NBER WORKING PAPER SERIES

CODIFICATION, TECHNOLOGY ABSORPTION, AND THE GLOBALIZATION OF THE INDUSTRIAL REVOLUTION

Réka Juhász
Shogo Sakabe
David Weinstein

Working Paper 32667
http://www.nber.org/papers/w32667

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024, Revised December 2025

We give special thanks to Chris Meissner and John Tang for sharing their trade data for Belgium and Japan, and to Davide Coluccia and Gaia Dossi for sharing their data on British patents from the second half of the 19th century. We thank Benjamin Eyal, Diti Jain, Isaac Loomis, Zachary Marcone, Ojaswee Rajbhandari, Roshan Setlur, Alex Zhang, and especially Michael Duarte, Hriday Karnani, Verónica C. Pérez, Angela Wu, and Dongcheng Yang for excellent research assistance. We also thank Treb Allen, Andrew Bernard, Kirill Borusyak, Serguey Braguinsky, Florian Caro, Davin Chor, John Fernald, Kyoji Fukao, Walker Hanlon, Shizuka Inoue, Takatoshi Ito, Debin Ma, Joel Mokyr, Chiaki Moriguchi, Tetsuji Okazaki, Charly Porcher, Jared Rubin, Robert Staiger, Jón Steinsson, John Tang, Dan Trefler, and Weiwen Yin for their comments. Réka gratefully acknowledges support from the Alfred P. Sloan Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Codification, Technology Absorption, and the Globalization of the Industrial Revolution
Réka Juhász, Shogo Sakabe, and David Weinstein
NBER Working Paper No. 32667
July 2024, Revised December 2025
JEL No. F14, F63, N15

ABSTRACT

This paper examines the global adoption of technology in the late nineteenth century. We construct several novel datasets to test the idea that the codification of technical knowledge in the vernacular was necessary for countries to absorb the technologies of the First Industrial Revolution. We find that comparative advantage shifted to industries that could benefit from these technologies in countries and colonies with access to codified technical knowledge, but not in other regions. Using the rapid and unprecedented codification of technical knowledge in Meiji Japan as a natural experiment, we show that this pattern emerged only after the Japanese government codified vast amounts of technical knowledge. Our findings shed new light on the frictions associated with technological diffusion and offer a novel explanation for why Meiji Japan was unique among non-Western countries in successfully industrializing during the first wave of globalization.

Réka Juhász
University of British Columbia
Vancouver School of Economics
and NBER
reka.juhasz@ubc.ca

David Weinstein
Columbia University
Department of Economics
and NBER
dew35@columbia.edu

Shogo Sakabe
LMU Munich
ss5122@columbia.edu

"At present, the learning of China and Japan is not sufficient; it must be supplemented and made complete by inclusion of the learning of the entire world... I would like to see all persons in the realm thoroughly familiar with the enemy's conditions, *something that can best be achieved by allowing them to read barbarian books as they read their own language.* There is no better way to enable them to do this than by publishing [a] dictionary."

Sakuma Shozan,[1] 1858, quoted in Hirakawa (2007, p. 442, emphasis added).

1 Introduction

Although recent econometric evidence finds that modern economic growth started in England around 1600 (Bouscasse et al., 2023), the spread of economic development has been highly uneven. For example, there are currently only four types of high-income countries in the world: English-speaking countries, countries close to England, resource-abundant countries, and Japan and its former colonies.[2] While economists have made enormous progress in understanding why English-speaking countries, Europe, and Petrostates are rich, data-driven studies of why the First Industrial Revolution (henceforth "Industrial Revolution" or IR) first spread to Japan and not to any other non-Western country are almost nonexistent. After centuries of resisting economic and social change, Japan transformed from a relatively poor, predominantly agricultural economy specialized in the exports of unprocessed, primary products to an economy specialized in the export of manufactures *in under fifteen years*.[3] Why did Japan in the Meiji Period (1868-1912) succeed in this structural transformation while so many other countries failed to develop in this period?

We bring several novel datasets to bear on this question and test one of the main theories proposed by Mokyr (2011): namely, that an essential component of the Industrial Revolution was the development of what Stevens (1995) calls "technical literacy," i.e., the *codification* of engineering, commercial, and industrial practices. We call this knowledge "technical knowledge." We hypothesize that codification reduces technology access costs by enabling entrepreneurs to read about technology in their own vernacular. While there is extensive evidence that what Mokyr (2016) refers to as the culture of Enlightenment created vast amounts of codified technical knowledge in Western European languages, our understanding of codification beyond Europe is much more limited. For example, we have no idea how many books containing technical knowledge a literate person in China could have read in 1870 or the extent to which the number of books containing codified knowledge changed over time. As a result, prior work has been unable to explore how access to technical knowledge in the vernacular contributed to the *spread* of the Industrial Revolution.

An ideal experiment would require both cross-sectional and time-series variation in technical knowledge supply. Cross-sectional variation allows us to examine whether industries that stood to gain the most from the supply of codified knowledge grew faster in countries that supplied this knowledge. Time-series variation lets us explore whether the timing of this faster growth coincides

[1]We follow the convention of referring to Japanese, Korean, and Chinese historical figures using their surname first, followed by their given name.

[2]We define high-income countries as those with a purchasing-power-parity adjusted GDP per capita of 50 percent or more than the US level in 2022, as measured by the World Bank. See Appendix H.1 for more details.

[3]We also see this sudden transformation in the efficiency of modern industry. For example, Clark (1987) finds that Japan transitioned from having no modern textile and weaving mills in 1870 to having modern mills that achieved levels of output per unit of capital that were 96 and 98 percent of those in Britain by 1910. By contrast, the Chinese textile and weaving industries had 79 and 66 percent output efficiency relative to Britain's in 1910.

with the provision of this knowledge. The experiences of Meiji Japan and the other codifiers of knowledge in the late nineteenth century provide precisely this empirical setting.

We test the link between codified knowledge and productivity growth in Meiji Japan and the late 19th-century global economy by constructing the first dataset that enables us to quantify the extent of codification by language, the usefulness of this codification by industry, and industry-level export and productivity growth in 37 "regions" (some of which are countries) in the late nineteenth and early twentieth centuries. We build this dataset by scraping the catalogs of libraries for every major language, digitizing technical books for every major tradable industry, digitizing the synopses of British patents issued between 1780 and 1852, digitizing bilateral industry-level trade data for Japan and the U.S., merging these trade data with extant trade datasets to create the first multicountry, bilateral, industry-level, trade dataset for the nineteenth century.

We establish three novel stylized facts about the global spread of the Industrial Revolution and the uniqueness of Japan's nineteenth-century industrialization, "the Meiji Miracle." The first stylized fact is that Meiji industrialization was exceptional in comparison to other regions in the periphery. We find that late 19th-century Japan experienced a surge in the share of manufacturing exports, surpassing that of any other region in our sample. Notably, this surge in manufacturing exports happened rapidly and much later than well-known events in nineteenth-century Japanese economic history. In particular, twenty-five years after Japan had jumped to free trade (1858) and fifteen years after the Meiji Restoration (1868), eighty percent of Japanese exports were still primary products. Similarly, average annual real per capita GDP growth between 1870 and 1883 was only 0.6 percent. In other words, there is no evidence that Japan's natural comparative advantage or state reforms were gradually shifting export patterns in the manufacturing sector. However, in a brief 13-year period from 1883 to 1896, Japan's manufacturing export share tripled and then remained at around 60 percent of total exports until the Second World War. Thus, the puzzle of Japan's development is why Japanese manufacturing exports grew so suddenly, and so much. It was as if Japanese entrepreneurs had suddenly learned how to do modern manufacturing.

The second stylized fact is that in 1870, entrepreneurs in most regions—including Japan—had almost no technical books to read in their vernacular languages. We document this by scraping thousands of libraries containing books in 33 major languages and find that in 1870, 84 percent of all *technical* books were written in just four languages: English, French, German, and Italian. People who could not read these four languages were therefore technically illiterate. For example, a person who could only read Arabic would have been able to read only 71 technical books in 1870. Libraries for other major non-European languages, such as Chinese, Hindi, and Turkish, have extensive collections of books but contain similarly small numbers of technical books. By contrast, speakers of major European languages would have had access to thousands of technical books. This puts the achievements of the Enlightenment, with its emphasis on the "coding, storing and transmission of technical knowledge" (Berg, 2007, p. 125) in comparative perspective for the first time. In short, outside of the languages of the Enlightenment, literacy in the vernacular was a ticket to reading the humanities and history, not science.

The third stylized fact is that the Japanese language is unique in that it started at a low base of codified knowledge in 1870 but experienced explosive growth in the publication of technical books, catching up with the West in the middle of the 1880s. By 1890, there were more technical books in Japanese than in any other language except English and French. This catch-up in codification coincides with Japan's sudden industrialization. Japan suddenly began exporting manufactured products shortly after Japanese entrepreneurs could read *in Japanese* how to make these products.

How did Japan achieve such remarkable growth in the supply of technical books? We show that the Japanese government was instrumental in overcoming a complex public goods problem, which enabled Japanese speakers to achieve technical literacy in the 1880s. We document that Japanese publishers, translators, and entrepreneurs initially could not translate Western scientific works because the Japanese language lacked the words needed to describe the technologies of the IR. The Japanese government solved this coordination problem by creating a large dictionary that contained Japanese jargon for many technical words. Indeed, we find that new word coinage in the Japanese language grew suddenly after a massive government effort to subsidize translations produced technical dictionaries and, subsequently, a large number of translations of technical books.

Beyond producing technical dictionaries, the Meiji government also made substantial investments in codifying knowledge by paying for the large-scale translation of technical knowledge from the West (Montgomery, 2000). Our analysis of the institutional affiliations of these translators reveals that 74 percent of them were government employees, indicating the relative importance of the government in funding this public good. This created two sub-periods in Meiji Japan: a period before the 1880s, in which Japan had completed substantial economic reforms but had codified less than half as much as had been codified in Spanish (and a small fraction of what had been codified in French, English, Italian, and German), and a period afterward in which Japanese people could read Western technical knowledge at a level equal to or exceeding that in the West.

Together, our stylized facts show that the sudden increase in codified knowledge and the sudden surge in manufacturing specialization, which occurred shortly thereafter, were unique to Japan in the 19th century. Thus, Japanese manufacturing export growth rates didn't gradually increase as institutions improved; rather, a rapid increase in manufacturing occurred only after Japan codified approximately the same amount of knowledge as Germany had in 1870. Together, we interpret these three stylized facts as suggestive evidence that access to technical knowledge may have been a necessary (although not sufficient) condition for the spread of the IR.

In the second part of the paper, we exploit the natural experiment of Japan's rapid codification of knowledge to test this hypothesis rigorously. This requires both time series variation and cross-sectional variation in technical knowledge; thus, we move our empirical analysis to the industry level. In particular, we develop a method to quantify the supply of useful, codified knowledge generated by the IR for each industry. We use a text-based approach that closely follows how codified technical knowledge was disseminated in this period: through the publication of technical manuals. For example, "The American Cotton Spinner, and Managers' and Carders' Guide," published in 1851, contains a description of every aspect of operating a cotton spinning mill from the dimensions of the building, to setting up the gearing which distributes power through the building, as well as the operation, and maintenance of each machine used in production.

For each industry, we calculate the similarity of text from these historical technical manuals (in English) to the text of British patents using cosine similarity, the standard metric in natural language processing. We call this measure "British Patent Relevance" or BPR. Our BPR measure rises in the similarity of the word use in an industry's technical manuals to that in British patents. Thus, it serves as a metric for assessing the usefulness of the knowledge codified in British patents for a particular industry. Reassuringly, industries such as textiles, which benefited the most from the new technologies of the IR, have descriptions of production processes, including flagship technologies such as spinning machinery and steam engines, that also feature prominently in patent texts. As such, the contents of patent texts are relevant for manufacturing textiles. On the other hand, the cosine similarity between word use in manuals and patent descriptions is smaller for industries like charcoal, which suggests that the makers of charcoal benefitted little from IR technologies. Importantly, BPR is a measure of the *supply* of new technical knowledge; it does not use *any* (potentially endogenous) information on what was translated in Japan or elsewhere.

To measure outcomes at the industry level, we use our novel, bilateral, industry-level trade dataset to compute industry-level export growth from 1880 to 1910. In robustness checks, we also use productivity growth estimates based on Costinot et al. (2012) and Amiti and Weinstein (2018). These methods are well-suited to data-scarce environments such as ours.

Armed with these data, we examine the relationship between the supply of technical knowledge and export/productivity growth in Japan and around the world. We show that Japanese export and productivity growth was higher in the industries where the *supply* of technical knowledge produced by the IR was greater. The estimates point to a large effect of access to technical knowledge on growth. Our estimates imply that a Japanese industry with a one-standard-deviation higher British Patent Relevance measure experienced annual export and productivity growth rates that were 12 and 1.2 percentage points faster, respectively.

We use several features of our setting to make the case that this relationship is likely to be causal. First, we exploit the fact that Japan was unique among periphery countries in codifying knowledge. If our findings for Japan are causal, this would suggest that other periphery countries, which did *not* have access to codified knowledge, should not have a similar association between industry growth and BPR. Indeed, we find that, on average, other regions do not exhibit a similar effect. Interestingly, other low-income regions and Asian regions tend to exhibit a negative relationship with BPR, suggesting divergence, although this negative effect is not always statistically significant. In contrast, for regions speaking one of the other major codifying European languages (English, French, German, and Italian), we find a similar positive effect of BPR on growth, though one that is smaller in magnitude. In summary, in the cross-section of countries, we find supporting evidence that only codifying countries have industry growth patterns systematically related to BPR.

Second, we exploit the sharp timing of codification in Japan. In most settings, identifying the effect of codification on growth is challenging, as codification typically proceeds slowly, making it difficult to rule out confounding factors. However, our third stylized fact is that Japan can be separated into two relatively well-demarcated periods: one in which Japan resembled other technically illiterate periphery economies and one in which codification in Japanese is comparable to that in the most codified Western European languages. Consistent with our hypothesis, we find a positive and statistically significant effect of BPR on industry growth in Japan *only after* Japan became technically literate. Indeed, until 1890, Japan looked remarkably similar to the rest of the global periphery, and Asia in particular, in which comparative advantage shifted away from industries that heavily used British technology.

Taken together, our results lend support to the idea that low-cost access to technical knowledge, which at the time usually meant access *in the vernacular*, was a necessary condition for the diffusion of IR technologies and modern manufacturing growth more broadly. Moreover, our results suggest that for regions not strongly affected by Enlightenment efforts to reduce what Mokyr (2011) terms "access costs," the codification of technical knowledge was a complex public good that required state provision. In the final section of the paper, we show that the Meiji model of technology policy had a lasting influence beyond Japan. In particular, we demonstrate that policymakers in postwar Korea and China, who had studied in Japan and were heavily influenced by the Meiji model of economic development, adopted similar codification policies.

2 Related Literature

This paper contributes to three strands of the literature. First, our results inform the technology adoption lags literature (e.g., Griliches 1957; Rosenberg 1972; Hall 2004; Comin and Hobijn 2010). In particular, our finding that state-led codification lowered technology access costs in Meiji Japan helps us understand the strikingly large cross-country technology adoption lags documented in

Comin and Hobijn (2010) for the 19th century. Our results suggest that in the absence of similar state-led technology policies in other parts of the global periphery, the costs of technology adoption were too high. As such, the paper offers a new explanation for a longstanding puzzle in economic history: why IR technologies failed to take hold beyond a few technologically advanced countries during the long 19th century (Clark, 1987; Allen, 2011; DeLong, 2022). Our explanation builds on Mokyr (2011)'s pioneering work on the importance of "technical knowledge" for European industrialization, though with a Gerschenkronian (Gerschenkron, 1962) twist.[4] In particular, for regions that were physically distant from Europe and linguistically distant from major European languages, the provision of technical knowledge required the state's involvement due to its public good-like attributes. This points to a novel arena where the Gerschenkronian argument of the state as a critical agent in late industrialization may apply.

Second, our results inform our understanding of the sources of Japan's unique industrialization. Previous work has examined the introduction of new institutions (Sussman and Yafeh, 2000), modern banking (Tang and Basco, 2023), railroads (Tang, 2014), subsidized firms (Morck and Nakamura, 2007, 2018) and trade (Bernhofen and Brown, 2004, 2005). This careful work has not found large positive impacts of these policies on economic outcomes and sometimes finds the policies were counterproductive. For example, Sussman and Yafeh (2000) conclude that "the great majority of the Meiji reforms—including the establishment of the Bank of Japan and the introduction of 'modern' monetary policy, the promulgation of the Meiji Constitution, and the introduction of parliamentary elections—produced no quantitatively significant market response." In the end, they conclude that only land tax reform and Japan's adoption of the gold standard mattered to investors.

Our findings offer a resolution to the puzzle of what drove the Meiji Miracle. Importantly, our result that BPR mattered for industry growth from the 1890s, but not before is consistent with the latest evidence suggesting that Japanese convergence to Britain in terms of GDP per capita only started in 1890 and was driven by productivity growth in manufacturing (Broadberry et al., 2025). Moreover, our findings also explain why cross-country technology adoption lags begun to decrease in Japan during the Meiji period (Comin and Hobijn, 2010).

These results are particularly helpful in placing the "Meiji Miracle" in a comparative perspective. That is, while the more standard modernization efforts of the Meiji government, such as the introduction of banking and railroads, certainly contributed to industrialization, given their fairly widespread adoption in other parts of the global periphery, which were characterized by more modest growth and longer cross-country technology adoption lags, it is unlikely they can give a full account. In contrast, our paper provides empirical support for the long-standing tradition in Japanese economic history that has emphasized the more distinctive aspects of the Japanese government's efforts to adopt Western technology. In fact, our results suggest that the Japanese state may have been uniquely successful in relaxing key constraints to adopting Western technology.

Third, our paper is related to prior research that has explored the effects of knowledge codification. Previous work has examined the effects of new codification technologies (Dittmar, 2011), the development of written language (Brown, 2024), and ideological barriers to knowledge diffusion (Abramitzky and Sin, 2014). We contribute to this literature by showing what is, to the best of our knowledge, the first evidence supporting the hypothesis that codified knowledge in the vernacular reduces technology access costs and benefits industries most reliant on that knowledge. This helps explain why prior work has found developmental effects of knowledge codification on

[4]As such, our paper is related to recent studies that examine the role of Enlightenment ideals in Europe and their effect on industrialization (Squicciarini and Voigtländer, 2015; Almelhem et al., 2023). Our focus is on the spread of technology outside of Europe and the context of the Enlightenment.

more indirect measures of development such as city growth (Dittmar, 2011), education, and health outcomes (Brown, 2024).[5]

The remainder of the paper is structured as follows. Section 3 discusses why Japan adopted its technology policy and provides details on the historical context and what the Japanese government did. Section 4 discusses the data we use and how we measure codification and British Patent Relevance. Section 5 presents our main results. Section 6 conducts a number of robustness exercises; Section 7 considers the lasting impact of Meiji technology policy beyond Japan; and Section 8 concludes.

3 Japanese Industrialization and Technology Policy: Historical Context

Nineteenth-century Japan presents an interesting study of late industrialization. In a very broad sense, Japan in the 1870s was similar to other poor, predominantly agricultural areas of the global periphery that had missed out on the first wave of industrialization. The most recent estimates of GDP per capita and wages for Japan confirm that it was a low-growth, low-wage society diverging from northwestern Europe until at least the 1870s (Bassino et al., 2019; Kumon, 2022).[6] In this section, we introduce and provide novel evidence for the main stylized facts that motivate our empirical analysis. We first examine Japan's unique industrialization amongst economies in the periphery. Second, we discuss Japan's technology policies and show how they contributed to Japan's rapid and unprecedented codification of technical books in the periphery.

3.1 Japan's shifting export composition

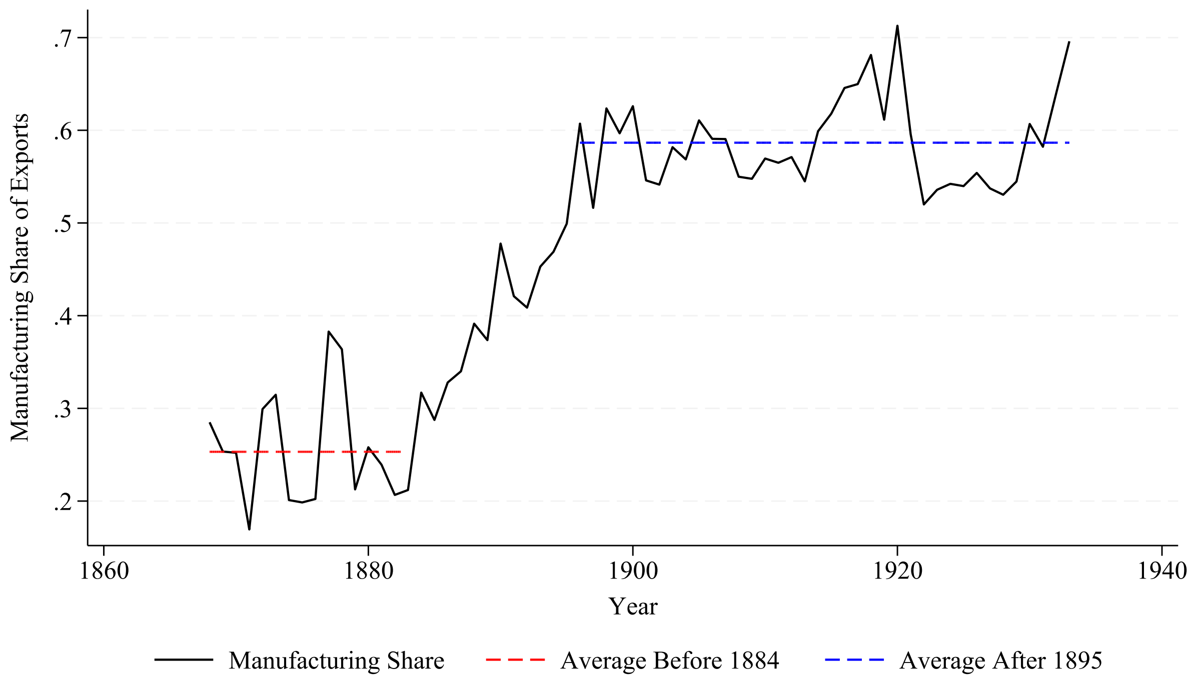
Starting in the closing decades of the nineteenth century, Japan began to industrialize (e.g., Yamamura 1997). The Maddison data suggest a growth acceleration with average annual per capita GDP growth rising from 0.6 percent between 1870 and 1885 to 1.4 percent between 1885 and 1900. Fukao et al. (2020) estimate that the average annual growth rates of TFP and labor productivity were even higher (between 1885 and 1899), coming in at 1.5 percent and 1.8 percent, respectively. They find that by 1913, Japanese TFP was 44 percent higher than it had been in 1885, and labor productivity was 64 percent higher. Japan was clearly escaping from the Malthusian equilibrium.

These aggregate changes obscure the rapid industrialization occurring in the small but rapidly growing manufacturing sectors. To understand the timing, speed, and scale of this change, Figure 1 plots the manufacturing share of exports in Japan during this period. Consistent with other evidence of a stagnating economy cited above, the trade data indicate that the economy was highly specialized in the exports of primary products until the early 1880s. Between 1868 and 1883, manufacturing exports as a share of total exports hovered around 25 percent with no discernible upward trend. The share suddenly began rising after 1883. Starting in 1884 and continuing through 1896, the share of Japanese manufacturing exports rose steadily before stabilizing at an average level that was more than double its earlier average for over forty years. Since this shift

[5] There is less work on how tacit knowledge affects development, with the exception of a recent paper by Bekkers et al. (2022) examining whether technological catch-up is slower in sectors that are more tacit-knowledge intensive.

[6] There is some debate about whether proto-industrialization was experienced in the late Tokugawa period. Building on Saito and Takashima (2016), Bassino et al. (2019) estimate annual GDP per capita growth of 0.26 percent for the period 1721-1874, driven by growth in the secondary and tertiary sectors. Based on these findings, the authors argue that Japan may have improved its relative position within Asia even before the Meiji Restoration in 1868. However, Kumon (2022) casts doubt on these findings, arguing that there is no direct evidence of growth in the secondary and tertiary sectors during this period.

Figure 1: Manufacturing Share of Exports (Japan)



Note: Data sourced from Oriental Economist (1935) *Foreign Trade of Japan: A Statistical Survey.* Tokyo: Toyo Keizai Shinposha.

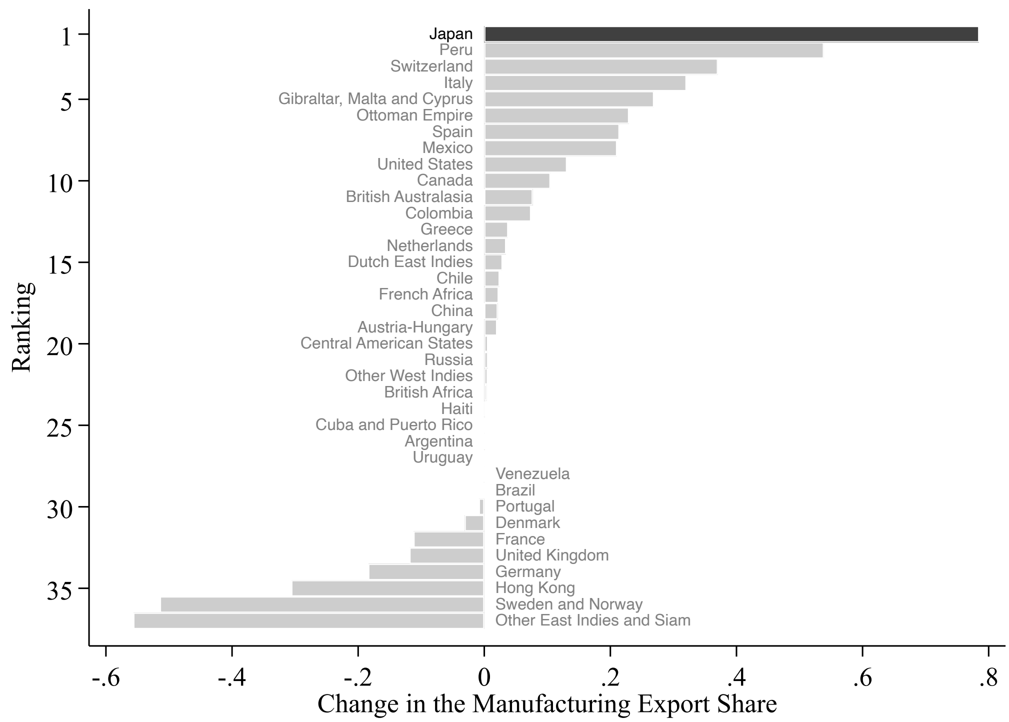
happened twenty-five years after Japan opened to trade, it is hard to explain the shift in terms of Japan simply having a comparative advantage in manufacturing.[7] As a result, by the turn of the twentieth century, Japan was specialized in the export of manufactured products. Moreover, Figure 2 shows that no other economy displayed a similar change in its export composition, which raises the question of what happened in Japan and why it didn't happen elsewhere.

3.2 Meiji technology policy

We hypothesize that a state-led technology absorption effort, unique in scale, provided widespread access to the technical knowledge needed to adopt the technologies of the IR. After centuries of self-imposed isolation, the U.S. forcibly opened Japan to foreigners in 1854 and to trade with Western countries in 1858. The Tokugawa shogunate, which had ruled Japan since the 1600s, was overthrown in the 1868 "Meiji Restoration" and replaced by an oligarchy of rival nobility ruling in the name of Emperor Meiji. From its inception on April 5, 1868, the Meiji Government declared that the assimilation of Western knowledge would be a central tenet of its policy. The Charter Oath, Emperor Meiji's five-sentence statement of the objectives of the fledgling government, stated that "knowledge shall be sought throughout the world so as to strengthen the foundations of imperial rule." (Hirakawa, 2007, p. 338) Thus, all members of the new government were required

[7]Similarly, Japan's Meiji reforms began in 1868 and many of the most important ones, like tax reform, foreign missions, peak hiring of foreigners, educational reform, postal reform, telegraph construction, banking reform, military reform, judicial reform, etc., were implemented by 1875, so it is not obvious why these reforms should have caused the manufacturing export share to stagnate only to sharply increase starting in the 1880s.

Figure 2: Change in the Manufacturing Share of Exports by Country (1880-1910)



Note: Percentage point change in the share of manufacturing exports relative to total exports of the country. For non-reporting countries, we use imports to reporting countries to estimate exports. We include the following SITC categories in manufacturing: code 6 (Manufactured goods classified chiefly by material), 7 (Machinery and transport equipment), 8 (Miscellaneous manufactured articles), 95 (Armoured fighting vehicles, war firearms, ammunition, parts, n.e.s.), 96 (Coin (other than gold coin), not being legal tender). See Section 4.2 for a description of how the data was constructed.

to support strengthening Japan by absorbing Western ideas.

However, Japanese historians argue that many members of the Tokugawa shogunate had already realized in the aftermath of China's ignominious defeat in the First Opium War (1839-1842) that Japan needed a strategy to absorb Western science (Bolitho, 2007, p. 157). Early Japanese reformers, most prominently Sakuma Shozan, began developing plans for how Japan could co-exist with the West. Sakuma developed a strategy for modernizing Japan, which he summarized with the slogan "Eastern morality, Western technology." While there was little concrete action until U.S. warships entered Edo Harbor in 1853, the arrival of the Americans prompted the shogunate to spring into action. Almost immediately after the Americans arrived, the Japanese government established the Institute of Barbarian Books (*Bansho Torishirabesho*), which was tasked with developing English-Japanese dictionaries to facilitate technical translations. This project was the first step in what would become a massive government effort to codify and absorb Western science.

This section discusses the three components of Japan's technology absorption effort. First, we describe the effort to codify Western technical knowledge. Second, we demonstrate that by investing in elementary and university education, the government ensured the population had the necessary skills to absorb and apply the technical knowledge it provided. Third, we discuss how the government raised enough tax revenue to finance these costly policies.

3.2.1 The effort to codify Western technical knowledge

The key component of the Meiji technology policy studied in this paper is the large-scale state-led *codification* of technical knowledge in the Japanese language. Codification refers to the creation of a means of transmitting knowledge, through "language creation and the writing of messages," that does not require direct contact between the originator of the knowledge and the recipient (Cowan and Foray, 1997, p. 595). Technical manuals, textbooks, and scientific papers are all examples of codified knowledge, as the knowledge contained in these publications can be accessed without personal contact with the author.

Codified knowledge is cheap to reproduce and disseminate (conditional on available technologies such as printing), making it a powerful tool for knowledge diffusion (Abramovitz and David, 1996). Yet its non-rival nature gives it public good attributes, implying that the market will typically undersupply codified knowledge (Foray, 2004, p. 73). Modern scholarship on the economics of knowledge has identified multiple factors that are important for codified knowledge to be successfully absorbed. First, codification often requires prior language development (Cowan and Foray, 1997). Intuitively, it is necessary to first develop the jargon used to express new ideas. Second, accessing codified knowledge requires absorptive capacity such as literacy or basic scientific training. Cowan and Foray (1997, p. 605) describe this as follows, "Diffusion and use of codified knowledge are thus dependent on the irreversible investment required to build a community of agents, a 'clique' or a network the members of which can 'read' the code." Third, complementary investments in tacit knowledge are required (Cowan and Foray, 1997; Mokyr, 2011). Unlike codified knowledge, tacit knowledge is deeply embedded in personal experience and context; it cannot be detached and fully made explicit by the person who holds it (Polanyi, 1966). For example, consider the master spinner demonstrating the operation of a spinning machine. It is not possible to break down and codify every movement of the spinner's hands (some of which even the spinner may not be aware). As we discuss below, Meiji technology policy made all of these investments.

Language development to facilitate translation. Consistent with the discussion above, language development was a significant barrier to codifying Western science and technology in the Japanese language. Linguists and lexicographers have written extensively on the difficulty of scientific translation between dissimilar languages (c.f. Clark 2009; Kokawa et al. 1994; Lippert 2001; Montgomery 2000). Technical translation is relatively easy in languages that share the same Greek and Latin roots. Thus, a speaker of French or German can easily guess that an English technical word like "telegraph" should be translated as "télégraphe" or "telegraf," respectively, and it would be easy for readers of all three languages to remember from their knowledge of Greek root words that telegraphs involve the transmission of words across distance.

Translation of English jargon into languages with root words not based on Greek and Latin is much harder because it requires language creation. For example, a typical speaker of Arabic would be hard-pressed to guess the meaning of the Arabic word "tiligraaf," from its spelling any more than a typical English speaker could guess the meaning of "algebra" (which comes from the Arabic word "al-jabr") from its spelling. People who did not understand languages closely related to English needed to translate vast amounts of English jargon to understand modern production techniques. Consistent with this, in Appendix Table A.1, we provide suggestive evidence that linguistic distance was a barrier to technology diffusion and, ultimately, economic growth during this period. Specifically, we show that GDP per capita in 1870 and 1913 tended to be lower in countries and regions speaking a plurality language that was more linguistically distant from English, *conditional on physical distance*.[8] While we do not interpret this negative relationship causally, we take it as suggestive empirical evidence consistent with scientific translations increasing technology access costs and inhibiting technology diffusion.

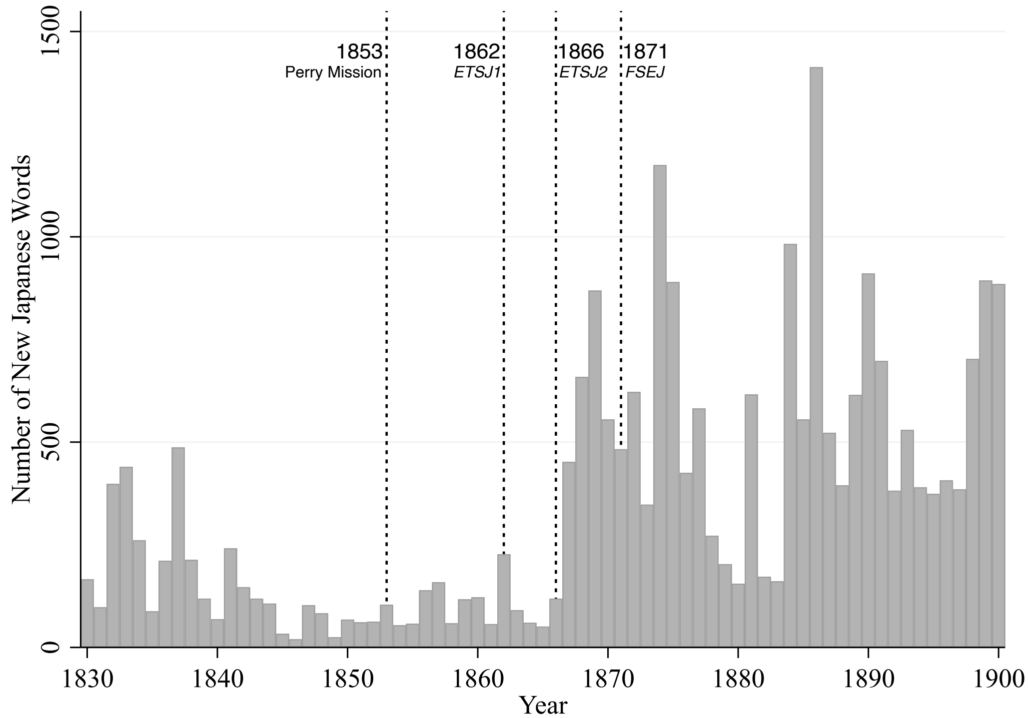
The language creation problem Japan faced in translating Western science was two-fold. First, there were no words in Japanese for IR products such as the railroad, steam engine, or telegraph, and using phonetic representations of all untranslatable jargon in a technical book resulted in transliteration, not translation. Second, translations needed to be standardized so that all translators would translate a given foreign word into the same Japanese one, a classic example of a coordination problem.

Solving these two problems became one of the Institute of Barbarian Books' main objectives. After carefully studying how to solve these problems, Japanese translators decided to base the translation of English jargon on Chinese glyphs whose meaning was known to all literate Japanese and played a similar role as Latin and Greek words in English (c.f Kokawa et al. 1994; Lippert 2001; Clark 2009). For example, in order to translate the word "telegraph," Fukuzawa Yukichi, who would go on to become one of the most famous government translators, created the Japanese word "*denshin*" in 1866 using glyphs that combined the Chinese characters for "electric" (*den*) and "message" (*shin*). While a Japanese reader encountering the term "electric message" might not recognize that it means telegraph on the first reading, it is easy to remember it once one learns the definition. Lippert (2001) argues that Japanese government translators' decision to create Japanese jargon based on Chinese root words is an important factor in making foreign scientific texts much easier to understand in Japanese than in alphabetic languages not based on Latin and Greek, where jargon is just transliterated. In other words, the Japanese government's investment in modifying the Japanese language to accommodate new words may have been a public good that lowered the cost of accessing technology for Japanese people.

The importance of this strategy for codification was not lost on Japanese reformers. For example, Sakuma wrote of the first English-Japanese dictionary, the ETSJ (*Eiwa-Taiyaku-Shuchin-Jisho* or "A Pocket Dictionary of the English and Japanese Language"), which was published by the Institute in 1862, "I would like to see all persons in the realm thoroughly familiar with the enemy's conditions, *something that can best be achieved by allowing them to read barbarian books as they read their own language.* There is no better way to enable them to do this than by publishing this dictionary" (Hirakawa,

[8]Appendix Figures A.5 and A.6 show the scatterplots for this relationship.

Figure 3: Word Creation in Japanese



Note: Number of new words created in Japanese from *Nihon Kokugo Daijiten*. The dictionary contains information on the first known time a word was used in a document, which we use to construct this graph. Dashed lines refer to the Perry Mission and publication dates of English-Japanese dictionaries.

2007, p. 442, emphasis added). A much larger dictionary supplanted this small dictionary, the FSEJ (*Fuon-Sozu-Eiwa-Jii* or "An English and Japanese Dictionary") in 1871, which contained two to three times as many words and a significant amount of English jargon.[9]

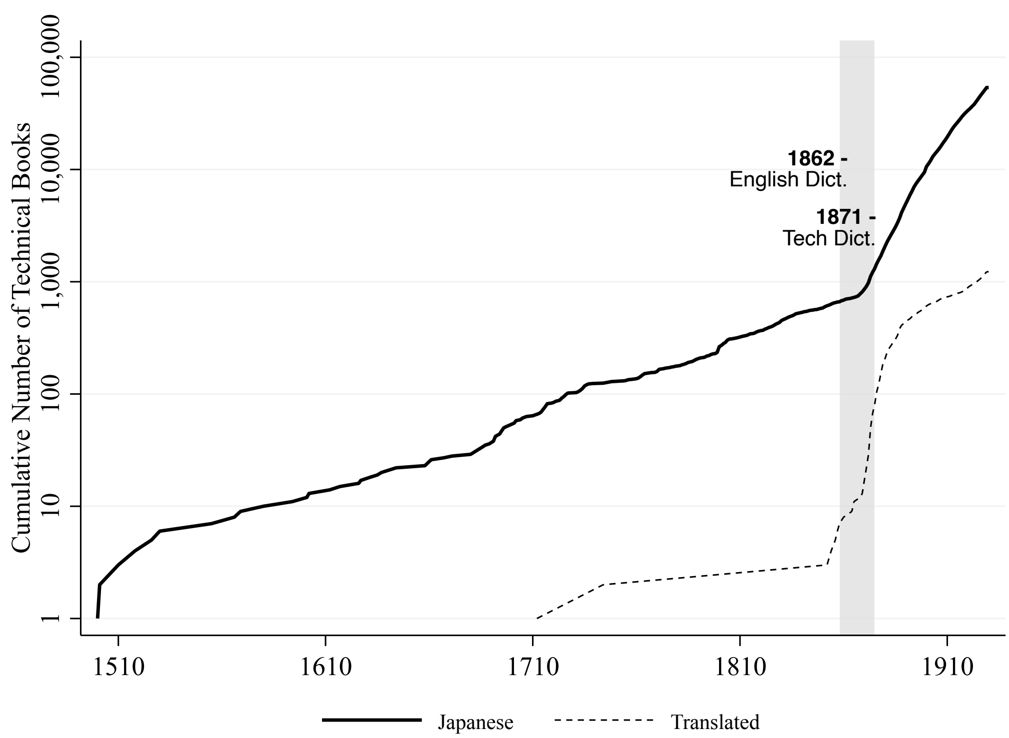
In Figure 3, we present suggestive evidence that solving the coordination problem associated with translating jargon facilitated the emergence of new words in Japanese. We obtained the first recorded use of Japanese words based on the revised edition of the *Nihon Kokugo Daijiten* ("The Unabridged Japanese Dictionary"), published by Shogakukan, encompassing 300,000 Japanese words. Word creation in Japan before the 1860s was surprisingly low—typically, only around 100 new Japanese words were created each year. Even in the first decade after Japan opened to the West following the Perry Mission, the rate of new-word creation in Japan remained essentially unchanged. This result is quite surprising given that in 1854, the Americans brought many pieces of new technology to show to the Japanese, such as a working locomotive, telegraph machine, cameras, etc. Exposing the Japanese to Western technology did not, in and of itself, lead to the emergence of new words. However, starting around the creation of the first English-Japanese dictionary (ETSJ1) in 1862 and accelerating with the large print run of this dictionary in 1866 (ETSJ2), the number of new words in Japanese rose to around 500 per year. Word creation

[9] Kokawa et al. (1994, pp. 80-119) is the source for our information on dictionaries. Publication and release dates are difficult to pinpoint exactly in this period. The Pocket Dictionary was first released in 1862 with a print run of only 200 copies but was reprinted and distributed much more widely in 1866. Similarly, the FSEJ was printed on a linotype machine in 1871 but officially published in 1873.

accelerated to over 1000 words per year following the release of the extensive English-Japanese dictionary, the FSEJ. Thus, to the extent that new word creation tracks new ways of codifying and conceptualizing the world, this evidence suggests that the government-led dictionary creation efforts in the 1860s helped solve the coordination problem inherent to introducing new ideas.

From words to books: state-led codification of technical knowledge. Alongside the public provision of dictionaries, the public sector played an outsized role in translating technical books. A search through the biographies of every person we could identify who translated a technical book between 1870 and 1885 reveals that 74 percent of the translators were government employees.[10] This number is likely a lower bound because Japanese biographical dictionaries do not necessarily list every job a person held. The failure of the private sector to generate many technical translations likely reflects the fact that it was impossible for translators to prevent other authors from paraphrasing their work and publishing it independently. Thus, government intervention was required to solve the public-good problem of codified knowledge creation.

Figure 4: Codified Technical Knowledge in Japan



Note: Codified technical knowledge for each year refers to all technical books written in Japanese in the NDL catalog or any other Japanese library linked to the NDL. A book is considered "Translated" if the NDL flags the book as a translation. The y-axis is on a log scale.

What were the effects of this effort to codify technical knowledge? In Figure 4, we examine the evolution of technical knowledge in the Japanese language by scraping the catalogues of the

[10]The NDL catalog specifies the translator for over 200 technical books translated to Japanese in the 1870s and 80s. We searched for the names of all translators on *JapanKnowledge Lib*, an online database, and made extensive use of Ueda et al. (2003), a biographical dictionary containing entries for more than 75,000 Japanese people and Heibonsha (1974), a biographical dictionary of 30,000 people.

National Diet Library and 81 additional Japanese libraries to construct a time series of all technical books from 1500 to 1930. Between 1600 and 1860, the number of technical books in Japanese grew by 1.6 percent per year.[11] The rate almost sextupled to 8.8 percent per year between 1870 and 1900, starting just as staff at the Institute for Barbarian Books produced the 1862 and 1871 English-Japanese dictionaries. After centuries in which the number of technical books written in Japanese doubled every 44 years, the number suddenly began to double every eight years. In other words, Japan's emergence from its Malthusian equilibrium is associated with a massive increase in the growth rate of codified technical knowledge. We observe a significantly sharper increase in translated technical books. Japanese translators had only succeeded in translating 8 Western technical books between 1500 and 1860; by 1900, they had translated 608 books. As the figure shows, the growth rate of new technology entering Japan changed suddenly and sharply after the government produced English-Japanese dictionaries and subsidized technology absorption.

How important was the supply of codified knowledge for absorbing Western technology in Japan? Historical evidence suggests that the translation of Western technical books played a central role in developing one of Japan's most important nineteenth-century industries: cotton textiles. Consider the story of Ishikawa Masatatsu, who established Japan's first cotton textile mill. Horie (1960) reports that "while employing [Ishikawa as an advisor], the lord [Shimazu Nariakira] showed him a book. Because it was in English, he sent it to Nagasaki for translation into Dutch, and it turned out to be a book on the cotton spinning industry. The attention of the lord, who had been previously interested in machine spinning, was abruptly caught by the book, and the plan for building a cotton spinning mill was made... [Thus began] the Kagoshima Cotton Spinning Mill, the forerunner of the modern spinning industry in Japan, which began operation in 1867." Braguinsky (2015) reveals that translating the book from English, a language Ishikawa never learned, into Dutch, a language Ishikawa understood, took a whole year. As one can tell from the passage, without English-Japanese dictionaries, technical books could often not be translated directly into Japanese. This roundabout means of learning technology meant that it took eleven years from the time that Ishikawa was hired by Shimazu as a technical advisor before he could establish the mill. It is a clear case of language differences raising technology access costs and translation lowering them.

One can find numerous other examples of Japanese entrepreneurs utilizing books to inform their investment and production decisions. For example, Tamagawa (2002) writes, "Many large mills such as Osaka, Kanegahuchi and Kurashiki and others founded their own training schools for male workers, teaching fundamental spinning technology. The textbooks and the study aid books on cotton spinning were mainly translated versions of the Platt Bros.' catalogues and instructions."[12]

[11] Although "Dutch learning" was considered significant during the Tokugawa period, translating Dutch technical books appears to have been rare. Japanese libraries typically do not specify the original language of their translated books, but we found 463 foreign technical books published before 1870 in their collections. This list likely includes books purchased both before and after 1870. Among these, only 23 are in Dutch and one in Chinese. Since the importation of books in other languages was prohibited during most of the Tokugawa period under penalty of death, these figures give us an upper bound on the number of foreign technical books that entered Japan. As we will see in Figure 5, part of this is likely explainable by the paucity of technical books written in Dutch before 1870.

[12] Meade (2022) provides details on the specific machines and methods described in translated books on textiles: "Translated works tended to be used in industry and in other training institutions where teaching was not carried out by foreign employees. Reflecting its importance to the Japanese economy at the time, textiles were the focus of a considerable number of translations during the Meiji Period. Among translations on textiles were Meriyasu Orikata (1873), a translation of Dana Bickford's Illustrated Instructions for Setting Up and Running the Bickford Family Knitting Machine (1871), Senko Shinsho Kagaku Jikken (Chemical

3.3 Codification in the West, the periphery, and Japan

Although Japan was not unique in translating technical books, Japan was unique among countries in the periphery of the IR in the scale of its translation project. We demonstrate this by constructing a novel database of the amount of codified technical knowledge available in local vernaculars each year for 33 languages, encompassing the 20 languages with the most speakers. We define the set of books containing technical knowledge as those with a subject that can be classified as applied sciences, industry, technology, commerce, and agriculture. We exclude books on theoretical technical knowledge, such as books in the hard sciences or subjects that do not directly benefit firms (e.g., medicine). After defining a common set of subject codes, we scraped the catalogs of national or other major libraries for books in the vernacular published in each year and report cumulative totals for each language (See Appendix K for details).

For many major European and Asian languages (e.g., English, French, German, Chinese, and Japanese), we scraped the national libraries of countries where the language is the native tongue of a substantial fraction of the population (so we restrict ourselves to only using the National Diet Library collection for Japan in this plot). For many other languages (such as Arabic and Russian), we could not find a scrapable national library. Instead, we scraped WorldCat, an online catalog of over 15,000 libraries worldwide covering dozens of languages. Using the publication year of each book in our sample, we construct the time series of codified knowledge by spoken language. This yields what, to the best of our knowledge, is the first systematic dataset on codified technical knowledge available in the vernacular for major languages.

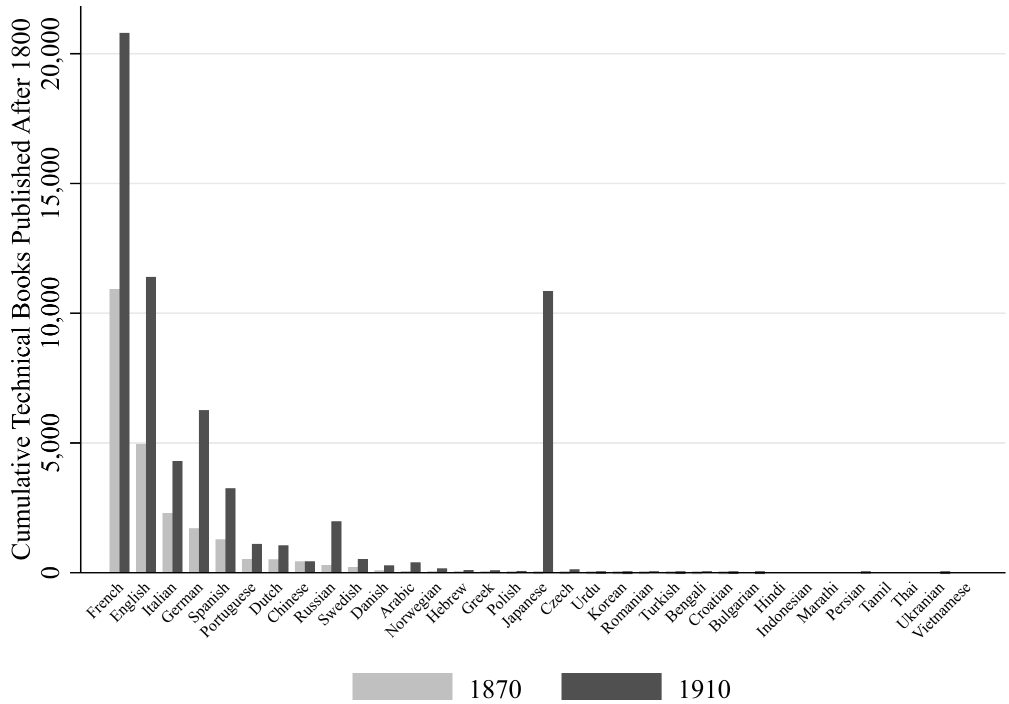
Despite Japan's rapid progress in codification at the end of the nineteenth century, the level of Japanese codification in 1870 was quite ordinary. Figure 5 presents the extent of codification of technical knowledge in 1870 and 1910. Two features of the data stand out. First, in 1870, 84 percent of all technical books were written in four languages: English, French, German, and Italian.[13] This puts into comparative perspective the achievements of the Enlightenment. There was little to no codification in any non-European language, meaning that people who could not speak European languages were technically illiterate. Second, this figure also puts into comparative perspective the achievements of Japan's public provision of technical knowledge *after* 1870. Starting from a level in 1870 that was comparable to other languages spoken in the global periphery, by 1910, Japan had amassed a body of codified technical knowledge comparable to that of the languages of the Enlightenment. Moreover, nowhere else in the global periphery do we see a similar increase in codified technical knowledge. This suggests that outside of the set of countries contributing to the Enlightenment, codified technical knowledge only appeared with state intervention.

These findings motivate our empirical strategy. Understanding the effects of codifying technical knowledge on economic outcomes such as productivity growth is typically difficult—codification in most settings happened gradually over time, making it hard to rule out other explanations. The swift creation of codified technical knowledge in Japan, however, creates two relatively well-demarcated periods: one in which Japan had no access to technical knowledge, and one in which it did.

Experiments in Dyeing, 1878), Seiyo Senshoku Ho (Western Dyeing Methods, 1878) and Seiyo Sarasa Senhosho (Western Cotton Dyeing Methods, 1879)." [pp. 13-14]

[13] We probably underestimate codification in English because we could not scrape the British Library, which suffered a cyber attack in 2023, resulting in security measures that prevented us from scraping it. Our data for English is from the Library of Congress, which provides a measure of technical books written in English and available outside of England.

Figure 5: Codified Technical Knowledge in Major World Languages



Note: Technical knowledge is measured as the number of books in the following subjects: agriculture, applied sciences, commerce, industry, and technology. The languages are ordered based on the total number of technical books published up to and including 1870.

3.4 Investment in Absorptive Capacity: Education Policies

Beyond spending on technology policies directly, the Meiji government also deployed education policies. As discussed above, these investments increased the absorptive capacity of the Japanese economy by creating a large network of individuals who could read and understand the codified knowledge. Compulsory elementary school education began in 1872, although most Japanese parents refused to send their children to government schools because, in the words of an 1877 Ministry of Education report, the "people do not yet see education as useful and parents are complaining" (Rubinger, 2000, p. 170). Government pressure quickly overcame the anti-education attitude of non-elite Japanese. The fraction of boys and girls attending school rose from 39.9 percent of eligible boys and 15.1 percent of eligible girls in 1874 to 58.2 percent of boys and 22.6 percent of girls by 1879. By 1890, 90.6 percent of boys and 71.7 percent of girls were enrolled in elementary school (National Institute for Educational Policy Research, 2011). Since child labor was common during this period, many of these elementary school graduates would have been in the labor force by the time they were teenagers.

These schools offered high-quality education by the international standards of the day. Notably, the national curriculum stipulated that over eight years of elementary education, no less than forty percent of class time should be allocated to scientific subjects – more than any other country at the time (Itakura, 2009; quoted in Meade, 2024, p. 230). Rubinger (2000) argues that data from mandatory intake examinations for Japanese army conscripts provide a representative sample of young Japanese males that we can use to assess educational levels. If one defines literacy as being able to write a formal letter in Japanese as judged by the Imperial Japanese Army, new conscripts in all but one of Japan's forty-seven prefectures in 1909 had literacy rates above 90 percent. Mathematics education was equally impressive. Conscripts who had completed six years of education were expected to answer word problems that required them to know algebra in order to solve, and those with eight years of education were expected to be able to compute bond yields. In other words, by the 1880s, most young Japanese men could have read technical books.

The Meiji government also pursued a conscious policy of building a modern, Western-oriented technical higher education system, which was to aid industrialization by producing technically and scientifically educated technicians and engineers (Pauer and Mathias, 2022). The Japanese government faced a complex problem in building a technical higher education system because there were almost no Japanese with advanced knowledge of STEM fields. To alleviate this problem, the Japanese hired foreign instructors to design the curriculum and teach. For example, the Japanese Imperial College of Engineering, founded in 1871, was headed by a young Scottish engineer, Henry Dyer, who had studied under William J. M. Rankine in Glasgow. The language of instruction was English, and pupils studied from a canonical set of engineering manuals published by Rankine in Britain in the middle of the 19th century (Meade, 2024) (See Table A.2). The new universities specialized in practical disciplines like foreign languages, medicine, engineering, mathematics, physics, law, chemistry, and management (Ministry of Education, 1980). Additionally, there was an emphasis on practical training with students spending long internships in mines and other companies (Pauer and Mathias, 2022).

Educated engineers appear to have had an impact on the production technologies used in Japan. Braguinsky (2015) provides econometric evidence that technical training led cotton mills to adopt different production techniques. He compares methods used in cotton mills that employed at least one college-educated engineer (who would have had to read technical manuals as part of his training) with those that did not. Using Japanese data from 1893 and 1898, he finds that Japanese mills with engineers used more Indian and US cotton, implemented different cotton mixing techniques, and had higher capital-to-labor ratios, a higher female-to-male operative ratio, many more spindles, higher TFP, and greater returns on assets.

3.5 Investments in Tacit Knowledge: "Live machines" and study-abroad missions

Alongside Meiji Japan's extensive efforts to codify technical knowledge in Japanese, the government also pursued a set of complementary policies that were well-suited to acquiring tacit knowledge. As a Tokugawa government foreign purchasing agent noted, it was important not just to ship back "a dead machine but [also] a live machine [i.e., a foreign instructor]" (Jones, 1980, p. 125). This practice became the start of a major program in which the Japanese government hired 2,400 foreigners to come to Japan as instructors or advisors. The foreigners hired by the government provided Japan with 9,506 person-years of technical training, of which over half was deployed either in educational institutions or in ministries that oversaw the building of Japan's transportation, telegraph, postal, gas, electrical, sewage, and water-supply networks. Fifty-nine percent of the training was conducted by individuals whose native language was English, 17 percent by those from France and Belgium, and 13 percent by those from Germany. The revealed preferences of the Japanese government in choosing instructors suggest that they viewed instructors whose native languages were English, and to a lesser extent, French and German, as the primary sources of advanced Western technology. As we showed above, these languages contained the most codified technical knowledge. The government not only financed foreigners to come to Japan but also paid for Japanese to study abroad. For example, in 1871, the government sent around one hundred officials and students abroad for two years on the "Iwakura Mission" to study Western society and technology and return with a wealth of knowledge and books. In addition to this high-profile study mission, the government subsidized many other Japanese to study abroad. Foreign study trips accounted for up to 0.20 percent of annual government expenditures in the 1870s (Jones, 1980, Table 7). By facilitating direct exposure to foreign expertise, foreign production facilities, and foreign technologies, the Japanese government lowered the access costs of tacit knowledge.

3.6 Paying For The Technology Transfer Policies

One may wonder how the Meiji government was able to raise enough government revenue to fund these policies. Real government expenditures tripled between 1871 and 1874 (See Appendix A for details on the sources for numbers presented in this section). Paying for the foreign workers alone required substantial expenditures—equalling about 2 percent of total government expenditures in 1876, one-third of the University of Tokyo budget, one-half of the Ministry of Education budget, and in 1879, two-thirds of the public works budget (Jones, 1980, p. 13). The key to Japan's newfound ability to pay for these programs was the 1873 Land Tax Reform Ordinance, which Japanese economic historians have called "the single most important reform of the Meiji Restoration," (Hayami, 1975, p. 47). Interestingly, the idea of instituting a land tax had its origin in the work of government translators in the 1860s. As Yamamura (1986) discusses in detail, Kanda Takahira, a high-ranking Meiji official who had translated a book on economics in the Tokugawa period, realized that Japan could raise enormous amounts of tax revenue with limited efficiency loss by instituting a heavy land tax, as opposed to the earlier output tax.[14]

One way to gauge the magnitude of Japan's investment in learning is by comparing it to China. By 1884, the land tax measures had given Japan an eight-to-one advantage in per-capita taxation relative to China, which enabled Japan to finance human capital investments and public goods at a rate that Chinese reformers could only dream about.[15] Put differently, Japanese government expenditures (taken from Ohkawa et al. (1965)) on education alone amounted to 11 percent of the budget in 1880. In other words, if China had attempted to implement only the education component of the Meiji reform package, it would have had virtually nothing left over for any other government functions.

4 Data

One of our main contentions is that, relative to other regions in the periphery, Japan made a uniquely large investment in codification that enabled its firms to rapidly assimilate British technology and increase productivity in the sectors that stood to gain from IR technologies. Testing this hypothesis requires us to construct several novel datasets. In this section, we describe the primary datasets used in the empirical analysis. First, we discuss how we quantify the British supply of codified technical knowledge by sector. Second, we create a bilateral industry export dataset that enables us to examine whether codification shifted comparative advantage towards sectors that benefited most from IR technologies. The appendix contains a complete discussion of

[14] Although agricultural taxes before the Meiji period were called "land taxes," they were closer to output taxes in implementation. As Ohno (2018, p. 39) writes, "A new land tax at the initial rate of 3 percent of the assessed land value replaced the old rice tax that was levied on the annual yield of rice."

[15] Wong (2012) reports that Chinese tax revenue in 1884 was 77 million silver taels. We converted it into yen in two ways. The number in the text uses the exchange rate series from Fouquin and Hugot (2016) of 1.39. We obtain a similar estimate if we convert silver taels into yen by noting that an 1867 Shanghai silver tael contained 36.0 grams of silver and an 1876 silver yen coin contained 24.3 grams of silver, according to https://en.numista.com. This implies an exchange rate of 1.48 yen per tael.

all data used, including all data construction steps and sources.

4.1 Constructing the British Patent Relevance measure

A key challenge for this paper is to quantify the supply of codified technical knowledge available to Japan and other regions by industry. We utilize textual information from one of the primary channels through which codified technical knowledge was disseminated during this period: the translation and publication of technical manuals. Nineteenth-century technical manuals give detailed, practical descriptions of the technological and organizational aspects of an industry. Their audience was the practitioner, the entrepreneur setting up a plant, or the manager overseeing production. Their value lay in the fact that they contained precisely the type of technical knowledge entrepreneurs would need to familiarize themselves with for the setting up of modern, factory-based manufacturing, as well as for its day-to-day operation.

Our aim is to quantify the amount of new, IR technical knowledge contained in these technical manuals that could be used in different industries. The main textual source we use to identify *codifiable* IR technologies is the corpus of British patent synopses (1780-1852). Using natural language processing (NLP), we compare the textual similarity between IR patents and technical manuals within each industry. If IR technologies are more relevant for an industry, then the jargon used in the corpus of IR patents should be more similar to the jargon used in manuals used to describe production techniques in that industry.

Before describing each data source and the NLP pipeline below, we first make four broader comments about our approach. First, given industry differences in the propensity to patent (Moser, 2005), our choice of patent synopses as the corpus of codifiable innovations requires some justification. We focus on patents as a proxy for innovation because we are interested in capturing the supply of *codified* innovation, which can diffuse at a distance, as opposed to all innovation. Previous work has argued that technology diffuses when the patent system creates a market for innovation – in fact, one rationale for the patent system is to diffuse technologies (Lamoreaux and Sokoloff, 1999). There is some supporting causal empirical evidence showing that patent protection leads to greater geographic diffusion of innovation (Moser, 2011).

Consistent with patents being an important direct source of knowledge for technological follower countries, the Japanese Patent Office collected summaries of British, U.S., French, and German patents (Smethurst, 2007).[16] Importantly, measuring the supply of codifiable innovations in this way excludes two types of innovation: tacit knowledge and knowledge protected by secrecy. We view both omissions as features of our approach to measuring codifiable knowledge, rather than bugs. Tacit knowledge cannot, by definition, be codified; hence, it cannot be transmitted via technical translation. By its very nature, knowledge protected via secrecy is also much more difficult to diffuse at a distance (as shown for the chemical industry in Moser (2011)).

Second, our focus on *First* Industrial Revolution technologies (as opposed to newer technologies that were available by 1880) is motivated by the technology adoption lags literature, which documents substantial cross-country lags in technology adoption. On average, countries adopt technologies with a 45-year lag relative to their invention (Comin and Hobijn, 2010). For Japan and the rest of the global periphery, the period between 1880 and WWI (our sample period) was a time when entering and mastering IR technologies remained the primary goal (e.g., DeLong (2022)). To take just one prominent and well-studied example, in 1911, Britain and the US alone accounted for

[16]For example, Doi (1980, p. 2) reports that a Japanese commissioner of patents visiting Washington in 1900 to study the U.S. system stated: "We have looked about us to see what nations are the greatest, so that we can be like them... We said, 'What is it that makes the United States such a great nation?' and we investigated and found that it was patents."

61% of installed factory spindles in cotton spinning, while the entire global periphery accounted for only 22% (calculations based on the U.S. House of Representatives, Tariff Board (1912)). That is, getting factory-based, mechanized textile manufacturing off the ground (which England achieved early in the 19th century) remained a central goal for most periphery economies well into the 20th century. In Appendix B, we discuss additional historical evidence that shows why Japanese industry was not ready to absorb newer technologies until it had mastered IR technologies.

Third, we focus on *British* patents because Britain is generally perceived as the technological leader during the IR (Broadberry, 1994; Crafts, 1998; Rosenberger et al., 2024).[17] The historical record also supports these findings for Japan. For example, Meade (2022) finds that the textbooks Japanese engineering students were reading in the 1870s and 1880s were heavily skewed towards British textbooks published in the middle of the nineteenth century. In addition, the Japanese government largely hired British instructors to teach in their engineering universities, and these instructors used British books.

Fourth, we focus on manuals written in English due to Britain's centrality in IR technologies, Japan's choice of instructors, and to prevent endogeneity bias from arising from Japanese translators translating more jargon in sectors particularly important to Japan. In addition, Shimizu (2010) reports that where data is available, a large majority of Japanese students studying foreign languages in public and private schools were learning English. Finally, the absence of *any* large French-Japanese or German-Japanese dictionaries published before the twentieth century would have made it relatively difficult for Japanese translators to translate technical documents written in French or German.[18]

4.1.1 Source Data For Technical Manuals

We hand-curated a sample of English-language technical manuals covering SITC3 Rev 2 industries published in the middle of the 19th century from *HathiTrust Digital Library*.[19] We tried to select a few books (or sections of books) for each industry that contained a detailed description of the techniques. We faced two issues when selecting books: sometimes they were more general than the SITC code, and other times they were narrower. We attempted to balance these considerations by using only relevant chapters when books were general, or selecting multiple books when manuals were narrow.[20]

Table 1 shows a random sample of the 460 books we selected from the *HathiTrust Digital*

[17]More precisely, while recent empirical evidence suggests Britain did not have the technological lead in *all* sectors (Hallmann et al., 2021), the sectors Britain did lead in were the ones more central in the innovation network, such as steam engines (Rosenberger et al., 2024).

[18]Garnier (2013) notes that "Between 1888 and 1905 no less than seven bilingual [French-Japanese] dictionaries were published: four French-Japanese and three Japanese-French. The three dictionaries published prior to 1900 remained small in size, containing between two and three hundred pages. Despite calling themselves 'dictionaries,' they more closely resembled glossaries which sought, according to the term used in their title, to present 'common words."' The first German-Japanese dictionary we could find in the NDL that contained more than 200 pages was published in 1904.

[19]Ninety percent of all technical books in our sample were written before 1883, and ninety percent of the books were published after 1838. We made exceptions only when we were unable to find a book within this time period. None of our books was published before 1806 or after 1900.

[20]For example, a treatise on metalliferous minerals and mining by D. C. Davies (1880) was relevant to both the copper and tin industries. In this case, we chose the chapter on copper mining for the copper industry and the chapter on tin mining for the tin industry. For copper, these chapters were then supplemented with books like Piggot's (1858) "The chemistry and metallurgy of copper" and Goodyear's 1865 translation of "A treatise on the assaying of lead, copper, silver, gold, and mercury," which was originally written in German.

Table 1: Random Sample of Book Titles from the *HathiTrust Digital Library*

SITC	Industry Description	Book Title
232	Natural rubber latex; rubber...	India rubber and gutta...
786	Trailers, and other vehicles,...	A complete guide for coach...
112	Alcoholic beverages	Hops; their cultivation,...
023	Butter	Butter, its analysis and...
764	Telecommunication equipment,...	The speaking telephone,...
882	Photographic and...	On the production of positive...
263	Cotton	Cotton in the middle states :...
274	Sulphur and unroasted iron...	A theoretical and practical...
271	Fertilizers, crude	American manures; and...
897	Gold, silver ware, jewelry...	Diamonds and precious stones,...
098	Edible products and...	Peterson's preserving,...
898	Musical instruments, parts...	Musical instruments ...
553	Perfumery, cosmetics, toilet...	A practical guide for the...
212	Furskins, raw	The trapper's guide: a manual...
046	Meal and flour of wheat and...	The American miller, and...
844	Under garments of textile...	Garment making a treatise,...
641	Paper and paperboard	Paper & paper making ancient...
664	Glass	The art of glass-blowing, or,...
268	Wool and other animal hair...	Sheep husbandry; with an...
061	Sugar and honey	The Chinese sugar-cane; its...

Note: This table provides a sample of books we used describing the technology in each industry. We randomly picked 20 industries, and for each industry, we randomly picked one of the books assigned to it.

Figure 6: Word Clouds for Textile Yarn

(a) Unigrams (b) Bigrams

Library.[21] We use the full text of these technical manuals to represent frontier knowledge of codifiable production techniques. Figures 6 and 7 illustrate the type of information we collect, using word clouds for unigrams and bigrams in two industries: textile yarn, and fuel wood and charcoal. Reassuringly, high-frequency unigrams and bigrams contain words associated with the production processes. The most common unigrams in books explaining textile yarn production include words like "spindle," "shaft," and "card," and common bigrams are "front roller," "driven pulley." In contrast, the unigrams and bigrams used in technical manuals about fuel wood and charcoal production are words and phrases like "billet," "hearth," "coal process," and "smoke vent." Finally, words used in patents, presented in Figure 8 capture the fact that patents discussed innovations like steam "engines," "weaving," and "spinning." As one can see from comparing the word clouds for textile yarn and patents, unigrams like "silk," "spin," "flax," and "wool" appear frequently in both textile yarn manuals and patents, but jargon used in fuel wood and charcoal are rare in patent texts.

4.1.2 Source Data for Patents

We digitize the synopses of all British patents issued between 1780 and 1852 from Bennet Woodcroft's (1857) "Subject Matter Index of Patent of Invention" (see Appendix I for details). This period covers the majority of what is generally considered to be the period of the IR in Britain (Allen, 2011).

We check the robustness of our results to using British patents from later periods (1853-1879) using the full text of British patents 1853-1879 from Coluccia and Dossi (2025). We also conduct robustness exercises using U.S. patent text from the period 1836 (the earliest date available) to 1879.[22] We webscraped the patent text from *Google Patents*, which provides digitized versions of all U.S. patents building on the tool developed by Kelly et al. (2021). Appendix I contains a detailed

[21]We built our BPR measure by collapsing the year dimensions in the book sample, treating the data as a cross-section.

[22]We also tried to include German patent data to explore another potential technological frontier. Unfortunately, German patent descriptions are only reliably available starting from 1877, which means there is insufficient coverage before 1880 (the beginning of our trade data). Additionally, the available German patent data showed large yearly fluctuations, raising questions about their reliability.

Figure 7: Stemmed Word Clouds for Firewood and Charcoal

(a) Unigrams

(b) Bigrams

Figure 8: Stemmed Word Clouds of Patent Synopses

(a) Unigrams

(b) Bigrams

description of each data source.

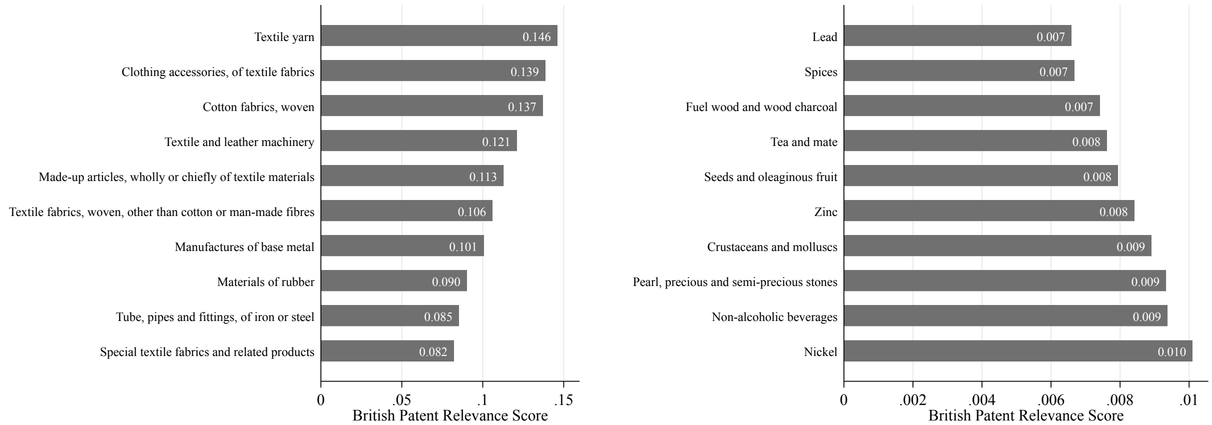
4.1.3 Quantifying the supply of technical knowledge by industry

We use NLP to quantify the amount of new technical knowledge created by the IR in each industry. This consists of two main steps (see Appendix I for a complete description).

First, we represent the textual information contained in technical manuals and patent synopses as data by taking a vector representation of the text. This involves stemming the words, dropping stop words, and dropping common words (see Appendix I for details). Each set of manuals associated with industry i or the set of patents is represented by a vector of length n, where n is the vocabulary size across the entire corpus. The vocabulary includes unigrams and bigrams and employs the term frequency-inverse document frequency (TF-IDF) weighting to account for the fact that some words appear more frequently across all documents.

Second, we quantify the relevance of IR technologies for the production processes described in the technical manuals for an industry. We assume that if an industry's manuals use words and

Figure 9: Industries Ranked by British Patent Relevance



(a) Top 10

(b) Bottom 10

phrases that are similar to those in British patent synopses, IR technologies are likely relevant for the industry. The standard metric for measuring the similarity of two texts (documents) in NLP is cosine similarity, which motivates our adoption of this measure. In our setup, it equals the cosine of the angle between the vector representation of the word frequencies in manuals and the frequencies in patent synopses. Formally, the cosine similarity between the vectorized Bennett Woodcroft patent text $\boldsymbol{BW}$ and the vectorized technical manuals $\boldsymbol{TM_i}$ for industry i is

$$BPR_i \equiv \frac{\boldsymbol{BW} \cdot \boldsymbol{TM_i}}{\|\boldsymbol{BW}\| \|\boldsymbol{TM_i}\|} = \frac{\sum_{j=1}^{n} BW_j TM_{ij}}{\sqrt{\sum_{j=1}^{n} BW_j^2} \sqrt{\sum_{j=1}^{n} TM_{ij}^2}}, \tag{1}$$

which we call *British Patent Relevance (BPR)* because it measures the relevance of knowledge contained in British patents for each industry. Figure 9 plots the bar chart for the industries with the ten highest and lowest cosine-similarity scores. Reassuringly, high BPR industries include textile, footwear, machinery, and manufactured intermediate-input sectors, whereas low BPR industries contain mostly unprocessed raw materials, which were largely unaffected by IR technology.

Our measure has several advantages for our setting. Most importantly, focusing on knowledge codified in technical manuals captures one of the key channels through which nineteenth-century Japan acquired Western knowledge: the translation of these documents. Moreover, this measure naturally accounts for how a given technology benefited different industries through input-output linkages. For example, since industries that make use of steam engines will likely have technical manuals that use the bigram "steam engine," our cosine-similarity measure will naturally quantify which industries benefited more from steam engines—a distinct advantage relative to using the industry classification of patents as a measure of relevance which only match the final output sector with patents about making that final output.

4.2 Cross-region, bilateral, industry level trade flows

We construct the first cross-region dataset of harmonized, bilateral, industry-level trade flows quinquennially for 1880-1910 using detailed historical trade records for Japan, the United States, Belgium, and Italy ("reporting countries," henceforth).[23]

We combine existing, region-specific data sources and add newly digitized trade data from various sources. Specifically, we digitized data on US trade flows (exports and imports), Japanese exports for 1875, and quinquennial Japanese imports between 1875 and 1910. We use existing data on Belgian exports and imports in manufacturing from Huberman et al. (2017); on Japanese exports from Meissner and Tang (2018); and on Italian exports and imports (for major trading partners only) from Federico et al. (2011).[24] An observation in this dataset, x_{ijkt}, refers to an export flow in sector k from origin i to destination j reported in year t. Using the fact that an export flow from i to j is equivalent, in theory, to imports from j to i, we can use import flows from reporting countries for unobserved regions' export flows.

Japanese trade data does not include its colonies, so Japanese territorial expansion over this period does not affect our results. We define the set of non-Japanese Asian Regions (ASIA) as French East Indies, Hong Kong, China, Korea, Portuguese East Indies, Siam, Straights Settlements, and India. We used the Maddison data to divide the set of *non-Japanese* exporters into three terciles—High (H), Medium (M), and low (L)—according to estimated GDP per capita in 1870. For regions that do not correspond to modern countries, we use the average GDP per capita of the countries in that region.

We harmonized product lines in a manner consistent with the other pre-existing data sources used in the dataset. We conducted extensive validation exercises to ensure that similar product lines were consistently concorded to the same three-digit SITC category across all datasets (see Appendix G.1 for details). Region names (and boundaries) were harmonized within and across datasets. All trade values were converted to yen (at current exchange rates) using historical exchange rates from Fouquin and Hugot (2016). Our dataset consists of export values for 37 regions in 93 industries. Appendix Table A.3 contains the summary statistics.

5 Codification and Development

The previous sections established that 1) Japan experienced unique shifts in its export composition between 1880 and 1910, mainly driven by its manufacturing sectors, and 2) Japan was unique among peripheral economies in providing its citizens with access to codified technical knowledge in their vernacular. This section presents empirical evidence consistent with a causal relationship between these two aspects of Meiji Japan's economy.

Our empirical approach relies on industry-level variation in the extent to which the codification of technical knowledge affected export growth. Intuitively, would-be entrepreneurs of textile yarn, which had undergone enormous changes in production methods during the FIR, had large productivity benefits to reap from access to technical knowledge. In contrast, producers of raw commodities such as nickel, zinc, or lead—the production of which was barely affected by IR technologies—had far fewer productivity benefits to reap from reading technical knowledge. We

[23]Recent years have seen a proliferation of high-quality, cross-country, bilateral trade datasets (see, e.g., Fouquin and Hugot 2016; Pascali 2017; Xu 2022). Yet because these data are not disaggregated by industry, they cannot be used for our purposes.

[24]We do not include Germany's digitized trade data in the combined dataset because Germany's historical trade statistics before 1906 present several distinct methodological challenges that make comparisons over time and across countries difficult (Hungerland and Wolf, 2022). First, until 1888, some parts of the German Empire were not part of the German customs union and maintained their own records, making it challenging to construct a single dataset covering all German trade. Second, during our sample period, the classification scheme for products was repeatedly revised: at different points in time, between 400-1,200 distinct products were listed, making it difficult to construct a consistent classification over time (see Hungerland and Wolf (2022, Figure 6A)).

operationalize the extent to which an industry could benefit from access to technical knowledge using the British Patent Relevance (BPR) measure introduced in Section 4.1.

5.1 Cross-Sectional Evidence

We test this relationship by estimating regressions of the form

$$g_{ik} = \alpha_i + \beta_J * BPR_k \times I_{iJ} + \beta_r * BPR_k \times I_{ir} + \epsilon_{ik}, \tag{2}$$

where g_{ik} is the average annual export growth in region i and industry k; α_i is an exporter fixed effect; BPR_k is the British patent relevance measure for sector k; I_{iJ} is a dummy that equals one if i is Japan; I_{ir} is a dummy that equals one if i is part of some other regional grouping r; β_J and β_r are estimated parameters; and ϵ_{ik} is an error term. We partition the regions in our sample into mutually exclusive regions to probe potential confounders.

Since BPR_k is *not* Japan-specific, our measure of British Patent Relevance captures the world *supply* of technical industry-level knowledge. This is important, as our measure is not based on what was written in Japanese, which would be endogenous if the government or entrepreneurs strategically generated knowledge for sectors more likely to succeed.

We hypothesize that $\beta_J > 0$; that is, Japanese industries that benefited more from knowledge codification experienced faster export growth in Japan. Column (1) in Table 2 shows the results from estimating this regression for just the sample of Japanese industries. Appendix Figure A.7 presents the corresponding scatterplot. Consistent with our hypothesis, industries with a higher BPR_k experienced faster export growth during the sample period. The coefficient is both economically meaningful and highly statistically significant. Since we standardize BPR in all specifications, our estimates imply that a Japanese industry with a one-standard-deviation higher British Patent Relevance measure experienced export growth that was 12 percentage points per year faster. These large effects help account for the sudden shift in Japanese exports from primary products to manufactured goods.

A causal interpretation of the parameter of interest, β_J, requires that BPR_k is uncorrelated with the error term ϵ_{Jk}. The main concern in this context is omitted-variable bias: unobserved factors correlated with BPR_k drive the pattern of productivity growth in Japan. For example, it may be that British Patent Relevance mattered not just in Japan but in all regions during the nineteenth century. Alternatively, it is conceivable that BPR_k is correlated with distance to the technology frontier. In this case, Japan's experience might be similar to that of medium- or low-income countries. Finally, it is also possible that some other Japan-specific factors, such as fundamental comparative advantage or institutional reforms implemented during the Meiji Restoration, are correlated with BPR_k. We tackle the various threats to identification using three strategies.

First, we exploit the fact that we know which countries codified and can measure industry performance around the world to examine the relationship between BPR_k and export growth in a pooled sample that includes all industry-region pairs for which we have data. Column (2) shows that, on average, other regions in our sample did not exhibit the same pattern of export growth that we found for Japan. Export growth in this period is *negatively* and statistically significantly correlated with BPR_k in countries other than Japan. In other words, the association between Japan's export growth pattern and the amount it stood to learn from British patents is not common to all regions.

Second, if codification was the key difference across countries, we should expect a positive relationship for regions with access to codified knowledge. Indeed, in columns (3) and (4), we see that British Patent Relevance is significantly correlated with faster export growth rates in regions speaking languages with the largest numbers of technical books (English and French) relative to

Table 2: Annualized Export Growth and British Patent Relevance

	Export Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BPR × Japan	0.121***	0.121***	0.121***	0.121***	0.121***	0.121***	0.121***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
BPR × Not Japan		-0.030***	-0.037***	-0.037***	-0.070***		
		(0.009)	(0.010)	(0.011)	(0.016)		
BPR × English-Speaking			0.042**				
			(0.020)				
BPR × French-Speaking				0.032**			
				(0.016)			
BPR × Top-4 Codified					0.078***		
					(0.018)		
BPR × High-Income						-0.460	-0.460
						(0.307)	(0.307)
BPR × Medium-Income						-1.166	-1.024
						(0.748)	(0.764)
BPR × Low-Income						-2.093***	-1.410*
						(0.680)	(0.850)
BPR × Asia							-1.443
							(1.094)
Observations	71	1395	1395	1395	1395	1395	1395
R^2	0.112	0.233	0.234	0.234	0.243	0.236	0.236
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for a region i's industry k between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance" and captures how relevant British patents are to the vocabulary used in the manuals of industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. "Japan" is a dummy variable that equals one if the region is Japan and zero otherwise. "Not Japan" is analogously defined. "English-speaking" is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the 4 most codified languages: French, English, German, or Italian. {High, Medium, Low}-Income are dummies that equal one when a region is in the top third of the income distribution (High), middle third (Medium), or bottom third (Low) according to the Maddison per capita GDP data; we set these dummies to 0 for Japan. The Asia dummy equals one if the region is in Asia and zero if it is Japan or not in Asia. Robust standard errors in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

those that do not. In column (5), we try to deal with the imprecision of estimating the impact of codification by pooling across the four languages with the most codification in 1870 (French, English, Italian, and German) and find that Japan and the other top-4 codifiers all experienced faster export growth in sectors that stood the most to gain from codified technical knowledge relative to regions that did not codify. Taking these results together, we conclude that Japan and European codifiers' export growth shows similar patterns.[25]

Third, we explore the possibility that Japan's experience was a product of its income level or geography in columns (6) and (7). We group countries by income tercile (column 6) and isolate Asia (column 7). No region group displays a similar productivity pattern. On the contrary, the poorest countries, particularly in Asia outside of Japan, show a negative correlation, though the patterns are never consistently statistically different from zero. In summary, the pooled specifications suggest that Japan's export growth pattern was unusual for a peripheral economy but closely aligned with other codifiers. Regional trends or structural factors, such as distance to the technology frontier, are unlikely to explain the relationship.

5.2 Time Series Evidence

The cross-sectional evidence established that Japan's pattern of technological acquisition was similar to that of other codifiers but differed from that of other countries in the periphery, but it leaves open the question of whether this difference was due to codification *per se* or reflects an unobserved Japan-specific characteristic. For example, perhaps some unobserved common geographic characteristic, like coal deposits, explains the commonalities observed among codifying countries. Alternatively, Tokugawa literacy, culture, living standards, or any of a host of unique Japanese characteristics might explain their ability to absorb Western technology at this time.[26] Finally, many of the Meiji government reforms were gradually transforming the Japanese economy. Perhaps it was one of these and not codification that explains Japan's transformation.

Section 3 established that the change in the composition of exports towards manufactures happened rapidly and immediately after Japanese entrepreneurs had access to technical knowledge in their vernacular. While we have argued that the sudden timing of this change suggests it was not the result of a slow-moving process like the share of literate Japanese, the results of the previous section suggest that we could implement a more powerful test of the hypothesis that codification was crucial for understanding Japanese economic development. We showed that British patenting caused exports to grow *faster* in industries that heavily used IR technologies in codifying regions and *slower* in non-codifying ones. If codification is the key to Japanese development, then we should expect to see the same pattern in Japanese data as the country shifted from being a non-codifying country to a codifying one. However, if some uniquely Japanese characteristic explains the shift in Japan's export patterns, we should expect Japan's export pattern to shift at a constant or gradually changing rate throughout the sample period.

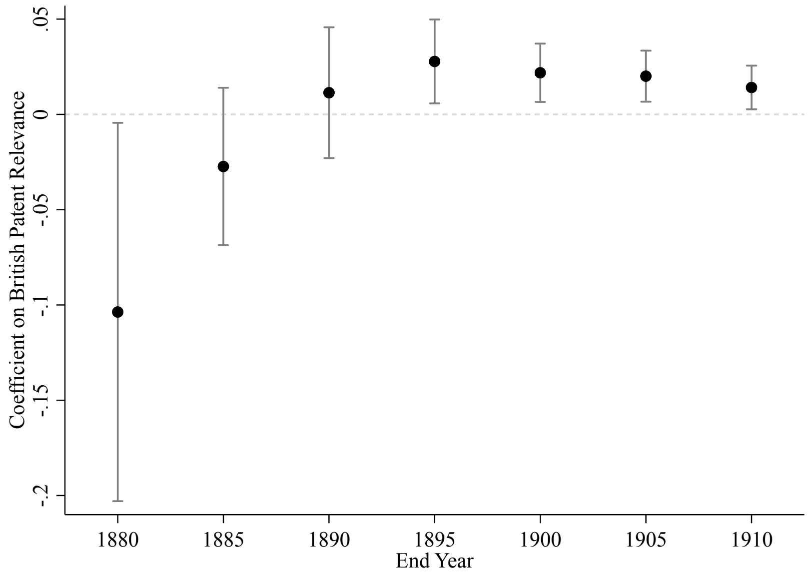
We now test whether Japan's sudden achievement of technical literacy coincides with a reversal in the relationship between export growth and BPR. It is not possible to pinpoint a specific year when Japan became technically literate, but the data suggest it likely occurred in the 1880s. For example, in 1880, there were only half as many technical books in Japanese as had existed in Spanish in 1870. Thus, in 1880, Japan was still a minor codifier compared to the major European codifiers. Between 1880 and 1890, however, the number of Japanese technical books in the National Diet Library grew from 706 to 2,823, surpassing the 1870 number of codified books in every European language except English and French.

We test how British Patent Relevance affected shifts in Japanese comparative advantage by

[25]The estimated coefficients for Japan are larger (although not always significantly so) than those for regions in which a plurality speaks a codified European language. This makes sense given that Japan's late industrialization implies it had more to learn from British patents.

[26]For example, some authors have pointed to Tokugawa literacy rates as a possible explanation for the ease with which Japan adopted foreign technologies in the Meiji period (Koyama and Rubin, 2022).

Figure 10: Coefficients from Regressing Japanese Export Growth on BPR by End Year



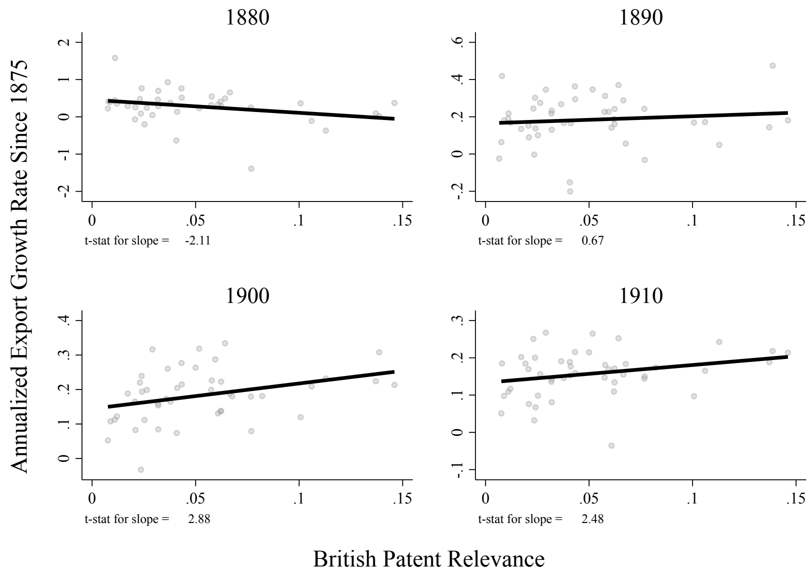
Note: We plot the estimated coefficient on BPR as well as the 95% confidence interval in a regression of Japanese industry export growth from 1875 to the year displayed in the figure on BPR.

regressing annualized Japanese industry export growth rates from 1875 to an end year that varies in 5-year increments starting in 1880. Figure 10 plots the estimated coefficients from these regressions, along with the 95% confidence intervals.[27] We interpret the specification for export growth between 1875 and 1880 as a placebo exercise that examines the relationship between export growth and BPR in the years before Japan achieved technical literacy. We obtain a negative and significant coefficient on BPR, indicating that Japanese export growth was slower in industries that benefited the most from the IR. This result is similar to what we saw in Table 2 for non-codifying regions. In other words, before Japan became technically literate, its export growth patterns resembled those of other countries in the global periphery: lower export growth in sectors with the highest potential to learn from the West.

This pattern flips around 1890, a point at which Japan was becoming a major codifier. By 1895, the coefficient of British patent relevance is positive and significant at the one percent level. Although the coefficient remains positive and significant at this level throughout the rest of our sample period, the point estimates slowly decline (though not significantly), consistent with the conjecture that by 1910, Japan was shifting to Second Industrial Revolution technologies and becoming less reliant on older IR ones. Figure 11 shows the scatterplots for 1880, 1890, 1900, and 1910 to show that outliers are not driving these results. The timing of this effect is hard to explain with conventional stories. We do not detect a significant impact of BPR on exports until 37 years after Japan opened to trade and 27 years after the Meiji Restoration. Neither of these stories can explain this basic pattern of Japanese industrialization—namely, why Japan's export composition shifted away from sectors that benefited most from technological progress before Japan became technically literate, and then shifted back towards them after.

[27] Appendix Table A.4 reports the estimated coefficients from the same specifications.

Figure 11: Japanese Export Growth and BPR by Decade, 1880-1910



Note: These graphs plot the annualized growth rate between 1875 and year X against British Patent Relevance.

6 Robustness

Alternative Measures of Patent Relevance One possible concern about our results is that British patents issued between 1780 and 1852 may not accurately capture the set of codifiable technologies Japan and other periphery economies were interested in adopting during the sample period. While our baseline choice is justified by the technology adoption lags literature and the historical record for Japan, here we consider the robustness of our results to using patents that capture the most novel technologies as of 1880, and those from a different country at the technology frontier by the late 19th century. Specifically, we construct similar measures for British patents issued during 1853-1879, and for U.S. patents issued 1836-1879.

We explore how our choice of which sample of patents to use affects our results in Table 3. Column 1 is identical to column 1 in Table 2, and we report it again for reference. Next, in column 2, we show that our results are very similar if we use the full text of British patents 1853-1879. Column 3 repeats the exercise, first summarizing each patent using a generative AI model so that the summaries are similar in length to the British patent synopses used for 1780-1852; the results are almost unchanged.[28] Finally, columns 4-5 show that the results are also very similar if we use early (1836-1860) or later (1861-1879) US patent text to construct the cosine similarity measure.[29] Note that while using later British patents, as well as U.S. patents, does not significantly affect our coefficient of interest, it does lower the R^2, suggesting that these innovations have less explanatory power for understanding the evolution of Japanese export growth. This is consistent with the

[28] Appendix I.5 contains details of how the patent text was summarized.

[29] Patent reforms in both Britain and the U.S. make it impossible to pool patents within a country for the full period we wish to consider. Specifically, we do not pool the entire corpus of British patents 1780-1879 because of a patent reform in 1852 that substantially decreased the cost of filing patents, and we do not pool U.S. patents 1836 (the first year they are available) - 1879 because of a patent reform in 1861. Appendix I.5 contains further information.

Table 3: Annualized Export Growth and Different Patent Relevance Measures

	Export Growth				
	(1)	(2)	(3)	(4)	(5)
BPR (1780-1852) × Japan	0.121***				
	(0.033)				
BPR (1853-1879) × Japan		0.121***			
		(0.036)			
BPR (Summarized, 1853-1879) × Japan			0.116***		
			(0.036)		
USPR (1836-1860) × Japan				0.111***	
				(0.036)	
USPR (1861-1879) × Japan					0.115***
					(0.038)
Observations	71	71	71	71	71
R^2	0.113	0.082	0.080	0.063	0.068
Country fixed effects	✓	✓	✓	✓	✓
Sample	Japan	Japan	Japan	Japan	Japan

Note: The dependent variable, "Export Growth," is the annualized export growth rate for a Japanese industry k between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance" and captures how relevant British patents are to the vocabulary used in the manuals of industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. USPR stands for "United States Patent Relevance", constructed in a similar way to BPR. The "Summarized" version used in Column (4) is explained in Appendix I. Robust standard errors are in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

notion that our baseline BPR measure captures the set of technologies most important for Japanese export growth.

Productivity We have shown that patents affected the export behavior of codifying and non-codifying countries differently, but we have not shown that it mattered for the productivity of industries directly. For example, it is possible that BPR is correlated with the demand for Japanese exports, rather than the supply, and therefore our results may not accurately reflect movements in Japanese productivity. Fortunately, Costinot et al. (2012) build a multisector Eaton and Kortum (2002) model that provides a mapping from industry productivity into exports. In their setup, if one regresses a country's bilateral industry exports on country-pair, importer-industry, and exporter-industry fixed effects, the coefficient on the exporter-industry fixed effect will be proportional to country-industry productivity. Following Costinot et al. (2012), we show in Appendix C.1 that these productivity estimates can be converted into measures of comparative advantage by netting out productivity shifts that affect all industries for a given exporter and industry productivity shifts that are common across all countries.

Using these productivity estimates, we show in Appendix C.3 that not only did Japan have exceptional manufacturing export growth between 1880 and 1910, but it also had exceptional national productivity growth, and this productivity growth was biased towards manufacturing sectors. Table 4 replicates Table 2 using productivity-based comparative advantage growth instead of export growth as the dependent variable (Figure A.8 shows the corresponding scatterplot for Japan). The results are qualitatively similar to what we found when using export growth as the

Table 4: Annualized Productivity Growth and British Patent Relevance

	Γ_{ik}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BPR × Japan	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
BPR × Not Japan		-0.000	-0.001	-0.001	-0.003		
		(0.001)	(0.001)	(0.001)	(0.002)		
BPR × English-Speaking			0.003				
			(0.002)				
BPR × French-Speaking				0.002			
				(0.002)			
BPR × Top-4 Codified					0.004**		
					(0.002)		
BPR × High-Income						-0.003	-0.003
						(0.034)	(0.034)
BPR × Medium-Income						0.050	0.079
						(0.079)	(0.081)
BPR × Low-Income						-0.111	0.022
						(0.083)	(0.096)
BPR × Asia							-0.277**
							(0.126)
Observations	56	1244	1244	1244	1244	1244	1244
R^2	0.074	0.010	0.011	0.011	0.013	0.012	0.016
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable is the annualized growth rate in comparative advantage for a region i's industry k between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures the extent to which the synopses of British patents are relevant to industry k. Japan dummy equals one if the region is Japan and zero otherwise. Not Japan is analogously defined. "Top-4 Codified" is a dummy for countries that speak one of the 4 most codified languages: French, English, German, and Italian. {High, Medium, Low}-Income are dummies are one when a region is in the top third of the income distribution (High), middle third (Medium), or bottom third (Low) according to the Maddison per capita GDP data; we set these dummies to 0 for Japan. The Asia dummy equals one if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors are in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

dependent variable. We find that BPR led to greater growth in comparative advantage in Japan and to deteriorations in comparative advantage in non-codifying countries. Between 1880 and 1910, we estimate that a Japanese industry with a BPR score one standard deviation higher than another experienced productivity growth that was 1.2 percentage points higher per year. Countries that codified also experienced significantly faster growth in comparative advantage in industries with higher BPR, although the point estimates for codification in English and French are not precisely estimated. In other words, the results are qualitatively the same as those we obtained for export growth.

Controlling for Additional Confounders We also tested the robustness of our results by checking whether they could be driven by correlations between BPR and colonization and industry steam intensity. In Appendix Table A.5, we control for whether the region was a British colony (as opposed to being a plurality English-speaking country), but the coefficient is not significant. An alternative confounder is that BPR may be correlated with an industry's steam-energy intensity, and our regressions are driven by Japan being a late adopter of steam power, which caused it to grow relatively quickly in steam-intensive sectors. We measure industries' steam-power intensity using French data from the 1860s, defined as the amount of steam power used in an industry divided by the wage bill (details of the calculation can be found in Appendix H.6). Table A.5 shows that controlling for steam power does not affect our results.[30]

Sample Selection Issues Appendix Table A.6 drops non-manufacturing sectors to show that the impact of BPR on Japanese export and productivity growth rates did not arise simply because of differences in manufacturing export growth relative to non-manufacturing export growth rates. We also demonstrate that our results for Japan are not driven by Japanese exports to any particular geographic region in Appendix Table A.7. Notably, our results hold excluding Asian destination markets which were the primary markets served by Japan (Meissner and Tang, 2018). Similarly, our results are not driven by individual high-growth sectors like textiles or iron and fabricated metal products (Table A.8). Together, these results indicate that broad-based changes were underway in Japan. These patterns are even visible within the manufacturing sector and are detectable even if we drop Japan's major export destination markets or its major export products.

7 External Validity

Our paper began with the observation that there are only four types of high-income countries: English-speaking countries, countries close to England, resource-abundant countries, and Japan and its former colonies. While it is beyond the scope of this paper to assess the external validity of our findings by conducting a similar analysis of codification on other countries, we can ask whether the Meiji model of codification was influential beyond Japan, and whether it could have had similar effects in other contexts.

In terms of its influence, there are two countries where we have direct evidence of the Meiji model of technology absorption being studied and transplanted: South Korea, under the Japonophile President Park Chung Hee, and China, under Premier Zhou Enlai (China's second-highest government official). Both men studied in Japan in their youth, and each was inspired by the Meiji economic model. Kohli (1994) writes that Park, who graduated fifth in his class from the prestigious Imperial Japanese Army Academy (during the period when Korea was a Japanese colony), was "fascinated by the 'Meiji model,' and bent on steering Korea along the Japanese path to modernity." Similarly, Zhou had spent two years studying economic development in Japan, and his views on technology policy were similar to those of Sakuma (presented at the beginning of this article), close to one hundred years earlier:

[30] Steam-engines were used outside of manufacturing too, where British Patent Relevance tends to be low. For example, the mining of coal and other natural resources is steam-intensive, but these industries do not have a high BPR_ks. This likely explains why our results are robust to the inclusion of a measure of steam-intensity.

> Doing science is like fighting a war... How can you fight this war? The foundation of our country's scientific and technological information work is very weak. The main task of intelligence work is to quickly establish institutions, train experts in intelligence work, comprehensively and on a timely basis collect, research and report on the development and new achievements of science and technology at home and abroad, especially in advanced scientific countries, so that national scientific work can understand these developments and achievements promptly. The specific method is to prepare for the establishment of specialized institutions, organize forces, engage in extracting papers from scientific and technological journals around the world, and compile, print, and publish. (Hannas and Chang, 2021, pp. 9-10)

Upon coming to power, each leader implemented technology policies which bore clear resemblance to the Meiji technology policies studied in this paper. For example, it is commonly argued that "the regime of Park Chung Hee, which lasted from 1961 to 1979, changed the course of modern South Korean science and technology." (Kim, 2011). Starting in 1962, Park developed a five-year technology plan based on the idea that "the promotion of science and technology is the most pressing issue for South Korea... Neither economic growth nor modernization can be achieved without progress in science and technology." [Park as quoted in Kim (2011)]. One of the cornerstones of his policy was the establishment of the Korea Institute of Science and Technology (KIST), which had a budget exceeding total university expenditures on science and engineering. The first director of KIST explained its mission: "At that time, no such institutes had researched the manufacturing technologies that companies demanded... In other words, there should be an intermediary unit for selecting, introducing, acquiring, and applying technologies." [Hyung Sup Choi as quoted in Thang (2018)]

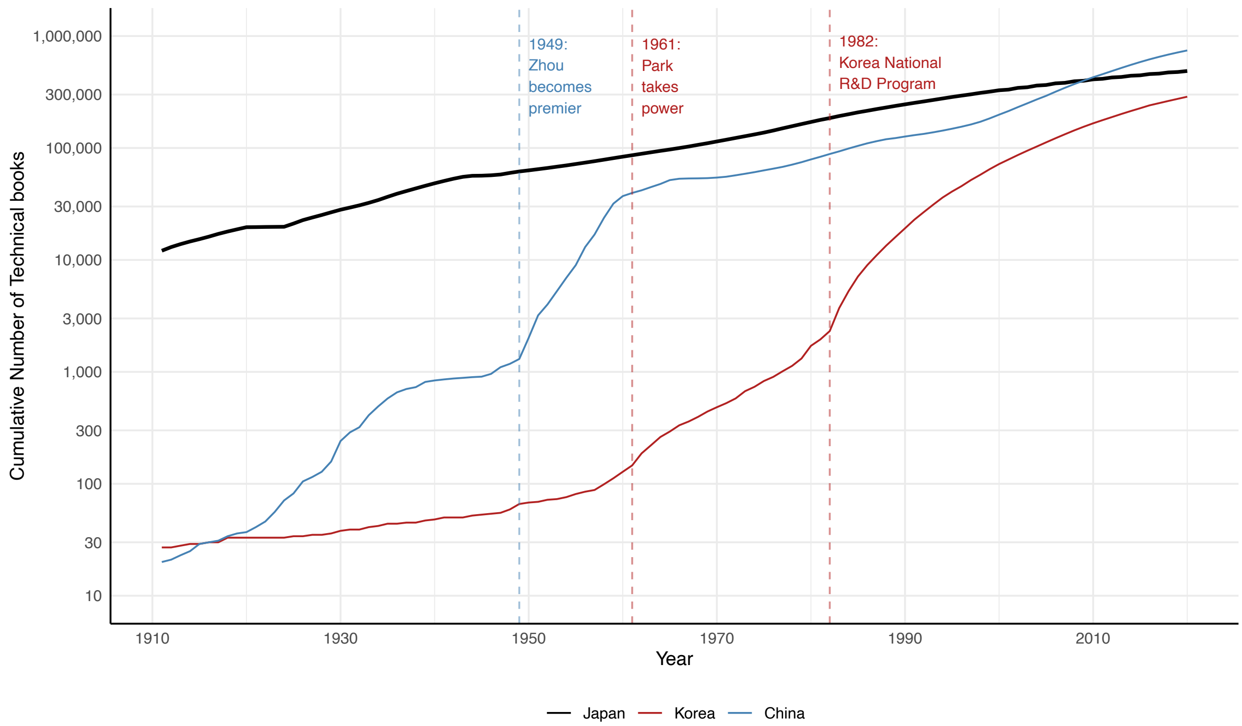
Fulfilling this mission involved an enormous expansion in the number of researchers in government research institutes and universities. Between 1965 and 1970, the number of researchers in government institutes rose from 2,135 to 5,628, and the number of university researchers rose from 352 to 2,011. After his assassination in 1979, Park's policy was supplemented in 1982 by the National R&D Project, which created government research institutes aimed at improving the level of domestic technology. As a result, the number of researchers in government research institutes rose from 4,598 to 7,542 between 1980 and 1985 and then doubled over the next decade (Chung, 2020).

Similarly, China's approach to technology adoption in the 1950s closely mirrored Meiji Japan's, with the Soviet Union playing the role of Britain. For example, Mengzhi (1999) writes that "after the foundation of the People's Republic of China, the government took emergency measures to train people with ability in foreign languages, especially Russian, to meet the needs of large-scale economic production... A large number of translated Russian textbooks were published to meet the requirements of various specialties." [p. 189]

We examine the impact of these policies on the publishing of technical books by scraping the national libraries of China, Japan and Korea.[31] In Figure 12, we plot the cumulative number of technical books available in the vernacular for each country. In 1950, Japan had about 70,000 technical books, while China had about 1,000, and Korea had fewer than 100. In China, the publication of technical books rose sharply the moment Premier Zhou Enlai became Premier, and by the early 1960s, there were over 30,000 technical books. The growth in technical books is also significant in Korea, and there is a clear acceleration in growth rates in both 1961, when President Park came to power, and 1982, when the National R&D plan was adopted.

[31]We start in 1911 because Maddison data on real per capita income in Korea in earlier years is based on

Figure 12: Cumulative Number of Technical Books in Japan, Korea, and China



Note: Codified technical knowledge for each year refers to all technical books written in Japanese in the NDL catalog, Korean in the National Library of Korea, or Chinese in Shanghai's Library. The y-axis is on a log scale.

Figure 13 shows per capita income growth for these three economies. We highlight two points. First, GDP per capita in South Korea clearly rises when Park comes to power and fosters codification efforts. While we do not interpret this relationship causally, it is at least plausible that codification played a role in the take-off of economic growth in Korea.[32] Second, Figure 13 reveals that codification in China failed to deliver long-run economic growth: by 1961, Chinese income was barely above where it had been in 1950. Moreover, despite robust growth in codification from 1960 onwards, Chinese growth rates did not pick up until after 1976.

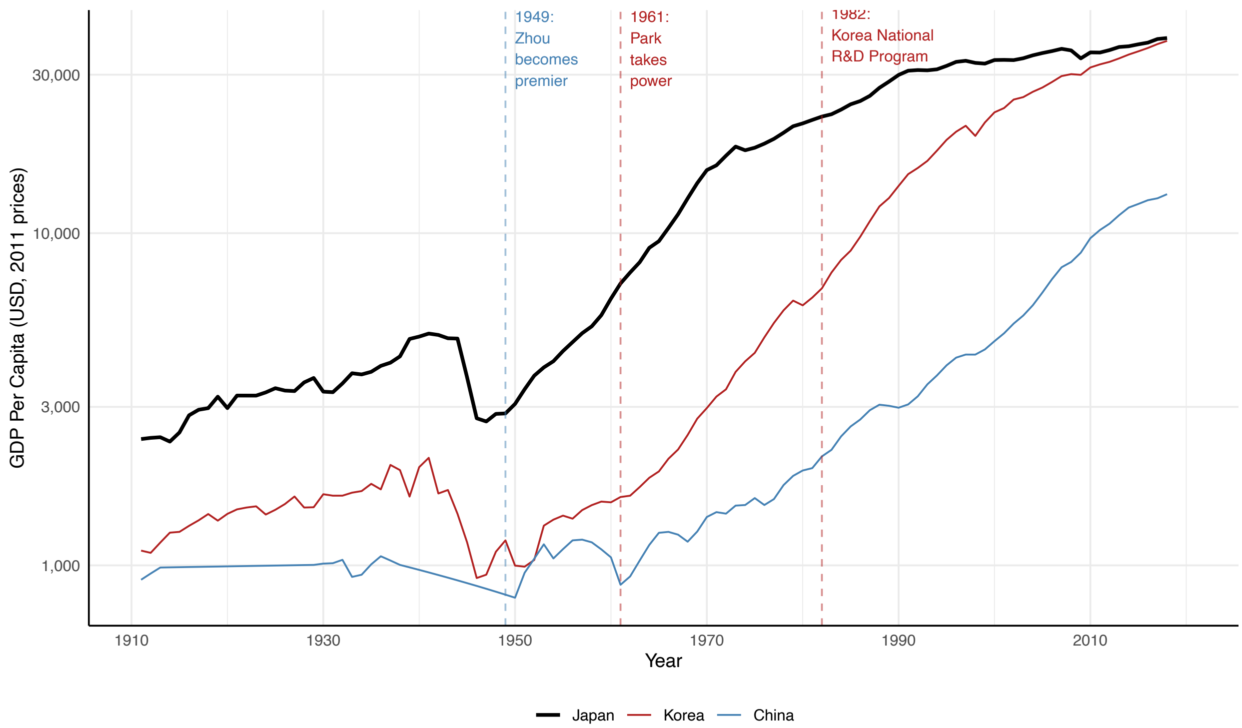
Mao's disastrous policies, such as the Great Leap Forward (1958-1962) and the Cultural Revolution (1966-1976), are likely the culprits behind why the Chinese people failed to benefit from their technical knowledge. Once the Cultural Revolution ended, it is plausible that codification and the gradual introduction of market forces into the economy led to a sustained rise in growth rates that persisted for 40 years.[33] In other words, China is the exception that proves the rule—its experience is consistent with our contention that codification is a necessary but not sufficient condition for development: totalitarianism can eliminate the benefits of codification.

extrapolating backwards.

[32]Recent work by Lane (2025) shows evidence for the growth-enhancing effects of other *sectoral* industrial policies adopted in Korea at this time.

[33]In related work, Giorcelli and Li (2021) find evidence that firms which directly received Soviet technical support during this period outperformed firms which did not.

Figure 13: Real GDP per capita for Japan, Korea, and China



Source: Maddison Project Database (2020). The y-axis is on a log scale.

The implication we take from this discussion is that the Meiji technology policies studied in this paper have been without a doubt influential beyond the borders of Japan. In fact, the discussion above shows that these policies were studied and adopted by two other East-Asia "miracle economies" (though China's economic miracle came much later than the initial state-led codification effort under Zhou Enlai). While there has been much debate and controversy surrounding the role of more standard *sectoral* industrial policies for these economies' growth miracles (see Juhász et al. (2024) for a recent discussion), our analysis suggests that the literature may have overlooked a different form of industrial policy adopted in a number of these countries: the public provision of technical knowledge in the vernacular. A fruitful direction for future research would be to estimate the causal effect of these technology policies in other contexts.

Notwithstanding the eventual influence of the Meiji model of codification, one may also wonder how our findings apply to other countries for the period examined in this paper. In Appendix D, we discuss how our findings inform the experience of two other late-industrializing economies: British India and Late Imperial Russia.

8 Conclusion

This paper presents evidence supporting the argument that the public provision of technical knowledge in the vernacular reduced an important friction that impeded the absorption of Western technology in Meiji Japan. Our results reveal an empirical pattern unique to Japan and other codifiers: industries that stood to benefit more from Western technology experienced faster growth in exports and productivity in relevant sectors. The cross-region evidence and the timing of when technical knowledge became predictive of industry growth in Japan provide a strong case for a causal interpretation of codification on Japanese export and productivity growth. By making technical knowledge widely available in the vernacular, the Meiji government relaxed a critical bottleneck for industrialization. This suggests that regions hoping to emulate European industrialization in the nineteenth century, particularly those linguistically or geographically distant from Western Europe, needed to provide complex public goods, such as access to technical knowledge, to emulate Britain successfully. Any alternative explanation of Japanese export and productivity growth must account for both its distinctive pattern across regions and its timing in Japan.

There is also suggestive evidence that these forces mattered for aggregate growth in Japan and other Asian nations. Both Japan and Korea experienced sustained increases in their average per capita growth rates only after embarking on significant codification efforts. However, our results also suggest that these public goods were unlikely to be sufficient in and of themselves to foster modern industrial development. China embarked on a similar codification strategy in 1949, but it did not experience sustained fast growth until after it abandoned Maoist economic policies. This buttresses our contention that codification may have been necessary but not sufficient for development.

The obvious question is why the Japanese government was unique among peripheral regions in first providing these public goods. Our reading of the historical record suggests that it was the severe, existential threat to the Japanese regime caused by the arrival of Western powers that aligned the elite in support of a strategy of aggressive defensive modernization.[34] A central part of this effort, as we have shown, was the absorption of Western science and technology. Historians argue that one important advantage Japan had was that its contact with the West happened late, meaning that it could observe and learn from what happened to China. Senior members of the shogunate correctly anticipated that Japan would be the next target of Western imperialism, and efforts were underway even prior to the Meiji Revolution to learn from the West. Importantly, however, Japan did not need to discover the policy tools itself. State support of technology absorption, particularly the translation of technical books, was a common strategy for regions hoping to emulate Britain. This has been observed from Bourbon France in the late eighteenth century to the Self-Strengthening Movement in China in the nineteenth century (Juhasz and Steinwender, 2024). Meiji Japan thus took the state-led technology adoption playbook developed elsewhere and deployed it at an unprecedented scale.

[34] See Koyama et al. (2018) for a geographic explanation for why Chinese and Japanese responses to the West in the nineteenth century differed.

References

Abramitzky, R. and I. Sin (2014). Book translations as idea flows: The effects of the collapse of communism on the diffusion of knowledge. *Journal of the European Economic Association 12*(6), 1453–1520.

Abramovitz, M. and P. A. David (1996). Convergence and deferred catch-up: productivity leadership and the waning of American exceptionalism. *The mosaic of economic growth*, 21–62.

Allen, R. C. (2011). *Global economic history: a very short introduction*, Volume 282. Oxford University Press.

Almelhem, A., M. Iyigun, A. Kennedy, and J. Rubin (2023). Enlightenment ideals and belief in progress in the run-up to the industrial revolution: A textual analysis.

Amiti, M. and D. E. Weinstein (2018). How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data. *Journal of Political Economy 126*(2).

Bassino, J.-P., S. Broadberry, K. Fukao, B. Gupta, and M. Takashima (2019). Japan and the great divergence, 730–1874. *Explorations in Economic History 72*, 1–22.

Bekkers, E., M. Landesmann, and I. Macskasi (2022). Complexity, tacit knowledge and the scope for technological catch-up. *The World Economy 45*(4), 1179–1212.

Berg, M. (2007). The genesis of 'useful knowledge'. *History of science 45*(2), 123–133.

Bernhofen, D. M. and J. C. Brown (2004). A direct test of the theory of comparative advantage: the case of Japan. *Journal of Political Economy 112*(1), 48–67.

Bernhofen, D. M. and J. C. Brown (2005). An empirical assessment of the comparative advantage gains from trade: evidence from Japan. *American Economic Review 95*(1), 208–225.

Bolitho, H. (2007). The Temp? crisis. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 116–167. New York: Cambridge University Press.

Bouscasse, P., E. Nakamura, and J. Steinsson (2023). When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870. Working Paper 28623, National Bureau of Economic Research.

Braguinsky, S. (2015). *Industrial and Corporate Change 24*(4).

Broadberry, S., K. Fukao, and T. Settsu (2025). How did japan catch-up with the west? some implications of recent revisions to japan's historical growth record. *Asia-Pacific Economic History Review*.

Broadberry, S. N. (1994). Technological leadership and productivity leadership in manufacturing since the Industrial Revolution: Implications for the convergence debate. *The Economic Journal 104*(423), 291–302.

Brown, G. (2024). The persistent effects of bible translations in africa. *Available at SSRN 4695488*.

Chanut, J.-M. (2000). *L'industrie française au milieu du 19e siècle: les enquêtes de la statistique générale de la France*.

Chung, S. (2020). Excelsior: The Korean innovation story. Accessed 15 August 2025.

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

Clark, P. H. (2009). The Kokugo Revolution. *Japan Research Monograph*.

Coluccia, D. M. and G. Dossi (2025, May). Return innovation: The knowledge spillovers of the british migration to the united states, 1870-1940. Working Paper.

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

Costinot, A., D. Donaldson, and I. Komunjer (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies 79*(2), 581–608.

Cowan, R. and D. Foray (1997). The economics of codification and the diffusion of knowledge. *Industrial and corporate change 6*(3), 595–622.

Crafts, N. (1998). Forging ahead and falling behind: the rise and relative decline of the first industrial nation. *Journal of Economic Perspectives 12*(2), 193–210.

DeLong, J. B. (2022). *Slouching towards Utopia: An economic history of the twentieth century*. Hachette UK.

Department of Finance (1916). *Annual Return of the Foreign Trade of the Empire of Japan. Part I*. Department of Finance.

Dittmar, J. E. (2011). Information technology and economic change: the impact of the printing press. *The Quarterly Journal of Economics 126*(3), 1133–1172.

Doi, T. (1980). *The Intellectual Property Law of Japan*. Alphen aan den Rijn, The Netherlands and Germantown, Md.: Sijthoff & Noordhoff.

Eaton, J. and S. Kortum (2002). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Federico, G., S. Natoli, G. Tattara, and M. Vasta (2011). *Il commercio estero italiano: 1862-1950*. Laterza.

Foray, D. (2004). *Economics of Knowledge*. MIT press.

Fouquin, M. and J. Hugot (2016). Two centuries of bilateral trade and gravity data: 1827-2014.

Fukao, K., T. Makino, and T. Settsu (2020, 08). Structural Change, Capital Deepening, and TFP Growth in Japan: 1885-1970. *Social Science Japan Journal 23*(2), 185–204.

Garnier, C. (2013). Two centuries of japanese linguistics in france: 1825–1995. *Cipango – French Journal of Japanese Studies* (2). Online since 10 June 2013. Accessed 15 October 2025.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective: A Book of Essays*. Cambridge, MA: Harvard University Press.

Giorcelli, M. and B. Li (2021). Technology transfer and early industrial development: evidence from the Sino-Soviet alliance. Technical report, National Bureau of Economic Research.

Government of India (1893). *General report on the census of India, 1891*. Eyre and Spottiswoode.

Government of India (1913). *Census of India, 1911. Volume I: India, Part I-Report*. Superintendent Government Printing, India.

Gregg, A. G. (2020). Vertical and horizontal integration in Imperial Russian cotton textiles, 1894–1900. *European Review of Economic History 24*(1), 157–191.

Griliches, Z. (1957). *Hybrid corn: An exploration in economics of technological change*. Ph. D. thesis, The University of Chicago.

Gupta, B. and T. Roy (2017). From Artisanal Production to Machine Tools. *The spread of modern industry to the periphery since 1871*, 53.

Hall, B. H. (2004). Innovation and diffusion.

Hallmann, C., L. Rosenberger, and E. E. Yavuz (2021). Invention and Technological Leadership During the Industrial Revolution. Technical report, Working Paper.

Hannas, W. and H.-M. Chang (2021, January). China's sti operations. Center for Security and Emerging Technology.

Hayami, Y. (1975). *A Century of Agricultural Growth in Japan*. University of Minnesota Press and University of Tokyo Press.

Heibonsha (1974). *Nihon Jinbutsu Bunken Mokuroku [Bibliography of Japanese Biographies]*. Tokyo: Heibonsha.

Hirakawa, S. (2007). Japan's turn to the West. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 432–498. New York: Cambridge University Press.

Hirschman, A. (1958). *The Strategy of Economic Development*. Yale paperbound. Yale University Press.

Horie, Y. (1960). Business pioneers of modern Japan—Ishikawa Masatatsu and Oshima Takato. *Kyoto University Economic Review 30*(2 (69)), 1–16.

Huberman, M., C. M. Meissner, and K. Oosterlinck (2017). Technology and geography in the second industrial revolution: new evidence from the margins of trade. *The Journal of Economic History 77*(1), 39–89.

Hungerland, W.-F. and N. Wolf (2022). The panopticon of Germany's foreign trade, 1880–1913: New facts on the first globalization. *European Review of Economic History 26*(4), 479–507.

Iriye, A. (2007). Japan's drive to great-power status. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 721–782. New York: Cambridge University Press.

Itakura, K. (2009). *Zodho Nihon rika kyoiku-shi [Japanese History of Science Education, Supplemented ed., with chronology]*. Tokyo: Kasetsu-sha.

Jones, H. J. (1980). *Live machines: hired foreigners and Meiji Japan*. Vancouver: University of British Columbia.

Juhász, R., N. Lane, and D. Rodrik (2024). The new economics of industrial policy. *Annual Review of Economics 16*.

Juhasz, R. and C. Steinwender (2024). Industrial Policy and the Great Divergence. *Annual Review of Economics 16*.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021, September). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights 3*(3), 303–320.

Kim, G. B. (2011). An Anatomical Chart of South Korean Science and Technology in the 1960s: Their Relationships with Political Power. *East Asian Science, Technology and Society: An International Journal 5*(4), 529–542.

Kohli, A. (1994). Where do high growth political economies come from? The Japanese lineage of Korea's "developmental state". *World Development 22*(9), 1269–1293.

Kokawa, T., Y. Shigeru, and Y. Masuda (1994). *Historical Development of English-Japanese Dictionaries in Japan*.

Koyama, M., C. Moriguchi, and T.-H. Sng (2018). Geopolitics and Asiaâs little divergence: State building in China and Japan after 1850. *Journal of Economic Behavior & Organization 155*, 178–204.

Koyama, M. and J. Rubin (2022). *How the world became rich: The historical origins of economic growth*. Cambridge, UK: Polity Press.

Kumon, Y. (2022). The labor-intensive path: Wages, incomes, and the work year in Japan, 1610–1890. *The Journal of Economic History 82*(2), 368–402.

Lamoreaux, N. R. and K. L. Sokoloff (1999). Inventors, firms, and the market for technology in the late nineteenth and early twentieth centuries. In N. R. Lamoreaux, D. M. G. Raff, and P. Temin

(Eds.), *Learning by Doing in Markets, Firms, and Countries*, pp. 19–57. University of Chicago Press.

Lane, N. (2025). Manufacturing revolutions: Industrial policy and industrialization in South Korea. *The Quarterly Journal of Economics*.

Lippert, W. (2001). Language in the Modernization Process: The Integration of Western Concept and Terms into Chinese and Japanese in the Nineteenth Century. In M. Lackner, I. Amelung, and J. Kurtz (Eds.), *New Terms for New Ideas. Western Knowledge and Lexical Change in Late Imperial China*. Netherlands: Koninlijke Brill.

Maddison Project Database (2020). The Maddison Project: Maddison style estimates of the evolution of the world economy. A new 2020 update.

Masanori, W. (2022). *The Role of the Ministry of Public Works in Designing Engineering Education in Meiji Japan: Reconsidering the Foundation of the Imperial College of Engineering (Kobu-dai-gakko)*, pp. 88–113. Amsterdam University Press.

Mass, W. and W. Lazonick (2013). The British cotton industry and international competitive advantage: the state of the debates. In *International Competition and Strategic Response in the Textile Industries Since 1870*, pp. 9–65. Routledge.

Mayer, T. and S. Zignago (2011). *Notes on CEPII's distances measures: the GeoDist Database*. CEPII Working Paper 2011-25.

Meade, R. (2022). *The Translation of Technical Manuals from Western Languages in Nineteenth-century Japan: A Visual Tour*, pp. 1–18. Amsterdam University Press.

Meade, R. (2024). Science across the meiji divide: Vernacular literary genres as vectors of science in modern japan. *History of Science 62*(2), 227–251.

Mehta, S. D. (1954). *The Cotton Mills of India: 1854 to 1954*. Bombay, India: The Textile Association, Bombay. Accessed 2025-09-09.

Meissner, C. M. and J. P. Tang (2018). Upstart industrialization and exports: evidence from Japan, 1880–1910. *The Journal of Economic History 78*(4), 1068–1102.

Mengzhi, F. (1999). Sci-tech translation and its research in China. *Meta 44*(1), 185–197.

Ministry of Education (1980). *Japan's modern educational system: a history of the first hundred years*, Chapter Higher Education in the Early Meiji Era.

Mitchell, B. (1975). *European Historical Statistics, 1750-1970* (1 ed.). Palgrave Macmillan London.

Mokyr, J. (2011). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.

Mokyr, J. (2016). *A culture of growth: The origins of the modern economy*. Princeton University Press.

Montgomery, S. L. (2000). *Science in translation: Movements of knowledge through cultures and time*. University of Chicago Press.

Morck, R. and M. Nakamura (2007). Business Groups and the Big Push: Meiji Japan's Mass Privatization and Subsequent Growth. *Enterprise and Society 8*(3).

Morck, R. and M. Nakamura (2018). Japan's ultimately unaccursed natural resources-financed industrialization. *Journal of the Japanese and International Economies 47*(C), 32–54.

Moser, P. (2005). How do patent laws influence innovation? Evidence from nineteenth-century world's fairs. *American economic review 95*(4), 1214–1236.

Moser, P. (2011). Do patents weaken the localization of innovations? evidence from world's fairs. *The Journal of Economic History 71*(2), 363–382.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989). Industrialization and the Big Push. *Journal of Political Economy 97*(5), 1003–1026.

National Institute for Educational Policy Research (2011). Primary Schools in Japan. https://www.nier.go.jp/English/educationjapan/pdf/201109BE.pdf. [Online; accessed 15-June-2024].

Ohkawa, K., M. Shinohara, and M. Umemura (Eds.) (1965). *Estimates of long-term economic statistics of Japan since 1868*. Tokyo: Toyo Keizai Shinposha.

Ohno, K. (2018). *The History of Japanese Economic Development: Origins of Private Dynamism and Policy Competence*. Oxon: Routledge.

Otsuka, K., G. Ranis, and G. Saxonhouse (1988). *Comparative Technology Choice in Development: The Indian and Japanese Cotton Textile Industries* (1 ed.). London: Palgrave Macmillan London. Accessed 2025-09-09.

Pascali, L. (2017). The wind of change: Maritime technology, trade, and economic development. *American Economic Review 107*(9), 2821–2854.

Pauer, E. and R. Mathias (2022). Accessing technical education in modern Japan.

Pearse, A. S. (1930). *The Cotton Industry of India: Being the Report of the Journey to India*. Manchester: Taylor, Garnett, Evans. Accessed 2025-09-09.

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge & Kegan Paul.

Rosenberg, N. (1972). Factors affecting the diffusion of technology. *Explorations in economic history 10*(1), 3.

Rosenberger, L., W. W. Hanlon, and C. Hallmann (2024). Innovation networks in the industrial revolution. Technical report, National Bureau of Economic Research.

Rubinger, R. (2000). Who Can't Read and Write? Illiteracy in Meiji Japan. *Enterprise and Society 55*(2), 163–198.

Saito, O. and M. Takashima (2016). Estimating the shares of secondary-and tertiary-sector outputs in the age of early modern growth: the case of japan, 1600–1874. *European Review of Economic History 20*(3), 368–386.

Santos Silva, J. and S. Tenreyro (2010). On the existence of the maximum likelihood estimates in poisson regression. *Economics Letters 107*(2), 310–312.

Shimizu, M. (2010). Japanese english education and learning: A history of adapting foreign cultures. *Educational Perspectives 43*(1-2), 5–11.

Smethurst, R. J. (2007). *From Foot Soldier to Finance Minister: Takahashi Korekiyo, Japan's Keynes*. Cambridge, Massachusetts and London: Harvard University Asia Center. Distributed by Harvard University Press.

Squicciarini, M. P. and N. Voigtländer (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics 130*(4), 1825–1883.

Stevens, E. (1995). *The Grammar of the Machine: Technical Literacy and Early Industrial Expansion in the United States*. Yale University Press.

Sussman, N. and Y. Yafeh (2000). Institutions, Reforms, and Country Risk: Lessons fom Japanese Government Debt in the Meiji Era. *Journal of Economic History 60*(2).

Suzuki, J. (1999). The Machine Industry in the Meiji Period. *Japanese Yearbook on Business History 16*, 113–133.

Tamagawa, K. (2002). The role of cotton spinning books in the developments of the cotton spinning industry in Japan. https://www.jsme.or.jp/tsd/ICBTT/conference02/KanjiTAMAGAWA.html (Accessed: 2024-07-10).

Tang, J. P. (2014). Railroad expansion and industrialization: evidence from Meiji Japan. *The Journal* *of Economic History 74*(3), 863–886.

Tang, J. P. and S. Basco (2023). Banks, credit supply, and the life cycle of firms: Evidence from late nineteenth century Japan. *Journal of Banking & Finance 154*.

Thang, B. T. (2018, Jan.). The KIST in South Korea and the issues on the establishment of the V-KIST in Vietnam. *Journal Science and Technology Policies and Management 6*(2), 32–42.

Treasury Department's Bureau of Statistics (1900). *Foreign Commerce and Navigation of the United States*. U.S. Government Printing Office.

Ueda, M., J. Nishizawa, I. Hirayama, and S. Miura (Eds.) (2003). *Nihon Jinmei Daijiten [The Biographical Dictionary of Japan]*. Tokyo: Kodansha.

U.S. Census Bureau (2022, August). Language Use in the United States: 2019. American Community Survey Reports, ACS-50. Accessed 2025-09-09.

U.S. House of Representatives, Tariff Board (1912). Cotton manufactures: Report of the tariff board on schedule i of the tariff law. Technical report, U.S. Government Printing Office, Washington, D.C.

Wong, R. B. (2012). Taxation and good governance in China, 1500-1914. pp. 353–377. Cambridge, MA: Cambridge University Press.

Woodcroft, B. (1857). *Subject-matter index (made from titles only) of patents of invention, from March 2, 1617 (14 James I.), to October 1, 1852 (16 Victoriae)*. Great Seal Patent Office.

World Bank (2024). World Development Indicators.

Xu, C. (2022). Reshaping global trade: the immediate and long-run effects of bank failures. *The Quarterly Journal of Economics 137*(4), 2107–2161.

Yamamura, K. (1986). The Meiji Land Tax Reform and its Effects. In M. Jansen and G. Rozman (Eds.), *Japan in Transition: From Tokugawa to Meiji*, pp. 382–399. Princeton: Princeton University Press.

Yamamura, K. (1997). *The economic emergence of modern Japan*. Cambridge University Press.

Zhuravskaya, E., S. Guriev, and A. Markevich (2024). New Russian economic history. *Journal of Economic Literature 62*(1), 47–114.

Online Appendix

A Government Expenditures

Figure A.1 shows that the imposition of the land tax enabled the early Meiji government to finance enormous investments in codification and technical absorption.

Figure A.1: Japanese Government Expenditure

Note: Government expenditure and revenue data are from Toyo Keizai Shimposha (1926) *Meiji Taisho Zaisei Shoran [Meiji and Taisho Financial Details]*, Toyo Keizai Shimposha: Tokyo, pp. 2 and 640. Before adopting the Gregorian calendar in 1873, Japanese fiscal years varied in duration and did not align perfectly with Western ones, but the mapping to Western years is approximately correct. These are deflated by the Wholesale Price Index from Ohsato, Katsuma (ed., 1966) *Hundred-Year Statistics of the Japanese Economy*, Statistics Dept., The Bank of Japan: Tokyo, p. 76.

As a result of Japan's impressive ability to raise government revenues, by 1884, Japanese government revenues equaled 83.1 million yen. By contrast, the Chinese government in 1884, still recovering from the chaos of the Opium Wars and Taiping Rebellion, could only raise 114 million yen even though China had ten times Japan's population.[1]

[1]Wong (2012) reports that Chinese tax revenue in 1884 was 77 million silver taels. We performed the

B Mastery of IR Technologies Required for Developing Newer Technologies: Historical Evidence From Japan

In Section 4.1, we argued that Japan needed to absorb IR technologies before it could master newer technologies available at the technology frontier by 1880. Here, we present additional historical evidence for Japan, which suggests that industrial development around 1880 was not nearly sufficiently mature for machine building and the other related sectors to emerge (Suzuki, 1999; Masanori, 2022). For machine-building in particular, interchangeable parts were a complex technical feat requiring a high level of precision and quality from related sectors (e.g., castings, steel). Until 1910 (when our sample period ends), Japanese industry did not possess these capabilities. In fact, consistent with the literature on the big-push (e.g., Murphy et al. (1989)) and sectoral linkages (Hirschman, 1958), it was necessary for Japan to master IR technologies before it could become competitive in sectors such as machine-building that required high-quality inputs (such as bolts, fittings, and standardized parts) and the knowledge acquired from mastering the first set of technologies.

For example, technicians from the cotton spinning industry assisted in the development of Toyoda's power loom (a mechanized machine for weaving) in 1909 (Suzuki, 1999). The knowledge acquired in mastering cotton spinning allowed the Japanese industry to move into machine building. Finally, we note that this discussion provides a micro-foundation for the technology adoption lags literature (Comin and Hobijn, 2010). Japan adopted interchangeable parts with a substantial lag relative to the West (where interchangeable parts were an integral part of the American System of Manufacturing that emerged in the early to mid-19th century), because the Japanese domestic economy was missing complementary capabilities until after the turn of the 20th century.

C Productivity Growth

C.1 Estimating Productivity Growth

In this section, we demonstrate how to utilize trade data to construct a global database that enables us to estimate productivity growth at the region-industry level. Here, we explain how we estimate productivity growth for our set of regions. The basic intuition for this procedure is based on the Ricardian model of trade. In the canonical two-country, two-good version of this model, knowing the relative labor productivities of the cloth and wine industries in England and Portugal tells us which country will export which product. The simple Ricardian model cannot be applied to data because the prediction that a country cannot import a good it exports is patently false. Costinot et al. (2012) solve this problem using the theoretical setup of the Eaton and Kortum (2002) model. In their model, each industry (k) in exporter (i) is composed of a continuum of varieties (goods) each produced based on a random productivity draw (z), whose mean rises with the "fundamental" productivity in the industry, z'_{ik}, where average industry productivity is a linear function of z'_{ik}. Thus, if a country has a high average productivity in some industry, it will tend to be the low-cost supplier of more varieties in that industry and therefore export more. Since there is a monotonic relationship between productivity and the value of exports, we can invert this relationship to obtain an estimate of productivity by observing the level of exports. Costinot et al. (2012) show that the relationship between exports from i to j in any period t (x_{ijkt}) and fundamental productivity in an

currency conversion in two ways. The number in the text uses the exchange rate series from (Fouquin and Hugot, 2016) of 1.39. We obtain a similar estimate if we convert silver taels into yen by noting that an 1867 Shanghai silver tael contained 36.0 grams of silver and an 1876 silver yen coin contained 24.3 grams of silver, according to https://en.numista.com. This implies an exchange rate of 1.48 yen per tael.

industry at time t (z'_{ikt}) is linear in logs and can be written as

$$\ln x_{ijkt} = \gamma'_{ijt} + \gamma'_{jkt} + \theta \ln z'_{ikt} + \epsilon'_{ijkt}, \tag{A.1}$$

where γ'_{ijt} is an importer-exporter fixed effect; γ'_{jkt} is an importer-industry fixed effect; $\theta > 0$ is the Fréchet scale parameter; and ϵ'_{ijkt} is an error term that captures how trade costs deviate at the industry-exporter-importer level from the exporter-importer average. The intuition for this formula is that the amount trade between two countries will depend on bilateral factors captured by γ'_{ijt} (such as bilateral distance, the relative sizes of the exporter and importer, etc.), industry demand conditions in the importer captured by γ'_{jkt}, and relative productivity of the exporter in the sector (x_{ijkt}). We could estimate $\theta \ln z'_{ikt}$ by regressing log bilateral exports on an ijt, jkt, and ikt fixed effects, but given the large number of zero trade flows, this would be biased.

Our path into solving this problem is to first note that our objective is to estimate not the level of productivity, but the change: $\gamma_{ikt} \equiv \theta \Delta \ln z'_{ikt}$. We estimate it by noting that we can first-difference equation (A.1) and rewrite it in terms of fixed effects:

$$\Delta \ln x_{ijk} = \gamma_{ij} + \gamma_{jk} + \gamma_{ik} + \epsilon_{ijk}, \tag{A.2}$$

where we have suppressed the time subscript and $\gamma_{\ell,m} \equiv \Delta \gamma'_{\ell,m}$ for any index (ℓ, m). Estimating this equation enables us to identify γ_{ik} and therefore $\theta \Delta \ln z_{ik}$ up to the choice of a normalization that pins down the reference exporter productivity, importer demand, and industry productivity.[2] This equation can be rewritten to yield

$$\Delta \ln x_{ijk} = \gamma_{jk} + \gamma_{ik} + \tilde{\epsilon}_{ijk}, \tag{A.4}$$

where variables without primes correspond to the first differences of variables with primes and $\tilde{\epsilon}_{ijk} \equiv \gamma_{ij} + \epsilon_{ijk}$.

Estimation of equation (A.4) requires us to drop observations whenever the initial bilateral export flow in a exporter-importer-industry tuple is zero, which is problematic because a large amount of nineteenth-century export growth was due to exporters expanding their set of export destinations over time. This can bias estimates of productivity growth based on a log-difference specification downwards because it cannot account for growth due to the extensive margin. Amiti and Weinstein (2018) [AW] propose an alternative estimation approach that corrects this problem.

Their estimator is closely related to weighted least squares. In particular, if there are no zeros in the export data, the AW estimates will match those obtained using weighted least squares with lagged export weights. A unique property of the AW estimates of γ_{jk} and γ_{ik} is that they aggregate to match the growth rate of total exports in every region-industry in which the industry's *aggregate* growth rate is well defined: i.e., the region initially has positive exports to at least one country in the industry. Similarly, the estimates aggregate to match region-industry import levels as long as a region has positive imports from at least one country in the industry in the initial period. Thus, an export-weighted average of the γ_{jk} and γ_{ik} will match *total* export growth in each country

[2] One can see this by noting that equation (A.2) can be rewritten as

$$\Delta \ln x_{ijk} = (\gamma_{ij} + \gamma_i + \gamma_j) + (\gamma_{jk} + \gamma_k - \gamma_j) + (\gamma_{ik} - \gamma_i - \gamma_k) + \epsilon_{ijk}, \tag{A.3}$$

where γ_i, γ_j, and γ_k are arbitrary normalization constants that define the baseline exporter productivity, importer demand, and industry productivity.

and industry.[3] One can formally see that the AW estimator will have this property by writing down the moment conditions used to obtain the estimates. In particular, the estimates will satisfy two types of moment conditions. First, the estimates aggregate to match total exports in every exporter-industry observation i:

$$\frac{\sum_j x_{ijk,t} - \sum_j x_{ijk,t-1}}{\sum_j x_{ijk,t-1}} = \gamma_{ik} + \sum_j \frac{x_{ijk,t-1}}{\sum_\ell x_{i\ell k,t-1}} \gamma_{jk}, \tag{A.5}$$

where we have added a time subscript, t, to be clear about how time differences are constructed from changes in levels. The left-hand side of the moment condition equals the growth rate of *total exports* in sector k from exporter i, and the right-hand side is the sum of the exporter fixed effect (γ_{ik}) and a bilateral export weighted average of the importer fixed effects (γ_{jk}). This condition, therefore, ensures that an export-weighted average of the parameters aggregates to match total exports. Second, the estimates will aggregate to match total imports in every importer-industry observation j because they impose a second moment condition:

$$\frac{\sum_i x_{ijk,t} - \sum_i x_{ijk,t-1}}{\sum_i x_{ijk,t-1}} = \gamma_{jk} + \sum_i \frac{x_{ijk,t-1}}{\sum_\ell x_{\ell jk,t-1}} \gamma_{ik}. \tag{A.6}$$

Here, the left-hand side of this moment condition is the growth rate of *total imports* in sector k by importer j, and the right-hand side is the sum of the importer fixed effect (γ_{jk}) and a bilateral export weighted average of the exporter fixed effects (γ_{ik}). Since the estimates satisfy these two moment conditions, the AW estimates aggregate to match the growth of exports and imports in every region for each industry.

Once we obtain the estimates of γ_{ik} and γ_{jk}, we run the following regressions to impose normalizations that lead to a meaningful decomposition of global trade patterns:

$$\gamma_{ik} = \gamma_i + \gamma_{1k} + \tilde{\gamma}_{ik}, \tag{A.7}$$

and

$$\gamma_{jk} = \gamma_j + \gamma_{2k} + \tilde{\gamma}_{jk}, \tag{A.8}$$

where $\tilde{\gamma}_{ik}$ and $\tilde{\gamma}_{jk}$ are regression residuals. This normalization choice has several useful properties. First, γ_i tells us the growth in exports resulting from shifts in exporter characteristics (e.g., productivity or size). Second, $\tilde{\gamma}_{ik}$, the "comparative-advantage" component of export growth, corresponds to the growth in exports due to shifts in productivity that are orthogonal to changes in exporter factors (i.e., γ_i) and changes in industry factors (γ_{1k}).[4] Since the former captures shifts in productivity at the national level and the latter captures the impact of comparative advantage

[3]We also considered using the Poisson pseudo-maximum likelihood (PPML) estimator. However, one well-known issue with PPML is that it often fails to converge in datasets with many zeros like ours (Santos Silva and Tenreyro, 2010). While the AW estimator only required us to drop country-industry observations where there were no exports to or imports from *any country* in the initial period, the PPML estimator did not converge unless we used data for countries with at least two export destinations or two import sources in each industry. As a result, while the AW procedure produced 1,358 productivity estimates based on 6,216 observations, the PPML estimator only converged on a subsample that was 36.5 percent as large. The PPML estimator only produced 38 percent as many productivity growth estimates as the AW estimator.

[4]Although we do not use the other normalization constants, we can recover them. $\hat{\gamma}_k \equiv \hat{\gamma}_{1k} + \hat{\gamma}_{2k}$ is the shift in exports that can be attributed to movements in industry k's characteristics (e.g., global productivity growth in k or global demand for k). Similarly, γ_{ij} can be recovered by regressing $(x_{ijk,t}/x_{ijk,t-1} - 1 - \hat{\gamma}_{ik} - \hat{\gamma}_{jk})$ on ij fixed effects.

on export growth, $\tilde{\gamma}_{ik}$. In the Costinot et al. (2012) model, $\gamma_{ik} \equiv \theta \Delta \ln z'_{ikt'}$, which enables us to define define $\Gamma_{ik} \equiv \tilde{\gamma}_{ik}/\theta$ as the change in exporters i's comparative advantage in industry k (i.e., the shift in productivity that cannot be explained by relative growth in industry k's productivity in all countries or relative productivity growth in the exporting country.[5]

In the following sections, we estimate γ_i and Γ_{ik} to understand patterns of productivity growth worldwide. We implement this methodology on annualized trade growth rates for the sample period (1880-1910), so our estimates correspond to averaged annual productivity growth rates. We show how to construct annualized rates in appendix C.2. All results reported below refer to annualized estimates.

C.2 Constructing Annual Growth Rates

We build the bilateral global trade data by merging bilateral industry export flows from different source countries (Belgium, Japan, Italy, or the U.S.). These data source countries sometimes only report exports in an industry in one of the early years (1880 or 1885) or one of the later years (1905 or 1910). Rather than throw out the industry for all countries when 1880 or 1910 is not reported by one source region, we adopt a procedure to let us be flexible about the start and end dates by computing the average annual export growth rates between any of two potential start years at the beginning of our sample (1880 or 1885) and any of two potential end years at the end of our sample (1905 or 1910).

We set the start year equal to 1880 if the source region reports data in that year or 1885 if data is not available for 1880 but is available for 1885. Similarly, we set the final year equal to 1910 if the source region reports data for that year or 1905 if data is not available for 1910 but is available for 1905. Since this means that the start and final years for bilateral trade growth rates can vary by data source region, we annualize the data so our export and productivity growth rates can be interpreted as average annual growth rates.

We use two procedures to annualize the data. If the reporting region exports the product in 1880 or 1885 (i.e., $\sum_j x_{ijks} > 0$ for $s = 1880$ or 1885), we set s equal to the first year that satisfies $\sum_j x_{ijks} > 0$. We drop the sector if $\sum_j x_{ijks} = 0$ because industry growth rates are undefined if a country does not export anything in the industry in the first period. Similarly, we set f equal to the last year ($f \in \{1905, 1910\}$) that satisfies $\sum_j x_{ijkf} > 0$. We compute the annual growth rate for all bilateral exports satisfying $x_{ijks} > 0$ as

$$g^C_{ijk} \equiv \left(\frac{x_{ijkf}}{x_{ijks}}\right)^{\frac{1}{f-s}} - 1$$

For this sample of exports, we define the implied level of exports in year $s+1$ as $x_{ijk,s+1} \equiv \left(1 + g^C_{ijk}\right) x_{ijk,s}$.

We face a different problem if a country exports the product in year s, i.e., $\sum_j x_{ijks} > 0$, but no bilateral exports are reported between two regions in the industry in the start year, i.e., $x_{ijks} = 0$ for some $\{i, j, k, s\}$. To deal with this problem, we define the average growth rate in exports due to new export destinations as

$$g^N_{ik} \equiv \left(1 + \frac{\sum_{j \in \mathcal{N}_i} x_{ijkf}}{\sum_j x_{ijks}}\right)^{\frac{1}{f-s}} - 1, \tag{A.9}$$

[5] We follow Eaton and Kortum (2002) and set $\theta \equiv 8.28$. The choice of θ does not qualitatively affect any of our results; it just raises or lowers all countries' productivity growth proportionally.

where $\mathcal{N}_i$ is the set of new export destinations, which are defined to be the observations satisfying $x_{ijks} = 0$ and $x_{ijkf} > 0$. In this case, we set the annualized level of exports to new destinations in $s+1$ as $x_{ijk,s+1} \equiv \left(1 + g_{ik}^{N}\right)^{-(f-s-1)} x_{ijkf}$. In other words, we set the counterfactual amount of exports to new destinations in year $s+1$ equal to the observed amount of exports in year f (x_{ijkf}) deflated by the growth rate in exports due to extensive margin growth between years $s+1$ and f. With these annualized values for exports in hand, we can write the left-hand side of equation A.5 as

$$\frac{\sum_j x_{ijkf} - \sum_j x_{ijks}}{\sum_j x_{ijks}} = \frac{\sum_j x_{ijk,s+1} - \sum_j x_{ijks}}{\sum_j x_{ijks}}, \tag{A.10}$$

and the left-hand side of equation A.6 as

$$\frac{\sum_i x_{ijkf} - \sum_i x_{ijks}}{\sum_i x_{ijks}} = \frac{\sum_i x_{ijk,s+1} - \sum_i x_{ijks}}{\sum_i x_{ijks}}. \tag{A.11}$$

We then can apply the AW estimation procedure in equations A.5 and A.6 to estimate the γ_{ik}.

C.3 Productivity Growth Results

Section 3 examined Japan and other regions' economic performance using the raw trade data. Here, we utilize the methodology developed in the Section C.1 to provide the first systematic estimates of productivity growth for many regions in the late nineteenth and early twentieth centuries. Our normalization choice implies that productivity or anything that shifts exporter i's exports conditional on demand conditions will be captured by our estimate of γ_i. We can interpret $\hat{\gamma}_i - \hat{L}$, where $\hat{L}$ is the annual population growth rate, as a measure of exporter productivity, i.e., how much exports in country i grew after controlling for demand conditions and population growth. Figure A.2 plots the annualized per capita shift in export supply net of population growth relative to the value for the US, i.e., $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$.[6] shows that the patterns are similar if we do not account for differences in population growth.

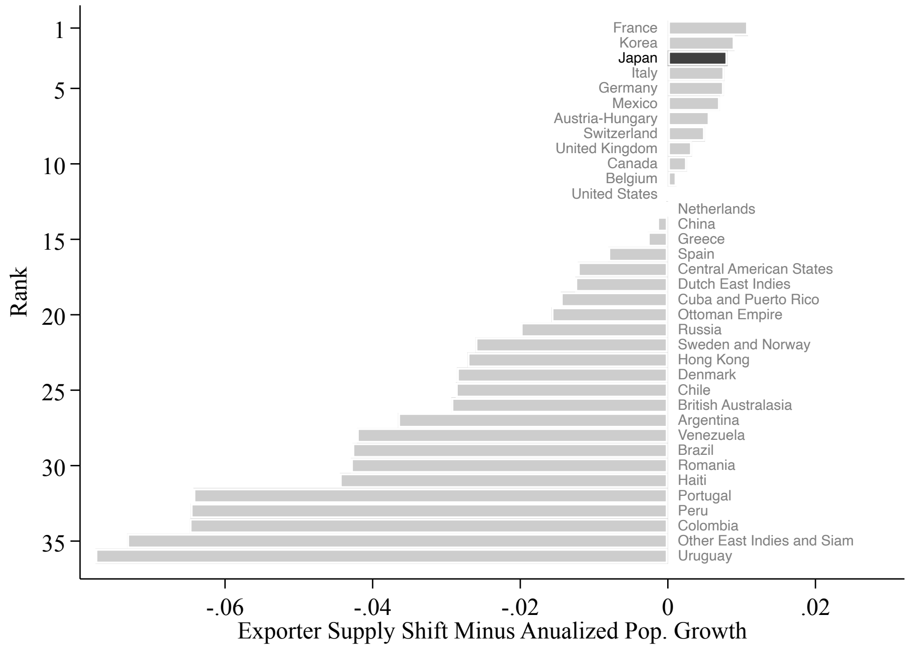
Reassuringly, the ranking of economies broadly aligns with what economic history teaches us about this period. France, Korea, Japan, Germany, Mexico, Italy, Austria-Hungary, Switzerland, the United Kingdom, Canada, Belgium, and the US show robust growth in their export supply shifter. In contrast, economies such as those of Portugal, Peru, Colombia, and Uruguay exhibit weak performance. Notably, Japan's export-supply shifter ranks third, confirming that its economy experienced some of the highest export productivity growth globally during this period. Notice that our estimates also suggest that Korea had high productivity growth (alongside Japan), which may be related to the fact that Japan forcibly opened Korea in 1876, and although nominally independent, the Japanese "reform[ed]" the Korean government and military administration by introducing to the country the kinds of measures that Meiji Japan itself had undertaken" (Iriye, 2007, p. 769)). Our result is consistent with the idea that the Meiji reforms may have also raised productivity in Korea.

Next, we examine the extent to which productivity growth was biased towards manufacturing. We regress the comparative-advantage component of productivity growth, Γ_{ik}, on broad industry dummies:

$$\Gamma_{ik} = \beta_i^{\text{Agg}} \times I_k^{\text{Agg}} + \beta_i^{\text{Mfg}} \times I_k^{\text{Mfg}} + \beta_i^{\text{Min}} \times I_k^{\text{Min}} + \epsilon_{ik} \tag{A.12}$$

[6] Appendix Figure A.3

Figure A.2: Relative Annualized Per Capita Exporter Supply Shifter by Exporter

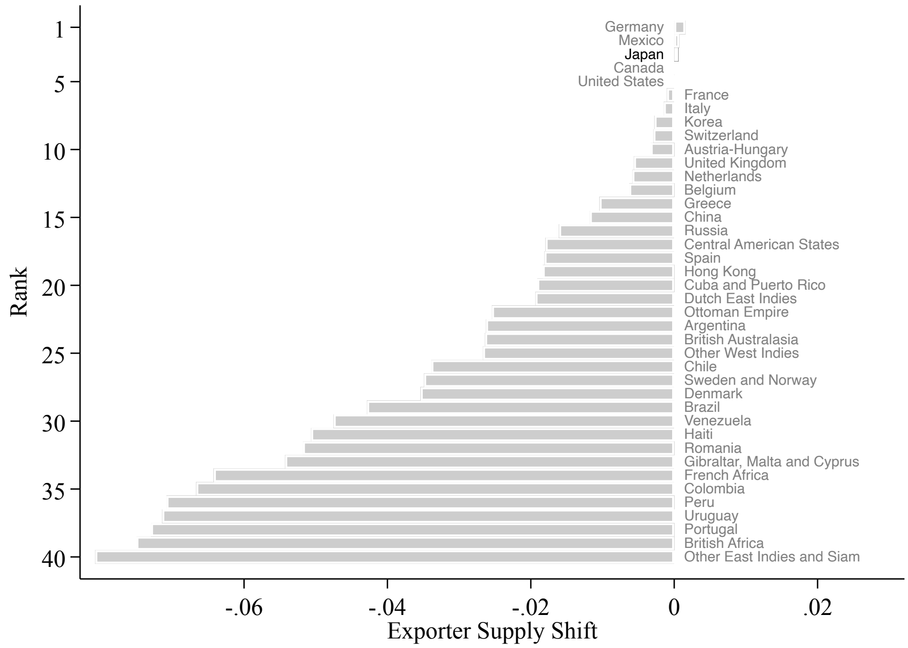


Note: Annualized per-capita exporter supply shifts are defined relative to the US, i.e., they are defined as $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$. Annual population growth is computed between {1870,1880} and 1913 using the Maddison data (see Appendix H.5 for details).

where I_k^{Agg}, I_k^{Mfg}, and I_k^{Min} are dummies that are one if sector k is in agriculture, manufacturing, or mining, respectively; and β_i^{Agg}, β_i^{Mfg}, and β_i^{Min} are parameters that measure the average growth rate of comparative advantage for exporter i in agriculture, manufacturing, and mining. In words, $(\beta_i^{\text{Mfg}} - \beta_{US}^{\text{Mfg}})$ tells us how fast productivity in manufacturing grew in exporter i relative to the US after controlling for its average growth and the average growth in world manufacturing. Figure A.4 reports the results from this exercise for countries in which the manufacturing share of exports in 1880 was not trivial. While Portugal and Hong Kong exhibit strong shifts in comparative advantage towards manufacturing, the results in Figure A.2 indicate that these economies had low overall rates of productivity growth, implying that their relatively strong performance in manufacturing was offset by their low overall productivity growth. The next seven countries (Japan, Belgium, Mexico, Italy, the UK, the US, and Canada) are all examples of regions that industrialized during this period, exhibiting rapid productivity growth and exceptionally high relative productivity growth in manufacturing.

Our structural estimates of industry productivity growth in this period confirm that Meiji Japan's economic performance was exceptional. Average productivity growth was high in international comparison and shifted strongly towards manufacturing. This result supports the idea that Japan's unparalleled shift towards specialization in manufacturing (Figure 2) was driven by productivity growth biased towards manufacturing—that is, shifting Ricardian comparative advantage.

Figure A.3: Relative Annualized Exporter Supply Shift by Exporter



Note: Annualized per-capita exporter supply shifts are expressed as relative to the US, i.e., they are defined as $\hat{\gamma}_i - \hat{\gamma}_{US}$. See text for details on variable construction.

D External Validity: Codification and Economic Performance in Other Periphery Economies Before WW1

In Section 7, we examined the influence of the Meiji model for Korea and China in the postwar era. Here, we are interested in further exploring the assertion that codification in the vernacular was a necessary, but not sufficient, condition for development in the late 19th century, consistent with other periphery economies' experience at this time.

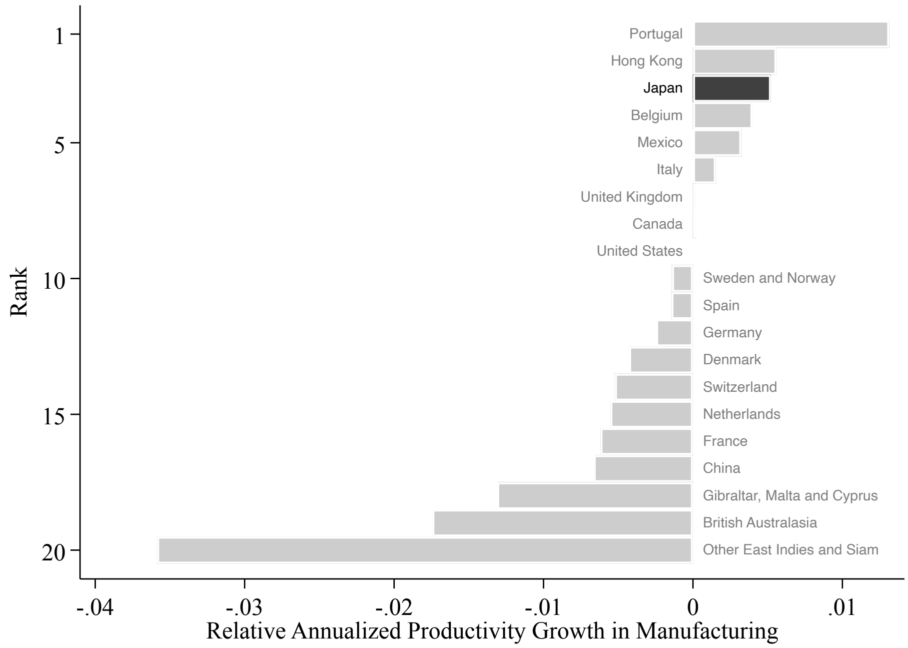
In Section 5, we showed econometric evidence that suggested that other periphery economies did not experience similar patterns of development. Here, we complement this evidence with historical evidence contrasting the experience of Japan with that of British India and Late Imperial Russia. Both have been the subject of influential case studies in industrial development (e.g., Gerschenkron (1962); Clark (1987), and each built up a sizeable modern, factory-based manufacturing sector by the eve of World War 1.[7]

We begin with British India, where—despite an early start compared to other periphery economies—Indian industry was quickly outcompeted by Japan in key sectors. By the 1930s, India had become the largest market for Japanese cotton cloth, even under a protective import tariff (Mass and Lazonick, 2013). Evidence shows that labor productivity and total factor productivity in Indian factories were especially low (Gupta and Roy, 2017), suggesting persistent difficulties in operating new technologies efficiently.

Turning to data on codification, Figure 5 shows that in 1910, there were virtually no technical

[7]For example, Russia and British India had the largest installed capacity in mechanized cotton spinning among periphery countries (U.S. House of Representatives, Tariff Board, 1912).

Figure A.4: Relative Annualized Productivity Growth in Manufacturing



Note: The plot presents our estimates of productivity growth in manufacturing relative to the US, i.e., $(\beta_i^{\text{Mfg}} - \beta_{US}^{\text{Mfg}})$. β_i^{Mfg} is estimated in equation A.12 for regions in which the manufacturing sector's export share in 1880 is at least 0.5% and for regions in which we can estimate productivity growth in at least five non-primary and five primary sectors.

books written in the major spoken languages: Hindi, Tamil, or Urdu. At first glance, British India therefore seems to be consistent with our assertion that codification in the vernacular was a necessary condition for development. While literate Japanese could read technical books in their language, Indians literate in local languages had no access to such material.

However, as a British colony, English was the lingua franca for higher education and technical instruction following the 1835 English Education Act. The real question, then, is whether Indians could access knowledge in English. Indian census data offers significant insight into this issue. The 1891 Census of India states that 537,811 people could read English, which was only 0.19 percent of the population (Government of India, 1893, p.224). Twenty years later, the 1911 Census reports that just 0.54 percent of Indians were literate in English (Government of India, 1913, p. 299). However, these figures likely overstate the actual number of Indians who could read English because they include foreign English speakers (e.g., British expatriates) who mostly lacked the ability to explain technical material in Hindi, Tamil, or Urdu. For example, the 1891 Census notes that only 386,032 Indians, or 0.14 percent of the population, could read English (Government of India, 1893, p.224), roughly the same as the percentage of Americans today who can speak Japanese: 0.15 percent (U.S. Census Bureau, 2022, p.3).

One can put the number of bilingual Indians into perspective by comparing it to the number of Japanese who could read Western technical manuals translated into Japanese. Given that Japan's population in 1891 was 41 million, and assuming a literacy rate of 40 percent, we estimate that approximately 16 million Japanese could read technical manuals. These numbers suggest that there were more than 40 Japanese people who could read technical manuals in Japanese for every

Indian person who could read them in English.

Of course, having forty times more people able to read engineering in Japan than in India might not have mattered if English speakers could easily share their knowledge across the language barrier in India. Although we lack concrete data on how difficult this was, Western and Indian historical accounts suggest that it was challenging for English speakers to communicate with those who could not understand them. Indeed, they argue that differences in training and literacy significantly contributed to the productivity gap between the Japanese and Indian textile industries. For example, Pearse (1930) study of the development of the Asian cotton textile industries notes that

> "Each [Japanese] firm has at least one engineer with university education and special textile engineering training. Some of the mill managers have passed through similar educational institutions, but all have at least graduated from one of the technical schools One notices everywhere the result of a good general education; the inside managers and foremen have had a sound training in technical schools, they have not grown up empirically in the mill; every mill girl reads and writes, and possesses general education quite on par with that of European countries. The foreman and general supervisors are specially trained in classes run by the combines. We are not dealing with labor as it exists in India, China, or South America" quoted in (Otsuka et al., 1988, pp. 84-85).

Similarly, Mehta (1954)'s 100-year history of the cotton textile industry in India emphasizes the communication and staffing challenges that arose from trying to use technology whose descriptions were written in English.

> "The difficulties of language [faced by English engineers] were unusually great, not only in relation to the workers but frequently also in relation to the employers and other members of the latter's office. The growth of other professions, namely, law, medicine and government service, generally precluded from the industry the extremely small number of Indians who had access to schools where English was taught. An exceedingly small number of Indians received their training in English technical institutes and factories. The capacity of the managing agents to ensure a high level of production on the basis of an informed judgment was extremely low in the first fifty years (i.e., from 1854 onwards). For one, the top technicians were Englishmen on whom direct control was extremely limited. Secondly, the managing agent was himself a novice in many cases in the art of management, not only of machines but also of men, and he was hardly fitted to achieve a proper control of production functions. (Mehta, 1954, pp. 101-108)

In light of the evidence above, the failure of British India to develop an internationally competitive industry aligns well with our narrative. With neither access to codified knowledge in spoken languages such as Hindi, Urdu, or Tamil, nor widespread literacy in English, Indians did not have access to codified technical knowledge.

Imperial Russia is another context which has been the subject of influential studies on late industrialization. Figure 5 shows the limited availability of technical books in Russian in 1910. Our theory would thus predict that Russia would struggle to develop an internationally competitive, modern industrial economy. Unfortunately, given the present state of knowledge, there is substantial debate in the literature about exactly how successful Russia's industrialization was during this period (see e.g., Zhuravskaya et al. (2024) for a recent overview). This makes it essentially impossible to draw definitive conclusions about whether Imperial Russia industrialized without access to a level of codified knowledge comparable to Japan's.

However, if one examines the comparative performance of a flagship industry such as mechanized cotton spinning, there are important differences between the two countries. In particular, while both Japan and Russia had a sizeable domestic cotton textiles industry, a key difference was that Russia's industry developed behind a high tariff wall and predominantly served the domestic market (Gregg, 2020).[8] In fact, Gregg (2020, p. 162) characterizes the Russian cotton industry as having achieved "a worldwide intermediate case of industrial development." Consistent with the narrative of modest progress in the industry, Clark (1987) argues that Russian textile workers who migrated to New England were 54 percent as productive as English textile workers. Contrast this with Japan, where cotton textiles became an important exported commodity during our sample period. That is, while Japanese cotton textile producers were sufficiently productive to compete in international export markets, there is no evidence that this was the case for Russian producers on the eve of World War I.

In summary, the qualitative evidence for British India paints a consistent picture with the econometric evidence presented in the main text. The lack of codified knowledge in major spoken languages, combined with a low proportion of English speakers among the native population, kept knowledge access costs high in British India relative to Japan. This may be an important reason why, despite a generous head start, Japan rapidly outperformed British India in modern industries. We know much less about both codification and development in Imperial Russia. While future scholarship on Russian industrial performance before World War 1 may lead us to revise our conclusions, given the current state of knowledge, the evidence suggests that Russia is also consistent with a country in which high technology access costs precluded the emergence of an internationally competitive industrial sector.

[8]Japan was prohibited from enacting protective tariffs during this period due to the unequal treaties it was forced to sign.

E Additional Tables

Table A.1: Linguistic Distance from English and GDP

	Log GDP per Capita					
	(1) 1870	(2) 1913	(3) 2018	(4) 1870	(5) 1913	(6) 2018
Log Physical Distance between Country and the UK	-0.170***	-0.207***	-0.237***	-0.248***	-0.315***	-0.323***
	(0.058)	(0.064)	(0.066)	(0.054)	(0.065)	(0.072)
Number of Weeks Required to Learn the Plurality Language	-0.010***	-0.013***	-0.008*	-0.005**	-0.007***	-0.003
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.005)
Observations	61	61	61	55	55	55
R^2	0.395	0.428	0.208	0.369	0.426	0.198
Includes English-speaking Countries	✓	✓	✓			

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: GDP per capita is from the Maddison Project. The physical distance between the region and the UK is from *CEPII* database using the great circle formula. The number of weeks an English-speaking native will take to attain "Professional Working Proficiency" in the country's plurality language is estimated by the U.S. Department of State's Foreign Service Institute. See Appendix H for data construction and sources. Robust standard errors are in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Manuals with the Most Copies Held by the Imperial College of Engineering Library

Category	**Author**	**Title**	**Copies**
Mathematics	Wilson	*Elementary Geometry*	340
	Todhunter	*Trigonometry for Beginners*	234
	Wilson	*Algebra for Beginners*	192
Civil Engineering	Unwin	*Elements of Machine Design*	71
	Rankine	*Applied Mechanics*	55
	Rankine	*Manual of Civil Engineering*	55
	Perry	*Treatise on Steam*	48
	Goodeve	*Elements of Mechanism*	34
Mining and Mineralogy	Egleston	*Hydraulic Mining in California*	62
	Milne	*Notes on the Ventilation of Mines*	47
	Lyman	*Reports of Progress for the First Year of the Oil Surveys*	30

Source: Reproduced from Meade (2022), Table 1, p. 12.

Table A.3: Summary Statistics

Variable	N	Mean	SD	p25	p50	p75
Change in exporter's i comparative advantage in industry k (Γ_{ik})	1246	0.00	0.04	-0.01	0.00	0.02
Change in Japan's comparative advantage in industry k ($\Gamma_{\text{Japan},k}$)	56	0.00	0.05	-0.01	0.01	0.02
Exporter's Industry Growth Rate	1397	-0.10	0.38	-0.05	0.03	0.09
Exporter's Industry Growth Rate in Japan	71	-0.05	0.37	-0.02	0.04	0.15
Britsh Patent Relevance	125	0.05	0.09	0.02	0.04	0.06

Note: The estimation of Γ_{ik} is detailed in Appendix Section C. Exporter's Industry Growth Rate is the annualized export growth rate for each industry between {1880, 1885} and {1905, 1910}. The details on the construction of British Patent Relevance are in Appendix Section I.

Table A.4: Japanese Export Growth and British Patent Relevance 1875-1910

	Annualized Export Growth Between 1875 and						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1880	1885	1890	1895	1900	1905	1910
British Patent Relevance	-0.104**	-0.027	0.011	0.028**	0.022***	0.020***	0.014**
	(0.049)	(0.020)	(0.017)	(0.011)	(0.008)	(0.007)	(0.006)
Observations	40	45	46	47	45	46	47
Constant	✓	✓	✓	✓	✓	✓	✓

Note: The dependent variable is annualized Japanese export growth for the year reported relative to 1875. The number of observations changes across specifications because of the different number of traded sectors in different years. Robust standard errors in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Annualized Export Growth and British Patent Relevance - British Colonies and Steam Intensity

	Export Growth	
	(1)	(2)
BPR × Japan	0.121***	0.123**
	(0.033)	(0.052)
BPR × Not Japan	-0.036***	-0.003
	(0.010)	(0.011)
BPR × British Colony	0.029	
	(0.020)	
Steam Intensity		-0.744**
		(0.297)
Observations	1395	690
R^2	0.234	0.309
Country fixed effects	✓	✓
Sample	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. The Japan dummy equals one if the region is Japan and zero otherwise, "Not Japan" is analogously defined. "British Colony" is a dummy for whether a region was a British colony in the 1880-1910 window. Steam Intensity is constructed as Steam Engine Horsepower/Wage Bill by industry using French manufacturing census data from the 1860s (see Appendix H.6 for details about the data construction). Robust standard errors are in parentheses. $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Table A.6: Annualized Export Growth and British Patent Relevance - Manufacturing Sectors

	Export Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BPR × Japan	0.124**	0.124**	0.124**	0.124**	0.124**	0.124**	0.124**
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
BPR × Not Japan		-0.014	-0.015	-0.023	-0.041*		
		(0.012)	(0.013)	(0.014)	(0.022)		
BPR × English-Speaking			0.001				
			(0.025)				
BPR × French-Speaking				0.038			
				(0.023)			
BPR × Top-4 Codified					0.050**		
					(0.024)		
BPR × High-Income						-0.347	-0.347
						(0.335)	(0.335)
BPR × Medium-Income						-0.041	0.079
						(0.885)	(0.911)
BPR × Low-Income						-0.856	-0.318
						(0.760)	(0.967)
BPR × Asia							-1.180
							(1.153)
Observations	31	661	661	661	661	661	661
R^2	0.133	0.362	0.362	0.364	0.366	0.363	0.364
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. Japan dummy equals one if the region is Japan and zero otherwise, "Not Japan" is analogously defined. "English-speaking" is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the four most codified languages: French, English, German, and Italian. {High, Medium, Low}Income are indicator variables which use 1870 GDP per capita from the Maddison Project to identify if a region is in the top third of the income distribution (high), middle third (medium), or in the bottom third (bottom); we set these dummies to 0 for Japan. Asia dummy equals 1 if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.7: Annualized Export Growth and British Patent Relevance: Dropping Regions

	Export Growth, Dropping Exports to						
	(1) English-Speaking	(2) British Colonies	(3) Languages Similar to English	(4) High-Income	(5) Medium-Income	(6) Low-Income	(7) Asian
British Patent Relevance	0.112***	0.112***	0.112***	0.089***	0.111***	0.108***	0.108***
	(0.031)	(0.031)	(0.032)	(0.031)	(0.032)	(0.036)	(0.036)
Observations	71	71	71	70	67	61	61
R^2	0.107	0.107	0.108	0.065	0.107	0.076	0.076
Constant	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	Japan	Japan	Japan	Japan	Japan	Japan

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. Each column drops exports to a different subset of countries/regions. (1) Drops English-Speaking countries. (2) Drops British Colonies. (3) Drops countries with a language similar to English, defined as those where it takes six or fewer months for an English speaker to become proficient. (4), (5), and (6) drop High, Medium, and Low-income countries, respectively. (7) Drops exports to Asian countries. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

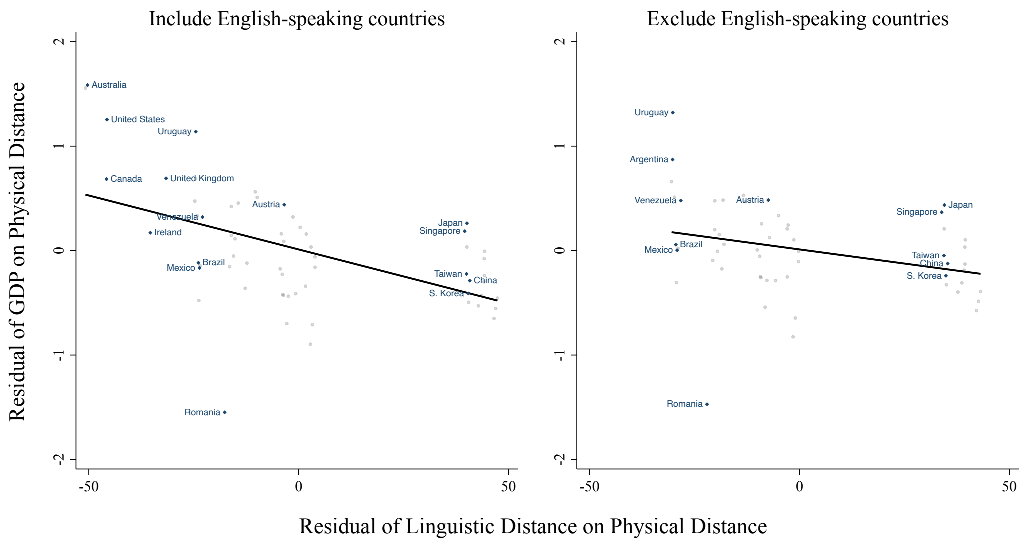
Table A.8: Annualized Export Growth and British Patent Relevance: Dropping Sectors

	Export Growth, Dropping		
	(1) Cotton-Textiles	(2) All Textiles	(3) Iron and Fabricated Metals
British Patent Relevance	0.121***	0.111**	0.123***
	(0.036)	(0.045)	(0.034)
Observations	69	63	69
R^2	0.112	0.086	0.111
Constant	✓	✓	✓
Sample	Japan	Japan	Japan

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. Each column drops exports to a particular industry or group of industries. (1) drops cotton textile-related industries. (2) drops all industries related to textiles. (3) drops industries related to producing iron. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

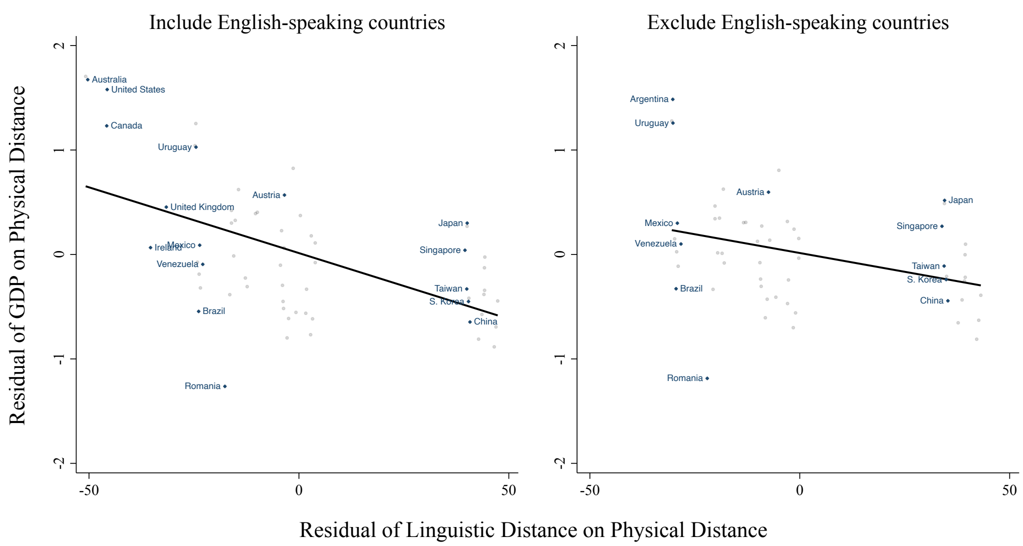
F Additional Figures

Figure A.5: Linguistic Distance Partial Regression Plot for 1870



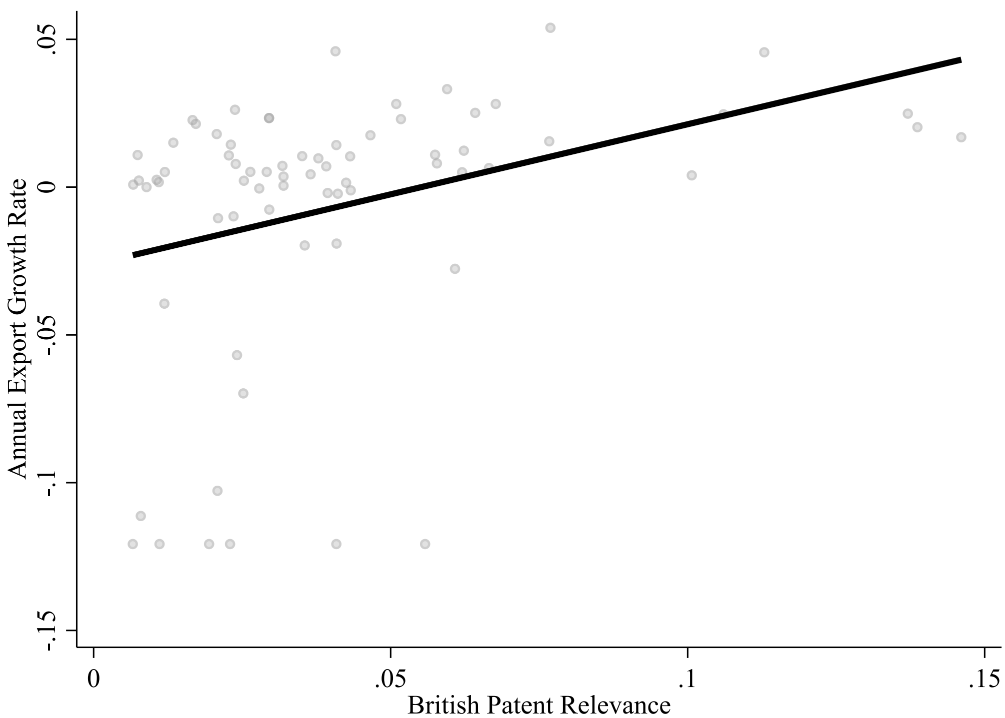
Note: This figure plots the relationship between log GDP per capita in 1870 and linguistic distance after controlling for log physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.6: Linguistic Distance Partial Regression Plot for 1913



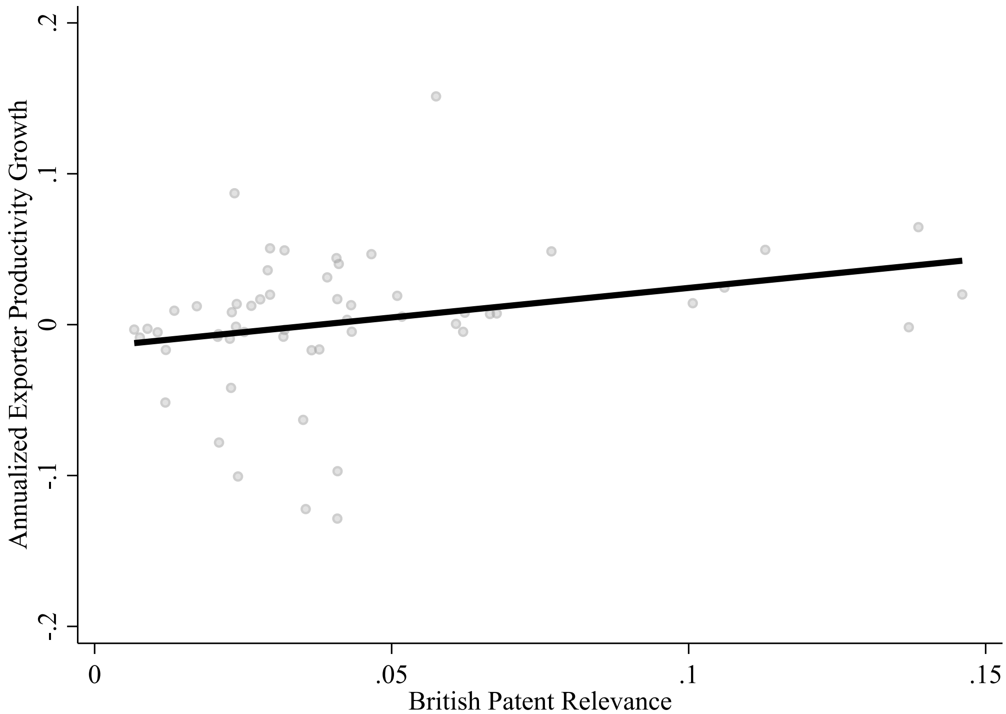
Note: This figure plots the relationship between log GDP per capita in 1913 and linguistic distance after controlling for log physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.7: Annualized Export Growth and British Patent Relevance for Japan



Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry *k* between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry *k*. BPR is standardized to have a mean of 0 and a standard deviation of 1. See text for details on variable construction.

Figure A.8: Annualized Prod. Growth Γ and British Patent Relevance for Japan



Note: The dependent variable, Γ_{ik}, is the annualized growth rate in comparative advantage for industry k in region i between {1880,1885} and {1905,1910}. BPR stands for "British Patent Relevance", it captures how relevant British patents are to the vocabulary used in manuals of an industry k. BPR is standardized to have a mean of 0 and a standard deviation of 1. See text for details on variable construction.

G Bilateral Trade Dataset

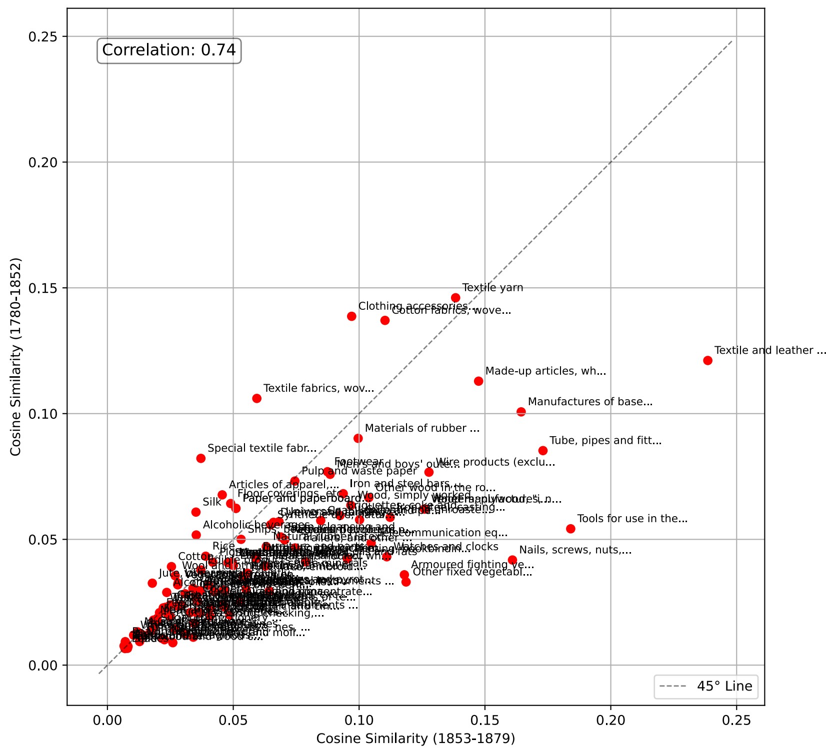
Our master bilateral, product-level trade dataset is constructed from four main sources:

1. **Belgian manufacturing exports and imports in 1880, 1885, 1905, and 1910.** We obtain the Belgian bilateral manufacturing product-level trade data from Huberman et al. (2017). They use the *Tableau générale du commerce extérieur* published by the Belgian government as their primary source and concord product lines to SITC Revision 2 codes. The authors record trade *in manufacturing* at five-year intervals between 1870 and 1910. In 1900, 50% of Belgian exports and 20% of imports were in manufacturing.

2. **Italian exports to and imports from top trading partners in 1880, 1885, 1905, and 1910.** We obtain Italian trade data from Federico et al. (2011). This dataset harmonizes historical trade records from Italian customs between 1862 and 1950 by reconciling the different product lines to SITC Revision 2 codes. The source reports bilateral trade at the product level between Italy and its ten biggest trading partners.

3. **American exports and imports in 1880, 1885, 1905, and 1910.** The U.S. data are digitized from yearly volumes of *Foreign Commerce and Navigation, Immigration, and Tonnage of the United States* published by the Treasury Department's Bureau of Statistics (1900). We digitized and concorded these data to SITC Revision 2 codes.

Figure A.9: Cosine Similarities between 1780-1852 and 1853-1879



Note: This plot compares cosine similarities constructed using British Patents from 1780-1852 (y-axis) against cosine similarities using British Patents from 1853-1879 (x-axis).

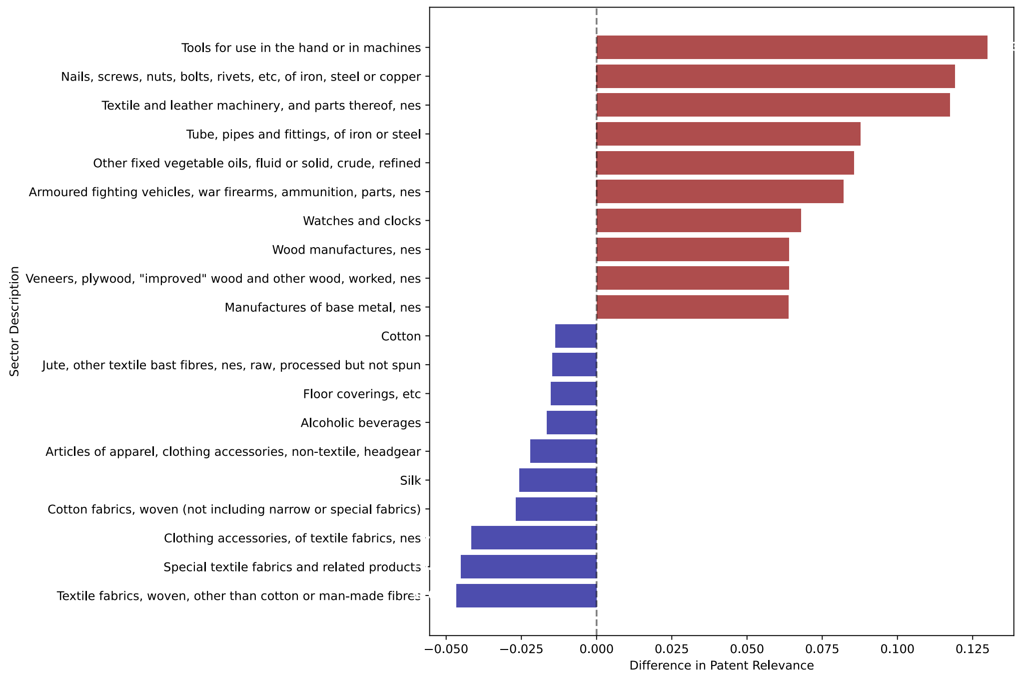
4. **Japanese exports and imports in 1875, 1880, 1885, 1905, and 1910.** We obtained bilateral product-level Japanese export data at five-year intervals between 1880 and 1910 from Meissner and Tang (2018). We digitized and concorded the Japanese export data for 1875. The Japanese trade data were sourced from the yearly volumes of *Annual Return of the Foreign Trade of the Empire of Japan*, published by the Department of Finance (1916). From these volumes, we use only the tables from the "Quantity and Value of Commodities Imported/Exported from Various Countries" sections.

Japan and the U.S. kept detailed records of their trade with foreign countries between 1880 and 1910. We used the Meissner and Tang (2018) product-SITC mapping wherever possible for Japan and the U.S. to ensure consistency. Each entry provides the name of the product, its quantity, units, transaction value, and year, as well as the names of the exporting and importing countries. The construction of these data involves digitizing the records and harmonizing products and country names. To construct a harmonized dataset across different reporting countries, we convert all data to a common currency, harmonize country names, and address issues of double reporting. The protocols we adopted are described in detail in the subsections below.

G.1 Harmonization of Countries

Country names are not standardized across reporters (Belgium, Italy, Japan, and the U.S.) and years. In order to make comparisons across years and countries, we standardized country names

Figure A.10: Sectors with Highest Positive and Negative Changes in Cosine Similarities



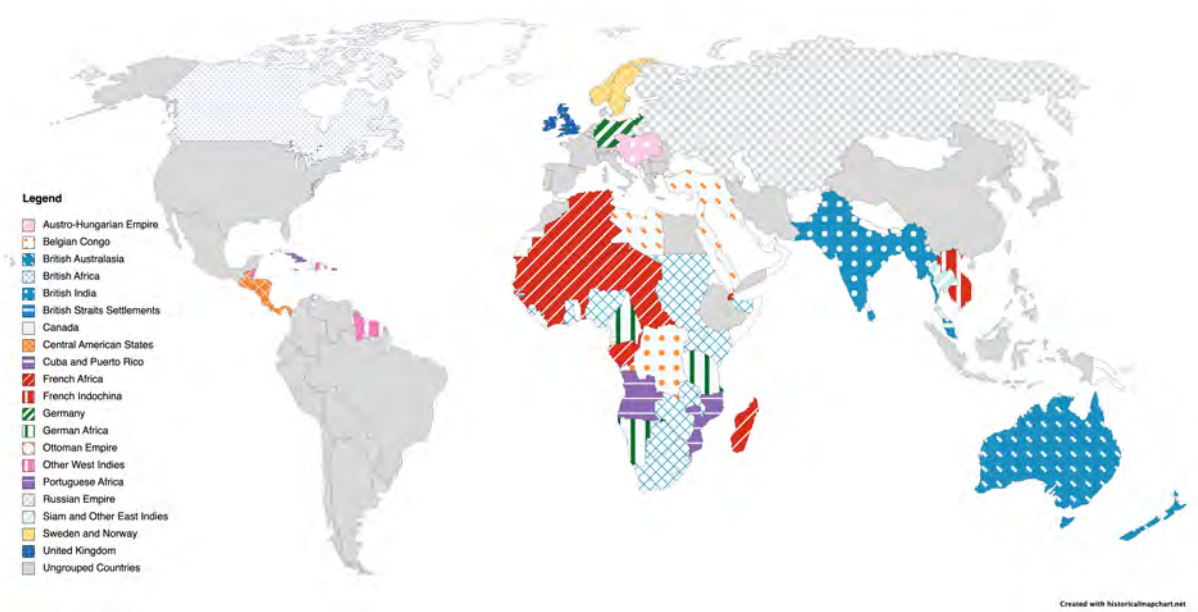
Note: This plot compares sectors with the highest positive and negative changes in cosine similarities based on the 1780-1852 sample and the 1853-1879 sample. A positive change means that the cosine similarity was higher in the 1853-1879 sample than in the 1780-1852 sample.

as follows:

1. We made a list of all the country names that appear in all of the trade books from the four reporters.
2. We grouped names that refer to the same country: e.g., Vietnam and French Indo-China both refer to the same political entity at the time.
3. We kept the group if it is used by at least three reporters in the 1880 or 1885 books *and* the 1905 or 1910 books for each reporter.
4. If the country group did not meet the previous requirement, then we built a regional group that did. For example, Honduras, Nicaragua, and Costa Rica do not have three reporters in all the required years. If we group all Central American States together, this larger regional group meets our requirements.
5. If a country could not be grouped and did not meet the reporter-year requirement, then we dropped it.
6. If a region was too disaggregated, we dropped it. For example, Singapore and Hong Kong are distinct entities, each with substantial trade volumes in our dataset. If one country, in one year, reported "Hong Kong & Singapore," we dropped this observation.

Appendix Figure A.11 illustrates how we grouped countries. We use the map of the world on the eve of World War I (1914) as a baseline for our country groups.

Figure A.11: Country Groups



Note: Colonies are grouped by imperial power and region (e.g., British Africa, French East Indies). All small, remote islands (e.g., Falklands) were dropped. Countries in white are missing from the dataset, and countries in gray were not modified. The remainder of the footnote reads from West to East on the map. The West Indies are grouped together, with the exception of Cuba and Puerto Rico. British Honduras (although technically in Central America) is considered part of the West Indies due to its political affiliation with other British colonies in the Caribbean. The Ottoman Empire includes Libya, but not Algeria (which fell to the French in 1881). Taiwan is never directly mentioned in any trade statistics and is not included in Japanese trade for the time period. Since each book either mentions French India or French Indochina, we conclude that French India refers to French Indochina, not to the French port cities in India. Thailand (then Siam) is grouped with other minor East Indies colonies such as Timor-Leste and British Borneo.

G.2 Double Reporting

Trade between reporting countries appears twice: once as exporters from the origin and secondly as imports by the destination. For all reporting countries except Belgium, we use their export data for their exports to reporting and non-reporting regions. Because Belgium did not report any trade data for non-manufacturing sectors, we use the reporting country's import data from Belgium to fill in these gaps. We use imports by reporting countries from non-reporting countries to construct the exports of non-reporting countries.

H Other Variables from External Sources

This section documents the variables we obtained from secondary sources and any changes we made to them. We discuss data from primary sources in the next sections.

H.1 Defining current high-income countries

We make a reference to "high-income" countries in the Introduction. We define a country as high income if its GDP per-capita (PPP adjusted) in 2022 is 50% or more of the US GDP per-capita, based on data from the World Bank (2024). Specifically, we use the variable "GDP per capita, PPP (current international dollars)."

H.2 Identifying the plurality language by country: Ethnologue (2023)

<u>Reference</u> *Ethnologue*, https://www.ethnologue.com/.

We identify the plurality language spoken by each country for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.5 - A.6. To do so, we obtain the modern (2023) plurality language spoken in each country from "Ethnologue".

H.3 Weeks to Learn a Language: Foreign Service Institute (2023)

<u>Reference</u> "Foreign Language Training - United States Department of State," U.S. Department of State, 03-May-2023. [Online]. Available: https://www.state.gov/foreign-language-traning/.

The Foreign Service Institute of the U.S. Department of State estimates the number of weeks required for an English native speaker to reach "General Professional Proficiency" in the language (a score of "Speaking-3/Reading-3" on the Interagency Language Roundtable Scale. We use this measure to proxy linguistic distance for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.5 - A.6.

H.4 Distance to U.K.: GeoDist Database (Mayer and Zignago, 2011)

We control for physical distance in the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.5 - A.6. To do so, we use data from *Centre d'Etudes Prospectives et d'Informations Internationales* (CEPII) which report different measures of bilateral trade distances (in kilometers) for 225 countries. Our measure of the distance between any two countries is the "dist" variable, which is calculated using the great circle formula. They compute internal distances by using the latitudes and longitudes of the most important cities/agglomerations (in terms of population). This means that the distance of a country to itself will never be zero; rather, the distance measure captures how far away major population centers within a country are from each other.

H.5 Historical income and population data: Maddison Project Database

The Maddison Project Database provides information on comparative economic growth and income levels over the very long run. We use the 2020 version of this database (Maddison Project Database, 2020), which covers 169 countries up until 2018. We use data on GDP per capita from this source for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.5 - A.6. Further, we also use this source to assign regions into income groups in the main analysis (Section 5).

Classifying regions as high-, medium- and low-income

We classify regions in our dataset by income level using the GDP per capita data from Maddison for 1870. To obtain this variable, we adopt the following steps:

1. The Maddison data uses modern country borders. We first map modern countries to the historic states they were part of in 1880-1914, which will match our trade data (e.g., Hungary and Austria map to Austria-Hungary).

2. The GDP per capita of a historical state that spans two or more modern countries is the simple mean of the GDP per capita of its constituent modern countries.

3. We rank regions by GDP per capita in descending order. Countries in the top third of this distribution are considered high income, countries in the middle third, middle income, and countries in the bottom third, low income.

Finally, we also use the Maddison data to estimate annualized population growth needed for constructing Figure A.2.

Estimating annualized population growth

We use the 1870 and 1913 population data to estimate a region's population growth according to the following protocol:

1. Concord the modern countries in the Maddison database with the historic regions we use in this paper.

2. The population of a historic region for a given year is the sum of the population of the modern states that make it up.

3. Compute annualized population growth

$$\text{Annualized Population Growth}_i = \left(\frac{\text{Population}_{i,1913}}{\text{Population}_{i,1870}}\right)^{\frac{1}{1913-1870}} - 1$$

The Maddison Project does not report data for the Russian Empire during this time period; we complement the database by using the Russian population estimates for 1880 and 1910 from Mitchell (1975).

H.6 Steam Intensity Usage: Chanut (2000)

In Table A.5, we control for the intensity of steam usage of industries in our regressions. We measure this variable based on 19th-century French energy data that comes from Chanut (2000). We manually map French industries to SITC codes. We define the steam intensity of an industry as the ratio between the steam engine horsepower of the industry over its Wage Bill, where the wage bill is defined as:

$$\begin{aligned}\text{Wage Bill} = \; & \text{\# of Male Workers} \times \text{Avg. Male Hourly Wage} \quad + \\ & \text{\# of Female Workers} \times \text{Avg. Female Hourly Wage} \quad + \\ & \text{\# of Child Workers} \times \text{Avg. Child Hourly Wage}\end{aligned}$$

H.7 Historical Exchange Rates: Fouquin and Hugot (2016)

Our bilateral-product level trade data converts the value of exports and imports (reported in local currency) into current yen. We use data on annual exchange rates from the *Historical Bilateral Trade and Gravity Dataset (TRADHIST)* from which we obtain the yearly exchange rates for the 1870-1915. Specifically, they provide us the value of one unit of the local currency in pounds.
We calculate the exchange rate from Yen to Belgian francs, Italian lira and US dollars as follows:

$$\frac{£_t / X_t}{£_t / ¥_t} = \frac{¥_t}{X_t}$$

where t refers to year and X to the local currency. The value that we obtain is the value of one unit of the local currency in yen.

I Constructing the British Patent Relevance measure

I.1 Overview

In our empirical analysis, we develop a method to quantify the supply of codified knowledge generated by the IR for each industry. We use a textual approach that follows how codified technical knowledge was disseminated in this period: through the publication of technical manuals. For each industry, we measure the textual similarity from historical technical manuals (in English) and patents. We call this measure British Patent Relevance (BPR). We also construct an analogous measure using U.S. patents, which we call U.S. Patent Relevance (USPR) measure. To implement this, we assign at least one technical manual describing production techniques to each SITC industry code and compute the similarity of its text to either British or U.S. patent texts.

We construct unigrams (e.g., *steam*) and bigrams (e.g., *steam engine*) from both patent text and technical manuals. These terms are stemmed (e.g., *steam engine* → *steam engin*) and aggregated into an industry-level corpus, with one corpus for each industry k. Patent text forms a separate corpus. For each corpus, we compute a TF-IDF (Term Frequency-Inverse Document Frequency) vector that characterizes its vocabulary. Patent relevance for industry k is then defined as the cosine similarity between the TF-IDF vector of industry k's technical manuals and that of the patent corpus. We describe each step in detail below.

I.2 Building the Terms

We construct terms from the raw text by generating n-grams. The procedure is as follows:

1. Split the raw text into sentences.
2. Convert all words to lowercase, stem them, and standardize spelling (UK spelling → US spelling).
3. Represent each sentence as an ordered list of words.
4. Generate n-grams from each sentence word list.
5. Count the frequency of each n-gram within a sentence and aggregate across sentences.
6. Remove n-grams that contain at least one stop word (e.g., “a,” “the”).
7. Produce a dataset containing all n-grams in the document and their corpus-level frequencies.

Example

1. **Text** "A stemmer for English operating on the stem cat should identify such strings as cats, catlike, and catty."
2. **Sentence** "A stemmer for English operating on the stem cat should identify such strings as cats" "catlike" "and catty"
3. **Processed Word List** "a stemmer for english oper on the stem cat should identifi such string as cat" "catlik" "catti"
4. **Unigrams** "a" "stemmer" "for" "english" "oper" "on" "the" "stem" "cat" "should" "identifi" "such" "string" "as" "cat" "catlik" "catti"
5. **Unigrams without Stopwords** "stemmer" "english" "oper" "stem" "cat" "should" "identifi" "string" "cat" "catlik" "catti"
6. **Final Unigrams with Count** "stemmer" 1 "english" 1 "oper" 1 "stem" 1 "cat" 2 "should" 1 "identifi" 1 "string" 1 "catlik" 1 "catti" 1

I.3 Focusing on Jargon

Many unigrams and bigrams are not technical jargon. In order to focus our analysis on jargon, we drop unigrams and bigrams that are commonly used. We use the Bible to identify commonplace non-technical words that are necessary to write a coherent text but are not helpful in defining an industry’s technical vocabulary. We use the 1885 King James Bible because it uses the common, non-technical nineteenth-century words and phrases. We define Biblical words as the 1,000 words occurring with the highest frequency in the Bible. However, if one of these words is used in the description of an SITC keyword, we do not count it as a Biblical word. For example, the stemmed word “brea” is a top 1,000 word in the Bible, but it also happens to be a keyword in the SITC description for cereal products.

I.4 Formally Defining TF-IDF

The term frequency (TF) measure is the count of instances a term appears in a corpus, divided by the number of terms in the corpus. The formula for the TF of term τ in corpus c is:

$$\mathrm{TF}(\tau, c) \equiv \frac{F_{\tau,c}}{\sum_{\tau' \in c} F_{\tau',c}} \tag{A.13}$$

where $F_{\tau,c}$ is the raw count of term τ in corpus c; and $\sum_{\tau' \in c} F_{\tau',c}$ is number of terms in the corpus. The inverse document frequency (IDF) is a measure of how common or rare a word is across all documents. The rarer the word, the higher the IDF score. We define the IDF for term τ in all corpora C (i.e., the complete collection of the corpus) as:

$$\mathrm{IDF}(\tau, C) = \log\left(\frac{N}{N_\tau + 1}\right) \tag{A.14}$$

where N is the total number of documents (books and patent[9]) in C; N_τ is the number of books in the corpora where the term τ appears.

The TF-IDF is then

$$\text{TF-IDF}(\tau, c, C) = \mathrm{TF}(\tau, c) \cdot \mathrm{IDF}(\tau, C) \tag{A.15}$$

We remove any n-grams that include words in the description of the SITC categories from the sample before estimating the cosine similarities. For example, removing the unigram "cotton" ensures that books describing how to grow cotton are not coded as part of the technology to spin cotton yarn.

Comparing the Vocabulary of Industries and Patents

We define the Patent Relevance of industry k as the similarity between the TF–IDF vector of its technical manuals and the TF–IDF vector of patent texts. The intuition is that if industry manuals use vocabulary similar to that found in patents, then patents contain knowledge relevant to that industry. We measure similarity using cosine similarity, the standard NLP metric for comparing text representations.

Cosine similarity corresponds to the cosine of the angle between two vectors. In the case of our baseline results, it compares the vector of word frequencies in the Bennett Woodcroft patent collection (British patents), $\boldsymbol{BW}$, with the vector of word frequencies in the technical manuals for industry i, $\boldsymbol{TM_i}$. Formally,

$$BPR_i \equiv \frac{\boldsymbol{BW} \cdot \boldsymbol{TM_i}}{\|\boldsymbol{BW}\| \|\boldsymbol{TM_i}\|} = \frac{\sum_{j=1}^{n} BW_j TM_{ij}}{\sqrt{\sum_{j=1}^{n} BW_j^2} \sqrt{\sum_{j=1}^{n} TM_{ij}^2}}, \tag{A.16}$$

where BPR_i denotes the *British Patent Relevance* of industry i. By construction, BPR_i lies between 0 and 1. A value of 1 indicates that industry manuals and patents use exactly the same vocabulary in the same proportions, while a value of 0 indicates no overlap in vocabulary.

I.5 Data Sources

Industry For each industry k (defined by SITC-3 Revision 2 codes), we hand-curated a list of nineteenth-century books describing the production process of the goods produced by k from

[9] The whole set of patents counts as one document.

HathiTrust. We picked the technical books that best matched the knowledge an entrepreneur would have had access to if they had studied Western knowledge before Japan began to industrialize, i.e., before the 1880s.

British Patents (1617-1852): The patent text from British patents between 1617-1852 comes from the second edition of *"Subject-Matter Index of Patent of Invention From March 2, 1617, to October 1, 1851 Parts I (A to M) and II (N to W)"*, published by Woodcroft (1857). These documents contain a synopsis of each patent published between 1617 and 1852. The document is divided by categories, where each patent can be categorized into one or more categories. We digitize the text of these documents and drop duplicated patents (i.e., patents that are in more than one category). Our baseline analysis uses only patents published between 1780-1852. This data was obtained through *HathiTrust*.

British Patents (1853-1899): For this period, we rely on the digitized collection of British patents compiled by Coluccia and Dossi (2025). Their data contains the full text of all British patents published between 1853 and 1899. We treat this period separately from 1617-1852 because of the major patent reform of 1852, which reduced filing costs by roughly 75% and triggered a fivefold increase in patenting within a single year. Moreover, while the pre-1852 data consists only of short synopses, the 1853-1899 dataset provides full patent descriptions. To avoid concerns about the comparison between full patent descriptions and patent synopses, we present a version of BPR (1853-1879) summarizing the full patent descriptions so that they have a similar length to patent synopses. To do this, we used OpenAI's API with the following prompt:

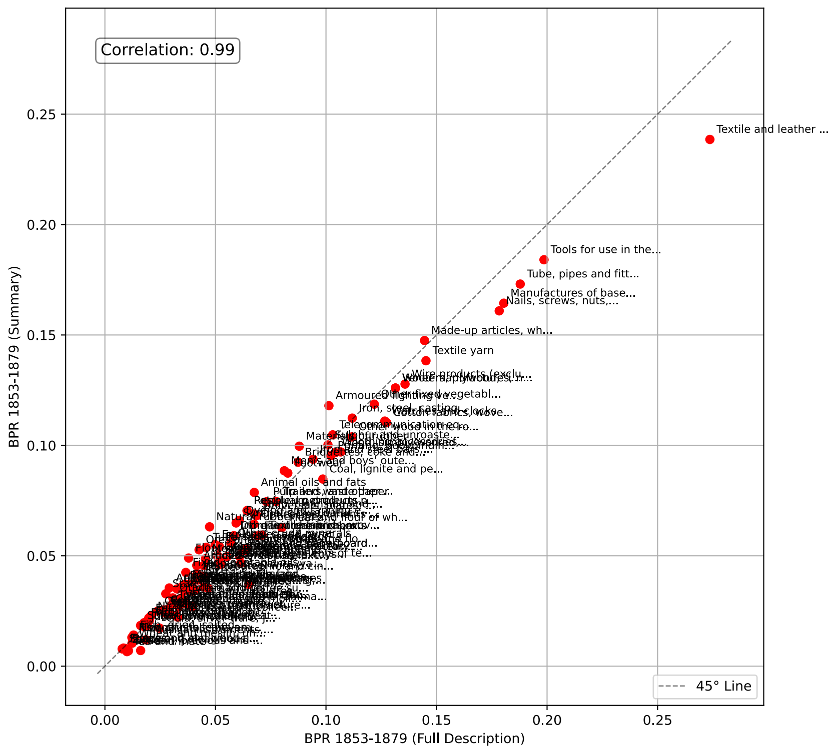
> *Summarize the following 19th-century British patent in MAX. 15 words. Focus strictly on the technical content, state what the invention is, and describe the mechanism or process. Use only vocabulary found in the patent itself or in common use at the time of application. Omit the author and date from the summary. Do not start with phrases like 'This invention describes'.*

The 15-word limit mirrors the average length of the synopses between 1617 and 1852, ensuring comparability across periods. We construct BPR for the 1853-1879 period (right before our analysis with trade data starts). To address potential concerns about the process of summarizing the patent descriptions, we also computed our BPR measure for the 1853-1879 period without summarizing the patent descriptions. The cosine similarities using full descriptions or summaries are very similar, with a correlation of 0.99, as can be observed in Figure A.12.

U.S. Patents (1836-1910): We collect U.S. patent descriptions from 1836 (the earliest year available) through 1910 by web-scraping *Google Patents*, which provides digitized versions of all U.S. patents. Our scraper builds on the tool developed by Kelly et al. (2021) and extracts the patent number, title, date, and full description for each patent.

U.S. patent descriptions typically begin with formulaic phrases such as "To all whom it may concern, be it known that (. . .)". We identify the most common introductory phrases and remove them so that descriptions begin directly with the technical content. Google Patents digitization relies on Optical Character Recognition (OCR), which can introduce transcription errors. To mitigate this problem, we retain only words appearing in the Oxford English Dictionary (which has over 500,000 entries). Words that are not in the dictionary are treated as OCR errors and discarded. On average, this cleaning step removes about 3% of words in a typical patent description.

Figure A.12: BPR (1853-1879) Computed with Full Descriptions and Summaries



J New Japanese Words in the Meiji Period

We utilize the etymology of Japanese words based on the revised edition of *Nihon Kokugo Daijiten* [The Unabridged Dictionary of the Japanese Language], published by Shogakukan (2006). Importantly, it includes the title and year of publication of the Japanese document in which each word is believed to have been first used. We obtained the digitized data for this dictionary from Kotobank.[10] The number of new words by year can be seen on Figure 3.

K Technical Books in the Top World Languages (1800-1910)

K.1 Overview

We report the source libraries for our data on technical books in Table A.9. We tried, where possible, to scrape national libraries. If we could not find a scrapable national library for a language (such as Arabic and Russian), we scraped WorldCat, an online catalog of thousands of libraries worldwide covering dozens of languages. Scraping national library catalogs has an advantage over using WorldCat as the latter source sometimes overstates the number of books because different libraries sometimes report book titles differently (e.g., slight variations in titles or author names).

We minimized the number of possible duplicates by removing spacing and punctuation in book titles and dropping any duplicated book titles published in the same year. In order to minimize the role played by reprints of the same book, we also dropped any duplicates arising from books

[10]Kotobank: https://kotobank.jp/dictionary/nikkokuseisen/

(possibly published in different years) with the same book ID. Importantly, the number of books reported for four of our five top codifying languages, French, English, German, and Japanese (but not Italian), were from national libraries, so we can be confident that there is minimal double counting in these book totals.

If we could scrape a national library or WorldCat, we made a judgment call about which source was better. If we saw that for a non-top-4 codifying language, there were more *genuine* technical books than we could find in a national library, we opted for the number from WorldCat. For example, the national libraries of Portugal and Spain have very few technical books in their catalogs relative to the libraries in WorldCat, so we opted to use WorldCat for these languages. Because of the duplication issue in WorldCat and the fact that WorldCat allows us to scrape many libraries for each language, our use of national libraries for English, French, German, and Japanese likely causes us to understate the concentration of technical books in these languages.

We scraped the number of technical books for 33 languages, which include all of the 20 most spoken native languages on earth.[11] We define the set of books comprising technical knowledge as those with a subject classified as applied sciences, industry, technology, commerce, and agriculture. For our purposes, we exclude books on theoretical technical knowledge, such as books in the hard sciences or in medicine.

[11] We assume that if someone speaks Yue or Wu Chinese, they can read Mandarin Chinese, given that these languages all use the same characters.

Table A.9: Catalogs Scraped

Library	Catalog	Languages	Years	Classification System	Tech Topics
Bibliothèque Nationale de France	Link	French	1500-1930	Universal Decimal Classification	Applied Sciences and Technology (6)
Deutsche Nationalbibliothek	Link	German	1500-1930	Dewey Decimal Classification	Technology (600)
National Diet Library	Link	Japanese	1500-2023	Nippon Decimal Classification	Technology (500) Industry (600)
Korean National Library	Link	Korean	1500-2023	Dewey Decimal Classes	Technology and Engineering (600)
Library of Congress	Link	English	1500-1930	Keyword Search	Hand-constructed
National Library of India	Link	Bengali Hindi Marathi Tamil Urdu	1500-1980	Only has three options	Non-Fiction Manually picked tech books.
Shanghai Library	Link not accessible	Chinese	1500-2023	Chinese Library Classification System	Agriculture (S) Industry (T) Transportation (U)
National Central Library (Taiwan)	Link	Chinese	1500-2023	Keyword Search	Hand-constructed
WorldCat	Link	Arabic Bulgarian Croatian Czech Danish Dutch Greek Hebrew Indonesian Italian Norwegian Persian Polish Portuguese Romanian Russian Spanish Swedish Thai Turkish Ukranian Vietnamese	1800-1930	Subject filter in advanced search	Hand-constructed

K.2 Search Filters

$\boxed{1.}$ **Format:** We only search for books. No images, periodicals, articles, or newspapers.

$\boxed{2.}$ **Language:** We always specify the language of the text. For example, when searching the National Diet Library, we only look for books written in Japanese.

$\boxed{3.}$ **Publication Year:** 1500-1930

$\boxed{4.}$ **Subject:** We always search by subject.

- We search by subject code, if possible. Otherwise, we manually picked technical books.
- If subject codes are not available, we use subject keywords. To do this, we first find the underlying subject classification system used by the library (e.g., Dewey Decimal Classification) to get the descriptions of the subject codes we want.