pounds and Parker, 1957).

Starting in the late Middle Ages, the direct technology began to be gradually replaced by an 'indirect' technology, which consists of two steps: smelting and refining. Smelting is similar to the direct technology. A blast furnace – with temperatures high enough for the complete fusion of the metal – is used to produce an intermediate product – pig iron. However, as pig iron is too brittle to be used in many applications, it needs to be refined one more time on a hearth. This second stage is known as refining. The blast furnace first appeared in Europe in the 15th century, but it was not widely adopted until the 17th or 18th century (Pounds and Parker, 1957). Figure A.4 illustrates the blast furnace and the organization of an 18th century metallurgy plant (foundry).

18C Charcoal Iron Blast Furnace

Organization of 18C Metallurgy Plant

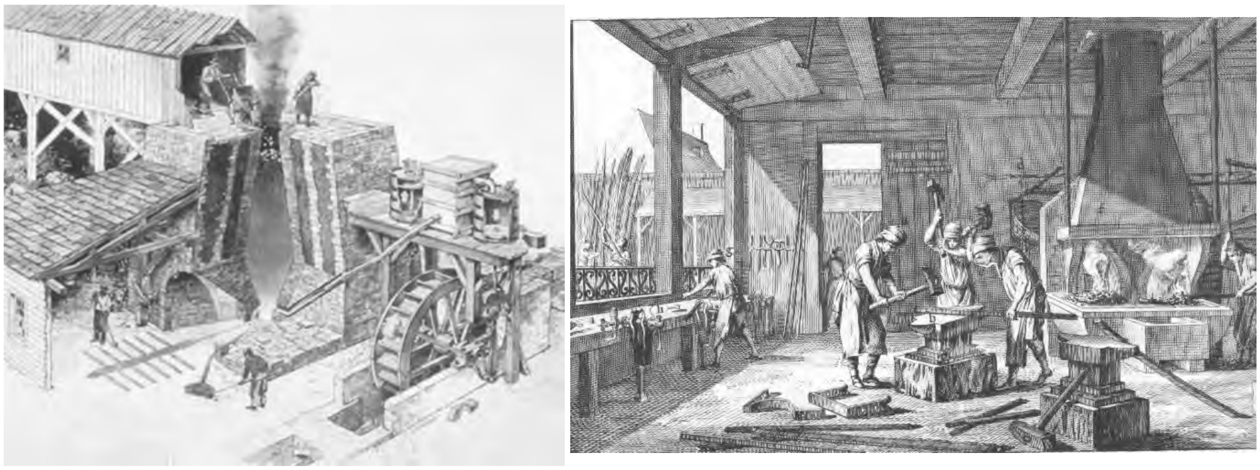


Figure A.4: 'Old' 18th Century Charcoal-Based Technology in Metallurgy

Source: https://www.nps.gov/articles/hopewell-furnace-a-pennsylvania-iron-making-plantation-teaching-with-historic-places.htm (left panel) and https://artflsrv04.uchicago.edu/images/encyclopedie/V26/plate_26_4_1.jpeg (right panel).

New technologies developed in Britain during the Industrial Revolution. Prior to the Industrial Revolution, both stages of the indirect process relied on charcoal as the source of fuel. The key innovation during the Industrial Revolution was the switch from charcoal to coal, through a series of gradual improvements in the period 1700-1850. The change in the type of fuel required modifications to the blast furnace, but the new technology "merely replaced earlier, recognizably similar, though less 'efficient' methods" (Pollard, 1965, p. 101). In particular, a coal-based blast furnace required larger furnace sizes and a switch from water to steam power. Such modifications could be made to existing blast furnaces (Pounds and Parker, 1957). Allen (2009) estimates that the cost of producing pig iron using coal decreased by 75% during this period in Britain. Figure A.5 presents illustrations of the 'new' coal-based technology. A comparison with Figure A.4 shows that the organization of metallurgy plants remained practically unchanged.

In refining, the move from charcoal to coal was achieved by the puddling process. Pig iron was melted in a reverbatory furnace fueled by coal while stirring, or 'puddling' the molten mass until the free carbon in pig iron was oxidized and the mass reduced to malleable form as bar iron.

19th Century Coal Blast Furnace

Organization of 19C Metallurgy Plant

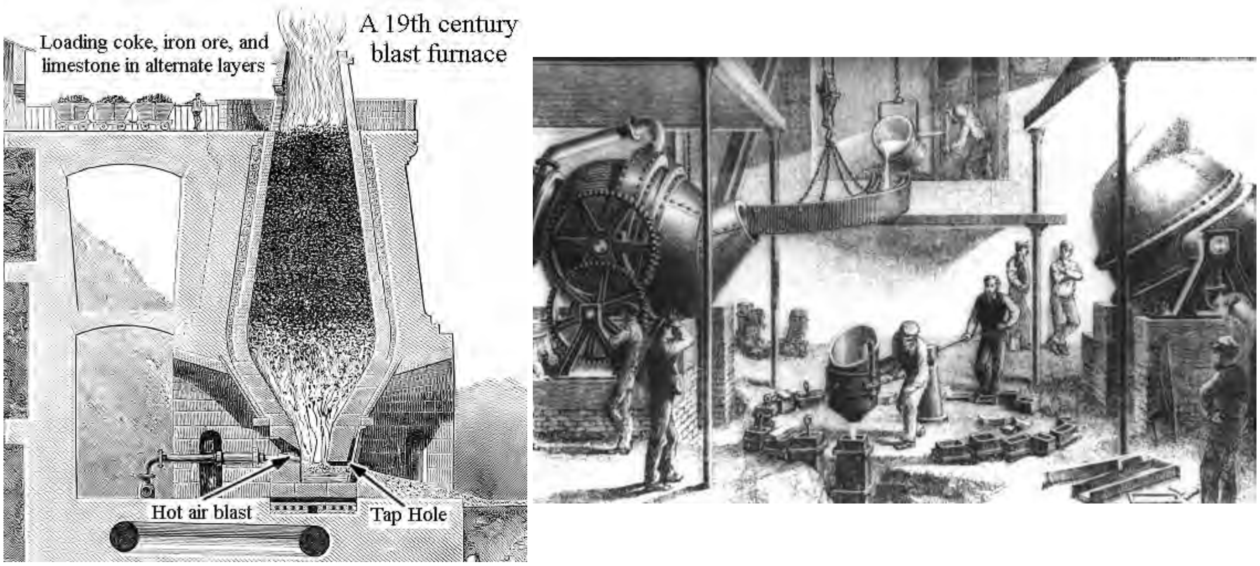


Figure A.5: ‘New’ 19th Century Coal-Based Technology in Metallurgy

Source: http://www.historywebsite.co.uk/articles/DarlastonIE/heavyindustry.htm (left panel) and https://www.sciencephoto.com/media/1288706/view (right panel).

Similar to smelting, the development of puddling proceeded gradually, and the technologies were not meaningfully different to others already in use. Pounds and Parker (1957, p. 35) characterize the development of puddling in the following way: “Cort’s [the inventor] invention was less the introduction of entirely new processes than a synthesis of practices that were already familiar. The reverberatory furnace, burning any fuel that might be available, had long been known. [...] Rolling mills were also known [...]. The lawsuits with which Cort was faced served to emphasize that some of his contemporaries regarded his claim to be an inventor as remarkably thin.”

Of course, adopting the new technology also entailed difficulties. Switching to coal as a source of fuel required changing or adapting machines, training workers, and modifying buildings (Gille, 1968). This is an important aspect of our setting, as in this regard the metallurgy sector is comparable to that of cotton spinning, where we also see dynamic innovation. We turn to this point after reviewing the process of adopting the new technologies in France.

Technology adoption in France. The switch from charcoal to coal as a source of fuel took place gradually throughout our sample period in France. However, a modernized metallurgy sector, characterized by large establishments producing for national markets, did not emerge until well after our sample period (Gille, 1968). By the end of our sample period, smelting had seen relatively little change; pig iron was still produced predominantly with charcoal using the old technology. In refining, technology adoption was more rapid, and the use of coal dominated charcoal three-to-one by 1847 (Gille, 1968).

The literature has put forward a number of explanations for the slow adoption of new technologies in smelting. One important factor was that, in contrast to Britain, iron ore and coal were not located in close proximity, making access to the necessary inputs expensive, particularly before the national rail network was established (Gille, 1968, p. 91). However, it was much easier to install puddling furnaces, which required less coal, and could use traditional water wheels to power the rolling mills (Gille, 1968).

In light of these constraints in France, a dual system began to emerge during our sample period. New technologies were adopted in existing forges. At the same time, new firms were set up near coal deposits. Our empirical findings confirm this. We see both an expansion of the industry near coal deposits (see Figure A.14 below) alongside relatively high survival rates of existing plants (Table 2 in the paper).

Re-organization of production in metallurgy. The organizational changes necessary to adopt new technologies in metallurgy were more modest relative to those described in the move to factory-based production in cotton spinning. This can be seen by comparing the before-vs. after illustrations for metallurgy (Figures A.4 and A.5) with those for spinning (Figures A.1 and A.2). One reason for the small changes in metallurgy is that plant-based production in this sector was already well-established historically in France. Data from the *Encyclopédie* (see Section A.5) confirms that the industry had well-established best practices regarding building layouts and the organization of production more generally.

Given that many of the new technologies were adopted within existing plants, major changes to structures or organizational practices were not required. To install a large English-style puddling furnace (which was the main margin of technology adoption during our sample period), the plant could continue to rely on water power, and only limited investments were necessary (Gille, 1968, p. 47). Where technology adoption relied on setting up new plants (as was often the case in refining), the primary impetus to do so was to locate close to coal deposits.

Paper Milling

In contrast to metallurgy and cotton spinning, paper milling was not a flagship industry of the Industrial Revolution, and its output was not particularly important for other sectors. However, paper making was one of the few manufacturing activities that was organized as plant-based production from well before the Industrial Revolution because of its reliance on water power. Moreover, it too underwent mechanization during our study period. For these reasons, paper making serves as a useful second comparison sector to cotton spinning.

Historical production process. In Europe, paper making had traditionally taken place in mills. Production consisted of several stages. First, the vegetable matter (the raw material) was broken down into cellulose fiber, which involved a water-powered stamping machine (see Figure A.6, left panel).

Next, it was formed into thin, wet sheets by a skilled worker, called a vatman (Figure A.6, right panel). It was then dried and – depending on its intended use – finished in different ways. Each of these steps was performed in a different section or room of the mill with a marked division of labor by function and gender. The only step of the production process that required water power was the washing, breaking, and beating (stamping) down of rags into fiber. The machine that performed all of these tasks (known as the washing, breaking or stamping engine) consisted of an oval, wooden tub containing a water-powered revolving roll and was operated by a skilled workman or an engineer. This stage of the production process and the work of the engineer determined to a large extent the quality of the paper that could be produced. Moreover, it was because of this technology that production was located in mills from very early on (McGaw, 1987).

New technologies developed during the Industrial Revolution. The important innovation that took place during our study period was mechanization of the forming of the paper, eventually fully replacing the tasks performed by the vatman with the Fourdrinier machine (Figure A.7). This technology is still at the core of modern-day paper production. The Fourdrinier machine was important not only because of the productivity improvements that it yielded, but also because it enabled the production of continuous rolls – something that had not been possible with the hand-based technology. The first vintage of the machine was patented in France in 1799 by Nicholas Louis Robert. In the 1800s, the idea behind the original machine was developed further by a British mechanic, Bryan Donkin, who developed a commercially viable machine with financing from the Fourdrinier brothers (André, 1996; McGaw, 1987).

Technology adoption in paper milling in France. The Foudrinier machine was gradually adopted during our sample period in France. André (1996, p. 253) claims that all large paper mills were mechanized by 1840, but the full assimilation of the new technologies in the industry was not completed until the 1850s and 1860s, after our sample period (André, 1996, p. 389).

Re-organization of production in paper milling. The organizational changes necessary to adopt the Foudrinier machine were minimal. One important point to note is that, different to mechanized cotton spinning, paper milling in this period never developed a standardized building layout. André (1996, p. 182) describes this explicitly: "In this [paper milling] industry, there are no stereotypical buildings that are easily identifiable like those of the large multi-story textile mills..." As we discuss in the main text, modifications of existing plants were often undertaken without having to substantially change other parts of the production process, and different parts of the paper milling process could be hosted in different buildings. For example, when plants adopted the Fourdrinier machine, they typically merely reconstructed the two sections hosting the cylinders and the machine, while re-using the buildings previously devoted to the other operations (André, 1996, p. 178).

Water-Powered Stamping

Handling by Vatman, Coucher, and Layer

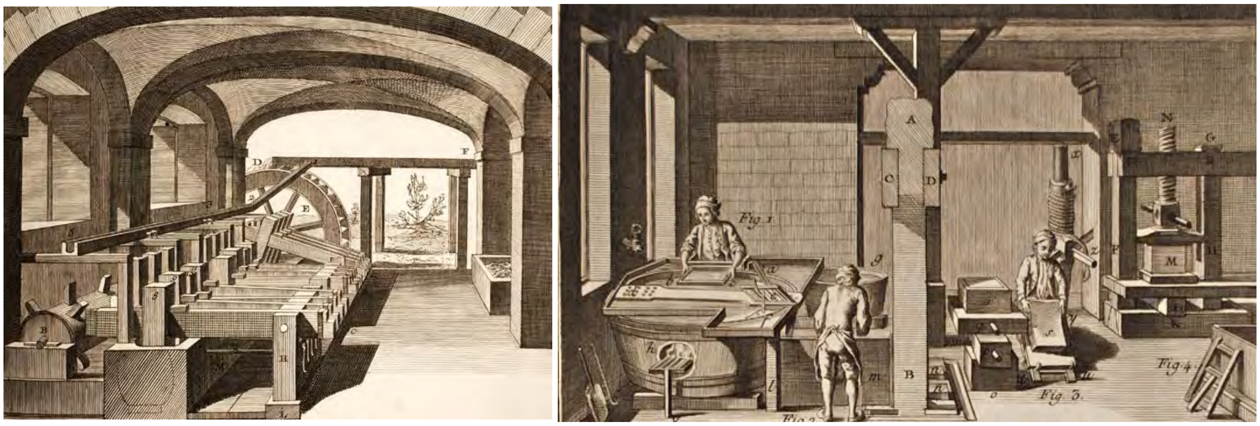


Figure A.6: Old Technology in a Paper Milling Plant

Source: http://paper.lib.uiowa.edu/european.php (both panels).

Sketch of Fourdrinier Machine

Fourdrinier Machine in a Plant

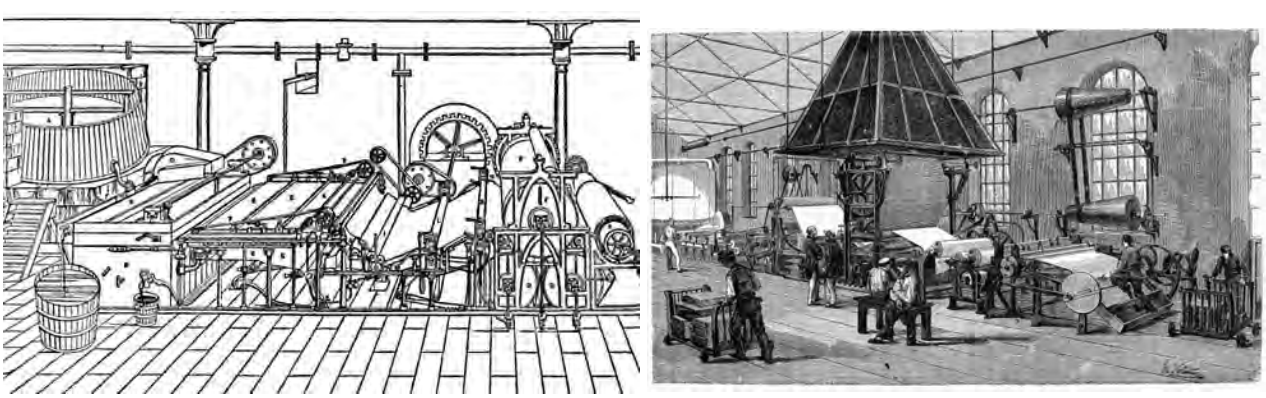


Figure A.7: New Technology in Paper Milling Plants: The Foudrinier Machine

Source: https://www.researchgate.net/figure/The-traditional-Fourdrinier-paper-making-machine-of-the-type-built-by-Bryan-Donkin_fig2_322872610 (left panel) and
https://www.granger.com/results.asp?image=0079612&screenwidth=1024 (right panel).

A.5 18th Century Encyclopédie Plates for the Three Sectors

This appendix complements the discussion of *differences* between our comparison sectors and mechanized cotton spinning in Section 2.3. Here, we examine illustrations from the 18th century on plant organization and production technology in the three sectors. These provide evidence that while standardized knowledge on plant organization and production technology existed for our two comparison sectors, it did not exist for mechanized cotton spinning. In particular, we use data on plates contained in the late 18th-century *Encyclopédie* of Diderot and d'Alembert from *the Encyclopedia of Diderot and d'Alembert: collaborative translation project*.[5] These plates were used to illustrate crafts, processes, and inventions from the time. They represent a unique source of information to study the amount and type of existing knowledge on manufacturing at the time. In total, there are 2,575 plates, accompanying 326 entries. Approximately half of them describe manufacturing technologies (Squicciarini and Voigtländer, 2015).

We identify all plates that illustrate plant organization or production technology for our three sectors. Overall, there are 28 such plates. Figure A.8 below provides two examples for plates on plant organization in metallurgy and paper milling. For cotton spinning, we further distinguish between home production and mechanized production.[6] Figure A.9 shows that for paper milling and metallurgy (and to a lesser extent for home spinning), there was a significant number of *Encyclopédie* plates specifically illustrating plant organization and production technology. In contrast, this type of codified knowledge was completely absent for mechanized cotton spinning, where we observe zero plates on technology or organization. This is in line with the historical evidence discussed in Section 2.2, suggesting that best-practice methods for mechanized cotton spinning did not yet exist in the late half of the 18th century. This absence of plates on mechanized cotton spinning is not surprising, as the technology had just been invented. Nevertheless, the *Encyclopédie* plates illustrate that codified knowledge was indeed available for our two comparison sectors in the late 18th century.

[5] This is available at http://quod.lib.umich.edu/d/did/index.html.

[6] We do not count other plates related to our three sectors that do not illustrate production technologies or plant organization. These include, for example, plates that describe products (e.g., metal products).

Metallurgy | Paper Milling

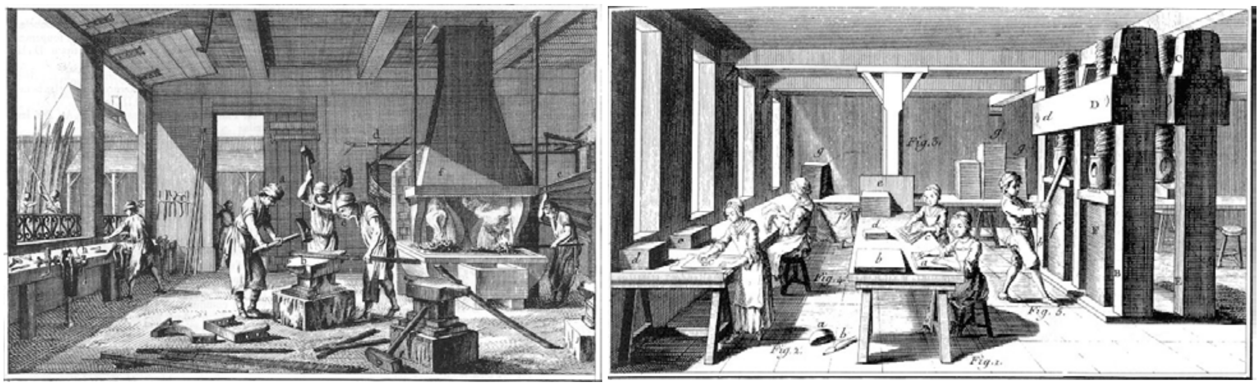
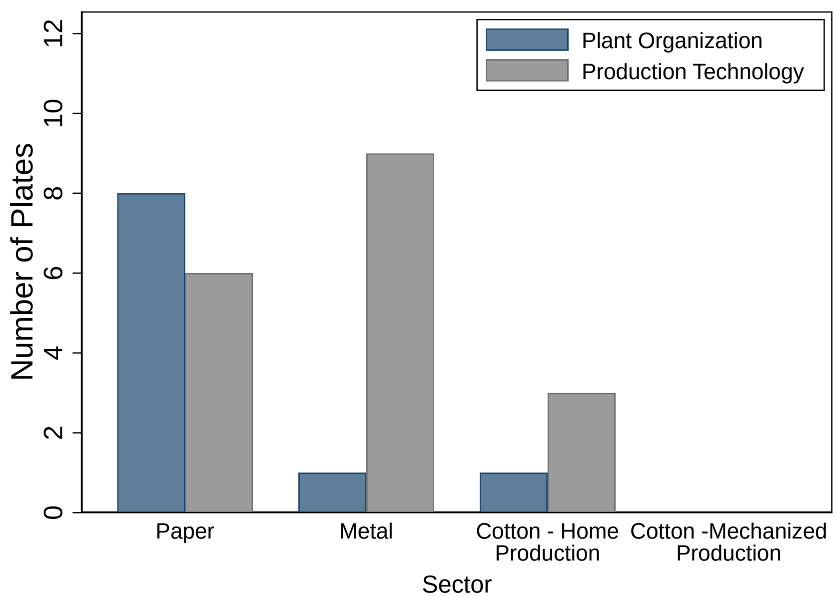


Figure A.8: *Encyclopédie* Plates on Plant Organization

Source: http://quod.lib.umich.edu/d/did/index.html.

Figure A.9: Number of Encyclopédie Plates about Plant Organization and Production Technology in the Three Sectors



Notes: Source: *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers* (1765). Available at https://quod.lib.umich.edu/d/did/ "Plant organization" refers to plates on plant layout/organization. "Production Technology" refers to plates on machinery and techniques for production. Plates relating to products are excluded.

A.6 Evidence on Innovation Across the Three Sectors Using British Patent Data

This appendix complements the discussion of *similarities* between our comparison sectors and mechanized cotton spinning in Section 2.3. Here, we provide evidence that all three sectors that we study were innovative during our sample period. In Figure A.10, we show that patenting activity was consistently high throughout the 1800-1840 period, using data on the number of British patents by category (as classified in the original source). Spinning was the third-most patent intensive industry among 146 categories, while metallurgy and paper milling were ninth and twenty-first respectively. These data were kindly shared by Walker Hanlon (2020), who digitized the data from Bennet Woodcroft's (1854) *Subject Matter Index of Patents of Invention*.

Figure A.10: Number of British Patents, 1800-1840

Notes: 'Spinning' refers to the patents related to all textile fibers, not just cotton. Data were kindly shared by Walker Hanlon based on work on patenting in Hanlon (2020).

Moreover, Table A.1 shows that in all three sectors, innovations were broad-based in the sense that they covered different parts of the production process. This is shown by the large number of patents in both 'core' (main part of the production process) and 'other' innovations in preparatory or finishing stages. In combination with the historical evidence on the nature of innovations, we conclude that the three sectors were undergoing technological change of a similar type *after* the adoption of mechanized cotton spinning. None of the sectors experienced major innovation that required reorganization of production.

Table A.1: Patents, 1800-1847

Sector	'Core'	Other	Total
Spinning	176	313	489
Metallurgy	100	143	243
Paper	89	33	122

Notes: The table reports the number of patents divided between different stages of the production process. 'Core' patents refer to innovations in the main part of the production process, while 'Other' refers to innovations in preparatory or finishing stages. Data were kindly shared by Walker Hanlon based on work on patenting in Hanlon (2020). Note that spinning patents include those for all textile fibers, not just cotton.

B Main Data: Surveys and Census of the Three Industries

This appendix complements Section 3 in the paper, where we introduced each of the four industrial surveys that we use. Here, we describe the data cleaning and construction steps and assess data quality. For each survey, we define the variables used in the paper and describe how they were constructed. We also provide further details on linking plants over time and define all control variables used in the analysis.

B.1 Mechanized Cotton Spinning, 1806

Primary data sources.

- J.-B. de Nompère de Champagny's survey of the cotton textile industry (1805/06).[7] *Archives Nationales, Series F12/1562-1564*.

- Price schedule by count of cotton yarn (price per kilogram in francs, 1806-07). *Archives Nationales, Series F12/533*.

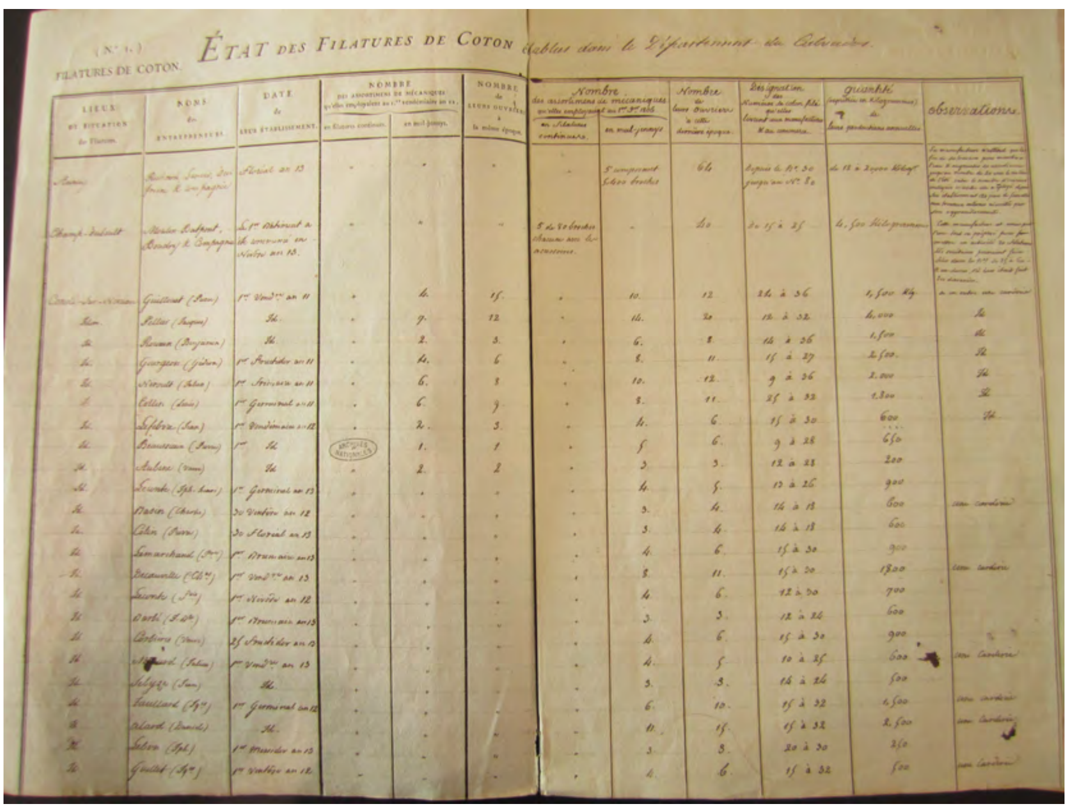

(N° 1.)
FILATURES DE COTON.
ÉTAT DES FILATURES DE COTON établies dans le Département du Calvados

Figure A.11: Sample Page from the Cotton Spinning Survey, 1806

[7]Champagny was the minister in charge of conducting the survey.

Source used in the paper. The data covering mechanized cotton spinning establishments were digitized and cleaned by Juhász (2018). We use the plant-level version of these data (which was not made public, but all cleaning procedures are as described in Juhász, 2018).

How was the survey administered? A standardized questionnaire was sent to each *département* (see Figure A.11). Returns were filled in by the *préfets*, the highest public official at the *département* level. For an extensive discussion and evaluation of the survey see Chassagne (1976).

Response rate. 107 of the 109 *départements* that were surveyed submitted a response.[8]

Missing data pattern. Of the 389 plants in the dataset, we are missing data on output for 37 (9.5%) plants, and on employment for 12 (3.1%) plants. In total, we are missing data on labor productivity for 49 (12.6%) plants in the dataset.[9] This leaves 340 cotton spinning plants with observed labor productivity in our sample. We refer to the latter as the baseline sample in our dataset.

Assessment of data quality. Beyond the high response rate, there are two further factors that suggest this survey was of a high quality. First, given the survey was administered five years after a previous inquiry into economic activity, *La Statistique des Préfets*, officials needed to update their existing knowledge, as opposed to starting from scratch (Chassagne, 1976). Second, these data have been used in qualitative work (e.g. Chassagne, 1991) as well as in recent empirical work Juhász (2018). In the latter case, mechanized cotton spinning capacity at the *département* level was compared to data on cotton textile manufacturing activity from around 1790 (using the *Tableaux du Maximum*) and from an industrial survey conducted in 1812 (which was typically submitted at the level of *départements* as opposed to individual plants). While there are some differences in the location of mechanized cotton spinning capacity, overall the data do line up fairly closely, suggesting that the 1806 survey of mechanized cotton spinning plants is of a high quality.

Plant locations. To geocode plant locations, we use information from the survey on the 'commune' in which the plant was located. We assign geocodes using a combination of automated packages and manual assignment.[10] Using the geocodes, we then assign each plant the present-day commune, *dèpartement* and region to which the commune belongs.

Variables directly reported in the 1806 cotton spinning survey.

- Number of employees
- Vintage of physical capital used and the number of machines[11]
- Location of the plant (up to the commune level)

[8]This number includes *départements* annexed to the French Empire during Napoleon's reign and is thus different from the modern-day number of *départements*, which we use for our analysis.

[9]In many instances where we are missing data on employment and/or output, it seems likely that the plant had shut down, or had temporarily ceased production at the time of the survey.

[10]For 19 plants out of the 340 for which we have labor productivity data, we could not identify a location because either there was no commune reported, or the historical name of the commune could not be located.

[11]The survey asked for throstles (water-frame) and mule-jennies, but many plants also reported spinning jennies. Juhász (2018) thus constructed a third, separate category to capture these.

- Name of the owner
- Output (reported in kilograms)
- Quality of yarn spun (measured by the count of the yarn)
- Date of foundation
- Number of spindles (including those imputed by Juhász, 2018)[12]

Constructed variables for cotton spinning plants in 1806.

- **Plant labor productivity**: This is defined as log(revenue per worker), where revenues are deflated as described in Appendix B.6. The survey reports the quantity of yarn spun in kilograms, as well as the minimum and maximum count of yarn spun. We use the (unweighted) average of the minimum and maximum count of the yarn produced by the plant as a proxy for its average output quality. We construct plant-level revenue by multiplying the quantity of plant-level output by the price of the average quality of yarn produced by the plant. We use a schedule of prices for different counts of yarn reported by the French government (see primary data sources above, at the beginning of this appendix section). We deflate revenues using the wholesale price index reported for 1806 in Mitchell (2003). See Appendix B.6 for detail.

  We also construct an alternative measure of labor productivity that does not adjust for the quality of yarn spun by the plant. In this case, we construct plant-level revenues by multiplying physical output at the plant level with the price of the (unweighted) average quality of yarn reported across *all* plants in 1806.
- **Capital:** We define capital as the log of the number of spindles at the plant level. Spindles are the standard measure of capital in the industry. Though these data were not explicitly asked for in the survey, many plants reported them, and Juhász (2018, Online Appendix pp. 22-24) imputed the remainder.
- **log(Employment)** This is defined as the log of the total number of workers. The 1806 cotton spinning survey does not provide a breakdown of male, female, and child labor. We interpret this measure as the sum of all forms of labor.
- **Plant total factor productivity:** TFP at the plant level is computed as the residual from regressing plant-level deflated log(total revenue) on log(total workers) and log(capital), as defined above.
- **Plant age:** This is defined as the number of years since the plant was established. We use log(plant age) in our regressions.
- **Vintage machine (Spinning jenny, Throstle, Mule jenny:** These are defined as binary indicators that take on value one when the plant has at least one machine of that vintage. The categories are not mutually exclusive.

[12] As discussed in Juhász (2018, Online Appendix p. 22–24), only a subset of plants reported the spindles used. The remainder were imputed using other plant-level information.

- **Log(Spindles per worker):** This measure of capital intensity calculates the log number of spindles per worker in the plant.
- **Young plant:** This binary indicator takes the value of one if the plant is younger than the median mechanized cotton spinning plant in 1806 (3 years).
- **Exit dummy:** This is a variable equal to one for plants that existed in 1806 and that had exited the market by 1840. This is defined using our baseline measure of survival (see Section 3.2). Any plant that we cannot match by owner name or by single-plant communes is classified as an exiting plant.

Identifying cotton spinning from other parts of the production process. In France, cotton spinning and weaving were generally not vertically integrated during this time period. Weaving, particularly in the early 19th century, was rurally organized. This implies less of an incentive to locate the workers in a common location, i.e, in a plant. Nevertheless, our dataset contains a few examples of vertically integrated spinning and weaving plants. We deal with these integrated plants in the following way. In the 1806 survey, enumerators were instructed to separately collect data for spinning and weaving activities (which is indicative of the lack of integration across these sectors in general). In the few cases where both took place under the same roof, we observe labor and output reported *by activity* and can thus estimate productivity separately for the spinning activities.

B.2 Metallurgy, 1811

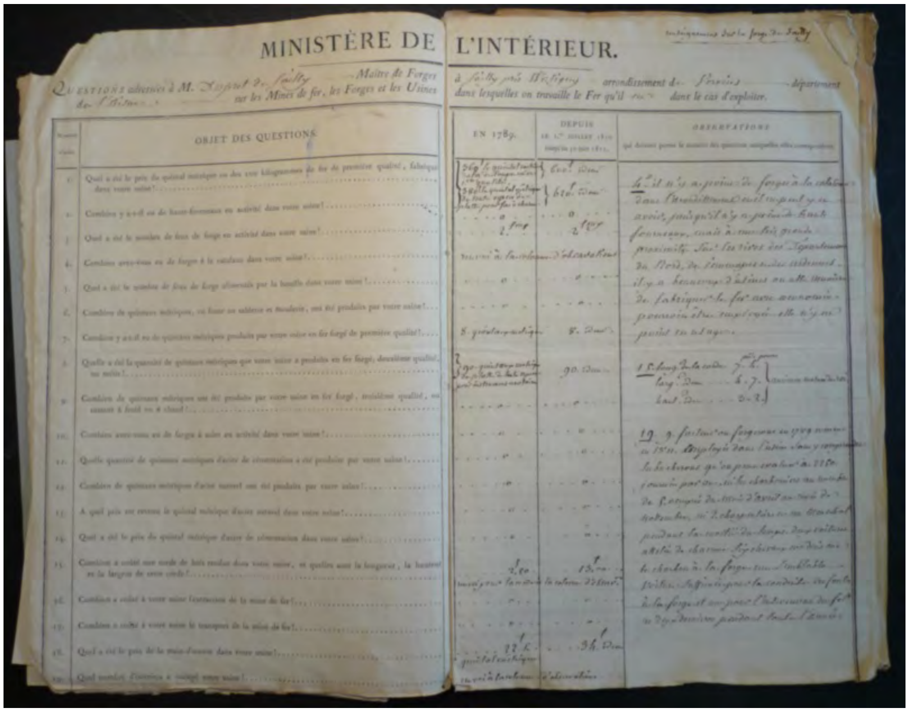
Primary data source. "Enquête sur les usines á fer de L'Empire (1811)" – Survey of iron manufacturers (1811). *Archives Nationales, Series F12/1603-1610.*

Source used in the paper. Primary data collected, digitized, and cleaned for this project.

How was the survey administered? A standardized questionnaire (see Figure A.12) consisting of 39 questions (8 pages in length) was sent to the *préfet* of each *département*. The officials were tasked with passing on the survey to owners (the *maîtres de forge* – forge masters), who needed to provide the necessary information. The *préfets* were also tasked with verifying that the data provided by the plants were accurate (Woronoff, 1984, p. 64).

Response rate. The response rate across *départements* was complete (Perrot and Woolf, 1984). In the case of metallurgy, we are able to assess not only compliance across *départements*, but also survey responses within them. As we describe in more detail below, this is because metallurgy plants had been surveyed multiple times since the 1770s, which meant that there was detailed knowledge about plant owners.

Woronoff (1984, p. 69) estimates that over two-thirds of the owners of active plants complied with the request for information. He suggests that compliance rates were inversely proportional to the number of establishments in metallurgy in the *département*. In the 20 or so *départements* with 1-5 plants, response rates were practically complete, whereas in regions with higher levels of


MINISTÈRE DE L'INTÉRIEUR.

OBJET DES QUESTIONS

EN 1789.

OBSERVATIONS

Figure A.12: Sample Page from the Metallurgy Survey, 1811

activity in the sector, the response rate was lower. This suggests that time constraints on officials were the main reason for non-compliance by plants.

Missing data pattern. Of the 576 plants surveyed, 44 reported no labor, 20 reported no output, and 34 plants reported neither of these variables. In many of these cases, the notes make it clear that the plant had shut down production. In total, labor productivity cannot be estimated for 98 (17%) of plants in the survey.[13]

Assessment of data quality. The survey covering metallurgy is a prime example of the 'statistical boom' that characterized the later empire between 1811-14 (Perrot and Woolf, 1984, p. 140). Beyond the high response rates at the *département* and plant level, other factors also contributed to this survey being of a particularly high quality. First, the level of existing knowledge about this industry was already high, given the information with which *préfets* could cross-check the returns. While the survey was addressed to the forge masters, it was cross-checked by *préfets* as well as by engineers of the mining agency. Second, Perrot and Woolf (1984, p. 161) note that the detailed questionnaire was designed so that it would be difficult for forge masters to manipulate their figures.

[13] It should be noted that as administrators had existing detailed information on forge masters from previous data collection efforts, it is likely that many of the plants with 'missing' data based on this definition had in fact been shut down for years. In this sense, we are likely overestimating the extent of missing data.

Variables directly reported in the 1811 metallurgy survey.

- Location of the plant (up to the commune level)
- Name of the owner (here: 'forge master')
- Quantity of output produced in metric quintals, by type (iron of first quality, iron of second quality, iron of third quality, steel using the cementation process, natural steel, and pig iron)
- Price of output produced (can be missing)
- Labor employed (does not always distinguish internal and external workers)
- Capital (number of blast furnaces, forges, catalan forges)
- Indicator variable if the firm was already in the market in 1788.

Data processing steps. In processing and cleaning the raw data, we performed the following steps:

1. **Data cleaning for numerical variables.** We cleaned strings (e.g., production reported as an interval) and converted observations where the unit of measurement was different (e.g., observation reported as "poids de mar" instead of reporting in quintals).
2. **Clean capital variables.** The number of machines reported under different categories (enumerated above) should add up to the total number of machines that the plant used; but in a small number of cases, it is likely that there is double-counting across categories. We manually correct these where possible.[14] Measurement error in the number of machines should not affect our results too much, as we only use capital in metallurgy for imputing internal labor (see next point).
3. **Harmonize labor variables.** Plants reported labor in various forms. Some reported labor by occupation. Others gave total by 'internal' and sometimes also 'external' labor[15]; or by male, female and child labor. We create a set of mutually exclusive categories based on whether the worker was an internal worker, an external worker, or of an unknown status. We classify occupations into internal and external categories based on a historical technical manual that describes the type of tasks performed by a particular occupation.[16]
4. **Impute internal labor.** As we mentioned in the text (Section 3.2), about 40% of the metallurgy plants reported either 'internal' labor only, or both 'internal' and 'external' labor, separately. The remainder of plants reported only total labor, with no indication of whether this includes external labor. To construct a consistent measure of 'internal' labor for all

[14]For example, plants were asked to report the number of forges and the number of Catalan forges (both are types of capital). Some plants report capital under each, which is unlikely, as these are two very different vintages of technology.

[15]Woronoff (1984, p. 138) describes external labor as only having very loose ties to the plant, performing tasks such as driving or collecting charcoal for the plant. Thus, external workers were unlikely to be considered formal salaried employees of the plant in the 1840 census.

[16]Le Blanc, V., Auguste, C., Walter de Saint-Ange, J. (1835). *Métallurgie pratique du fer, ou, Description méthodique des procédés de la fabrication de la fonte et du fer: accompagné de documents relatifs à l'établissement des usines, à la conduite et aux résultats des opérations.* Librairie Scientifique et Industrielle de L. Mathias, France.

plants, we estimate the size of the internal labor force for the 60% of plants that reported only total labor. We use a nearest neighbor matching algorithm to determine whether plants that only report total labor are more likely to be reporting internal labor only or the sum of internal and external labor. We match each plant that reports only total labor to its nearest neighbor that reports internal and external labor, where matching is based on capital, output, and the stage of production (see below, under *'Constructed variables for metallurgy plants in 1811'*, for the definition of this variable) the plant is involved in. We then classify a plant as "reporting only internal labor" if its reported total labor is closer to the matched plant's internal labor force. Likewise, we define a plant as "reporting total labor" when it is closer to the internal plus external labor force of the matched plant. When our algorithm suggests that the plant is reporting internal and external labor together, we estimate the number of internal workers by using the mean proportion of internal labor from all plants that report both types (the average internal labor share is 20%).

5. **Clean owner names.** In a small number of cases, the notes from the survey make it clear that the name entered under 'forge master' was the manager, not the owner. In these cases, we code the owner name as 'missing.'
6. **Drop plants outside of mainland present-day France** We drop 8 plants located on the island of Corsica. We do so because our spatial diffusion analysis does not apply to areas that were isolated from other plants. Thus, our baseline sample is restricted to plants within mainland France in its current borders. With this restriction, we are left with 470 metallurgy plants that have labor productivity data (478 out of 576 plants have the information needed to compute labor productivity).
7. **Plant locations.** To assign plant location geocodes, we use information provided in the survey on the 'commune' in which the plant was located, using a combination of automated packages and manual assignment. Using the geocodes, we then assign each plant the present day *dèpartement* and region to which the commune belongs. Using this procedure, only 44 (9.4%) of the 470 plants for which we have labor productivity data could not be geocoded.

Constructed variables for metallurgy plants in 1811.

- **Plant labor productivity**: This is defined as log(revenue per worker), where revenues are deflated as described in Appendix B.6. For all plants, the survey reports the quantity of output produced by product type: iron of first quality, iron of second quality, iron of third quality, steel using the cementation process, natural steel, and pig iron. In addition, some plants report also the prices by product type. To construct plant-level revenue, we multiply the quantity of each product type that the plant produced by the average price of this product, and then sum across all of the plant's products. We compute these average prices for each product type using the information from those plants that reported prices for the corresponding product.[17] We deflate revenues using the wholesale price index reported for 1811 in Mitchell (2003), as described in Appendix B.6.

We also construct an alternative measure of labor productivity that uses the plant-specific output prices for those plants that report this information, and the average product-specific price (as described above) for those plants that do not report output prices.

- **log(Employment):** This is defined as the log of the total number of internal workers.
- **Stage of production:** We assign plants to mutually exclusive stages of production based on the capital they use and the type of output they produce. Upstream plants produce only pig iron with a blast furnance. Downstream plants produce wrought iron or steel with a forge and have no catalan forge (old technology used for indirect production). Integrated plants do both upstream and downstream stages. Indirect producers use a catalan forge to produce wrought iron. These binary variables are only used for the matching algorithm used to impute internal labor.
- **Young plant:** Indicator variable that takes the value of one if the plant was not active in 1788 (and we consider it an 'entrant' in 1811).
- **Exit dummy:** This is a variable equal to one for plants that existed in 1811 and that had exited the market by 1840. This is defined using our baseline measure of survival. Any plant that we cannot match by owner-name or by single-plant communes will be classified as an exiting plant.

B.3 Paper Milling, 1794

Primary data sources.

- *"Enquêtes sur les papeteries en France, an II."* – Survey of the paper milling industry in Frnace, 1794 (year 2 according the the Revolutionary calendar). *Archives Nationales, Series F12/1482– 1485*.
- Price data for paper milling products from the *Tableaux du Maximum*. Images from the *Tableaux* for paper milling were kindly shared by Guillaume Daudin. Original source: *Archives Nationales, Series F12/1516–1544*. See Daudin (2010) for further details.

Source used in the paper. Primary data collected, digitized, and cleaned for this project.

How was the survey administered? The survey was administered by the Jacobin government using a standardized template that was given to local authorities. Given the level of detail asked for by the survey (name, birthplace, age, tenure and occupation of workers), it is likely that plants themselves had to provide this information. In some cases, the information is certified by the mayor or other local public officials.

[17] Overall, 308 out of 470 plants reported at least one price for their products. We show in Appendix E.2 that our results hold when we use the product-specific prices for those plants that reported them, while dropping the remaining metallurgy plants in 1811.

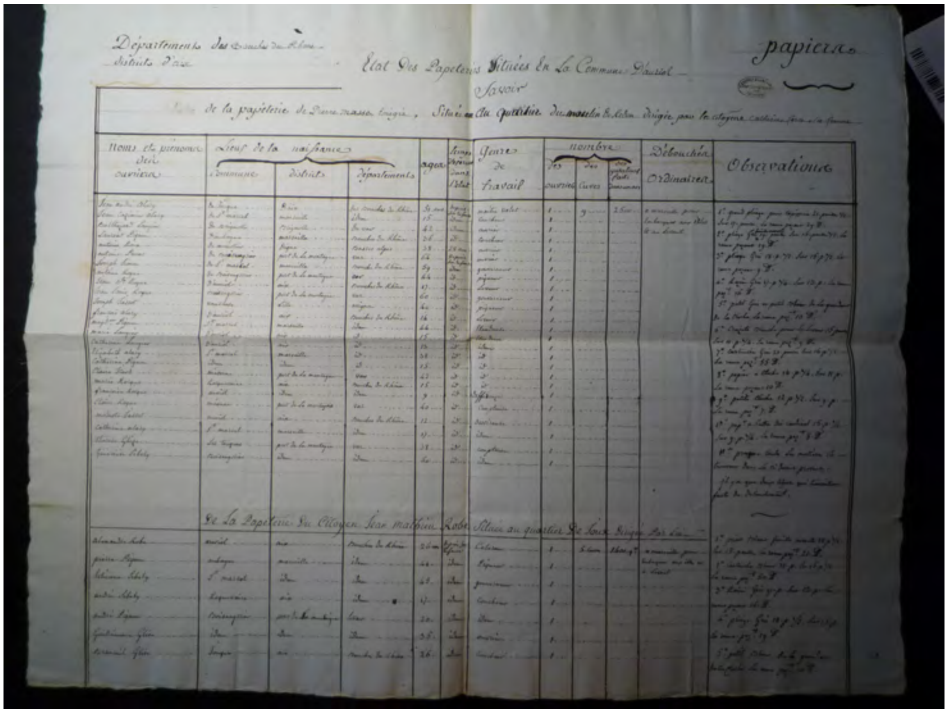


Figure A.13: Sample Page from the Paper Milling Survey, 1794

Response rate. All returns except for those from Corsica were submitted.

Missing data pattern. Of the 593 plants surveyed, 28 reported no labor, 30 reported no output, and 15 plants reported neither of these variables. In many of these cases, the notes from the survey enumerators make it clear that the plant had shut down production. In total, labor productivity cannot be estimated for 73 (12%) of plants in the survey.

Assessment of data quality. Bonin and Langlois (1987) use the data from the 1794 paper milling survey and argue that the geographic dispersion of production lines up fairly well with another survey of paper milling conducted later (in 1811, under the same Napoleonic regime that conducted the cotton spinning and metallurgy surveys).[18] This suggests that despite our survey being conducted earlier than the other two from the time period, the data quality is similar.

Variables directly reported in the 1794 paper milling survey.

- Name of the owner
- Location of the plant (up to the commune)
- Name (first and last) of workers
- Age of the worker

[18]While we collected the data from the 1811 paper milling survey, it proved to be of little use for our purposes in this paper, as 40% of plants surveyed gave no information at all on labor employed, making it infeasible to calculate labor productivity.

- Occupation of each worker
- Total number of workers employed by the plant
- Output (in quintals)[19]

Data processing steps. In processing and cleaning the raw data, we performed the following steps:

1. **Data cleaning for numerical variables:** We cleaned strings (e.g., production reported as an interval) and converted observations where the unit of measurement was reported incorrectly (e.g., observation reported as 'livres' instead of quintals). For a handful of observations, we could not convert the unit of measurement for output, when there was no obvious conversion rate (e.g. 'reams' of paper). These were dropped. For 27 observations, the output reported was so high, it suggested the unit of measurement was incorrect. For these plants, we suspect they reported output in 'livres' which is one-hundredth of a quintal. Using the original high values would lead to unrealistic labor productivity estimates. Rather than dropping these observations, we decided to keep them in the data and convert them to quintals from livres, assuming that they were erroneously reported in the latter. We note in passing that this data cleaning step is not crucial for our results: Omitting this step does not affect any of our results.
2. **Clean individual worker data:** We group workers into four categories: men, women, children and apprentices. We drop children younger than 7 years old (and only keep those between 7-10 that report an occupation). Any worker aged younger than 15 years is classified as a child worker. Apprentices are classified based on their reported occupation. Male and female workers were assigned a gender based on their first name.
3. **Create employment data:** All paper milling plants report employment in two forms. The returns list individual workers as well as total employment. We add the individual workers and compare the sum to the reported total. In 90% of the cases, the two match up perfectly, or differ by only one unit.[20] However, only male labor seems to be consistently reported across plants. Only 37% of plants report any female workers; only 38% of plants report any child labor; and only 22% of plants report employing apprentices. Given that proto-factories were characterized by family units working together, it is highly likely that the raw totals are undercounting employees in many plants. For these reasons, we use male labor as our baseline measure of employees. To construct a measure of total employment by plants that

[19]The metric system was adopted by the Revolutionary French government in 1795, *after* the paper milling survey was conducted. In processing this data, we assume that output is reported in 'old regime' quintals (which is 100 pounds (*livres*)). This is the most likely unit, given that the metric system had not yet been adopted when the paper milling survey was administered in 1794. In addition, the later cotton spinning survey in 1806 explicitly specified that output be enumerated in kilograms. Likewise, the 1811 survey in metallurgy requested that units of output should be in *metric* quintals (i.e., 100 kilograms).

[20]We create a consistent variable when there is a discrepancy between the two employment numbers by taking the larger reported number of the two.

we can compare to 1840, we impute total employment for 1794 by scaling male labor by the proportion of total labor to male labor in 1840 (2.29). As discussed in the main text, the validity of this method hinges on the assumption that the ratio of total employment to male employment remained constant over our sample period. We find that the proportions are consistent. The proportion of total employees to male employees is 2.11 in 1794 for the 18 plants that report all types of labor, while in 1840, it is 2.29 (averaged across all plants).

4. **Plant locations:** We use information provided in the survey on the 'commune' in which the plant was located to assign geocodes, using a combination of automated packages and manual assignment. Using the geocodes, we then assigned each plant the present-day *dèpartement* and region to which the commune belongs. Using this procedure, only 13 (2.5%) of the 520 plants for which we have labor productivity data could not be geocoded.

Constructed variables for paper milling plants in 1794.

- **Plant labor productivity**: This is defined as log(deflated revenue per worker). For all plants, the survey reports the quantity of output produced. We price this output using the mean price of paper products as reported in the *Tableaux du Maximum*. Employment is the total imputed employment as described above. We deflate revenues using the wholesale price index reported for 1811 in Mitchell (2003). See Appendix B.6 for detail. We also construct an alternative measure of labor productivity that uses only male employment in 1794.
- **Exit dummy:** This is a variable equal to one for plants that existed in 1794 and that had exited the market by 1840. This is defined using our baseline measure of survival. Any plant that we cannot match by owner-name or by single-plant communes is classified as an exiting plant.

B.4 The Manufacturing Census of 1839-47

Primary data source. The data are from the four-volume *Statistique de la France: Industrie* published in 1847 by the Ministry of Agriculture and Commerce. The volumes were scanned by the French National Library (BNF) and are available to view on their online catalogue (https://gallica.bnf.fr/ark:/12148/bpt6k857958?rk=64378;0).

Source used in the paper. The data were digitized by Chanut, Heffer, Mairesse, and Postel-Vinay (2000). We use this version of the data.

How was the survey administered? This was the first full industrial census conducted by the royal statistical agency, *"Statistique générale du royaume."* Its execution was similar to previous industrial surveys covering individual sectors in that the circular to collect the requested information was sent to regional officials (*préfects* and their subordinates). Plants themselves submitted the information. Local officials and technical experts were tasked with verifying the information provided by plants. However, different to the industrial surveys from the 1800s, the data were also checked, cleaned, and harmonized by the central statistical agency (Ministère de l'Agriculture et du Commerce, 1847). These data were then released by the Ministry of Agriculture and Commerce, and this publication formed the basis of the modern digitization efforts of Chanut et al. (2000).

While the timing of the census seems broad (1839-47), the data were actually collected in a relatively narrow window of time around 1845. Local administrators set to work on collecting the data in 1839, but the effort was halted after 18 months. The reason was that a concurrent survey was started by the Minister of Finance, and officials became concerned that the returns would be unreliable. Work began again in 1845, at which time *préfects* updated their records and submitted the returns (Ministère de l'Agriculture et du Commerce, 1847, pp. xxiv-xxv). The returns were centrally cleaned and organized. The results were released in 1847.

Response rate. All returns except those for Paris and Corsica were submitted.

Missing data pattern. Every observation included in the dataset for our three sectors of interest has a strictly positive value for labor and output. This is plausibly due to the fact that the released data were already cleaned.

Assessment of data quality. The census was conducted with a high degree of care. Local officials were instructed to declare to surveyed plants that the investigation had no fiscal purpose. The plant was tasked with writing a descriptive bulletin that was verified by local authorities and people with relevant technical knowledge. Figures were checked at the *département* level and centrally, by authorities with relevant technical knowledge. Where possible, figures were cross-checked against other sources (e.g., against tax records) (Ministère de l'Agriculture et du Commerce, 1847).

An important limitation is that the 1840 manufacturing census undercounted plants with fewer than 10 employees (Ministère de l'Agriculture et du Commerce, 1847, p. xviii.). We deal with this issue by conducting robustness checks where we omit plants with fewer than 10 employees from all sector-year pairs (see Appendix E.10).

Variables directly reported in the 1840 census for all three industries.

- Location of the plant (up to the commune level)
- Name of the owner
- Value of production (in francs)
- Total number of employees
- Employees by male, female and child labor
- Number of water-powered, steam-powered, animal-driven and wind-powered engines (separately).
- Spindles (including those imputed by Juhász, 2018)[21].

Data processing steps. In processing and cleaning the raw data, we performed the following steps:

[21] As discussed in Juhász (2018, Online Appendix p. 25), only a subset of plants reported the spindles used. The remainder were imputed using other plant level information.

1. **Data filtering for the three industries:** The three industries can be identified with a high level of precision. For cotton spinning we use all plants that report their main activity as cotton spinning (*CODB5000 = 5283*). For metallurgy, we classified all plants that reported metallurgy as their main activity (*CODBRAG = 3*). For paper milling, we use all plants that report paper and cardboard as their main activity (*CODB2000 = 2550*).
2. **Data cleaning:** As the raw census data were processed by the central statistical agency of France in 1847, we cannot rule out that observations with missing labor or output data were dropped in a previous cleaning step. We dropped observations where multiple establishments jointly reported their data. This issue affects 6.8% of establishments in metallurgy, 2.0% of establishments in cotton spinning, and 18% of establishments in paper milling.
3. **Plant locations:** We use information provided in the survey on the 'commune' in which the plant was located to assign geocodes, using a combination of automated packages and manual assignment. Using the geocodes, we then assigned each plant the present-day *dèpartement* and region to which the commune belongs. Using this procedure, only 9 (1.7%) of the 528 plants for which we have labor productivity data in cotton spinning could not be geocoded. In metallurgy, this number is 33 (3.7%) out of 896 plants; in paper milling it is 4 (1.1%) of 347 plants.

Constructed variables for the three sectors in 1840.

- **Plant labor productivity**: This is defined as log(revenue per worker). The census reports the value of output (in francs) and the total labor employed. We deflate output using the wholesale price index reported for 1840 in Mitchell (2003). The base year for the index is 1820; see Appendix B.6 for detail.

- **Capital (for cotton spinning only):** We define capital as the log number of spindles at the plant level. Spindles are the standard measure of capital in the industry. Some plants reported them and Juhász (2018, Online Appendix pp. 22-24) imputed the remainder.

- **log(Employment):** This is defined as the log of the total number of workers.

- **Plant total factor productivity (for cotton spinning only):** TFP at the plant level is computed as the residual from regressing plant-level deflated log revenue on log(total workers) and log(capital), as defined above.

- **Log(Spindles per worker) – for cotton spinning only:** This measure of capital intensity calculates the log number of spindles per worker in the plant.

- **Entrant:** Binary variable equal to one if the plant in 1840 did not have a name-based match in the 1800 survey wave.

B.5 Plant Linking and Plant Survival

This appendix section provides detail on our plant-linking procedure and on our computation of plant survival over the two sample periods.

Linking plants. When linking plants across survey periods, we use a fuzzy string match to allow for differences in spelling as well as for different first names of owners (e.g., in cases where the plant was passed on within a family). All matches were verified by hand. Ownership data indicate that only a small fraction owned more than one plant: Among the 528 cotton spinning plants in 1840, only 30 were part of multi-establishment firms with the same owner (in almost all these cases, one owner had two plants). Throughout the text, we thus refer to plants and firms interchangeably.

Our 'local matching' routine builds on locations that had only *one* plant in the respective sector in 1800, and at least one plant active in the same sector in 1840. Here, we provide a further way to validate the underlying assumptions that these were actually the same plants – or at least plants in the same sector and the same location. We examine how frequently communes with a single plant active in the sector in 1800 show up in 1840 with multiple plants active in the same sector. If this occurred frequently in the data, it would suggest that in fact there were often multiple suitable locations for production in that sector in the same commune. This is not the case in our data. It is rare (6.9% in cotton spinning, 5.2% in metallurgy, and 5.7% in paper milling) across all three surveys for single-plant communes to 'add' additional plants (despite the large increase in the overall number of plants in metallurgy and cotton spinning).

Plant survival. We can verify the methodology behind our computation of plant survival by using the rich data from the metallurgy sector. The 1811 metallurgy survey asked about each plant's activities in 1788. If our strategy of 'local matching' led to too many plant matches over time, we would expect an exaggerated survival rate. The contrary is true. Among the metallurgy plants in the 1811 survey, 77% reported that they had existed in 1788. Our plant-linking procedure for 1811-40 yields that among the 896 plants in 1840, 177 (20%) existed in 1811.[22] While the later period is longer, this alone cannot account for the substantially smaller number of initially existing plants in our matching procedure. This suggests that it is unlikely that we systematically *overestimate* plant survival.

B.6 Variables Used to Complement the Four Industry Surveys

Price Deflators.

Source: Mitchell (2003)

Methodology of deflating output prices: For all three sectors, and both time periods, we deflate revenues using the one wholesale price index reported in (Mitchell, 2003). The base year is 1820.

[22]Note that we cannot compute the actual survival rate for 1788-1811 because we do not have information on the initial number of plants in 1788. We thus compare the share of plants that existed in the earlier period, conditional on existing in the later period.

We use these to deflate revenues in cotton spinning (1806), in metallurgy (1811) in the corresponding exact years of the survey. For the paper milling survey from 1794, we need additional information, as the series starts in 1798. We use price data from the *Tableaux du Maximum*, which lists prices from 1790, before the Mitchell (2003) series begins.[23] We thus we need to make an assumption about how 1790 prices relate to prices we observe. Based on our reading of the literature on the timing of hyperinflation during the French Revolution and its subsequent reversion, we assume that 1790 prices were the same as prices in 1800.[24] For the census, we deflate revenues using the value for 1840.

Variables Common to all 3 Sectors:

- **Distance to high-productivity plants:** For each sector, this is computed as the log straight-line distance (in km) to the nearest plant with productivity in the 90th percentile (in the same sector) in the initial period. Any plant in the top decile of the productivity distribution is excluded from these regressions (which is why the number of observations in these regressions falls).
- **Distance to London:** This is the log straight-line distance (in km) from each plant's location in France to the city of London (England).

Control Variables:

- **Access to high stream-flow:**
  Data sources: EURO-FRIEND (2015). *European Water Archive, River Discharge Data*
  Data construction: The source data contain information on monthly mean streamflow rates (m^3/s) for 1,279 collection points across France for the years 1863 – 2012 (with data coverage best around 1960 – 1990). For each measuring station, we calculated the 95th percentile of streamflow across all months. That is, 95% of streamflow values are greater than this number at the given measuring station. This captures the year-round minimum streamflow that can be expected. We match each plant to its closest measuring station. Generally, this should

[23] As part of the revolutionary government's fight against inflation, a survey was conducted across the country, asking French districts to submit prices of goods produced or imported from aboard, along with their price (Daudin, 2010) These are reported in the *Tableaux du Maximum*.

[24] The high inflation during the revolutionary period was accompanied by the increasing issuing of *assignats* – a type of paper money. Initially, *assignats* were linked to gold – one *louis d'or* (a gold coin) was equal to 100 *assignats* in 1791. Thereafter, the value of *assignats* depreciated: In 1794, the same *louis d'or* corresponded to about 500 *assignats*. After a period of hyperinflation (1793-1797), the *assignats* were removed from the market (White, 1991; Sargent and Velde, 1995). In 1800, Napoleon founded the *Banque de France* and then eliminated paper money (Lefebvre, 2011), while trying to increase the stock of metallic money. While there is no systematic price index linking the 1790 and 1800 prices, it seems that by 1800, prices reverted back to their 1790 level (we are grateful to Eugene White for these insights on hyperinflation during the French Revolution). Moreover, for one town (Château-Gontier), we also observe a (wheat) price index every year from 1790 to 1800: This was 202 in 1790; it increased to 336 in 1794, and went back to 209 in 1800 (Hauser, 1985). This evidence from Château-Gontier further supports our assumption that 1790 prices were approximately the same as prices in 1800.

be a good proxy for the streamflow available locally. The median commune in France is located about 10km from its nearest measuring station. The commune at the 95th percentile of distance to its nearest measuring station is 46 km away. In other words, almost all plants are at a reasonable distance from their nearest measuring station.

Access to high streamflow is a binary variable that takes the value of one if a plant's nearest data collection point for river discharge has streamflow in the top quartile of the distribution (within a sector-year).

- **Proximity to coal:**
  Data sources: For coalfields within France: Guiollard P. – C. (1993) *Les Chevalements des Houilléres Françaises*. France, (2éme éd.)
  For coalfields in Europe outside of France: Coalfields of Europe, 1910. Available online at: https://etc.usf.edu/maps/pages/7200/7264/7264.htm
  Data construction: These maps were georeferenced by a Digital Cartography Specialist at the Harvard Map Collection. For each plant in our data, we then computed the distance to the nearest coalfield.
  Proximity to coal is a binary indicator that takes the value of one if a location is in the bottom quartile of plant locations (within a sector-year) in terms of distance to the nearest coalfield.
- **Share of forest area:**
  Data sources: Vallauri, Grel, Granier, and Dupouey (2012)
  Data construction: We calculate the share of each commune covered by forests as reported in the *Cassini* maps from the late 18th century. These maps were georeferenced and made available by Vallauri et al. (2012).
- **Production density:**
  Data construction: We sum the total revenue produced by commune in a sector-year net of the plant's own output. Production density is defined as log of 1 plus this sum.
- **Conscripts per capita by region:**
  Data sources: Vallée and Hargenvilliers (1936) and INSEE (https://www.insee.fr/fr/statistiques/2591293?sommaire=2591397)
  Data construction: We collected data on conscripted men from Vallée and Hargenvilliers (1936). This contains data by historical *département* on the number of men that were recruited for the years 1798-1805 (more specifically, between year 7 and year 13 according to the Republican Calendar). While this does not include recruited men for later years, it is arguably a good proxy, as recruitment rates across *départements* displayed high persistence (Forrest, 1989). We then compute the variable log(conscripts per capita), using data on population at the historical *département* level from INSEE. Since historical *département* do not map into their contemporaneous counterparts (which we use in our analysis), we aggregate

the data on conscripts per capita to the regional level.

- **Proximity to battles during the Napoleonic Wars:**
  Data sources: Wikipedia
  https://en.wikipedia.org/wiki/List_of_battles_of_the_War_of_the_Sixth_Coalition (for the War of the Sixth Coalition);
  https://en.wikipedia.org/wiki/List_of_battles_of_the_Hundred_Days (for the Hundred Days War)
  Data construction: We first geocode the location of all battles that took place on French territory. These comprise the final battles of the Sixth Coalition (1813-14) and of the Hundred Days War (culminating in Napoleon's defeat at Waterloo). We then construct a dummy (*Near Battles*) equal to one for plants located within 10km from a battlefield.
- **Market access:**
  Data sources: Özak (2018) and Nunn and Qian (2011)
  Data construction: We construct two measures of market access: i) within France and; ii) across Europe. Both measures are computed as the sum of inverse-distance-weighted urban populations j around each French commune i in 1800 (Market Access $MA_i \equiv \sum_j \frac{pop_j}{dist_{ij}}$). For commune i, we use the geocoded location (described above). We take data on cities j with a population of more than 1,000 inhabitants in 1800 from Bairoch, Batou, and Chèvre (1988), as reported in Nunn and Qian (2011), including geocodes for these cities. Then, for each plant in our data, we calculate the shortest travel time between the plant and each city, using the *Human Mobility Index with Seafaring* (Özak, 2018). This measure constructs minimum travel time (in hours) using constraints on human mobility and technological constraints from before the steam-age. Specifically, the data provides the cost of crossing each 1 x 1 km cell across the globe (land and sea). As such, the minimum travel time between any two points in space can be computed using GIS software. For the within-France measure, we only use population centers in France, whereas for the Europe-wide measure we use all European cities in the data.

C Descriptive Statistics

This appendix describes key descriptive statistics for the three industries covered in our data.

C.1 Plant Locations

Figure A.14 shows the spatial distribution of firms in the three sectors and in the two time periods.

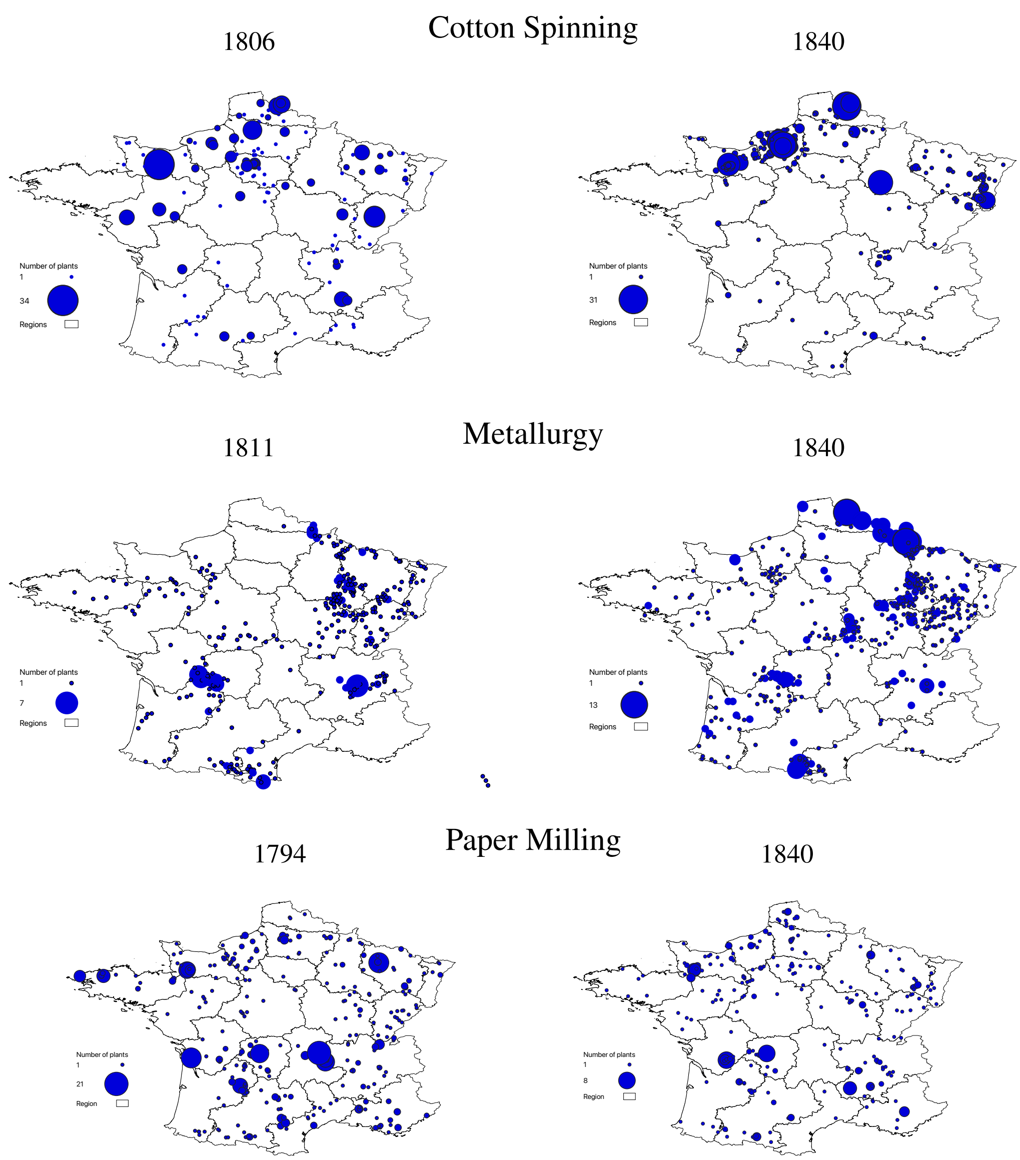


Figure A.14: Spatial Distribution of Plants Across France in the Three Sectors

Note: The figure shows the spatial distribution of plants in cotton spinning (top), metallurgy (middle), and paper milling (bottom). Dot sizes reflect the number of plants per commune.

C.2 Summary Statistics for the Three Sectors

Tables A.2, A.3, and A.4 show summary statistics for mechanized cotton spinning, metallurgy, and paper milling plants, respectively.

Table A.2: Summary Statistics – Mechanized Cotton Spinning

	Around 1800					1840					Sources
	#obs	Mean	Std. Dev.	Min	Max	#obs	Mean	Std. Dev.	Min	Max	& Detail
Firm-level characteristics											
Log (output per worker)	340	7.08	0.87	2.48	9.73	528	7.90	0.51	3.65	9.38	App. B.1
Number of workers	340	63.9	103	1.00	950	528	112	148	4.00	1,413	App. B.1
Plant Age	334	6.59	8.06	0.00	56						App. B.1
Exit dummy	340	0.95	0.22	0.00	1.00						App. B.1
Avg. quality of yarn	323	35.19	19.98	4.50	135						App. B.1
Total num. of machines	300	14.76	23.32	1.00	250						App. B.1
Num. spinning jennies	336	2.39	6.99	0.00	60						App. B.1
Num. water-frames	301	4.11	11.24	0.00	96						App. B.1
Num. of spinning-mules	302	7.91	19.37	0.00	200						App. B.1
Spindles	340	1,658	3,686	26.00	49,200	528	6,392	7,393	300	85,000	App. B.1
Spindles per worker	340	27.72	19.05	1.67	120	528	78.18	90.11	4.92	1,000	App. B.1
Water power						528	0.66	0.47	0.00	1.00	App. B.1
Steam power						528	0.39	0.49	0.00	1.00	App. B.1
Other power						528	0.02	0.14	0.00	1.00	App. B.1
Control Variables											
Access to high streamflow	321	0.25	0.43	0.00	1.00	519	0.22	0.41	0.00	1.00	App. B.6
Production density	321	0.25	0.43	0.00	1.00	519	0.26	0.44	0.00	1.00	App. B.6
Share of forest area	321	0.10	0.15	0.00	0.82	519	0.15	0.19	0.00	0.81	App. B.6
Production density	340	9.28	5.65	0.00	14.94	528	9.47	6.51	0.00	15.84	App. B.6
Market access, France	321	5.09	0.85	4.08	8.14	519	4.81	0.50	4.11	6.35	App. B.6
Market access, Europe	321	5.91	0.55	5.31	8.18	519	5.78	0.26	5.32	6.64	App. B.6
Access to overseas market	321	0.39	0.49	0.00	1.00	519	0.56	0.50	0.00	1.00	App. B.6
log(conscripts pc)	321	2.88	0.57	1.63	3.38	519	2.88	0.47	1.63	3.38	App. B.6
Near Battles	321	0.14	0.35	0.00	1.00	519	0.01	0.09	0.00	1.00	App. B.6
Distance Variables											
Dist to p90	290	87.8	86.2	0.00	428.8	467	34.8	54.82	0.00	288	App. B.6
Dist to p90 metal (1800)	321	110	57.0	10.8	207						App. B.6
Dist to p90 paper (1800)	321	61.6	33.3	0.00	148						App. B.6
Dist to London	321	433	202	192	943	519	359.8	182	195	1,005	App. B.6

Note: The table shows the summary statistics for the variables used in the paper and appendix.

Table A.3: Summary Statistics – Metallurgy

	Around 1800					1840					Sources
	#obs	Mean	Std. Dev.	Min	Max	#obs	Mean	Std. Dev.	Min	Max	& Detail
Firm-level characteristics											
Log (output per worker)	470	8.12	0.94	2.09	11.04	896	8.80	0.92	4.87	11.69	App. B.2
Number of workers	470	22.55	37.38	2.00	500	896	56.55	112.4	1.00	1,400	App. B.2
Exit dummy	470	0.62	0.49	0.00	1.00						App. B.2
Water power						896	0.64	0.48	0.00	1.00	App. B.2
Steam power						896	0.15	0.36	0.00	1.00	App. B.2
Other power						896	0.08	0.28	0.00	1.00	App. B.2
Control Variables											
Access to high streamflow	426	0.23	0.42	0.00	1.00	863	0.25	0.43	0.00	1.00	App. B.6
Proximity to coal	426	0.25	0.43	0.00	1.00	863	0.25	0.43	0.00	1.00	App. B.6
Share of forest area	417	0.24	0.21	0.00	0.88	863	0.24	0.22	0.00	0.90	App. B.6
Production density	470	3.85	5.92	0.00	15.56	896	5.57	6.45	0.00	16.47	App. B.6
log(conscripts pc)	426	3.05	0.41	1.39	3.38	863	2.99	0.50	1.39	3.38	App. B.6
Near Battles	426	0.01	0.11	0.00	1.00	863	0.03	0.16	0.00	1.00	App. B.6
Distance Variables											
Dist to p90	385	39.3	40.3	0.00	236.0	863	47.5	51.7	0.00	273.9	App. B.6
Dist to London	427	632.3	178.8	253.9	1,035	864	560.5	191.4	148.5	1,007	App. B.6

Note: The table shows the summary statistics for the variables used in the paper and appendix.

C.3 Plant Scale

We examine plant scale and the number of plants in each industry. Plant size is measured by the number of workers. A few points stand out. First, as early as 1806, cotton spinning plants were strikingly large. The average spinning plant in this period had 64 employees. Despite the recent introduction of mechanized cotton spinning in France, plants were already larger than in the two comparison sectors, both of which had a longer tradition of factory-based production. Plants in metallurgy (reported in 1811) had on average 23 workers; paper milling plants had on average 13 employees.[25]

We also observe that between 1806 and 1840, the number of mechanized cotton spinning plants

[25]One caveat with making this comparison is that the paper milling survey dates from 1794. Thus, plant size may have grown by 1806 – the year of the cotton spinning survey. In addition, we had to extrapolate the overall number of workers in paper milling in 1794 (including women and children – see Section 3.2). However, it is unlikely that the actual paper plant scale was very different in 1806. This is because even in 1840, the average plant size in paper milling was only 43 (including women and children, which are reported in this year). We can thus be confident that paper milling plants in 1806 were substantially smaller than cotton plants. Finally, as described below in Appendix B.4, there is a concern that the 1840 census did not enumerate all plants with less than 10 employees (which, however, does not affect our results – see Appendix E.10).

Table A.4: Summary Statistics – Paper Milling

	Around 1800					1840					Sources
	#obs	Mean	Std. Dev.	Min	Max	#obs	Mean	Std. Dev.	Min	Max	& Detail
Firm-level characteristics											
Log (output per worker)	520	7.28	0.77	3.55	10.52	347	7.61	0.71	4.85	11.51	App. B.3
Number of workers	520	12.56	17.81	2.00	317	347	42.61	58.52	1.00	507	App. B.3
Water power						347	0.85	0.36	0.00	1.00	App. B.3
Steam power						347	0.12	0.32	0.00	1.00	App. B.3
Other power						347	0.02	0.14	0.00	1.00	App. B.3
Control Variables											
Access to high streamflow	507	0.24	0.43	0.00	1.00	343	0.23	0.42	0.00	1.00	App. B.6
Proximity to coal	507	0.26	0.44	0.00	1.00	343	0.25	0.43	0.00	1.00	App. B.6
Share of forest area	507	0.11	0.15	0.00	0.76	343	0.10	0.15	0.00	0.76	App. B.6
Production density	520	6.62	5.56	0.00	13.45	347	5.46	5.77	0.00	14.32	App. B.6
log(conscripts pc)	507	2.96	0.44	1.39	3.38	343	2.91	0.51	1.39	3.38	App. B.6
Near Battles	507	0.00	0.06	0.00	1.00	343	0.01	0.12	0.00	1.00	App. B.6
Distance Variables											
Dist to p90	456	38.1	33.1	0.00	168.8	343	44.0	32.6	0.00	225.9	App. B.6
Dist to London	510	605.3	222.4	194.7	1,036	343	544.31	238.5	177.6	1,022	App. B.6

Note: The table shows the summary statistics for the variables used in the paper and appendix.

active in the market expanded markedly (from 340 in 1806, to 528 in 1840). This is important, as it suggests that our results, which show a disappearance of the lower tail of the productivity distribution, were driven by more than simply the 'shake-out' of unsuccessful plants. In fact, each exiting plant was replaced on average by more than one new entrant.

D Mechanism: Stylized Framework and Evidence

This section describes a stylized theoretical framework that generates the main pattern in our data: a lower-tail bias in productivity growth in mechanized cotton spinning. For simplicity, we focus on a partial equilibrium setting where the economy-wide expenditure for spun cotton yarn is given. Spinning firms produce differentiated products, which reflects differences in output varieties as well as spatially segmented markets. Firms randomly draw their productivity, based on a combination of complementary input tasks. The complementarity across individual input tasks leads to a fat lower tail in the initial productivity distribution.

We consider three periods over a firm's lifetime. In the first period, firms either establish themselves in the market or they exit (if they cannot pay the fixed cost of production). This weeds out firms with very low productivity draws. In the innovation period, established firms can decide whether they want to invest their time in learning organizational knowledge from other producers, or whether they want to continue producing. In the spirit of Perla and Tonetti (2014), searching comes at the cost of foregone output. In the final period, firms that searched adopt the improved organizational knowledge. This leads to the disappearance of the lower tail, as relatively unproductive firms endogenously sort into improving their productivity by learning from the high-productivity firms.

D.1 Production

We choose a Leontief production function that features strong complementarity across multiple inputs (tasks). When organizing production, a firm needs to coordinate $m = 1, ..., M$ production tasks, such as feeding raw cotton into its machines, collecting output, managing fire hazards, ensuring power supply, etc. Each task is performed by task-specific production labor l_m^P, and the corresponding task-efficiency γ_m is drawn from a uniform distribution with support $[0, 1]$. Firm output q is given by:

$$q = M \cdot \min\left\{\gamma_1 l_1^P; \gamma_2 l_2^P; ...; \gamma_M l_M^P\right\} \tag{D.1}$$

We refer to the set of γ_m draws as "organizational knowledge." Note that the maximum efficiency (i.e., the technology frontier) is reached if all tasks are performed with $\gamma_m = 1$, reflecting a 'perfect' organization of production. The strong (Leontief) complementarity implies that any particularly low γ_m draw has a disproportionate negative effect on overall productivity, leading to a fat lower tail of the productivity distribution.

Optimal choice of task-specific labor for given organizational knowledge (γ_m draws) implies $\gamma_m l_m^P = \gamma_1 l_1^P, \forall m$. This allows us to write firm output as $q = M \cdot \gamma_1 l_1^P$. In addition, defining total production labor $l^P = \sum_{m=1}^M l_m^P$ and substituting $l_m^P = \frac{\gamma_1}{\gamma_m} l_1^P$, we obtain: $l^P = l_1^P \cdot \sum_{m=1}^M \frac{\gamma_1}{\gamma_m}$, and therefore: $l_1^P = \frac{l^P}{\sum_{m=1}^M \frac{\gamma_1}{\gamma_m}}$. Substituting this expression in the above equation for q yields:

$$q = \psi \cdot l^P, \quad \text{where} \quad \psi \equiv \frac{1}{\frac{1}{M}\sum_{m=1}^M \frac{1}{\gamma_m}} \tag{D.2}$$

Note that ψ reflects the overall productivity of a firm, which in turn is a composite of its γ_m draws. In addition to production labor, firms also incur a fixed cost f each period which – following Melitz (2003) – is paid in units of labor.

Thus, the overall labor used to produce output q is given by

$$l = f + \frac{q}{\psi} \tag{D.3}$$

The fixed cost f affects a firm's decision to operate vs. shut down (as discussed below).

D.2 Demand and Profit Maximization

We assume that each mechanized cotton spinning firm i produces a differentiated variety, q_i. Differentiated varieties can reflect differences in the type of cotton that is spun (different counts of yarn), but also spatially segmented markets due to imperfect market integration. Overall demand for output from the cotton spinning sector is given by

$$Q = \left(\sum_{i=1}^{I} q_i^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} , \tag{D.4}$$

where $\epsilon > 1$ is the elasticity of substitution, and I is the total number of firms. We assume that there are sufficiently many firms so that each individual producer takes Q as given. We focus on a partial equilibrium setting, where the aggregate spending for cotton-spinning output, $R \equiv \sum_{i=1}^{n} p_i q_i = \mathcal{P}Q$ is given, with $\mathcal{P} \equiv \left(\sum_{i=1}^{I} p_i^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$ denoting the price index for cotton spinning output. The aggregate expenditure can reflect both domestic and international demand. Wages w are also given. Each firm sets its price p_i to maximize profits. In this setup, demand for individual varieties is given by

$$q_i = \left(\frac{\mathcal{P}}{p_i} \right)^{\epsilon} \cdot Q \tag{D.5}$$

Firms maximize profits, which are given by $\pi_i = p_i q_i - w \left(f + \frac{q_i}{\psi_i} \right)$. This yields the profit-maximizing price as a constant markup over firm i's marginal cost:

$$p_i = \frac{\epsilon}{\epsilon - 1} \frac{w}{\psi_i} \tag{D.6}$$

Substituting the optimal price in the profit equation, we obtain the following expression for firm i's profits:

$$\pi_i = \frac{1}{\epsilon} \cdot R \cdot \left(\frac{\epsilon - 1}{\epsilon} \cdot \mathcal{P} \cdot \frac{\psi_i}{w} \right)^{\epsilon-1} - f \tag{D.7}$$

This equation shows that firms with higher productivity draws ψ_i will make higher profits, which leads to the final step: firms' decisions to operate and innovate.

D.3 Firms' Decisions to Operate and Innovate

Following (D.7), profits depend on firm i's organizational knowledge draws $\gamma_{i,m}$, which are aggregated into ψ_i as in equation (D.2). We assume that firms receive these draws *after* committing to pay the fixed costs for the first period. For example, in order to get information about how productive they are, firms need to set up their plant and start producing. Given this setup, initially, all firms produce, and the productivity distribution exhibits a thick lower tail. Firms with low productivity draws that imply $\pi_i < 0$ exit in the first period, i.e., their revenues are not sufficient to cover their variable and fixed costs, and they go bankrupt.[26] We refer to all remaining firms (those with $\pi_i \geq 0$) as the operating firms who remain in business. We denote the productivity level that corresponds to zero profits (and thus the decision to operate) by $\bar{\psi}^O$.

Rather than examining productivity dynamics with an infinite time horizon as in Perla and Tonetti (2014), we simplify the setup by focusing on three time periods, which is sufficient to highlight the relevant productivity dynamics.

1. **Initial Period.** In the first period, entrepreneurs commit to paying the fixed cost f, receive their initial productivity draw, and start producing. Those with low productivity draws make negative profits and drop out of business at the end of the period. Thus, productivity dynamics in the first period are driven by unproductive firms exiting the market.

2. **Innovation Period.** In period 2, surviving entrepreneurs decide whether to innovate by observing and copying organizational knowledge from other firms. We follow the setup from Perla and Tonetti (2014) whereby firms that decide to search will forgo profits and encounter a randomly drawn *producing* firm, copying this firm's task-efficiency draws.[27] In equilibrium, relatively unproductive firms decide to search, while productive firms continue producing.[28] This gives rise to an endogenous productivity threshold, and firms below this threshold sample from the productivity distribution above the threshold. This process shifts mass from lower to higher productivity levels.

3. **Final Period.** Finally, in period 3, all firms are producing again: those that were searching in period 2, and those that remained in production. We compare the new productivity distribution to its counterpart in the initial period.

Assumptions. Before moving on to the simulation results, we state the implicit assumptions in our setup. First, market entry occurs only in the initial period: Firms receive their productivity draw and then decide whether or not to operate. There is no entry of new firms in the later periods. Second, only operating firms can decide to search for better organizational knowledge in period 2. Thus, only 'established' firms that survived the first period can innovate. Note that this assumption also implies that 'outsiders' who have never operated mechanized cotton technology cannot search and copy from producing firms. Third, if a searching firm i is matched to a producing firm j, it

[26]For example, think of the fixed costs per period as the payments to the bank for a loan taken out to finance the plant. If these payments cannot be met, the firm faces bankruptcy.

[27]In our setup, innovating entrepreneurs stop production and do not pay the fixed cost. They thus effectively go out of business and re-enter in the next period with their newly drawn productivity. This represents the historical findings that a large part of innovation occurred through churn, and that improvements in production facilities often meant that the old design had to be scrapped.

[28]Importantly, we simplify the structure from Perla and Tonetti (2014) by assuming that innovation (search) only occurs in the second period – as opposed to in each period. This simple structure is sufficient for our purpose – to study the productivity distribution before and after the innovation period.

copies the *full* set of organizational practices $\gamma_{j,m}$, $\forall m$.[29] A historical justification for this assumption is that many components of mechanized spinning plants were closely linked to each other. For example, in order to change the organization of machines within plants, buildings had to be modified or even re-built (see Section 2 in the paper). Finally, we assume that all firm activity ceases after the final period, thus abstracting from long-horizon dynamics in the decision to innovate.

D.4 Simulation Steps

To simulate the model, we use $M = 5$ production tasks with the corresponding organizational efficiency $\gamma_{it,m}$, drawn from a uniform $(0, 1)$ distribution for $I = 1,000$ firms. Using the wage rate $w = 1$, fixed cost $f = 1$, and aggregate spending $R = 10,000$, we compute firm profits using (D.7).[30] The final step in the initial period is to find the lowest-productivity firm that decides to operate. This establishes the operating threshold $\bar{\psi}^O_{ini}$, with the subscript indicating the initial period. We refer to the three periods as initial (ini), innovation ($innov$), and final ($final$).[31] All firms with $\psi_i \geq \bar{\psi}^O_{ini}$ survive the initial period and enter the innovation period.

Next, we move on to the second period, when firms decide whether to innovate (and stop production). Among all surviving firms, we compute the cutoff $\bar{\psi}^S$ above which firms produce, while those below the cutoff search for better efficiency draws. This procedure involves four steps:

1. Use an initial guess for the productivity cutoff $\bar{\psi}^S$ below which firms with $\bar{\psi}^O_{ini} \leq \psi_i < \bar{\psi}^S$ search (denoted by the set S of firms), while those with $\psi_i \geq \bar{\psi}^S$ produce and will (potentially) see their efficiency draws being copied by searching firms.[32]

2. Compute output prices $p_{i,innov}$ for the producing firms (with $\psi_i \geq \bar{\psi}^S$), the corresponding price index $\mathcal{P}_{innov}$, and profits $\pi_{i,innov}$.

3. Compute the expected productivity draw for searching firms, $E(\psi')$ for $i \in S$, where S denotes the set of searching firms. This is equal to the mean of ψ_i over all producing firms, i.e., $E(\psi') = E(\psi|\psi \geq \bar{\psi}^S)$. Based on this expected productivity draw, compute the expected profits of all firms in the next (final) period, $E(\pi_{final})$, assuming that all searching firms

[29]Without this assumption, the simple Perla and Tonetti (2014) framework cannot be applied because it builds on a productivity ranking based on overall firm productivity (ψ_i in our model), with low-productivity firms endogenously deciding to search. In our setting, even firms with a relatively high ψ_i could find it beneficial to search for better *individual* $\gamma_{i,m}$ draws, especially if just one of their $\gamma_{i,m}$ draws is particularly low, disproportionately affecting ψ_i because of the strong complementarity in (D.1).

[30]Our choice of aggregate spending R implies that about two-thirds of all firms make positive profits and thus operate.

[31]The subscript is needed because the threshold depends on the number of firms that produce output, which differs in the innovating period. Similarly, note that the price index $\mathcal{P}_{ini}$ depends on the number of firms that produce; and profits, in turn, depend on $\mathcal{P}_{ini}$. Thus, an iterative process is required to find the threshold $\bar{\psi}^O_{ini}$.

[32]The first term, $\bar{\psi}^O_{ini} \leq \psi_i$ reflects our assumption that only firms that survived the initial period can innovate by searching for better efficiency draws.

produce with productivity $\psi_{i,final} = E(\psi'), \forall i \in S$, while all producing firms continue with the same productivity as in the initial period, $\psi_{i,final} = \psi_{i,ini}, \forall i \notin S$.

4. Compute the expected overall profits (from periods 2 and 3) for a firm with initial productivity at the search cutoff $\psi_{i,ini} = \bar{\psi}^S$ for two scenarios: i) if the firm produces in the innovation period and thus keeps the same productivity in the final period: $\pi(\bar{\psi}^S|produce) = \pi_{innov}(\bar{\psi}^S) + \beta\pi_{final}(\bar{\psi}^S)$, where $\beta < 1$ is the discount rate;[33] and ii) if the firm searches in the innovation period and thus receives the expected productivity draw $\psi_{i,final} = E(\psi')$ for the final period, while foregoing profits in the innovation period: $\pi(\bar{\psi}^S|search) = 0 + \beta\pi_{final}(E(\psi'))$.[34] If $\pi(\bar{\psi}^S|\text{produce}) < \pi(\bar{\psi}^S|\text{search})$ then the initial guess for the threshold $\bar{\psi}^S$ was too low, as the threshold firm would still be better off searching instead of producing. We thus update $\bar{\psi}^S = \psi_{i+1}$, where ψ_{i+1} represents the next-higher initial productivity draw among all firms.

We order all operating firms by their productivity draws such that low i in ψ_i represent lower draws. We begin with a relatively low guess for $\bar{\psi}^S$ and repeat steps 1-4 until the firm with $\psi_i = \bar{\psi}^S$ is opting to search, while the firm with the next-higher productivity draw (ψ_{i+1}) finds it optimal to produce, rather than search. This yields the cutoff for searching in the innovation period, $\bar{\psi}^S$. Finally, we match each searching firm i (with $\bar{\psi}^O \leq \psi_i < \bar{\psi}^S$) at random to a producing firm j (with $\psi_j \geq \bar{\psi}^S$) and update i's productivity so that $\psi_{i,final} = \psi_{j,ini}$. This yields the productivity distribution in the final period.[35]

D.5 Simulation Results

The left panel in Figure A.15 illustrates the weeding-out of unproductive firms during the initial period. Because of the strong complementarity across production tasks, the original productivity distribution exhibits a fat lower tail. Firms with low productivity draws (below the cutoff $\bar{\psi}^O_{ini}$) exit the market. The remaining firms survive into the second period, where they decide between searching for better organizational knowledge and production. This leads to the productivity distribution shown in the right panel of Figure A.15. Among the surviving firms, those with productivity below the search threshold $\bar{\psi}^S$ forgo profits in the innovation period and are randomly matched to one of the producing firms, copying their efficiency draws. In the final period, searching firms proceed with their new copied productivity, whereas producing firms continue with their initial draws.

Figure A.16 illustrates the productivity dynamics that result from the search-and-innovation process, by comparing the productivity distributions before and after (i.e., in the innovation period

[33] We assume that periods 2 and 3 are of equal length and we choose $\beta = 0.6$, corresponding to cumulative discounting at a rate of 0.95 over ten years. Note also our assumption that all firm activity ceases after period 3 simplifies the setting, as we can abstract from longer-run dynamics in the decision to innovate.

[34] We assume that when searching, firms make zero profits.

[35] Matching occurs with replacement: Different searching firms can be matched to the same producing firm.

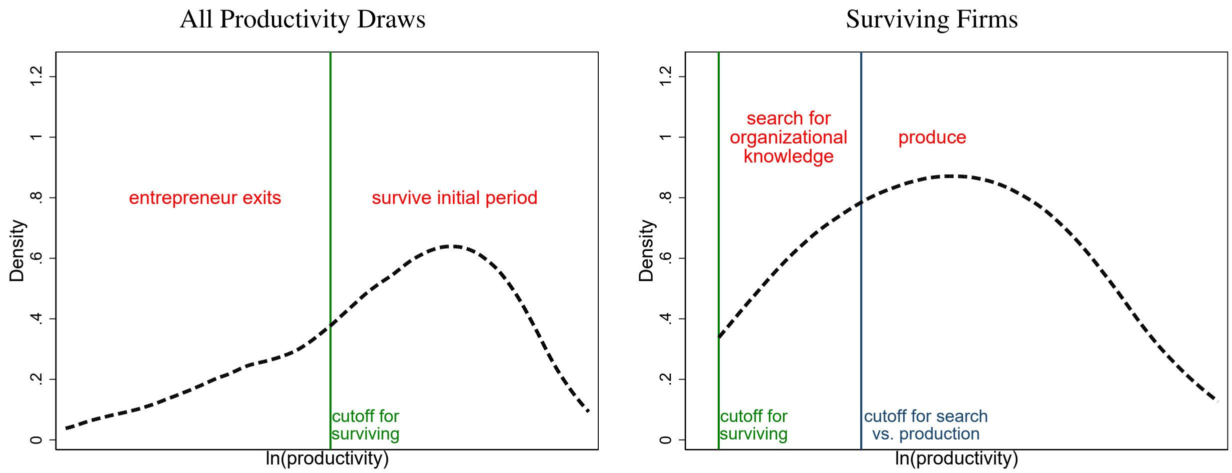


Figure A.15: Initial Period: Firm Survival

Notes: The figure illustrates two firm decisions in the model, as a function of the initial firm productivity draws (ψ_i): The left panel plots all productivity draws, showing that the distribution of ψ_i exhibits a thick lower tail due to the underlying strong complementarity across individual production tasks in equations (D.1) and (D.2). Firms with low productivity draws make negative profits and exit in the initial period. Firms with higher productivity draws decide to operate. The right panel illustrates the surviving firm' decision to produce vs. search for better organizational knowledge. Firms with relatively low productivity draws can raise their expected profits by searching for better draws among those firms that continue production. The cutoff for search vs. production is endogenously determined.

and in the final period): The lower part of the initial productivity distribution disappears, and the upper part becomes thicker. That is, productivity growth due to search exhibits a lower-tail bias. At the same time, the productivity frontier does not move out; instead, the distribution is tilted towards the frontier.

Overall, there are thus two mechanisms that lead to a lower-tail bias of productivity growth: i) the exit of unproductive firms and ii) the search of medium-productivity firms for better organizational knowledge. These features reflect the pattern that we document in the historical data for mechanized cotton spinning in France. The overall productivity dynamics that result from the three periods in combination are illustrated in Figure 3 in the paper.

D.6 Relationship to Comparison Sectors

The process described in our stylized framework (following Perla and Tonetti, 2014) eliminates the lower tail and shifts the productivity distribution closer to the frontier. The productivity distribution in cotton spinning in the final period resembles the one observed for our comparison sectors, where factory production had been adopted earlier, and the process of exit and organizational learning had already occurred around 1800. As we noted in Section 2.3, codified knowledge was already available for the two comparison sectors in the late 18th century (while it only became available after 1830 in cotton spinning – see Section 2.2).

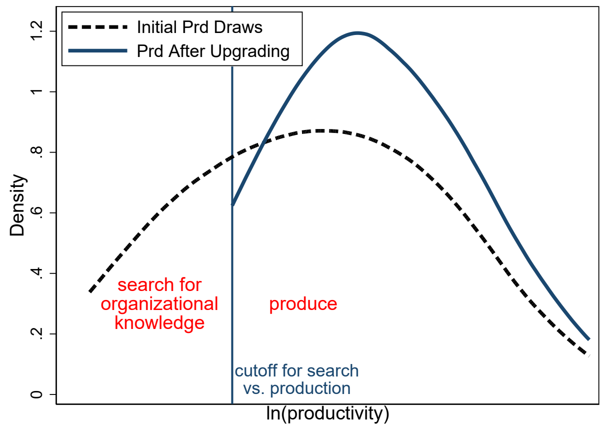


Figure A.16: Productivity Dynamics During the Search Period

Notes: The figure illustrates the productivity dynamics during the search period, focusing only on those firms that survived the initial period (see Figure A.15). Less productive firms decide to search for better organizational knowledge, and they adopt the productivity draws from more productive producing firms. Thus, the lower tail of the productivity distribution disappears, and the mass shifts towards higher productivity draws. At the same time, the technological frontier remains the same.

Why do we observe substantial variation in productivity even after best-practice knowledge became available? While there are many reasons for productivity dispersion in practice (Syverson, 2011), here we focus on one that follows from the model: The adoption of better practices is costly – it requires production to stop for one period. Thus, even if (in the model) plant owners knew where to observe the best practices, not all would adopt them (although the sudden availability of best practices would lead to a spike in plants adopting them).[36] In particular, being already closer to the frontier makes adoption less attractive, because the opportunity cost of halting production is higher. This is also in line with Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013), who show that even when knowledge about standardized practices is readily available, it may not be applied.

This discussion highlights the important historical difference between cotton spinning in 1800 and i) our comparison sectors in 1800, ii) all three sectors in 1840: The former faced a lack of standardized solutions, while in i) and ii), codified knowledge was available, at least to those plants that looked for it. Consequently, only the productivity process in cotton spinning in 1800 exhibited the additional feature of experimentation that is at the heart of our stylized framework, and that leads to the elimination of the fat lower tail.

[36] Here, we refer to our model with just three periods. In the original Perla and Tonetti (2014) model with an infinite horizon, plants would continue to approach the frontier in subsequent periods, and this process would be accelerated if best-practice knowledge becomes available.

E Additional Data and Results

This appendix presents the additional empirical analyses and robustness checks referenced in the main text.

E.1 Cotton Spinning: Output Quality and TFP

In Section 4.1 of the paper we showed that the productivity gains in mechanized cotton spinning were largely concentrated in the lower tail of the productivity distribution. The lower tail disappeared over our sample period, while increases in productivity at the upper tail of mechanized spinning were modest. Panel A of Table A.5 repeats the results of our main quantile regressions from the paper, using quality-adjusted prices.[37] Panel B of Table A.5 shows that these results are robust when we do not adjust for quality differences in the count of yarn spun by individual plants, instead using the *same* sector-level output price across all plants in cotton spinning. In particular, we use the price of the yarn count that the average plant produced.[38] This reduces the productivity dispersion in 1806 because more productive plants produced higher-quality cotton.[39] As a result, the difference in productivity growth across quantiles is somewhat muted as compared to Panel A (as we would expect), but the lower-tail bias remains striking.

Panel C of Table A.5 presents results for total factor productivity (TFP) instead of output per worker in the productivity regressions of mechanized cotton spinning. To estimate (revenue-based) TFP, we use data on the labor employed by the firm (measured as number of workers) and proxy for the capital stock with the number of spindles – a standard measure of production capacity in the industry. We regress the log revenue of the firm on a constant, log labor, and the log number of spindles of the plant, separately for 1806 and 1840. This allows for the capital and labor shares to be time-varying. Log TFP for each plant i in a given year t is thus the regression constant plus the residual of the regression. We find that the lower-tail bias of productivity growth in mechanized cotton spinning is robust to using TFP instead of output per worker.

E.2 Metallurgy and Paper Milling: Robustness Checks for Imputed Variables

For metallurgy and paper milling, we had to impute labor inputs in 1811 and 1794, respectively. In this section, we check the robustness of our results to alternative data construction choices.

In metallurgy, labor inputs are not consistently reported: about 40% of the plants reported either 'internal' labor only, or both 'internal' and 'external' labor, separately. The remainder of plants reported only total labor, with no indication of whether this included external labor. As explained

[37] The output price used for each plant corresponds to the market price of the quality of yarn reported by the plant, as described in Appendix B.1 (under "Plant labor productivity").

[38] To compute the not-quality-adjusted productivity, we first construct plant-level revenues by multiplying physical output at the plant level with the price of the (unweighted) average quality of yarn reported across *all* plants in 1806.

[39] The correlation between the not-quality-adjusted productivity measure used in Panel B of Table A.5 and the plant-specific (quality-adjusted) output price (that underlies the results in Panel A) is statistically highly significant, with a p-value below 0.01.

Table A.5: Alternative Productivity Measures in Cotton Spinning

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
PANEL A: Baseline (Table 1)							
Spinning (1806-1840)	2.420***	3.917***	3.293***	2.234***	1.651***	1.014***	868
	(0.154)	(0.204)	(0.229)	(0.151)	(0.167)	(0.297)	
PANEL B: Using prices not quality-adjusted							
Spinning (1806-1840)	2.373***	3.381***	2.828***	2.105***	1.829***	1.628***	868
	(0.138)	(0.285)	(0.199)	(0.193)	(0.160)	(0.188)	
PANEL C: Using TFP							
Spinning (1806-1840)	2.845***	3.233***	3.107***	2.834***	2.647***	2.317***	868
	(0.050)	(0.080)	(0.072)	(0.056)	(0.083)	(0.072)	

Notes: Panel A reproduces the specification in Table 1. In Panel B, the dependent variable is log(output per worker) computed using prices that are *not* adjusted for plant-specific output quality. In Panel C, the dependent variable is total factor productivity. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

in Section 3.2 in our main analysis, we imputed internal labor for those plants not reporting it (60% of the total number of plants in 1811), because internal labor is more consistent with the 1840 data. In Panel B of Table A.6, we check whether our results are driven by this imputation: We drop all plants for which we imputed labor. Despite the fact that this drops 292 out of 470 plants in 1811 (so that the overall observations decrease from 1366 to 1074), the results remain very similar to our baseline specification. If anything, productivity growth in metallurgy skews more in the direction of an *upper*-tail bias than in the baseline.

As described in Appendix B.2, to compute plant-level labor productivity in metallurgy in 1811, we computed average prices for each product type, by using the information from those plants that reported prices for the corresponding product. Overall, 308 out of 470 plants reported these product-specific prices. In Panel C of Table A.6 we show that our results hold when we use the product-specific prices for those plants that reported them, while dropping the remaining metallurgy plants in 1811.

In paper milling, many plants reported only male labor in 1794, while the 1840 survey reports both male and total labor for all plants. We thus imputed total labor in 1794 using a scaling factor between male and total employees as described in Section 3.2 (under "Constructing Consistent Labor Variables"). As our baseline variable, we used (imputed) total labor in 1794 and the reported total labor in 1840 (reported again in Panel A of Table A.7). Panel B shows that the productivity growth pattern in paper milling is robust to using only male labor in both periods to construct

Table A.6: Alternative Productivity Measures in Metallurgy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
PANEL A: Baseline (Table 1)							
Metallurgy (1811-1840)	2.328***	2.205***	2.068***	1.979***	2.285***	2.998***	1,366
	(0.183)	(0.530)	(0.317)	(0.247)	(0.193)	(0.232)	
PANEL B: Using plants in 1800 with non-imputed labor							
Metallurgy (1811-1840)	1.744***	0.974	1.383***	1.460***	1.998***	2.791***	1,074
	(0.259)	(0.634)	(0.342)	(0.292)	(0.288)	(0.265)	
PANEL C: Using plants in 1800 with plant-specific prices							
Metallurgy (1811-1840)	2.694***	2.695***	2.574***	2.542***	2.552***	3.286***	1,204
	(0.207)	(0.503)	(0.412)	(0.254)	(0.205)	(0.207)	

Notes: Panel A reproduces the specification of Table 1. In Panel B, the dependent variable is log(output per worker), using only plants with non-imputed labor. In Panel C, the dependent variable is log(output per worker), using only plants reporting plant-specific output prices. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

productivity: There is no clear pattern across the productivity distribution, and if anything, growth is concentrated in the mid-area. Note that the number of observations in this check (Panel B) remains the same as in our baseline results (Panel A) because all plants reported male labor in 1794.

Table A.7: Alternative Productivity Measures in Paper Milling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
PANEL A: Baseline (Table 1)							
Paper milling (1794-1840)	0.719***	0.697***	0.717***	0.846***	0.691***	0.542**	867
	(0.111)	(0.145)	(0.139)	(0.092)	(0.130)	(0.258)	
PANEL B: Using only male labor							
Paper milling (1794-1840)	0.791***	0.522***	0.569***	0.728***	1.157***	0.884***	867
	(0.115)	(0.161)	(0.164)	(0.131)	(0.133)	(0.229)	

Notes: Panel A reproduces the specification of Table 1. In Panel B, the dependent variable is log(output per worker) computed using male labor only. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.3 Evidence for Proposed Mechanism: Building Layout in Mechanized Cotton Spinning

Data. Our data on the layout of French cotton spinning mills are from Chassagne (1991), who provides details on building dimensions for 59 cotton spinning mills constructed across France between 1789-1845. While Chassagne does not give details for how these mills were chosen, we have been able to trace the source for some. In all cases, they come from notarial archives across different French *départements*. A note of caution is due here: Chassagne's sample of cotton mills is likely biased towards important, large plants for which design records have survived. We discuss how this type of bias may effect our results in footnote 16 in the paper.

We observe the number of floors as well as the dimensions of the factory floor (length and width). A limitation of these data is that they do not contain variables that would allow us to estimate productivity. As a second-best proxy, we examine plant survival. We match the plants to the 1840 census using the name of the owner and the location of the plant.[40] We construct a binary indicator that takes the value 1 if the plant shows up in the 1840 census (which collected plant data in 1839-47).

Results. We regress a binary indicator for plant survival on the number of floors and the squareness of the building (up to a quadratic term). Table A.8 presents the results of this exploratory exercise. The coefficients point to a statistically and economically meaningful relationship between survival and both dimensions of plant layout. Taking the estimated coefficients, the predicted optimal number of floors (i.e., the point where the odds of survival are maximized) is 3.67, and the predicted optimal squareness is $S = 0.49$. Comparing these numbers to Figure 4 in the paper shows that

[40]Chassagne (1991) reports the owner name for each of the 59 plants in his sample, making it straightforward to match this information with the 1840 census.

both are close to what the industry converged to after 1820.[41]

These results on building design suggest that plants were initially experimenting with a wide range of organizational practices, and as the industry matured, they converged to best-practice designs.

Table A.8: Survival and Building Layout

Dep. Var.: Indicator for Survival until 1840

	(1)	(2)
Number of floors	0.154***	
	(0.056)	
Num. floors squared	-0.021**	
	(0.009)	
Percent of encompassing square		1.343**
		(0.624)
Squareness squared		-1.314**
		(0.565)
Predicted optimal	3.67	0.49
R^2	0.06	0.05
N	46	55

Notes: Robust standard errors in parentheses. Number of floors represents the number of floors a building had. 'Squareness' is defined as $S \equiv \frac{length*width}{max\{length,width\}^2}$. Data on the length, width, and number of floors of each building are from Chassagne (1991). "Predicted optimal" corresponds to the value of the respective explanatory variable at which the odds of survival are maximized. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.4 Evidence for the Proposed Mechanism: Strikes in the Three Industries

Data. Our data on strikes are from Shorter and Tilly (1874). They provide information on the strikes occurring in France from 1830 until 1968, including details on the location and industrial sector. In particular, we focus on all strikes that took place until 1847 (the last year of the data collection of our 1840s census) in textile (*industrie textile*), paper (*papiers, cartons, industrie polygraphiques*), and metallurgy (*travail des métaux fins et ordinaires* and *métallurgie*). In total, there are 14 macro-sectors. Textile is one of them, but we do not have more detailed information on whether strikes occurred in cotton spinning or in other textile sectors. The data by Shorter and Tilly (1874) for the 1830-1847 period had originally been digitized by Jean-Pierre Aguet from archival sources of the Interior and Justice Ministries in Paris. He found information on almost 400 strikes,

[41] One potential concern with these results is that plant age may be correlated with both design and the odds to survive until 1840. To address this (at least partially, given the limited sample size), we include a second-degree polynomial in plant age. For squareness, the results are robust: we retain statistical significance, and the predicted optimal squareness is virtually unchanged. For the number of floors, we lose statistical significance – although the coefficients for age and age^2 are themselves statistically insignificant.

but these represent only part of all strikes taking place in the country in this period (likely the larger and most important ones).[42] This notwithstanding, "the Aguet data should reveal accurately the movement of strikes over time and give information on basic structural characteristics of these early conflicts." (Shorter and Tilly, 1874, p.354)

Results. Over the period 1830-47, strikes were more frequent in textile than in the other two sectors: there were 116 strikes in textile, 29 in metallurgy, and 24 in paper. Table A.9 examines this pattern more systematically: we regress log(1+number of strikes) on a dummy for the textile sector. We add the value 1 so that the outcome variable is defined for *département*-industry observations with zero strikes. The result in columns 1 and 2 suggest that, in an average *département*, strikes were 0.38 log points more frequent in textile than in metallurgy or paper milling, even when accounting for *département* fixed effects. An obvious concern is that the larger size of the textile sector is driving these results. In column 3 we thus control for male employment at the sector-*département* level. While the coefficient on textile becomes somewhat smaller in magnitude, it remains statistically highly significant: strikes (per worker) in textiles were approximately 0.30 log points higher than in the two comparison sectors. Finally, in column 4 we also include total manufacturing employment across all sectors in a given *département*, which accounts for possible scale effects (note that we have to drop *département* fixed effects for this coefficient to be identified). Again, we confirm the magnitude and significance of the coefficient on textiles. Overall, this evidence suggests that strike activity in textiles was higher than what one would expect based on the number of workers and *département* characteristics.

[42] The Justice Ministry counted 1,049 prosecutions for *coalition ouvriére* (worker coalitions, which were illegal until 1864). However, many of these were either not officially reported or subsequently considered benign (Shorter and Tilly, 1874).

Table A.9: Strikes in Textile vs. Comparison Sectors

Dependent variable: log(Strikes)

	(1)	(2)	(3)	(4)
Textile	0.377***	0.377***	0.301***	0.305***
	(0.064)	(0.079)	(0.068)	(0.064)
Log(Workers)			0.029**	0.027**
			(0.014)	(0.013)
Log(Total workers – dept)				0.168***
				(0.037)
Department FE		✓	✓	
R^2	0.10	0.68	0.69	0.26
N	258	258	258	258

Notes: Dependent variable is the log number of strikes in each of the three sectors (textile, metallurgy, paper), at the *département* level. There are overall 86 *départements*, but not all report strike data for each sector. 'Textile' is a dummy equal to one for the textile sector. 'Workers' include all workers in the respective sector and *département*. "Total workers - dept' include all workers in the *départements* (including from sectors other than the three used in our analysis). Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.5 Robustness: Spatial Diffusion of Knowledge

This appendix provides robustness checks for the results on spatial diffusion in Section 5.3 in the paper. There, we asked how information about best practice organizational methods diffused through the economy. We suggested that plants copied from successful experimenters and provided evidence supporting the spatial diffusion of knowledge. Figure 5 in the paper plots the estimated coefficients that show that proximity to high-productivity plants mattered in cotton spinning in 1800, while this pattern was much weaker for the later period and for the two comparison sectors. Table A.10 reports the corresponding regressions. Figure A.17 displays the spatial distribution of plants in mechanized cotton spinning, metallurgy, and paper milling, distinguishing those in the 90th percentile of the productivity distribution.

Table A.10: Proximity to High-Productivity Plants

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)	(5)	(6)
	Spinning		Metallurgy		Paper milling	
	1806	1840	1811	1840	1794	1840
$lnDist^{p90}$ (1800)	-0.841***		-0.304***		-0.233*	
	(0.135)		(0.081)		(0.130)	
$lnDist^{p90}$ (1840)		-0.174*		-0.065		-0.183
		(0.101)		(0.079)		(0.130)
Department FE	✓	✓	✓	✓	✓	✓
R^2	0.56	0.15	0.38	0.29	0.27	0.46
N	290	467	385	779	456	309

Notes: The table reports the regression results that are underlying Figure 5 in the paper. For each specification, we report the standardized beta coefficients on $lnDist^{p90}$, which measures the log distance to the closest plant with productivity in the 90th percentile (in the same sector and in the same period – 1800 and in 1840, respectively). The number of observations in these specifications is smaller than the full sample as plants that belong to the 90th percentile are excluded. Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Next, we show that our evidence on the spatial diffusion of knowledge is robust to a series of potentially confounding explanations. While our inclusion of *département* fixed effects across all specifications already captures *département*-level unobserved characteristics, it does not account for unobserved differences at a more disaggregated level. In what follows, we examine potential confounders at the local (e.g., commune) level.

Controlling for location fundamentals. Table A.11 controls directly for some key location fundamentals at the commune level: the availability of fast-flowing streams (as a source of water power), proximity to coal (which mattered for steam power), and the share of forest cover (which mattered

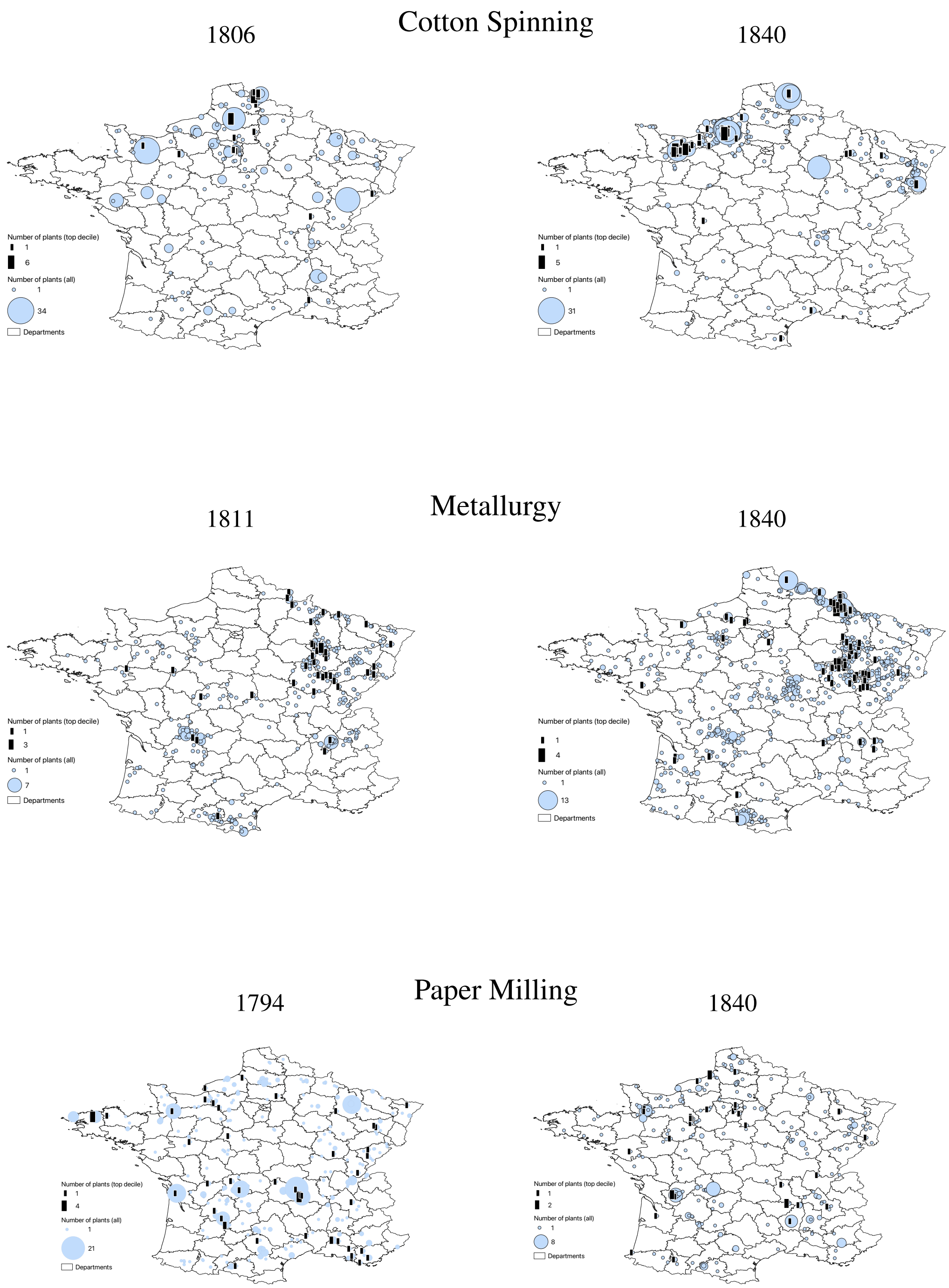


Figure A.17: Spatial Distribution of (High-Productivity) Plants Across France

Note: The figure shows the spatial distribution of plants in cotton spinning (top), metallurgy (middle), and paper milling (bottom). The figure distinguishes plants in the 90th percentile of the productivity distribution (black columns) from all other plants in a commune (light circles).

for access to charcoal – a major input in metallurgy).[43] The results in Table A.11 show that controlling for these location fundamentals does not affect the pattern of the coefficients of interest. The estimated magnitudes remain very similar to those in Table A.10. Moreover, the location fundamentals themselves are mostly small and statistically insignificant. This is probably driven by the fact that the *département* fixed effects already account for the most important spatial differences in location characteristics.

Table A.11: Proximity to High-Productivity Plants – Controlling for Location Fundamentals

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)	(5)	(6)
	Spinning		Metallurgy		Paper milling	
	1806	1840	1811	1840	1794	1840
$lnDist^{p90}$ (1800)	-0.882***		-0.314***		-0.214	
	(0.116)		(0.069)		(0.136)	
$lnDist^{p90}$ (1840)		-0.186*		-0.066		-0.194
		(0.105)		(0.077)		(0.118)
Access to high streamflow	-0.115	0.281**	-0.015	0.181	-0.142	-0.176
	(0.283)	(0.113)	(0.150)	(0.197)	(0.271)	(0.210)
Proximity to coal	-0.011	-0.069	-0.336*	0.036	0.151	-0.213
	(0.201)	(0.330)	(0.195)	(0.156)	(0.406)	(0.259)
Share of forest area	-1.337***	0.356	-0.171	-0.057	0.542	0.330
	(0.484)	(0.337)	(0.279)	(0.334)	(0.516)	(0.892)
Department FE	✓	✓	✓	✓	✓	✓
R^2	0.58	0.16	0.39	0.29	0.28	0.47
N	290	467	376	779	456	309

Notes: The table reports robustness checks of the results in Table A.10. For each specification, we report the standardized beta coefficients on $lnDist^{p90}$, which measures the log distance to the closest plant with productivity in the 90th percentile (in the same sector and in the same period – 1800 and in 1840, respectively). Access to high streamflow is a binary variable that takes the value of one if a plant's nearest data collection point for river discharge has streamflow in the top quartile of the distribution. Proximity to coal is a binary indicator that takes the value of one if a location is within the bottom quartile of plant locations in terms of distance to the nearest coalfield. Share forest area measures the forest area over the total area of the commune where the plant is located (using data on forest cover from the late 18th century). Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Controlling for agglomeration. Another possible concern is that our results may be affected by more general agglomeration externalities, as opposed to learning. In particular, our findings could be driven by high-productivity plants emerging (within *départements*) where the density of produc-

[43] Data sources and the construction of each variable are described in Appendix B.6.

tion was large due to agglomeration forces. To address this possibility, we control for the density of production at the commune level. This is measured as the log of total output in the sector, excluding a plant's own output. Table A.12 shows that controlling for the local density of production barely affects our results. The estimated coefficient on distance to high-productivity plants in cotton spinning in 1800 remains large and highly significant, and also the distance coefficients in the other sectors and in 1840 are essentially the same as in our baseline specification in Table A.10. The coefficient on local production density itself is generally small, positive, and never statistically different from zero.

Table A.12: Proximity to High-Productivity Plants – Controlling for Local Production Density

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)	(5)	(6)
	Spinning		Metallurgy		Paper milling	
	1806	1840	1811	1840	1794	1840
$lnDist^{p90}$ (1800)	-0.771***		-0.320***		-0.203	
	(0.175)		(0.080)		(0.139)	
$lnDist^{p90}$ (1840)		-0.144		-0.083		-0.179
		(0.113)		(0.073)		(0.136)
Production density	0.019	0.007	-0.009	-0.006	0.016	0.003
	(0.021)	(0.013)	(0.013)	(0.008)	(0.015)	(0.019)
Department FE	✓	✓	✓	✓	✓	✓
R^2	0.57	0.15	0.38	0.29	0.28	0.46
N	290	467	385	779	456	309

Notes: The table reports robustness checks of the results in Table A.10. For each specification, we report the standardized beta coefficients on $lnDist^{p90}$, which measures the log distance to the closest plant with productivity in the 90th percentile (in the same sector and in the same period – 1800 and in 1840, respectively). Production density is the log of total output in the sector in a given commune, excluding a plant's own output. Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Persistent unobservables? A placebo check. Next, Table A.13 performs a placebo exercise and examines whether plant productivity around 1800 was also related to the distance to high-productivity plants in the top-90th percentile of productivity in 1840 in the same sector. The estimated coefficient in cotton spinning is close to zero and statistically insignificant, implying that productivity in cotton spinning in 1806 was not related to high-productivity locations more than three decades later. This suggests that our results are not driven by persistent location fundamentals within *départements*. Our results in Table A.13 also imply that it is unlikely that our findings are driven by plant selection into (persistent) high-productivity locations.

Plant selection into the proximity of high-productivity plants? Next, we examine the extent to which

Table A.13: Proximity to High-Productivity Plants – Distance Placebo in 1840

Dependent variable: log(Output per worker)

	Spinning 1806	Metallurgy 1811	Paper milling 1794
	(1)	(2)	(3)
$lnDist^{p90}$ (1840)	-0.053	-0.234**	0.173
	(0.235)	(0.099)	(0.151)
Department FE	✓	✓	✓
R^2	0.55	0.32	0.21
N	321	426	507

Notes: The table reports a placebo check of the results in Table A.10. $lnDist^{p90}$ (1840) measures the log distance to the closest plant in the same sector with productivity in the 90th percentile in 1840. Dependent variable is log output per worker in the earlier period (around 1800). We report the standardized beta coefficients on the distance variables. Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

the estimated distance coefficient in cotton spinning in 1800 may be driven by plant selection. It is possible that we estimate a large (negative) coefficient in cotton spinning not because plants were learning from their high-productivity neighbors but rather because *ex-ante* high-productivity plants selected into locations near existing high-productivity plants. Given that we observe plant age in cotton spinning in 1806, we can examine this potential selection pattern. In Table A.14, we first report our baseline result in column 1 and then compare it to the restricted sample of plants that entered *before* the nearest high-productivity plant. In this subsample (column 2), the coefficient on distance to high-productivity plants remains statistically highly significant, although it is somewhat smaller than in the baseline sample (-0.425, se 0.144).[44] The timing of entry of the plants in this subsample rules out the type of selection described above: Our results cannot be entirely driven by selection of entering plants into locations that already featured high-productivity plants – simply because the latter were not there yet.

In combination, the results in Tables A.13 and A.14 address the possibility of selection both based on persistent location fundamentals and features that may have made locations more attrac-

[44]Note that it is not surprising that the coefficient on distance declines (in absolute value). In order to perform this check, the particularly restrictive subsample in column 2 also excludes plants that are in line with our mechanism: 'younger' plants that did *not* have high ex-ante productivity but instead learned about optimal mill design from nearby high-productivity plants during their construction phase. Since these plants entered after the nearby high-productivity plants, such cases are excluded from the subsample in column 2. Since the restrictive subsample excludes cases that are in line with our mechanism, it arguably biases the distance coefficient downward.

tive over time (i.e., the entrance of a high-productivity plant).

Table A.14: Testing for Spatial Selection in Cotton Spinning in 1806

Dependent variable: log(Output per worker)		
	(1)	(2)
	Baseline	Subsample‡
$lnDist^{p90}$ (1800)	-0.841***	-0.425***
	(0.135)	(0.144)
Department FE	✓	✓
R^2	0.56	0.66
N	290	175

Notes: The table shows that our results for proximity to high-productivity plants (see Figure 5) are not entirely driven by selection of entering plants into the proximity of high-productivity plants. $lnDist^{p90}$ (1800) is the log distance to the nearest plant in cotton spinning with productivity in the 90th percentile in 1800. We report standardized beta coefficients for all variables. Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.
‡ Subsample includes only plants that entered before the nearest high-productivity plant.

Learning across sectors? We examine whether there is evidence consistent with learning across sectors. In particular, we check whether proximity to high-productivity plants in the comparison sectors also mattered for mechanized cotton plants in 1800. Table A.15 shows that there is no consistent pattern in the data. The coefficient on distance to high-productivity metallurgy plants (column 1) is not statistically different from zero and noisily estimated.[45] For paper milling, on the other hand, the distance coefficient is actually positive and also insignificant. These cross-sector results are consistent with the historical record, showing no indication that early cotton spinning mills were able to learn from high-productivity plants in the more mature comparison sectors.

Spillovers from England. Finally, as many innovations in all three sectors (and the spinning machinery per se) were invented in England, it may have been easier to observe and adopt the best organizational practices for firms closer to the channel. This concern is partly addressed by the inclusion of *département* fixed effects in our regressions. Table A.16 also controls for log distance to London. Our results on distance to high-productivity plants are essentially unaffected, and the coefficients on distance to London are mixed – which is unsurprising, given that these are added on top of *département* fixed effects.

[45] A likely reason for the noisy results is that metallurgy and cotton spinning mills were located relatively far from each other (see the maps in Figure A.17). The median cotton spinning mill was located 130 km from its nearest high-productivity peer in metallurgy, but only 57 km (73km) from the nearest high-productivity plant in paper milling (cotton spinning).

Table A.15: Proximity of Cotton Spinning Plants to High-Productivity Plants in Metallurgy and Paper Milling in 1800

Dependent variable: log(Output per worker)

	Spinning-Metallurgy	Spinning-Paper
	(1)	(2)
$lnDist^{p90}$ metal (1800)	-0.410	
	(0.403)	
$lnDist^{p90}$ paper (1800)		0.169
		(0.106)
Department FE	✓	✓
R^2	0.55	0.56
N	321	321

Notes: $lnDist^{p90}$ (1800) measures the log distance of cotton spinning plants to the closest plant in metallurgy (col. 1) and in paper milling (col. 2) with productivity in the 90th percentile in 1800. We report the standardized beta coefficients on the distance variables. Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.16: Proximity to High-Productivity Plants – Controlling for Distance to London

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)	(5)	(6)
	Spinning		Metallurgy		Paper milling	
	1806	1840	1811	1840	1794	1840
$lnDist^{p90}$ (1800)	-0.812***		-0.302***		-0.165	
	(0.122)		(0.082)		(0.105)	
$lnDist^{p90}$ (1840)		-0.229**		-0.073		-0.208*
		(0.102)		(0.080)		(0.123)
Distance to London	0.468	-1.829**	-0.276	0.895	4.224*	-1.144
	(1.959)	(0.756)	(1.053)	(0.840)	(2.293)	(0.921)
Department FE	✓	✓	✓	✓	✓	✓
R^2	0.56	0.15	0.38	0.29	0.30	0.47
N	290	467	385	779	456	309

Notes: The table reports a robustness check of the results in Table A.10. $lnDist^{p90}$ ($\sim$ 1800) and $lnDist^{p90}$ (1840) measure the log distance to the closest plant in the same sector with productivity in the 90th percentile in 1800 and in 1840, respectively. We report the standardized beta coefficients on $lnDist^{p90}$. Distance to London is the log distance to London (UK). Standard errors (clustered at the *département* level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.6 Plant Survival, Exit, Age Profile, and Productivity

This appendix complements Section 5.4 in the paper, where we compared plant survival rates in the three sectors in order to distinguish learning about the new technology itself from learning about optimal plant design. Here, we examine alternative drivers of the differential survival rates, and we provide additional complimentary evidence for building design challenges as a mechanism.

Power sources in the three sectors as a confounding factor? In the main text, we showed that the survival rate in spinning was lower than in the comparison sectors. This is consistent with a mechanism in which entrepreneurs who invested in cotton spinning mills with poor layout had to exit the market, and the mill was not subsequently used by other cotton spinning entrepreneurs. However, it could also be driven by the cotton industry adopting steam power (and moving away from water power) more than the other sectors. The summary statistics (Tables A.2-A.4) suggest that this was not the case: Even in spinning, water remained the prominent source of power until the end of our sample period in 1840: 66% of cotton spinning plants still used water power, as compared to 64% in metallurgy and 85% in paper milling. The enduring dependence on water power is a well-known aspect of the French setting (see Cameron, 1985, for a discussion). Moreover, Table A.17 shows a *negative* association between labor productivity and the use of steam power in all three sectors. This confirms that in France, plants did not face a strong profit incentive to move away from water power (Cameron, 1985).

Table A.17: Productivity and the Use of Steam Power in Cotton Spinning (1840)

Dependent variable: log(Output per worker)

	(1)	(2)	(3)	(4)
Steam power	-0.090**		-0.082*	-0.087*
	(0.046)		(0.048)	(0.046)
Water power		0.060	0.017	
		(0.050)	(0.053)	
Other power				0.144
				(0.142)
R^2	0.01	0.00	0.01	0.01
N	528	528	528	528

Notes: Water power, steam power, and other (wind or animal) power are binary indicators equal to one for plants using the respective source of power. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Productivity handicap of exiting plants. Did plants that exit indeed have particularly low productivity? Table A.18 examines whether plants that eventually exited the market by 1840 had lower initial productivity around 1800, as compared to surviving plants. This pattern is particularly strong in cotton spinning, consistent with the large exit rates in the sector that we documented in Section 5.4. Exiting cotton plants were 46% less productive than survivors, and this difference is statistically significant. This pattern is much less pronounced in the comparison sectors: Exiting plants were about 15% less productive in metallurgy, and 6% less productive in paper milling.

Table A.18: Productivity of Exiting Relative to Surviving Plants

Dependent variable: log(Output per worker) around 1800			
	(1)	(2)	(3)
	Spinning 1806	Metallurgy 1811	Paper Milling 1794
Exit dummy	-0.458**	-0.145*	-0.055
	(0.205)	(0.085)	(0.137)
R^2	0.003	0.001	0.001
N	340	470	520

Notes: Exit is a dummy variable equal to one for plants that existed in the initial period and that had exited the market by 1840 (based on the baseline survival rate – see Section 3.2). In cotton spinning, there were 340 plants in 1806 with information on output and labor, and 317 of these had exited by 1840. In metallurgy, there were 470 plants with data to compute productivity in 1811, and 293 had exited by 1840. In paper milling, there were 520 plants with information on output and labor in 1794, 464 of which had exited by 1840. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Plant age and productivity. In Section 5.4 in the paper we also examined productivity over the plant age profile as a second piece of evidence that points to organizational methods as a mechanism behind the lower-tail bias of productivity. We documented that younger plants in mechanized cotton spinning were significantly more productive in 1806. Here, we investigate the productivity-age pattern in 1840. While the data for this second period are generally more comprehensive, we do not observe plant age. However, we can perform a similar – albeit weaker – test based on the comparison of surviving and entrant plants: In Table A.19 we regress log output per worker on an indicator for whether the plant was an 'entrant' in 1840 (as opposed to a surviving plant by our definition from Section 3.2). The coefficient on the 'entrant' dummy thus reflects the average productivity differential for plants that entered between the initial survey year (1806) and 1840. Best-practice mill design evolved over this period, and it had largely converged by 1840 (Pollard, 1965). Correspondingly, we find that 'young' plants were not more productive; the coefficient is small and statistically insignificant. Columns 2-5 show that this holds also when we control for the use of water power, steam power, any other power source (wind or animal power used by a small subset of plants), and for the number of workers.

Next, we investigate whether a similar pattern holds in metallurgy. Tables A.20 and A.21

Table A.19: Cotton Spinning in 1840: Productivity and Plants' Age Profile

Dependent variable: log(Output per worker) in 1840

	(1)	(2)	(3)	(4)	(5)
Entrant 1840	0.039	0.029	0.029	0.036	0.013
	(0.201)	(0.206)	(0.199)	(0.201)	(0.201)
Water power		0.060			
		(0.050)			
Steam power			-0.090*		
			(0.046)		
Other power				0.168	
				(0.140)	
log(Workers)					-0.153***
					(0.027)
R^2	0.00	0.00	0.01	0.00	0.06
N	528	528	528	528	528

Notes: The table shows that in 1840, when mechanized cotton spinning technology had reached maturity, new entrant plants did not have a productivity advantage over existing plants anymore. Entrant 1840 is a binary indicator equal to one for plants that entered the market after 1806. For this definition, we only use surviving plants that are linked based on commune and owner name (as this is almost always a one-to-one match). Water power, steam power, and other (wind or animal) power are binary indicators equal to one for plants using the respective source of power. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

examine the relationship between age profile and productivity for metallurgy plants. For this sector, in both periods, the best measure of plant age that we observe is a binary indicator of plant survival from 1788 to 1811, and from 1811 to 1840 (the latter being the same procedure as for cotton spinning in Table A.19). Thus, in 1811, we define a plant as 'young' if the survey's recall data do not report existence in 1788. The results do not point to younger plants in metallurgy having a strong productivity advantage. While column 1 in Table A.20 shows a positive raw correlation between 'young' metallurgy plants in 1811 and labor productivity, the coefficient becomes smaller in magnitude and statistically insignificant once we control for plant size (column 2).[46] For our second comparison sector, paper milling, information on plant age or recent entry is not available for the early period. We thus cannot perform the comparison for this sector.

Table A.20: Metallurgy in 1811: Productivity and Plants' Age Profile

Dep. variable: log(Output per worker) in 1811		
	(1)	(2)
Young 1811	0.201*	0.081
	(0.115)	(0.115)
log(Workers)		-0.298***
		(0.044)
R^2	0.01	0.09
N	470	470

Notes: Young 1811 is a binary indicator equal to one for plants that entered the market after 1788. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Finally, Table A.21 examines 'young' (entrant) metallurgy plants in 1840. In this later period, 'young' plants actually had a somewhat *lower* productivity, although this difference is not statistically significant.

[46]Note that the metallurgy survey has sparser information on this dimension; plant size is the only control that can be added in 1811.

Table A.21: Metallurgy in 1840: Productivity and Plants' Age Profile

Dependent variable: log(Output per worker) in 1840

	(1)	(2)	(3)	(4)	(5)
Entrant 1840	-0.109	-0.018	-0.106	-0.123	-0.025
	(0.119)	(0.120)	(0.120)	(0.118)	(0.106)
Water power		0.324***			
		(0.063)			
Steam power			-0.072		
			(0.077)		
Other power				-0.239***	
				(0.091)	
log(Workers)					-0.351***
					(0.028)
R^2	0.001	0.029	0.001	0.006	0.201
N	896	896	896	896	896

Notes: Entrant 1840 is a binary indicator equal to one for plants that entered the market after 1811. For this definition, we only use surviving plants that are linked based on commune and owner name (as this is almost always a one-to-one match). Water power, steam power, and other (wind or animal) power are binary indicators equal to one for plants using the respective source of power. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.7 Robustness: Region Fixed Effects

In Section 5.5, we examined a set of alternative explanations that may account for the lower-tail bias observed in mechanized cotton spinning. One key robustness check studies the extent to which our main result holds *within* regions. Table A.22 shows that the lower-tail bias, while more muted, remains striking when we add fixed effects for 22 regions in France. This suggests that regional fundamentals, institutions, market access, or differential access to input markets do not account for our results.

Table A.22: Productivity Growth by Quantiles – Controlling for Region FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	2.029***	2.941***	2.352***	1.987***	1.943***	1.668***	840
	(0.158)	(0.455)	(0.207)	(0.168)	(0.214)	(0.191)	
Region FE	✓	✓	✓	✓	✓	✓	
Metallurgy (1811-1840)	2.012***	2.038***	1.730***	1.879***	2.006***	2.137***	1,289
	(0.179)	(0.222)	(0.222)	(0.161)	(0.130)	(0.194)	
Region FE	✓	✓	✓	✓	✓	✓	
Paper milling (1794-1840)	0.776***	1.009***	0.746***	0.731***	0.616***	0.731***	850
	(0.118)	(0.137)	(0.122)	(0.096)	(0.127)	(0.213)	
Region FE	✓	✓	✓	✓	✓	✓	

Notes: The table reports the average annual productivity growth (in %) between the initial sample period (around 1800) and 1840 for the three sectors, as well as annual productivity growth estimated at various quantiles. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.8 Robustness: Market Integration

Can market integration explain the disappearance of the lower tail in cotton spinning after 1800? In the previous appendix section, we already showed that our main finding is robust to the inclusion of region fixed effects. This also addresses – at least to some extent – the concern that market integration could confound our results. Here, we present several additional pieces of evidence. In particular, for market integration to be the main driver of the lower-tail bias in cotton spinning, market access would have had to increase *disproportionately* in this sector between 1800 and 1840, relative to the comparison sectors. We present data that suggest the opposite. Figure A.18 uses data in 1794 from Daudin (2010) on within-country trade linkages across French districts by industry.[47] For each *département*, we sum the number of districts across France that reported consuming

[47] Districts were administrative units that stayed in place only for a short period between the French Revolution and 1800, when they were replaced by *arrondissements*. Each *département* included from a minimum of 3 to a maximum of 10 districts.

products (e.g., cotton textiles) produced in that *département*. The numbers in Figure A.18 show the count of districts that reported consuming a given product from the *département*.

Intuitively, higher market integration means lower price differentials across *départements*, which in turn implies that highly productive areas could dominate the market throughout France. Consequently, we can infer high market integration from the data if we observe that a few (presumably highly productive) *départements* sold to many other *départements*, while the majority of *départements* produced no output, or did so only for local consumption. Figure A.18 shows that this pattern is particularly strong in cotton textiles. Many *départements* produced mostly for themselves if at all (these are the zeros and small, positive numbers), while a few *départements* supplied cotton textiles to a large number of districts. The top tercile of *départements* exported cotton textiles to 30 or more districts. In the two comparison sectors, there is less specialization and less evidence for market integration: Fewer *départements* report not supplying to anyone (particularly in paper), and the top decile of *départements* supplied only to 6 (paper) and 7 (iron) districts in total. This suggests that cotton textile plants were already competing in a bigger market than the comparison sectors around 1800.

Given that cotton spinning already started off with more integrated markets, we would expect – if anything – that further market integration after 1800 played a *smaller* role than in our comparison sectors. This renders it unlikely that relatively tougher competition in cotton led to the disappearance of the lower tail.

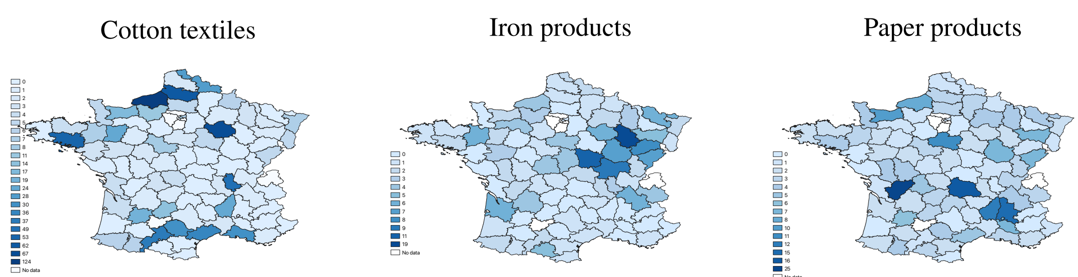


Figure A.18: Market Integration for the Three Sectors in 1794

Notes: Data source: Daudin (2010). The figure shows the extent of market integration in cotton textiles (left), iron (middle), and paper products (right). Market integration is measured as the number of districts (see footnote 47) across France that reported consuming cotton textiles, iron, or paper products from districts in the given *département*. A higher number for a *département* indicates that firms from that *département* sold their products to many distinct locations across France. Data are from Daudin (2010).

Next, in Table A.23 we control directly for market access. We construct two measures of market access: i) within France and ii) across Europe. Both measures are computed as the inverse distance-weighted sum of urban populations in 1800 (see Appendix B.6 for detail). We begin in

Panel A of Table A.23 by controlling for the log of market access within France. While we find that market access in France is correlated with productivity in cotton spinning, this relationship is relatively stable across the productivity distribution: Plants in the lowest decile of the productivity distribution benefited just as much from market access as those in the top decile. Thus, market access in France is unlikely to have had differential effects on low- vs. high-productivity plants. In addition, when adding market access as a control variable, the lower tail-bias in cotton spinning remains very similar, confirming that market integration does not confound our results.

To account for access to *foreign* markets, we perform two additional exercises. First, Panel B in Table A.23 controls for market access across Europe. Second, Panel C in Table A.23 includes a dummy for *départements* located on the coast (either the Channel, Mediterranean or Atlantic) as a proxy for access to foreign markets. In both Panels B and C we find very similar results as in Panel A: While foreign market access is also associated with higher productivity in cotton spinning, this relationship is relatively flat over the productivity distribution, and the inclusion of market access does not change our main result: The lower-tail bias remains substantial. In addition, in unreported results we also confirmed that the productivity growth patterns in our comparison sectors are robust to controlling for the three measures of market access.

Table A.23: Productivity Growth by Quantiles – Controlling for Market Access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
PANEL A: Controlling for Market Access in France							
Spinning (1806-1840)	2.569***	3.973***	3.126***	2.317***	1.996***	1.533***	840
	(0.157)	(0.266)	(0.208)	(0.176)	(0.155)	(0.305)	
Market access, France	0.718***	0.774***	0.760***	0.587***	0.664***	0.838***	
	(0.116)	(0.199)	(0.113)	(0.129)	(0.105)	(0.189)	
PANEL B: Controlling for Market Access in Europe							
Spinning (1806-1840)	2.535***	3.825***	3.073***	2.333***	1.863***	1.357***	840
	(0.154)	(0.232)	(0.219)	(0.154)	(0.144)	(0.276)	
Market access, Europe	1.252***	1.335***	1.321***	0.966***	1.040***	1.623***	
	(0.205)	(0.223)	(0.211)	(0.204)	(0.206)	(0.194)	
PANEL C: Controlling for Coastal *Départements*							
Spinning (1806-1840)	2.225***	3.699***	2.965***	2.315***	1.419***	0.809***	840
	(0.149)	(0.237)	(0.201)	(0.127)	(0.197)	(0.232)	
Coastal Department	0.810***	0.975***	1.010***	0.520***	0.736***	0.909***	
	(0.129)	(0.216)	(0.138)	(0.120)	(0.154)	(0.214)	

Notes: The table reports the average annual productivity growth (in %) between the initial sample period (around 1800) and 1840 for cotton spinning, as well as annual productivity growth estimated at various quantiles. Market access in France and Europe are computed as described in Appendix B.6. Access to overseas markets is a dummy equal to one for *départements* located on the coast. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.9 Robustness: Napoleonic Blockade and Napoleonic Wars

The period of our analysis coincides with the French Revolutionary and Napoleonic Wars (1792-1815). Here, we discuss potential channels through which these events may affect our results.

Productivity. By disrupting trade between Britain and France, the Napoleonic Blockade (1806-14) affected the spatial composition of economic activity in France (Juhász, 2018). This in turn may have affected plant-level productivity. While our robustness checks using region fixed effects partially addressed this concern, we now examine it in more detail. First, Figure A.19 splits the sample into plants in the northern and southern regions of France (corresponding to the main dimension along which protection varied). It shows that the productivity distributions in cotton spinning in the two areas are remarkably similar, suggesting that varying trade protection does not drive our results.

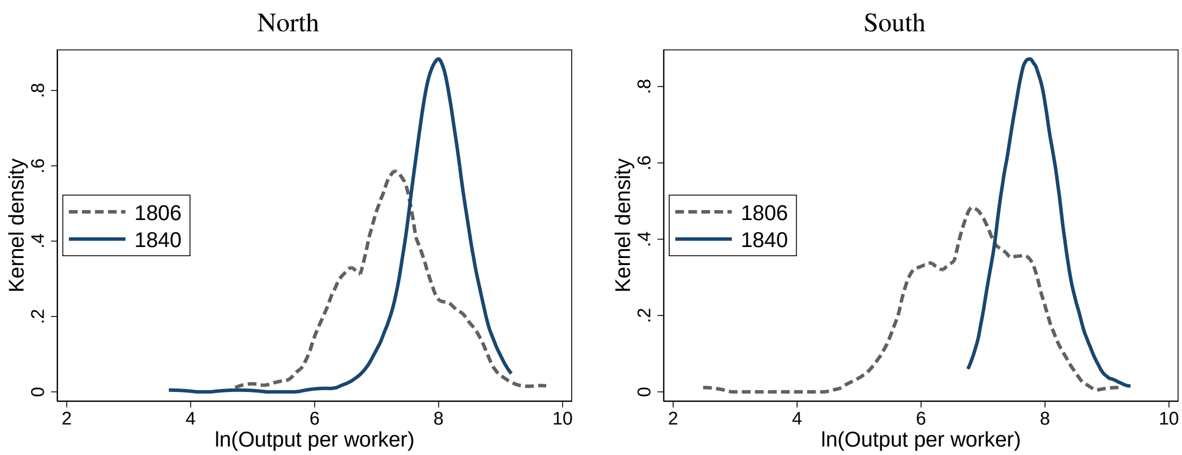


Figure A.19: Spinning: Productivity growth in the 'North' and 'South' of France

Notes: The figure shows the productivity distribution of cotton spinning plants in 1806 and 1840, separating France into North and South. Northern plants are those located in communes with above-median latitude. Southern communes are those located in below-median latitude communes. Median latitude is time invariant and defined based on the sample of 1806 plants.

Plant survival. Could the blockade explain the lower plant survival rates observed in cotton spinning relative to the other sectors? In Table A.24 we split mechanized cotton spinning plants into the same northern and southern regions and report survival rates separately. Indeed, consistent with Juhász (2018), survival rates are lower in the south than in the north (which experienced a relative increase in trade protection during the blockade). However, survival rates remain low relative to the other two sectors even when we narrow our sample to only northern plants.[48] That is, even

[48]Note that the restricted sample survival rate for northern cotton spinning plants is below, but close to that observed in paper milling (22% and 24.6%, respectively). Recall, however, that the cotton spinning survey was conducted much later than paper milling (1806 as opposed to 1794).

comparing only cotton spinning plants in the north of France to the comparison sectors (calculated for all of France), survival rates for northern cotton spinning plants were low. This suggests that an important part of low survival rates in mechanized cotton spinning are not driven by the uneven effects of the Napoleonic blockade across France.

Table A.24: Plant Survival Rates in Cotton Spinning in the North and South of France

	'North'	'South'
Survival rate	6.8%	2.7%
Number of plants	192	148
Restricted sample survival rate	22%	5.9%
Number of plants	36	51

Notes: The "survival rate" is defined as the percentage of plants from the initial period that survived to the later period based on matching either by name or location (see Section 3.2 for detail). The "restricted sample survival rate" adjusts for the fact that different sectors have single-plant communes to a varying degree. It is based on the subset of plants located in communes that had only one plant in the initial period. Northern communes are the set of plants with above-median latitude in 1806. Median latitude is time invariant and defined on the sample of 1806 plants.

We note that France also experienced high raw cotton prices (the input to producing yarn) during the blockade (which ended in 1814) as documented in Juhász (2018). However, high input prices are unlikely to explain the high exit rate of plants in the sector for the period 1806 – 1840, as mechanized cotton spinning activity increased dramatically at the aggregate level during the period of the Napoleonic Blockade (Juhász, 2018).

Napoleonic Wars. The Napoleonic Wars (1803-15) may have affected production in France over our initial sample period. However, to confound our results, war should have tilted the productivity distribution in cotton spinning disproportionately, relative to metallurgy and paper milling. Such a lower-tail bias could have been generated by war hitting some firms' production more than others' (by affecting inputs, the actual production, or output markets). We address this concern in several steps. First, region fixed effects partly accounts for this, as the effects of war would likely have been similar for firms in the same region.

Second, we use data on conscripts per capita by region during the Napoleonic wars (see Appendix B.6 for data construction details). As the Napoleonic Wars started in 1803, plants in cotton spinning and metallurgy may have been affected already in the early period – and plants in all three sectors may have still suffered the consequences of the wars in the 1840s. Thus, in Table A.25, we check the correlation between conscripts per capita and productivity in each of the three sectors in 1800 (columns 1-3) and in 1840 (columns 4-6). The coefficients are small and statistically insignificant throughout the different specifications – except for metallurgy in 1840, where the coefficient is marginally significant.[49] Overall, these results suggest that the conscription of soldiers does not confound our results.

Table A.25: Productivity in the Three Sectors and Conscripts During the Napoleonic Period

Dependent variable: log(Output per worker)

	1800			1840		
	Cotton	Metal	Paper	Cotton	Metal	Paper
	(1)	(2)	(3)	(4)	(5)	(6)
log(Conscripts pc)	-0.061	0.030	-0.079	0.041	0.366*	0.020
	(0.306)	(0.202)	(0.084)	(0.061)	(0.203)	(0.155)
N	321	426	507	519	863	343

Notes: log(Conscripts per capita) is the log of the number of men conscripted during the Napoleonic Wars in each French region per 1,000 inhabitants. Standard errors (clustered at the regional level) in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Third, while until 1813, battles in the Napoleonic Wars were fought outside French soil, the final battles of the Sixth Coalition (1813-14) and of the Hundred Days War (culminating in Napoleon's defeat at Waterloo) took place on French territory. We thus collect data on all battles fought in the 1813-15 period from Wikipedia and construct a dummy (*Near Battles*) equal to one for plants located within 10km of a battlefield (see Appendix B.6 for sources and data construction). Because these battles all took place after the initial survey years, we only examine their relationship with productivity in 1840. Table A.26 shows the relationship between productivity in 1840 in the three sectors and our *Near Battles* dummy. The coefficients for mechanized cotton spinning and metallurgy are quantitatively small and statistically insignificant. The positive and significant coefficient for paper milling is driven by five particularly productive plants that were within the 10km radius of battles. Excluding these from the paper milling sector does not affect the productivity growth pattern in paper milling (results available upon request).

[49] While we refrain from over-interpreting this one positive coefficient, one possible explanation is that demand for metal could have been higher in areas with higher conscription because military equipment also had to be provided to the soldiers.

Table A.26: Proximity to Battles during the Napoleonic Wars and Productivity in the Three Sectors

Dep. variable: log(Output per worker), 1840			
	Cotton	Metal	Paper
	(1)	(2)	(3)
Near Battle	0.005	0.023	1.146***
	(0.160)	(0.160)	(0.365)
N	528	896	347

Notes: Near Battle is a dummy equal to one for firms located within 10km of a battlefield during the Napoleonic Wars. Standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.10 Robustness: Excluding Small Plants and those that Used Spinning Jennies

This appendix section complements the discussion in Section 5.5, under 'early spinning workshops.' Table A.27 shows that the our results are robust to using only larger plants with more than 10 workers, in both periods and in all three sectors. The magnitudes remain similar to those in our baseline specification (Table 1), and the lower-tail bias of productivity growth remains unique to cotton spinning plants.[50]

Table A.28 implements an even more conservative approach, dropping all plants that report using a spinning jenny (even if they also used other, newer vintages of capital). This drops overall 76 cotton spinning plants in 1806. The lower-tail bias of productivity growth in cotton spinning remains striking.

[50]Dropping small plants (with fewer than 10 employees) mostly affects the distribution in 1806, because small plants were not supposed to be enumerated in the 1840 survey (the very problem that this robustness check is designed to address). We drop 84 out of 340 plants in 1806 and only 7 out of 528 plants in 1840. All of these small plants have below-median productivity in the respective year, and about 25% of the small plants in 1806 are in the lowest decile of the productivity distribution. Thus, dropping small plants shifts mass from the bottom to the top of the distribution in 1806, but has little effect on the distribution in 1840, due to the small number of plants that are dropped in that year. Consequently, growth in the top-90th percentile of the productivity distribution is less pronounced in Table A.27 as compared to our baseline results in Table 1 in the paper. In other words, our core result – the lower-tail bias of productivity growth in cotton spinning – is actually *stronger* when we drop small plants from the sample.

Table A.27: Productivity Growth by Quantiles – Plants with at Least 10 Workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	2.261***	3.917***	3.191***	2.179***	1.612***	0.309	777
	(0.177)	(0.227)	(0.258)	(0.170)	(0.240)	(0.292)	
Metallurgy (1811-1840)	2.418***	2.073***	2.546***	1.981***	2.442***	2.558***	967
	(0.231)	(0.567)	(0.374)	(0.308)	(0.226)	(0.300)	
Paper milling (1794-1840)	1.265***	1.012***	1.146***	1.369***	1.506***	1.312***	511
	(0.137)	(0.223)	(0.150)	(0.116)	(0.172)	(0.333)	

Notes: The table reports the average annual productivity growth (in %) between the initial sample period (around 1800) and 1840 for the three sectors, as well as annual productivity growth estimated at various quantiles. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table A.28: Productivity Growth by Quantiles – Excluding Plants that used Spinning Jennies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	1.983***	3.165***	2.678***	2.097***	1.213***	0.383	792
	(0.169)	(0.303)	(0.256)	(0.187)	(0.228)	(0.248)	

Notes: The table reports the average annual productivity growth (in %) between 1806 and 1840 in cotton spinning, as well as annual productivity growth estimated at various quantiles. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.11 Robustness: Controlling for Plant Scale

As mentioned in Section 5.5 in the paper, we address the concern that our results may be driven by increasing plant scale. For this reason, Table A.29 controls for the log number of workers at the plant level in all sectors and the two periods. Our results continue to hold: The lower-tail biases remains strong and unique to cotton spinning.

Table A.29: Productivity by Quantiles – Controlling for Number of Workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	2.427***	3.941***	3.427***	2.292***	1.836***	0.974***	868
	(0.162)	(0.231)	(0.243)	(0.165)	(0.185)	(0.304)	
log(Workers)	-0.006	-0.073	-0.072	-0.048	-0.169**	-0.257**	
	(0.063)	(0.092)	(0.090)	(0.058)	(0.071)	(0.115)	
Metallurgy (1811-1840)	3.079***	3.501***	2.859***	2.659***	3.195***	2.855***	1,366
	(0.176)	(0.407)	(0.283)	(0.214)	(0.193)	(0.195)	
log(Workers)	-1.171***	-1.304***	-1.211***	-1.191***	-1.053***	-0.994***	
	(0.080)	(0.133)	(0.092)	(0.098)	(0.085)	(0.090)	
Paper milling (1794-1840)	0.807***	0.670***	0.664***	0.915***	0.897***	1.554***	867
	(0.128)	(0.141)	(0.157)	(0.117)	(0.160)	(0.254)	
log(Workers)	-0.106*	0.157**	0.076	-0.049	-0.169**	-0.469***	
	(0.059)	(0.066)	(0.073)	(0.050)	(0.071)	(0.097)	

Notes: The table reports the average annual productivity growth (in %) between the initial sample period (around 1800) and 1840 for the three sectors, as well as annual productivity growth estimated at various quantiles. We control for the log number of workers at the plant level across all specifications. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.12 Robustness: Capital Deepening

Table A.30 examines whether capital deepening can account for our results. It controls for the capital-labor ratio in cotton spinning plants (measured as the log of the number of spindles per employee at the plant level). The table shows that the lower-tail bias of productivity growth remains robust and similar in magnitude to the pattern in the baseline specifications.

Table A.30: Spinning Productivity by Quantiles – Controlling for Capital Deepening

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	1.960***	3.555***	2.930***	1.966***	1.254***	0.755***	868
	(0.167)	(0.267)	(0.247)	(0.178)	(0.190)	(0.281)	
log(Spindles per worker)	0.522***	0.374***	0.467***	0.389***	0.379***	0.542***	
	(0.075)	(0.082)	(0.085)	(0.068)	(0.090)	(0.141)	

Notes: The table reports the average annual productivity growth (in %) between 1806 and 1840 in cotton spinning (column 1), as well as annual productivity growth estimated at various quantiles (cols 2-6). Column 7 reports the number of observations. K/L is the capital-labor ratio in the plant, measured as the log of the number of spindles per employee.) Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

E.13 Robustness: Quality Upgrading

Finally, we also considered quality upgrading as a potential confounder in Section 5.5. Table A.31 addresses the concern that even if quality is not directly driving the lower-tail bias, it could still affect it indirectly through higher sales and larger plant size. To address this, we estimate quantile regressions using prices *not* adjusted for quality in 1806, *and* controlling for plant size using the number of employees. The lower-tail bias of productivity growth in cotton spinning remains strong.

Table A.31: Annual Productivity Growth (in %) at Different Parts of the Distribution—Using prices not quality-adjusted and controlling for number of workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average	At the following quantiles:					N
		0.1	0.25	0.5	0.75	0.9	
Spinning (1806-1840)	2.517***	3.488***	2.886***	2.420***	1.994***	1.657***	868
	(0.143)	(0.295)	(0.223)	(0.198)	(0.141)	(0.218)	
log(Workers)	-0.140**	-0.055	-0.127	-0.210***	-0.200***	-0.211**	
	(0.058)	(0.103)	(0.087)	(0.068)	(0.057)	(0.090)	

Notes: The table reports the average annual productivity growth (in %) between 1806 and 1840 in cotton spinning, as well as annual productivity growth estimated at various quantiles. Column 7 reports the number of observations. Robust standard errors in parentheses. Notation for statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

F Summary of Evidence

Evidence for Learning About Organization of Factory-Based Production	
We argue that learning about the efficient organization of factory-based production can explain the lower-tail bias of productivity growth in mechanized cotton spinning. Here we summarize the evidence in line with our proposed mechanism.	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
Learning About Best-Practice Methods	
H: Section 2.2	Section 2.2 documents how early adopters of the factory system in mechanized cotton spinning needed to engage in trial and error along multiple dimensions related to the organization of spinning mills. This process led to the development of best-practice methods for operating the new technology efficiently.
Learning About Optimal Mill Design	
E: Figure 4 and Table A.8	Figure 4 shows that there was large variation in the layout of cotton mills and the number of floors in cotton spinning plants in the early period. Best practice converged to cotton mills with a more rectangular shape and around 3.5 floors. Table A.8 shows that there is a systematic relationship between these design elements of cotton mills and a proxy for productivity: plant survival.
Learning About Management Challenges and Strikes	
E: Table A.9	Using strikes as proxy for labor management challenges, Table A.9 shows that strikes were more frequent in textiles than in the comparison sectors.
Spatial Diffusion of Knowledge	
E: Figure 5 and Table A.10	Figure 5 and Table A.10 show that plants located closer to high-productivity plants (in the same sector) were themselves more productive. This relationship is strong only for cotton spinning plants and only during the initial period of technology adoption. This evidence is in line with spatial spillovers of knowledge in mechanized cotton spinning in 1800.
Plant Survival Across Sectors	
E: Table 2	Table 2 shows that plant survival rates in mechanized cotton spinning were lower than in our comparison sectors. This is consistent with a mechanism in which owners of a cotton spinning mill faced considerable challenges along the organizational dimension of factory design. Those that invested in mills with poor layout had to exit the market.

Evidence for Learning About Organization of Factory-Based Production (ctd.)	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
	Plant Exit and Productivity
E: Table A.18	Table A.18 shows that exiting plants in mechanized cotton spinning were much less productive than those that survived. In the comparison sectors, the productivity handicap of exiting plants is also present, but less pronounced. In other words, early cotton plants that 'got it wrong' were particularly unproductive, which can explain the fat lower tail of productivity in this sector. These plants eventually exited the market, and for many of them, the same building was not used by another plant in the industry. This pattern is consistent with large organizational challenges and low initial guidance in switching to factory-based production in cotton spinning.
	Age Profile of Plant Productivity
E: Tables 3, A.19, A.20, and A.21	Table 3 shows that around 1800, cotton spinning plants that entered the market later had higher productivity. This is in line with the argument that knowledge about the optimal organization of mechanized cotton spinning diffused slowly. On the other hand, if learning about technology was the dominant dimension, older plants that had accumulated more experience should have had a productivity advantage. In addition, we show that younger spinning plants were not more productive in 1840 (Table A.19) and that in metallurgy, young plants were as productive as older ones in both periods (Tables A.20 and A.21). This is consistent with the idea that later entrants in mechanized cotton spinning could draw from a better pool of knowledge about organizing production. This process was muted in cotton spinning in 1840, when best practice had diffused, and in metallurgy, where plant-based production methods had been established much earlier.

Alternative Explanations for the Spatial Diffusion of Knowledge	
We presented evidence for the spatial diffusion of organiziational knowledge in mechanized cotton spinning: Plants located closer to other high-productivity plants were themselves more productive, and this relationship is strong only for cotton spinning plants and only during the initial period of technology adoption. Here we show that this result holds when performing a series of robustness checks.	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
Location Fundamentals	
E: Table A.11	Our results could be affected by prominent location fundamentals, not captured by *département* fixed effects (such as the availability of fast-flowing streams, proximity to coal, or the share of forest cover). Table A.11 controls for these factors and shows that the pattern of proximity to high-productivity plants holds.
Agglomeration Externalities	
E: Table A.12	Our findings could be driven by high-productivity plants emerging (within *départements*) where the density of production was large due to agglomeration forces. Table A.12 controls for the local density of production and shows that our results hold.
Unobserved Location Fundamentals	
E: Table A.13	If there are unobserved location fundamentals (within *départements*) not captured by our controls (in Tables A.11 and A.12), they could still confound our results. Table A.13 performs a placebo exercise and studies whether plant productivity in 1800 was related to distance to plants in the top-90th percentile of productivity in 1840 (i.e., plants that emerged later). The coefficient is close to zero and statistically insignificant, suggesting that our results are not driven by persistent unobserved location fundamentals within *départements*.
Distance to London	
E: Table A.16	As many innovations in all three sectors were developed in England, firms closer to the channel may have been able to better observe British practices. This concern is partly addressed by the inclusion of *département* fixed effects in our regressions. Table A.16 also controls for distance to London. The results are virtually unchanged.
Plant Selection	
E: Table A.14	Another concern is that ex-ante high-productivity plants selected into 'productive locations' (i.e., chose to locate near existing high-productivity plants). Table A.14 shows that the results hold when limiting to the subsample of plants that entered *before* the nearest high-productivity plant.
Learning Across Sectors	
E: Table A.14	Did spatial diffusion of knowledge occur across sectors? Table A.15 shows that this was not the case, suggesting that early cotton spinning mills were unlikely to learn from high-productivity plants in the more mature comparison sectors.

Alternative Mechanisms	
We provided evidence that the lower-tail bias in mechanized cotton spinning was due to the reorganization of production. Here, we test for alternative (potentially confounding) mechanisms. Importantly, factors that affected the comparison sectors in similar ways (e.g., economy-wide trends, the introduction of innovations during our sample period, or improvements in power sources) are unlikely to explain our findings. Thus, we consider confounders that are either specific to cotton spinning or that may have affected this sector differentially.	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
A. Regional Differences	
E: Table A.22	Different growth potential of French regions, access to better domestic suppliers of machines, or plants sorting into areas with better location fundamentals (such as fast-flowing streams) could potentially be driving our results. Similarly, wars and revolts (we deal explicitly with the various effects of the Napoleonic wars below) affected some regions of France more than others. Table A.22 performs the quantile regressions including region fixed effects and shows that our results are robust to using only within-region variation.
A.1 Market Integration	
E: Figure A.18, and Table A.23	As the French economy became more integrated, lower-productivity firms may have faced stronger competition and exited the market. However, this can only explain our results if market integration affected mechanized cotton spinning differentially, as we do not observe the lower-tail bias in the comparison sectors. Figure A.18 shows the extent of market integration in the three sectors and suggests that cotton plants were already competing in a larger market around 1800. In addition, Table A.23 performs the quantile regressions for mechanized cotton spinning controlling for different measures of market potential (market access within France, market access in Europe, and access to overseas market). The results on the lower-tail bias hold.
A.2 Machine Quality	
H: Section 2.2 and Appendix Section A.2	The quality of machines available to producers could potentially explain our findings. However, the historical evidence (reported in Section 2.2 and Appendix Section A.2) suggests that machine production and maintenance was external to most plants. This suggests that plants within the same region had access to the same suppliers. As our results hold when using within-region variation, it is unlikely that differential access to machine producers can account for the lower-tail bias.
A.3 The Napoleonic Blockade	
E: Figure A.19 and Table A.22	As shown by Juhász (2018), varying trade protection during the Napoleonic Blockade affected the location of mechanized cotton spinning plants. This raises the concern that the blockade may explain the lower-tail bias. Table A.22 shows that the results hold within region, where the pattern of protection was very similar. In addition, Figure A.19 splits the sample into northern and southern regions (the main dimension along which protection varied) and shows that in both regions, productivity growth until 1840 was due to a disappearing lower tail.

Alternative Mechanisms (continued)	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
A.4 Napoleonic Wars	
E: Tables A.22, A.25, and A.26	The Napoleonic Wars may have affected plant productivity. This would be a concern if they affected the productivity distribution in cotton spinning disproportionately, relative to metallurgy and paper milling. First, region fixed effects partly accounts for this, as the effects of war would likely have been similar for firms in spatial proximity (Table A.22) Second, Table A.25, checks the correlation between conscripts per capita and productivity for each of the three sectors. The coefficients are small and statistically insignificant for cotton spinning. Finally, after 1813, some battles took place on French territory. Table A.26 shows the relationship between productivity in 1840 and a dummy for plants located within 10km of a battlefield. The coefficient for mechanized cotton spinning is statistically insignificant and small in magnitude.
B. Early Spinning Workshops	
E: Tables A.27 and A.28	Our results may be driven by the disappearance of small cotton spinning plants. It may be the case that there were systematic differences between small plants operating early vintages of spinning jennies (that did not necessarily need inanimate sources of power) and larger-scale factories. We test for this in two ways: first, we adopt a stricter definition of 'factory-production' and use only plants with more than 10 employees (see Table A.27); second, we drop plants that used the earliest vintage of machinery – the spinning jenny (see Table A.28). In both cases, the lower-tail bias remains striking.
C. Changes at the Plant Level	
C.1 Plant Scale	
E: Tables A.2-A.4 and A.29	Our findings could be driven by increasing plant scale. First, this is unlikely as all sectors witnessed an increase in plant scale (see Tables A.2, A.3, and A.4). In addition, Table A.29 shows that our quantile regression results are robust to controlling for the number of workers (at the plant level).
C.2 Capital Deepening	
E: Table A.30	Productivity growth in cotton spinning could be driven by technological improvements of the mechanized machinery after 1800. One important dimension of improving technology in cotton spinning was the increased capital per unit of labor (measured as spindles per worker). Table A.30 controls for the capital-labor ratio at the plant level (measured as the number of spindles per worker). The lower-tail bias of productivity growth remains robust.

Alternative Mechanisms (continued)	
Type of evidence: **H**istorical/**E**mpirical	Description of Evidence
	C.3 Quality Upgrading
E: Tables A.5 and A.31	The lower-tail bias could be driven by quality upgrading across plants over time. Table A.5 (Panel B) uses prices not quality-adjusted and shows that our results hold. In addition, if quality led to higher sales and thus larger plant size, it could still drive our results indirectly. Table A.31 estimates the quantile regressions not adjusting for quality differences in prices across plants and controlling for plant size. The lower-tail bias is robust to this specification.
	C.4 Age Profile of Plants
E: Tables 3	The median age of mechanized cotton spinning plants in 1806 was only 3 years. One concern is that our results could be driven by the fact that cotton spinning plants are very young and inexperienced. Table 3 shows that the opposite holds in the data: Younger plants were significantly *more* productive than older ones.

References

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*. Cambridge University Press.

André, L. (1996). *Machines à Papier: Innovation et Transformations de l'Industrie Papetière en France : 1798-1860*. Editions de l'Ecole des Hautes Études en Sciences Sociales.

Bairoch, P., J. Batou, and P. Chèvre (1988). *La Population des Villes Europeenees de 800 à 1850: Banque de Données et Analyse Sommaire des Résultats*. Geneva: Centre d'histoire economique Internationale de l'Université de Genève, Libraire Droz.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *Quarterly Journal of Economics 128*(1), 1–51.

Bonin, S. and C. Langlois (1987). *Atlas de la Révolution Française: Economie*, Volume 10. Editions de l'École des Hautes Études en Sciences Sociales, Paris.

Boston Mutual Fire (1908). *Report No. 5: Slow-Burning or Mill Construction*. Boston Manufacturers Mutual Fire Insurance Co. 3rd ed., rev. and enl. Boston: The Company. .

Bradley, M. (2010). Examples of Industrial and Military Technology Transfer in the Eighteenth Century. *Documents Pour L'Histoire des Techniques 19*, 87–95.

Cameron, R. (1985). A New View of European Industrialization. *Economic History Review 38*(1), 1–23.

Chanut, J., J. Heffer, J. Mairesse, and G. Postel-Vinay (2000). *L'Industrie Française au Milieu du 19e Siècle: Les Enquêtes de la Statistique Générale de la France*. Editions de l'École des Hautes Études en Sciences Sociales, Paris.

Chassagne, S. (1976). L'Enquête, dite de Champagny, sur la Situation de l'Industrie Cotonnière Française au Début de l'Empire (1805-1806). *Revue d'Histoire Économique et Sociale 54*(3), 336–370.

Chassagne, S. (1991). *Le Coton et ses Patrons: France 1760-1840*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Clark, G. (1994). Factory Discipline. *The Journal of Economic History 54*(1), 128–163.

Cookson, G. (1997). Family Firms and Business Networks: Textile Engineering in Yorkshire, 1780 - 1830. *Business History Review 39*(1), 1 – 20.

Crouzet, F. (2003). The Historiography of French Economic Growth in the Nineteenth Century. *Economic History Review 62*(2), 215–242.

Daudin, G. (2010). Domestic Trade and Market Size in Late-Eighteenth-Century France. *Journal of Economic History 70*(3), 716–743.

Forrest, A. (1989). *Conscripts and Deserters: The Army and French Society During the Revolution and Empire*. Oxford University Press.

Gille, B. (1968). *La Sidérurgie Française au XIXe Siècle: Recherches Histioriques*. Librairie Droz.

Hanlon, W. W. (2020). The Rise of the Engineer: Inventing the Professional Inventor During the Industrial Revolution. NBER Working Paper 29751.

Harley, C. (1998). Cotton Textile Prices and the Industrial Revolution. *Economic History Review 51*(1), 49–83.

Harris, J. (1998). *Industrial Espionage and Technology Transfer: Britain and France in the Eighteenth Century*. Aldershot: Ashgate.

Hauser, H. (1985). *Recherches et Documents sur l'Histoire des Prix en France de 1500 à 1800*. Slatkine.

Horn, J. (2006). *The Path Not Taken: French Industrialization in the Age of Revolution, 1750-1830*. Cambridge MA: MIT Press.

Huberman, M. (1996). *Escape from the Market*. Cambridge University Press.

Johnson, H. and A. Skempton (1955). William Strutt's Cotton Mills, 1793-1812. *Transactions of the Newcomen Society 30*(1).

Juhász, R. (2018). Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade. *American Economic Review 108*(11), 3339–76.

Khan, B. Z. (2020). *Inventing Ideas: Patents, Prizes, and the Knowledge Economy*. Oxford University Press, USA.

Lefebvre, G. (2011). *Napoleon*. Taylor and Francis.

Maddison, A. (2001). *The World Economy. A Millennial Perspective*. Paris: OECD.

McGaw, J. (1987). *Most Wonderful Machine: Mechanization and Social Change in Berkshire Paper Making, 1801-1885*. Princeton University Press.

Melitz, M. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Ministère de l'Agriculture et du Commerce (1847). *Statistique de la France: Industrie*, Volume 1. Imprimerie Royale.

Mitchell, B. (2003). *International Historical Statistics. Europe, 1750-2000*. Palgrave Macmillan.

Mokyr, J. (2005). The Intellectual Origin of Modern Economic Growth. *Journal of Economic History 65*(2), 285–351.

Mokyr, J. (2010). *The Enlightened Economy: An Economic History of Britain 1700-1850*. Yale University Press.

Nunn, N. and N. Qian (2011). The Potato's Contribution to Population and Urbanization: Evidence from an Historical Experiment. *Quarterly Journal of Economics 126*(2), 593–650.

Nuvolari, A., G. Tortorici, and M. Vasta (2023). British-French Technology Transfer from the Revolution to Louis Philippe (1791–1844): Evidence from Patent Data. *The Journal of Economic History*, 1–41.

O'Brien, P. and C. Keyder (1978). *Economic Growth in Britain and France, 1780-1914: Two Paths to the 20th Century*. London: Allen and Unwin.

Özak, O. (2018). Distance to the Pre-Industrial Technological Frontier and Economic Development. *Journal of Economic Growth 23*(2), 175–221.

Perla, J. and C. Tonetti (2014). Equilibrium Imitation and Growth. *Journal of Political Economy 122*(1), 52–76.

Perrot, J. and S. Woolf (1984). *State and Statistics in France, 1789-1815*. Taylor & Francis.

Pollard, S. (1965). *The Genesis of Modern Management: A Study of the Industrial Revolution in Great Britain*. Harvard University Press.

Pounds, N. and W. Parker (1957). *Coal and Steel in Western Europe: The Influence of Resources and Techniques on Production*. Bloomington: Indiana University Press.

Ridolfi, L. and A. Nuvolari (2021). L'histoire immobile? A reappraisal of French economic growth using the demand-side approach, 1280–1850. *European Review of Economic History 25*(3), 405–428.

Riello, G. (2013). *Cotton: The Fabric that Made the Modern World*. Cambridge University Press.

Rostow, W. (1975). *How it All Began: Origins of the Modern Economy*. Methuen and Co., London.

Sargent, T. and F. Velde (1995). Macroeconomic Features of the French Revolution. *Journal of Political Economy 103*(3), 474–518.

Shorter, E. and C. Tilly (1874). *Strikes in France: 1830–1968*. Cambridge University Press.

Squicciarini, M. and N. Voigtländer (2015). Human Capital and Industrialization: Evidence from the Age of Enlightenment. *Quarterly Journal of Economics 30*(4), 1825–1883.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature 49*(2), 326–365.

Szostak, R. (1989). The Organization of Work: The Emergence of the Factory Revisited. *Journal of Economic Behavior and Organization 11*(3), 343–358.

Vallauri, D., A. Grel, E. Granier, and J. Dupouey (2012). *Les Fôrets de Cassini. Analyse Quantitative et Comparaison Avec les Fôrets Actuelles*. Marseille: WWF/INRA.

Vallée, G. and A.-A. Hargenvilliers (1936). *Compte Général de la Conscription de A.-A. Hargenvilliers Publié d'Après le Manuscrit Original avec une Introduction et des Notes*. Librairie de Recueil Sirey.

White, E. (1991). Measuring the French Revolution's Inflation: the Tableaux de Dépréciation. *Histoire et mesure 6*(3–4), 245–274.

Williamson, O. (1980). The Organization of Work: A Comparative Institutional Assessment. *Journal of Economic Behavior and Organization 1*(1), 5–38.

Woronoff, D. (1984). *L'Industrie Sidérurgique en France Pendant la Révolution et l'Empire*. Paris: École des Hautes Études en Sciences Sociales.