Codification, Technology Absorption, and the Globalization of the Industrial Revolution*

Réka Juhász †
University of British Columbia, CEPR, and NBER

Shogo Sakabe ‡
LMU

David E. Weinstein §
Columbia University, CEPR, and NBER

June 29, 2024

Abstract

This paper studies technology absorption worldwide 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 Industrial Revolution. Using the rapid and unprecedented codification of technical knowledge in Meiji Japan as a natural experiment, we show that productivity growth in Japan was higher in industries that could potentially benefit from Western technical knowledge only after the Japanese government codified as much technical knowledge as what was available in Germany in 1870. We find no similar patterns in other parts of the world that did not codify knowledge. Our findings shed new light on the frictions associated with technology diffusion and offer a novel take on why Meiji Japan was unique among non-Western countries in successfully industrializing during the first wave of globalization.

JEL CLASSIFICATION: F14, F63, N15

KEYWORDS: Industrialization, Codification, Technology Adoption, Late Development, Productivity, Industrial Policy, Japan

*We thank Benjamin Eyal, Isaac Loomis, Zachary Marcone, Ojaswee Rajbhandari, Roshan Setlur, Alex Zhang, and especially Michael Duarte, Verónica C. Pérez, Angela Wu, and Dongcheng Yang for excellent research assistance. We also want to thank Treb Allen, Andrew Bernard, Kirill Borusyak, Davin Chor, John Fernald, Shizuka Inoue, Takatoshi Ito, Chiaki Moriguchi, Robert Staiger, Jón Steinsson, and Dan Trefler for their excellent comments.

†6000 Iona Drive Vancouver, BC V6T 1L4. Email: reka.juhasz@ubc.ca.

‡Email: s.sakabe@columbia.edu.

§420 W. 118th Street, MC 3308, New York, NY 10027. Email: david.weinstein@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."

Shozan Sakuma, 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, currently, there are 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.[1] While economists have made enormous progress in understanding why English-speaking countries, Europe, and Petrostates are rich, data-driven studies of why the Industrial Revolution 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*.[2] How and why did Meiji Japan 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." While there is extensive evidence about the codification of technical knowledge in English and French, there is little understanding of how codification levels varied across languages and time outside Western Europe. 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. In turn, this implies that we have not been able to explore how access to technical knowledge in the vernacular contributed to the spread of the Industrial Revolution. An ideal experiment would require both time-series and cross-sectional variation in technical knowledge so that a researcher could examine whether a country's relative productivity growth rose in industries where more technical knowledge existed only after entrepreneurs became technically literate. Meiji Japan in the late nineteenth century provides 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 growth in 39 countries and regions in the late nineteenth and early twentieth centuries. We use this novel dataset to construct the first estimates of industry-level productivity

[1] 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 the Appendix for more details.

[2] We also see this sudden transformation in productivity data. For example, Clark (1987) finds Japan transitioned from not having any modern textile and weaving mills in 1870 to having modern mills that had productivity levels that were 96 and 98 percent as productive as those in Britain by 1910. By contrast, the Chinese textile and weaving industries were 79 and 66 percent as productive as Britain's in 1910.

growth for many nineteenth-century countries and regions. 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 all British patents issued between 1617 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 four novel stylized facts about the global spread of the Industrial Revolution and the uniqueness of Japan's industrialization in the nineteenth century, "the Meiji Miracle." The first stylized fact is that books in most languages contained little codified knowledge. In 1870, ninety percent of all codified knowledge was done in three languages: English, French, and German. We find that people who could not read these three languages were likely technically illiterate because they had very few technical books in their vernaculars to read. For example, a person who could only read Arabic would only have been able to read 72 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 English, French, and German would have had access to thousands of technical books. Put simply, for most of the world at this time, literacy in the vernacular was a ticket to reading social science and humanities, not reading science.

The second stylized fact is that the Japanese language is unique in starting at a low base of codified knowledge in 1870 and catching up with the West by 1910. For example, The Japanese National Diet Library in 1910 held more technical books written in Japanese than the Library of Congress housed in English and 75 percent more than the Deutsche Nationalbibliothek held in German.

How did Japan achieve such a 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 Japanese words describing the technologies of the Industrial Revolution did not exist. The Japanese government solved the 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 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.[3] This created two sub-periods in Meiji Japan: a period before 1887, in which Japan had completed substantial economic reforms but lagged behind the 1870 level of codification in English, French, and German, and a period afterward in which Japan could read Western technical knowledge at a level equal to or exceeding that in the West.

The third stylized fact is that per capita income falls with linguistic distance from codifying languages. We document that even after controlling for physical distance, countries that spoke languages that were *linguistically* close to English had significantly higher per capita incomes in

[3]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 in Japanese history.

1870. While we do not interpret this relationship causally, it establishes the plausibility of the theory that interactions with England, either through physical distance-related activities like trade or linguistic distance-related activities like reading English or a close-cognate language, seem to have mattered for development in the nineteenth century.

The fourth stylized fact is that Japanese manufacturing grew suddenly and very fast after Japan succeeded in codifying knowledge. In 1868, the first year of the Meiji Restoration, less than 30 percent of Japanese exports were manufactured products; seventeen years later, in 1885, the share of manufactured exports had fallen to 20 percent. In other words, there is no evidence that the Japanese industrial structure shifted towards manufacturing almost three decades after Japan opened to the West and nearly two decades after the Meiji Restoration. However, ten years after Japanese authors and translators created substantial amounts of codified knowledge—publishing over a thousand technical books—the share of manufactured exports grew to 60 percent and stayed at this level for the next 40 years.

We further document that this sudden increase in codified knowledge and this sudden increase in manufacturing specialization was unique to Japan in the 19th century. Thus, Meiji development didn't gradually increase growth rates as institutions improved; rather, a very rapid increase in manufacturing happened only after Japan succeeded in codifying about as much knowledge as Germany had in 1870. Together, we interpret these four stylized facts as presenting evidence that access to technical knowledge may have been a necessary (although not sufficient) condition for the spread of the Industrial Revolution.

In the second part of the paper, we exploit the natural experiment of Japan's rapid codification of knowledge to test this hypothesis more 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 Industrial Revolution 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 America 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 to the text of British patents using cosine similarity, the standard metric in natural language processing (NLP). 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 the word use in British patents. Thus, it is a metric for how useful the knowledge codified in British patents is for a particular industry. Reassuringly, industries such as textiles, which benefited the most from the new technologies of the Industrial Revolution, have descriptions of production processes, including flagship technologies such as spinning machinery and steam engines, that also feature prominently in patent texts. As such, we say that 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 musical instruments, which suggests that the makers of musical instruments benefitted little from Industrial Revolution technologies.

To measure outcomes at the industry level, we use a novel, bilateral, industry-level trade dataset to back out industry-level productivity growth from 1880 to 1910. This methodology is well-suited to data-scarce environments such as ours. Importantly, no data from individual reporting countries is required in the sample. This allows us to examine industry-level productiv-

ity trends globally for the first time. Using these data, we provide empirical evidence that late nineteenth-century Japan had high country-level productivity growth rates in international comparison, which were concentrated in manufacturing sectors. We thus show that the Meiji Miracle is indeed "miraculous" in comparative perspective, and find that it is driven by an increase in manufacturing productivity consistent with the historical narrative.

Armed with these data, we examine the relationship between the supply of technical knowledge and productivity growth in Japan and around the world. Consistent with our hypothesis, we find that Japanese productivity growth was higher in industries where the supply of technical knowledge was greater, but importantly, *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. Furthermore, we find no relationship between productivity growth and the supply of technical knowledge in other countries, consistent with our mechanism. This lends support to the idea that broad access to technical knowledge, which at the time usually meant access *in the vernacular* if people spoke a vernacular that was linguistically distant from English, was a necessary condition for the diffusion of Industrial Revolution technologies and manufacturing growth more broadly. Moreover, our results suggest that for regions outside of Western Europe, the codification of technical knowledge was a complex public good that required state provision. We address in Section 2 why Japan succeeded in producing many technical translations and what prevented others from doing the same.

This paper contributes to three strands of the literature. First, our results shed light on why technology diffusion to the global periphery was slow in the 19th century. Economic historians have put forward a number of explanations ranging from the imperialist context of the period (Allen, 2012) to culture (Clark, 1987). Our explanation builds on Mokyr (2011)'s pioneering work on the importance of "technical knowledge" for European industrialization, though with a Gerschenkronian (Gerschenkron, 2015) twist.[4] In particular, our results show that codified technical knowledge was almost non-existent outside of Europe. Thus, moving outside of the European culture of Enlightenment, the provision of technical knowledge required the state's involvement due to its public good-like attributes. This points to a novel arena in which 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, 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 the land tax reform and Japan's adoption of the gold standard mattered to investors. Our findings thus offer a resolution to the puzzle of what drove the Meiji Miracle.

Our results point to the importance of certain public goods necessary for industrialization.

[4]As such, our paper is related to Squicciarini and Voigtländer (2015), who show for France that the presence of knowledge elites who could apply "technical knowledge" became a predictor of growth once technological progress during the Industrial Revolution became rapid. While this is an important explanation for why Europe developed, our focus is on the spread of technology outside of Europe and the English-speaking world.

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, it is unlikely they can give a full account. In contrast, our paper provides empirical support for the long strand of Japanese economic history that has emphasized the more unique 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 makes a methodological and data contribution to the quantitative study of economic history. In particular, the paper contributes to a long and rich tradition in economic history that uses country-specific data sources to reconstruct GDP and sectoral output (Bolt and van Zanden (2020)). We view our work as providing a complementary approach utilizing more widely available trade data. Our approach may be particularly useful in data-scarce environments, as is the case for many economies in the global periphery during the nineteenth and early twentieth centuries. As the geographic and temporal scope of detailed bilateral trade data widens, we hope this approach may yield new insights about economic growth at this critical juncture in economic history.

2 Historical Context

Nineteenth-century Japan presents an interesting study of late industrialization. There is wide-ranging consensus amongst economic historians that Japan was the only non-Western economy that managed to industrialize in this period. After centuries of self-imposed isolation, the U.S. forcibly opened Japan to foreigners in 1854 and to trade with Western countries in 1858. However, Japanese historians argue that many in the *shogun*'s government, the *bakufu*, 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). The fate of China in the wake of the Opium Wars loomed large in Japanese thinking. After the British imposed crushing indemnity payments, the Chinese government was thrust into a perpetual state of near-bankruptcy. Senior members of the *bakufu* correctly anticipated that Japan would be the next target. The indemnity payments, coupled with China's subsequent descent into a brutal civil war, meant that Chinese efforts to modernize through the "self-strengthening movement" received little government support and, at least for the reform-minded Taiping rebels, open opposition.

This section discusses the three components of Meiji Japan's state-led technology absorption effort. First, we describe the specific technology policies themselves, most notably the codification of Western technical knowledge. Second, we show that by investing in elementary and university education, the government ensured the population had the necessary skills to absorb and use the technology they supplied. Third, we discuss how the government was able to raise enough revenue to finance these costly policies.

2.1 Meiji Technology Policy

Early Japanese reformers, most prominently Shozan Sakuma, began developing plans for how Japan could co-exist with the West. The challenge for the *bakufu* was immense as nineteenth-century Japan was a poor, feudal, and agricultural society with little knowledge of Western science and technology. Nevertheless, 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 caused

the *bakufu* to spring into action. Almost immediately after the Americans arrived in Japan, 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. Linguists and lexicographers have written extensively on the difficulty of scientific translation, which explains why little codification of knowledge happened in languages other than English and its close cognates: French and German (c.f. Kokawa et al. 1994; Lippert 2001; Clark 2009). The linguistic problem was two-fold. First, no words existed in Japanese for canonical Industrial Revolution products such as the railroad, steam engine, or telegraph, and using phonetic representations of all untranslatable jargon in a technical book resulted in transliteration of the text, not translation. Second, translations needed to be standardized so that all translators would translate a given foreign word into the same Japanese one.

Solving these two problems became one of the Institute's main objectives. Lexicographers have recognized that the Japanese language has a major advantage over other languages that translators could leverage (c.f. Kokawa et al. 1994; Lippert 2001). Because Japanese extensively uses Chinese glyphs, lexicographers and historians of translation note that it is far easier to express jargon in Japanese than in alphabetic systems. For example, few English speakers could guess the meaning of "locomotive" from its spelling, but Japanese translators created a word using glyphs that combined the Chinese characters for "steam" and "cart." While a reader coming across the term "steam-cart" might not recognize that it means locomotive on the first reading, it is easy to remember that "steam-cart" means "locomotive" once one learns the definition. Thus, while many languages coin jargon using root words from another language (e.g., Greek, Latin, or, in Japan's case, Chinese), the use of glyphs makes it much easier for a reader to remember a word's meaning.

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, 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, like "locomotive."[5]

Although Sakuma would be assassinated two years after the completion of his dictionary for his advocacy of Western studies, his ideas had a profound impact on Meiji reformers. From its inception on April 5, 1868, the Meiji Government stated that the assimilation of Western knowledge would be a central policy tenet. The Charter Oath, Emperor Meiji's five-sentence statement of the objectives of the fledgling government, declared 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 to support strengthening Japan by absorbing Western ideas.

Alongside the public provision of the dictionary, the public sector played an outsized role in

[5]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 first published in 1873.

translating technical books. After looking up the biographies of every person who translated a technical book between 1870 and 1885, we found that 74 percent of the translators were government employees.

In addition, the government adopted several other policies that facilitated the transfer of technology. First, the Japanese government hired 2,400 foreigners to come to Japan as instructors or advisors (Jones, 1980). Only four countries supplied a hundred people or more—Britain, the US, France, and Germany—with almost all other Europeans only sending ten or fewer people. 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, and postal networks, as well as public works projects. Japan chose Britain as the most popular source country for instructors in Industrial Revolution technologies, accounting for 46 percent of the total person-years of training. Adding in person-years from the U.S. and Canada reveals that 59 percent of the training was by people whose native language was English, 17 percent was by people from France and Belgium, and 13 percent was by people from Germany. The revealed preferences of the Japanese government in choosing instructors suggest that they saw instructors whose native languages were English and, to a lesser extent, French and German as the key sources of advanced Western technology. Second, the government not only financed foreigners to come to Japan but also paid for Japanese to study abroad.

Third, the government also provided practical training in state-owned enterprises (SOEs), which arguably was more effective than the "model factories" themselves. While early historical research argued that Meiji SOEs were important in leading the way for Japanese manufacturing, careful forensic accounting work by Morck and Nakamura (2007) and Morck and Nakamura 2018 has challenged this view. Between 1868 and 1885, Morck and Nakamura (2007) show that the cumulative losses of Japanese SOEs amounted to 12 percent of cumulative Japanese government expenditures over the same period. One might be tempted to dismiss these losses as a needed investment in learning-by-doing sectors, but the data rejects this hypothesis. Morck and Nakamura 2018 find that the government's privatization efforts, which began in 1880 in response to the dismal performance of SOEs, resulted in the private sector only willing to pay on average 50.0 percent of the *book value* of manufacturing SOEs. This was especially true in future high-growth manufacturing industries like textiles. For example, Tomioka Textiles and Hiroshima Cotton Spinning only sold for 53 and 22 of the book values of their respective capital stocks. In other words, markets believed the capital choices made by Japan's SOEs were so poor that even new management would have to write off much of the existing capital stock. Moreover, Meiji government budget statistics indicate that subsidies to private mining and manufacturing firms were nearly nonexistent, never amounting to more than 0.04 percent of all industrial subsidies in any year before 1910 (Ohkawa et al., 1965, vol 7, p. 180). Industrial subsidies primarily flowed to construction firms and other firms involved in building public infrastructure. Jointly, these results suggest that Japan's early experiments with SOEs and conventional industrial policy did not produce viable firms, but it leaves open the possibility that they may have provided some benefit in transferring technology by familiarizing workers with modern production techniques.

2.2 Education Policies

Beyond spending on technology policies directly, the Meiji government also deployed education policies. This is important, as supplying technical knowledge in the vernacular would make little sense unless entrepreneurs and managers had the human capital to absorb it. 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). 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 at this time 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. Rubinger (2000) argues that data from mandatory intake examinations for Japanese army conscripts provides us with a representative sample of young Japanese males that we can use to assess education 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 that 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 Japanese young men could have read technical books.

The Japanese government faced a more complex problem in building a university system because there were almost no Japanese with advanced knowledge of STEM fields. For this reason, many foreign workers hired by the Meiji government were employed by newly founded universities.

2.3 Paying For The Technology Transfer Policies

One may wonder how the Meiji government was able to raise enough government revenue to pay for these policies. As shown in Figure 1, real government expenditures tripled between 1871 and 1874. Paying for the foreign workers alone required substantial expenditures—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). Foreign study trips accounted for up to 0.20 percent of annual government expenditures (Jones, 1980, Table 7).

The key to Japan's newfound ability to pay for these programs was the 1873 Land Tax, 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 translators in the 1860s. As Yamamura (1986) discusses in detail, Takahira Kanda, 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 deadweight loss by instituting a heavy land tax on farmers. Figure 1 shows that the imposition of the land tax enabled the early Meiji government to finance enormous investments in codification and technical absorption. 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 War and Taiping Rebellion, could only raise 114 million yen even though China had ten times Japan's population. This eight-to-one Japanese advantage in per-capita taxation enabled Japan to finance human capital investments and public goods at a rate that Chinese reformers could only dream about.[6] For example, Japanese government ex-

[6] Wong (2012) reports that Chinese tax revenue in 1884 was 77 million silver taels. We performed the currency conversion in two ways. The number in the text uses the exchange rate series from (Fouquin and

Figure 1: Japanese Government Expenditure

Note: Government Educational Expenditure from *Meiji Taisho Zaisei Shoran [Survey of Meiji and Taisho Public Finances]* (1926).

penditures on education (discussed below) alone amounted to 11 percent of the budget in 1880 (Ohkawa et al., 1965); if China had attempted to implement this one part of the Meiji reform package for its population, it would have had virtually nothing left over for any other government functions.

In summary, the absorption of Western technology was a central aim of the Meiji government. To achieve this goal, the government adopted a multitude of large-scale technology and education policies. The funding of these fiscally intensive policies was made possible by the land tax reform, which itself was a product of Western "technology transfer". Starting in the mid-1880s, the historical record points to a marked shift in the Japanese economy. Pockets of modern, private, factory-based manufacturing began to emerge, predominantly in textiles. These textile mills used British machinery, inanimate power sources, and a modern industrial labor force. We now examine whether this shift in industrial structure can be linked to the codification and absorption of foreign technology. We note that we view the technology and education policies of the Meiji government as one "package" that made technical knowledge accessible to large swaths of the population. Our interest lies in understanding the effects of these policies.

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.

3 Data

In this section, we describe the main datasets used. First, we show how we quantify the supply of technical knowledge by sector. Second, we describe how we construct our novel dataset of codified technical knowledge across twenty major languages. Third, we discuss the trade dataset we constructed, which allows us to measure export and productivity growth by industry for many regions in the late nineteenth and early twentieth century. The appendix contains a complete discussion of all data used, including all data construction steps and sources.

3.1 Constructing the British Patent Relevance measure

A key challenge for this paper is to quantify the supply of technical knowledge by industry available to Japan and other regions. Our approach uses natural language processing to quantify one of the main channels through which codified technical knowledge diffused during this period: the dissemination (and translation) of technical manuals. We first describe industrial manuals and show i) that they contain relevant information on production processes and ii) that they are a source of knowledge transmission for technical follower countries. We then describe how we use natural language processing to quantify the amount of new Industrial Revolution technologies contained in each industry.

3.1.1 Industrial manuals as repositories of codified technical knowledge

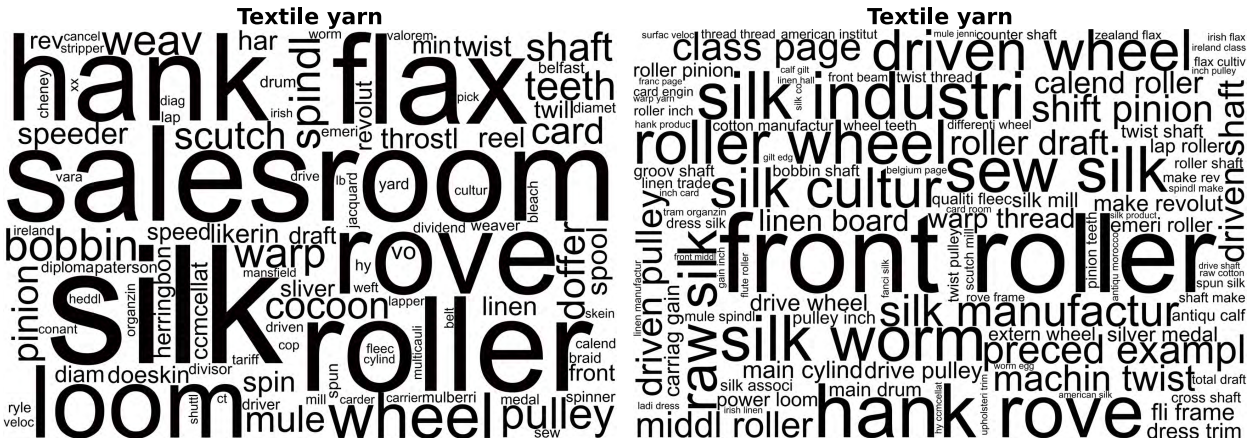
Industrial manuals initially emerged as a product of the industrial enlightenment in Europe (Mokyr, 2011), a movement which has been described as "access to knowledge created through an 'industrial public sphere' which included scientific societies, and *the coding, storing and transmission* of technical knowledge" (Berg, 2007, p. 125, emphasis added).

In the 17th and 18th centuries, contemporaries realized that knowledge could only be easily accessed if it was sorted and arranged systematically, giving rise to the publication of encyclopedias in multiple languages, the most famous of which was Diderot's *Encyclopédie*. By the end of the 18th century, " [m]anuals and books of instructions, often with excruciating detail and endless diagrams and minute descriptions of implements and processes, were published in every field" (Mokyr, 2011, p. 93).

Given that almost two-thirds of the foreign instructors brought to Japan were from Britain and other English-speaking countries, we assume that British technology and technology codified in English was the most common source of codified technical knowledge sought by the government. Nineteenth-century technical manuals in English 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. For example, "The American Cotton Spinner, and Managers' and Carders' Guide," published in 1851, provides guidance on the layout of the factory building, gearing, water wheels, the calculations of horsepower for propelling cotton spinning machinery, other preparatory machines (e.g., carding), and seventy pages dedicated to descriptions about the operation of spinning machines. A reader was able to attain a remarkable amount of useful, practical information on all aspects of modern mechanized, factory-based cotton spinning.

The perceived value of technical manuals is evidenced by the fact that their translation was part of states' technology absorption efforts. This was true in Revolutionary France (Horn, 2006, p. 176), in China during the Self-Strengthening Movement (Bo et al., 2023), and most importantly, in Meiji Japan (Montgomery, 2000), where public translation efforts were by far the most extensive we are aware of. 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

Figure 2: Word Clouds for Industries with the Highest British Patent Relevance



(a) Unigrams (b) Bigrams

of modern, factory-based manufacturing, as well as their day-to-day operation. Thus, we use industry-specific technical manuals to capture how intensively these technologies benefited each industry.

3.1.2 Quantifying the supply of technical knowledge by industry

We use a text-based approach to quantify how much new technical knowledge was created by the Industrial Revolution in each industry. This consists of four main steps (we provide a complete description in Appendix F). First, we use the text of technical manuals as a measure of production methods at the technology frontier. Second, we take patent text as a measure of technical knowledge created during the Industrial Revolution. Technical manuals and patent text thus constitute the data sources we rely on. The third step entails using standard natural language processing techniques to represent the text as data. Fourth, we compute the similarity of text in industry technical manuals and patent text for each industry. This similarity score captures how relevant Industrial Revolution technologies were for an industry's production processes. We now describe each step in turn.

The previous section established that technical manuals contained information about state-of-the-art production and organizational methods by industry. We use the full text of these historical technical manuals directly. Specifically, for each three-digit SITC industry in our data, we create a curated list of technical manuals from the *HathiTrust Digital Library*. Table 1 shows a random sample of the 460 books we selected from the *HathiTrust Digital Library*. We take the full text of these technical manuals to represent frontier knowledge of codifiable production techniques. Figures 2 and 3 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".

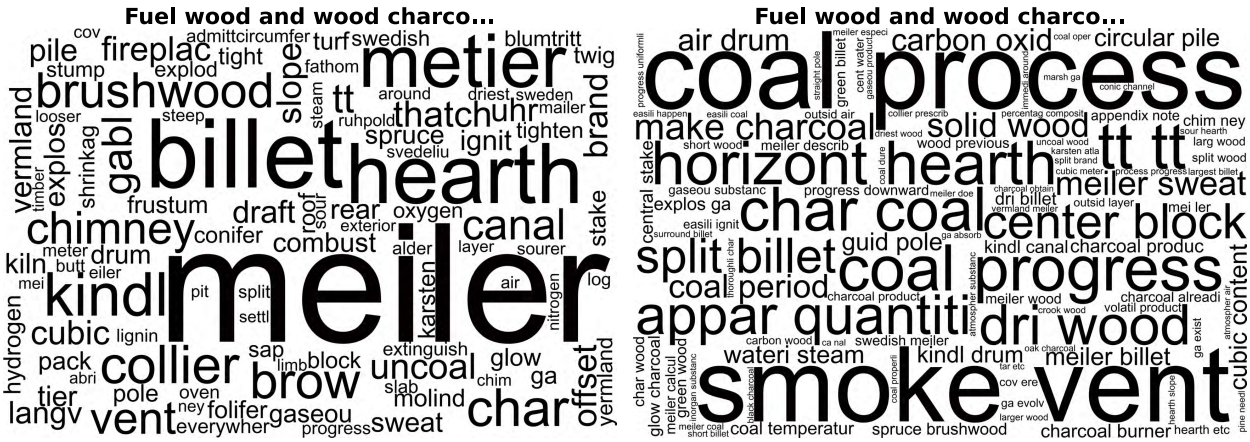
To proxy the full body of new technical knowledge created during the Industrial Revolution, we digitized the text of British patent synopses (1617-1851) from Bennet Woodcroft's "Subject Mat-

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

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

Note: This table provides a sample of the books we obtained from the *HathiTrust Digital Library*. We randomly picked 20 industries, and for each industry, we randomly picked one book.

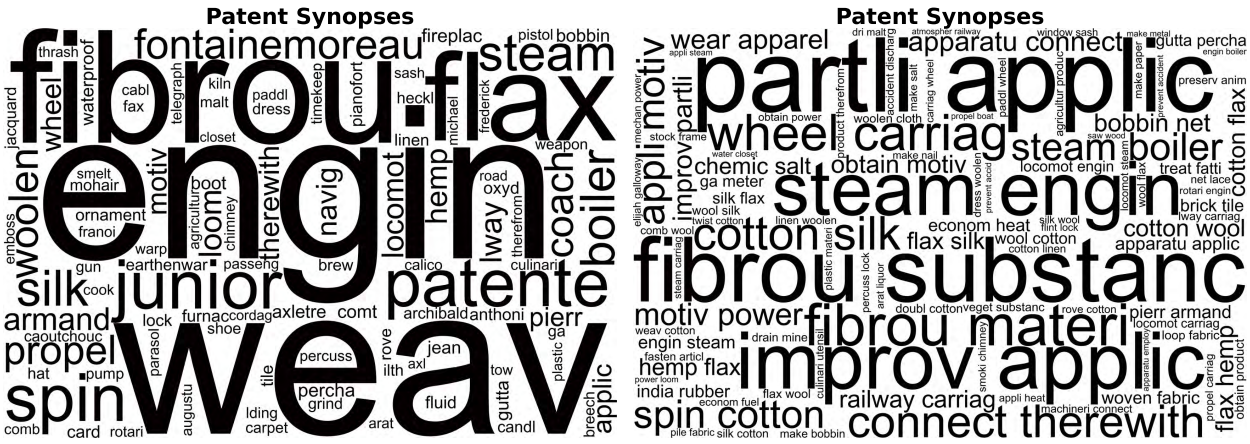
Figure 3: Stemmed Word Clouds for Industries with the Lowest British Patent Relevance



(a) Unigrams

(b) Bigrams

Figure 4: Stemmed Word Clouds of Patent Synopses



(a) Unigrams (b) Bigrams

ter Index of Patent of Invention". While patents capture only a subset of innovation (Moser, 2005), the part they do may be more relevant for measuring innovation that could diffuse at a (potentially very large) distance. The main alternative form of protecting intellectual property is secrecy, which inherently limits the ability of the technology to diffuse. Likewise, knowledge that cannot be codified is harder to diffuse and absorb and will be absent from technical manuals. For these reasons, the patent text should proxy new technologies of the Industrial Revolution that could most easily diffuse. Figure 4 plots the unigrams and bigrams for British patent synopses. High-frequency unigrams include "engine," "spin," "weave," "steam," "loom," and "boiler;" high-frequency bigrams include "steam engine", "fibrous substance", and "motive power". Thus, these unigrams and bigrams seem to capture the technologies and concepts used in the Industrial Revolution.

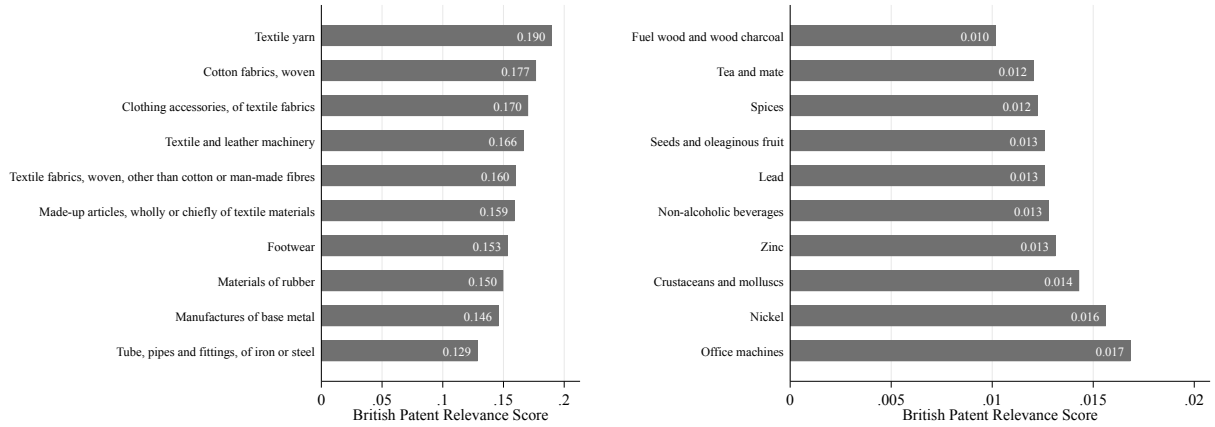
With these two data sources, we have measures of the state-of-the-art production methods by industry (the technical manuals) and new technologies of the Industrial Revolution (British patent text). The third step is to represent this textual information as data, which we do by taking a vector representation of the text. Each document (technical manual, patent text) n is represented by a vector of size 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.

The fourth and final step is to quantify the technological relevance of British patents for the processes described in the technical manual for an industry. We assume that if an industry's manuals use similar words and phrases as the patent synopses, the patents are likely relevant for the industry. The cosine similarity measure is the standard metric for measuring the similarity of two texts (documents) in natural language processing. This is 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 manual $\boldsymbol{TM_i}$ for industry i is

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

which we call the *British Patent Relevance (BPR)* measure. Figure 5 plots the bar chart for the industries with the ten highest and lowest cosine similarity scores. Reassuringly, high BPR industries

Figure 5: Industries Ranked by British Patent Relevance



(a) Top 10

(b) Bottom 10

include textile, footwear, machinery, and manufactured intermediate-input sectors, whereas low BPR industries contain mostly unprocessed raw materials, which were largely unaffected by Industrial Revolution technologies.

Our measure has several advantages for our setting. Most importantly, by focusing on knowledge codified in technical manuals, it captures one of the key channels through which Meiji 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 patent about making that final output.

3.2 Measuring the codification of knowledge around the world

We collected data on codified "useful" knowledge available in local vernaculars every year for 33 languages, encompassing all twenty 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 H for details).

For many major European and Asian languages, we scraped the national libraries of countries where the language is the native tongue of a substantial fraction of the population. 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 thousands of libraries worldwide covering dozens of languages. If we can scrape a language from a national library and WorldCat, we scrape this language from both sources and pick the source that yields the most books. We also supplemented data scraped from the National Diet Library (NDL) for Japan by scraping an additional 81 major libraries because we found that the NDL collection of books published before 1870 is limited.

We, therefore, present two samples of books for Japan. The "Japanese: NDL" sample is based on the holdings of just the NDL and is thus methodologically comparable to the methods used for other major languages; and the "Japanese: All" sample contains all books contained in any major Japanese library and is a more comprehensive measure for Japan that we use when international comparisons are not needed. Using the publication year of each book in our sample, we construct the time series of codified knowledge by spoken language. This yields what is, to the best of our knowledge, the first systematic dataset on technical knowledge available in the local vernacular for major languages.

3.3 Cross-region, bilateral, industry level trade flows

We construct what is, to the best of our knowledge, the first cross-region dataset of harmonized, bilateral, industry-level trade flows quinquennially for the years 1880-1910 using detailed historical trade records for Japan, the United States, Belgium, and Italy (*"reporting countries"*, henceforth).[7]

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 for Belgian exports and imports in manufactures from Huberman et al. (2017); for Japanese exports from Meissner and Tang (2018); and for Italian exports and imports (for major trading partners only) from Federico et al. (2011).[8] An observation in this dataset, x_{ijk}, refers to an export flow in sector k from origin i to destination j. 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.

We harmonized product lines using the three-digit Standard International Trade Classification (SITC) revision 2 in line 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 SITC category across all datasets. 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).[9]

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. Our dataset consists of export values for 39 regions in 91 sectors. Table 2 contains

[7]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 at the industry level, they cannot be used for our purposes.

[8]We do not include Germany's digitized trade data in the combined dataset. This is due to the fact that Germany's historical trade statistics before 1906 present a number of distinct methodological challenges that make comparison 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 kept their own records, making the construction of a single dataset accounting for all German trade challenging. Second, during our sample period, the classification scheme for products was revised multiple times: at different points in time, between 400-1,200 distinct products were listed, making it challenging to construct a consistent classification over time (see Hungerland and Wolf (2022, Figure 6A)).

[9]We use the term "region" instead of "country" because many trade flows were reported for colonies and other geographies that were not formally countries.

Table 2: Summary Statistics

Variable	N	Mean	SD	p25	p50	p75
Double-Relative Productivity Growth of Industry k in the Country i ($\tilde{\gamma}_{ik}$)	1245	0.00	0.04	-0.01	0.00	0.02
Double-Relative Productivity Growth of Industry k in Japan	56	0.00	0.05	-0.01	0.01	0.02
Exporter's Industry Growth Rate	1396	-0.10	0.38	-0.05	0.03	0.09
Exporter's Industry Growth Rate in Japan	71	-0.06	0.37	-0.02	0.04	0.15
Britsh Patent Relevance	125	0.07	0.09	0.03	0.05	0.08
Country's log of GPD Per Capita in 1870	61	7.33	0.61	6.86	7.25	7.76
Country's log of GPD Per Capita in 1913	61	7.77	0.72	7.16	7.54	8.43
Number of Weeks to Learn Language from English	32	3.71	0.42	3.29	3.78	3.78
Country's Distance to UK	61	8.17	1.14	7.37	8.39	9.13

Note: $\tilde{\gamma}_{ik}$ refers to the productivity growth of industry k in the country i. "Number of Weeks to Learn Language from English" as measured by the U.S. State Department. The GDP per capita data is from the Maddison Project. "Country's Distance to the UK" refers to the distance from country i to the UK using the Great Circle formula. The data was sourced from CEPII.

the summary statistics.

4 Stylized Facts

We use our data to document four novel facts about language, codification of technical knowledge, and economic development in the nineteenth century. First, we show that people speaking a few languages produced prodigious amounts of codified technical knowledge, but readers of most languages had almost no codified technical knowledge to read in their vernacular. Second, we document that Japan exhibited a massive increase in codified knowledge shortly after Japan invested heavily in translation technology, education, and technology transfer. Third, we document that difficulty in reading one of the main languages in which knowledge was codified (English) was associated with significantly lower per capita income, which establishes that the ability to read technical books is associated with higher per capita income. Fourth, we document that the rise in Japanese manufacturing exports did not happen immediately after opening to trade nor after the institutional reforms of the Meiji Restoration. Instead, it happened directly and swiftly after Japan codified large amounts of technical knowledge.

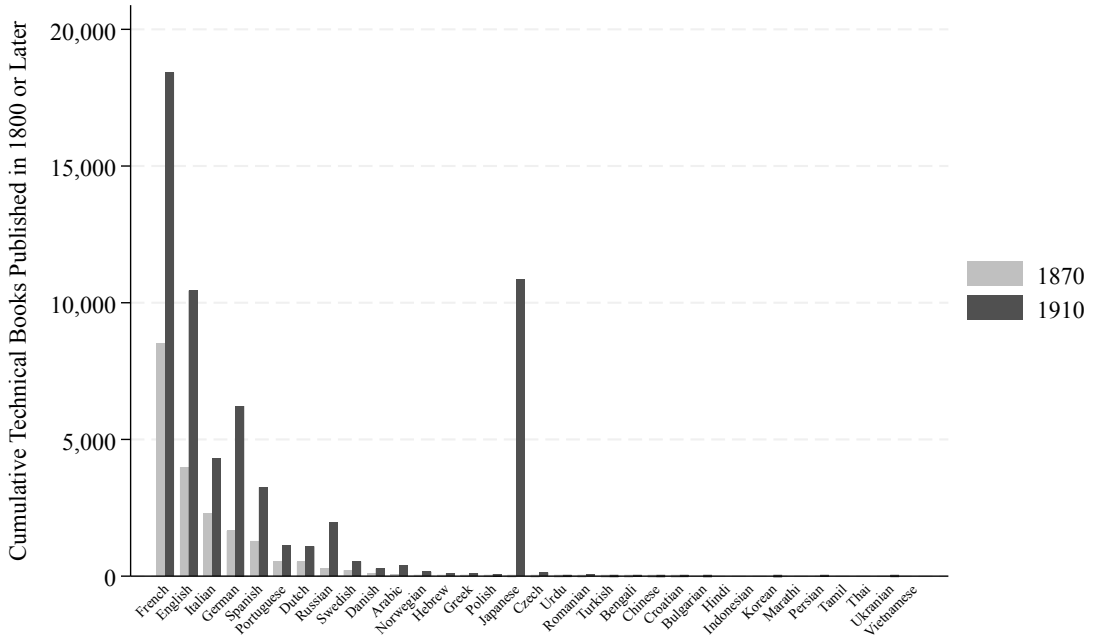
4.1 Books in most languages contained little codified knowledge

We use our novel data, assembled from the catalogs of national libraries, to examine access to codified technical knowledge across major languages in the late nineteenth century for the first time. There is striking inequality in codified knowledge worldwide. Figure 6 presents the extent of codification of technical knowledge in 1870 and 1910. France led the world in codification, with 8,753 volumes in their national library in 1870 and close to 18,678 by 1910.[10] In 1870, 70 percent of all technical books were written in three languages: French, German, and English. Forty years later, these three languages accounted for 58 percent of all technical books, with the fourth language, Japanese, accounting for 18 percent of technical books. Thus, people unable to speak English, French, German, and later Japanese were largely excluded from accessing codified technical knowledge. Put differently, while literacy in a vernacular opened the doors to vast amounts of codified knowledge for speakers of a few languages, it meant little to no access to codified knowledge in most languages.

[10]The historical record is consistent with this finding. France was the world leader in science (Gillispie, 2004, 2009), and led Britain in codifying knowledge (Mokyr, 2021).

These results also help explain the Japanese choice of source countries for foreign advisors that we discussed in Section 2. We have shown that 89 percent of all foreign instructors were native speakers of the three languages that accounted for 86 percent of all codified technical knowledge. Clearly, the Japanese government saw learning the technical information encoded in these languages, especially English, as the key to mastering Western technology. It also provides us with an estimate of the lower bound of codification needed to obtain mastery of Western technology in this period. If Germany was seen by the Japanese as close to the technical frontier with having 1,801 technical books written in German in 1870, then we can assume that Japan, too, would have mostly codified Western technology if it crossed this threshold.[11] Since this happened in 1887, we will choose this year as the point that Japanese attained the ability to read Western science in Japanese as easily as major Western industrial powers could read it in their vernaculars. By 1890, the Japanese NDL housed more technical books than any other national library in our sample except the Bibliothèque National de France and the Library of Congress.

Figure 6: Codified technical Knowledge in Major World Languages (num. books)

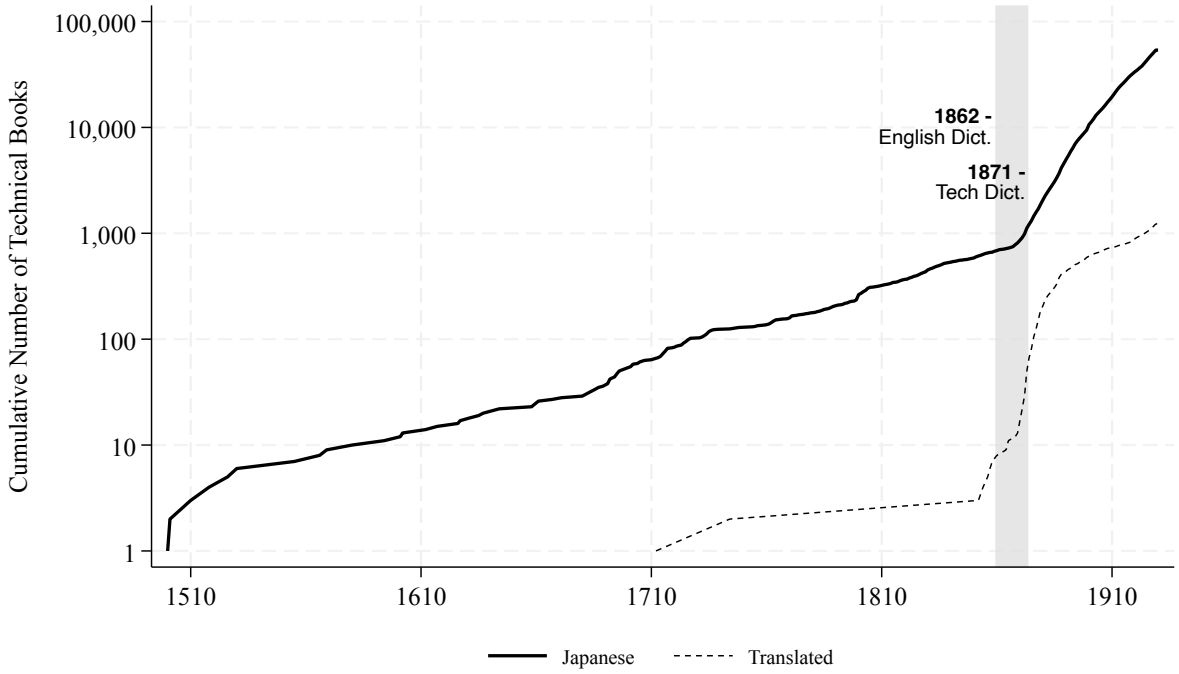


Note: Technical knowledge is measured as the number of books in the following subjects: agriculture, applied sciences, commerce, industry, and technology. Data on the number of books in these subjects was scrapped from the online catalogs of WorldCat (Italian, Spanish, Portuguese, Russian, Arabic, Turkish, Persian), the Bibliothèque National de France (French), the Deutsche Nationalbibliothek (German), the Library of Congress (English), all major Japanese libraries (Japanese: All) or the NDL (Japanese: NDL) and the National Library of Korea (Korean). The languages are ordered based on the total number of technical books published up to and including 1870.

[11] Examples of nineteenth-century German technical mastery can be garnished by considering contemporaneous innovations such as Hertz's proving the existence of electromagnetic waves (1880s), Röntgen's discovery of X-rays (1895), Otto Lillenthal's invention of the glider (1894), Philipp Reis's development of an early telephone (1861), Benz and Daimler's invention of gas-powered automobiles (1886), Rudolf Diesels' invention of the diesel engine (1896), Felix Hoffmann invention of aspirin (1897), and Friedrich Wohler's synthesis of urea (1828)

4.2 The Japanese language had uniquely high growth rates of codified knowledge

Figure 7: Codified Technical Knowledge in Japan (log scale)



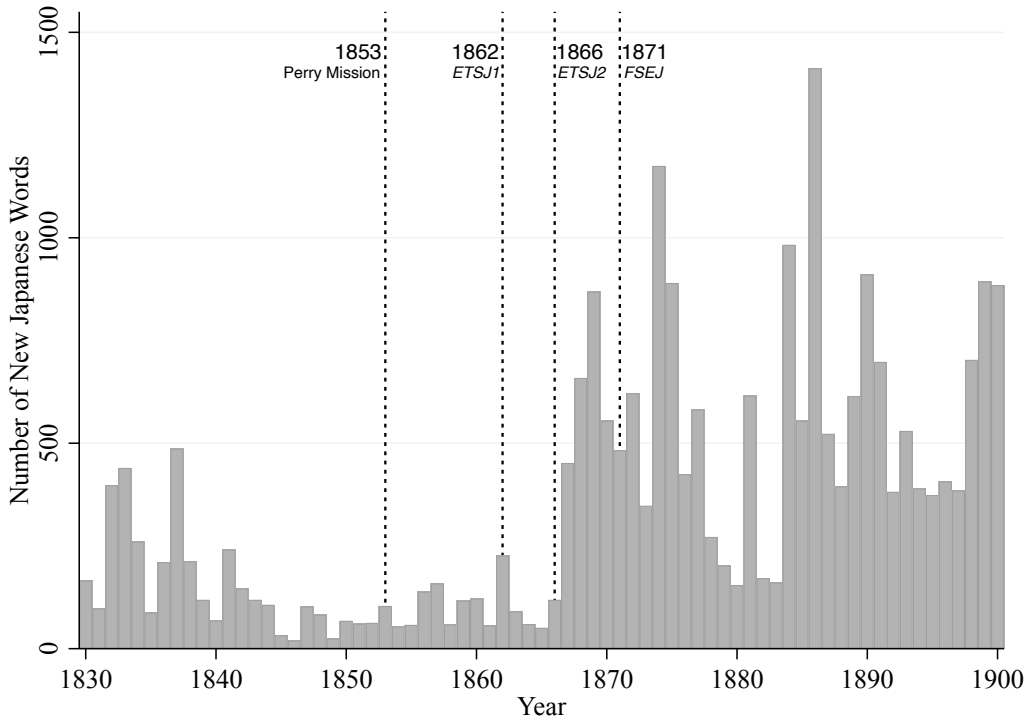
Note: Codified technical knowledge for each year refers to all technical books written in Japanese in the NDL catalog or any other Japanese library linked to the NDL. A book is considered "Translated" if the NDL flags the book as a translation.

An interesting feature of Japan's trajectory is that the increase in codified knowledge happened suddenly and almost immediately after Japan began producing English-Japanese dictionaries and investing heavily in technological transfer. We measure the universe of technical knowledge in Japan by supplementing the NDL holdings with those in 81 additional Japanese libraries to construct a time series of all technical books from 1500 to 1930. We present the result in Figure 7. Between 1600 and 1860, the number of technical books in Japanese grew by 1.6 percent per year. The rate almost sextupled to 8.8 percent per year between 1870 and 1900, starting just as staff at the *bakufu*'s 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 see an even 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 technological absorption.

One potential concern about this evidence is that the growth in books containing technical knowledge might reflect changes in library acquisition policies rather than the success of Japanese

lexicographers in inventing Japanese jargon for technical concepts. To ensure this possibility does not drive our results, we supplement these data with additional data on new word creation in the Japanese language to better understand the country's unique growth trends. We obtained the first recorded use of Japanese words based on the revised edition of the *Nihon Kokugo Daijiten* (Large Japanese Dictionary), published by Shogakukan, encompassing 300,000 Japanese words.

Figure 8: Word Creation in Japan



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 Japanese-English dictionaries.

As one can see from Figure 8, 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, a telegraph machine, cameras, etc. Obviously, exposing the Japanese to Western technology did not prompt the Japanese to create new words to describe new technology. 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.[12] Word creation 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

[12]Only 200 copies of the dictionary's first edition were published in 1862. Thousands were published in 1866. Similarly, the FSEJ was printed in small quantities in 1871 but was widely available for sale in 1873 (Kokawa et al., 1994).

Table 3: Linguistic Distance from English and GDP

	Log GDP per Capita					
	(1)	(2)	(3)	(4)	(5)	(6)
	1870	1913	2018	1870	1913	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. Physical distance between the region and the UK is from *CEPII* database using the Great Circle Formula. Number of weeks an English native speaker will take to obtain "Professional Working Proficiency" in the majority language of a country as estimated by the U.S. Department of State's Foreign Service Institute. Robust standard errors are in parentheses.

the introduction of new ways of codifying knowledge about the world, this evidence suggests that dictionary creation and other translation efforts in the 1860s helped foster the inflow of new ways of expressing ideas in Japan.

4.3 Per capita income falls with linguistic distance from languages that codified knowledge

We have documented that English, French, and German accounted for almost 90 percent of all codified technical knowledge in 1870. Thus, if access to technical knowledge were important for technology diffusion and development, we would expect a correlation between development and linguistic distance (a proxy for how costly or difficult it was for a region to access technical knowledge in English). We focus on linguistic distance from English because we can readily compute this number and because French and German are linguistically close to English. As such, languages that are very different from English are also very different from French and German. To understand how proximity to England was related to per capita income, we regress per capita income in several benchmark years from the Maddison data on linguistic distance from English, controlling for physical distance. We measure linguistic distance using the U.S. State Department's estimate of the number of weeks required for an English speaker to obtain "Professional Working Proficiency" in the plurality language in a region under the assumption that if it is difficult for an English speaker to learn a language, it will be difficult for a speaker of the other language to learn English. We control for physical distance from England because physical proximity facilitates technological transfer through travel, trade, investment, and communication.

Table 3 shows that linguistic distance from English is negatively associated with economic development. In each column, we regress the log of per capita income on physical and linguistic distance. Columns 1-3 report the results for 1870, 1913, and 2018 for a specification in which we include regions in which a plurality of residents speak English, and in columns 4-6, we drop English-speaking regions from the regression.[13] The negative relationship between per capita income and linguistic distance is significant in 1870 and 1913 and remains present, although not significant, in 2018.[14] Moreover, the magnitude of the coefficient in the early years is also econom-

[13]We use 1913 because it is a benchmark year in the Maddison data with more observations than 1910.

[14]This likely reflects the fact that there has been substantial globalization of knowledge over the last 100

ically significant. For example, an English speaker can become proficient in Spanish or Portuguese in 24 weeks if they are in class 25 hours per week, but it takes 88 weeks to reach the same level of proficiency in Arabic, Chinese, Japanese, or Korean. The parameter estimates in column 1 imply that this difference in linguistic distance is associated with a 46 percent lower GDP per capita.

Although much of the estimated impact appears to come from whether a region's native language is English, columns 4-6 indicate a substantial linguistic distance effect even if we restrict ourselves to regions whose vernacular is not English. In column 4 of the table, we repeat the exercise after dropping the six English-speaking regions in our dataset (Australia, Canada, England, Ireland, New Zealand, and the U.S.). The coefficient on linguistic distance falls by half, but even so, we identify a statistically and economically significant association. Our estimates indicate that speaking a linguistically distant language like Japanese is associated with 27 percent lower income per capita than speaking a linguistically close language like Spanish.

To ensure that outliers do not drive this result, Appendix Figures A.1 and A.2 present partial regression plots for 1870 and 1913, respectively. The generally negative association in the data is apparent, though there are some outliers associated with resource-abundant countries like Australia, Argentina, Canada, the U.S., and Uruguay. While we do not interpret these results as causal, we conclude that these patterns are consistent with access to codified knowledge playing a potentially important role in driving development in the nineteenth and early twentieth centuries.

4.4 Japanese manufacturing grew rapidly after codifying knowledge

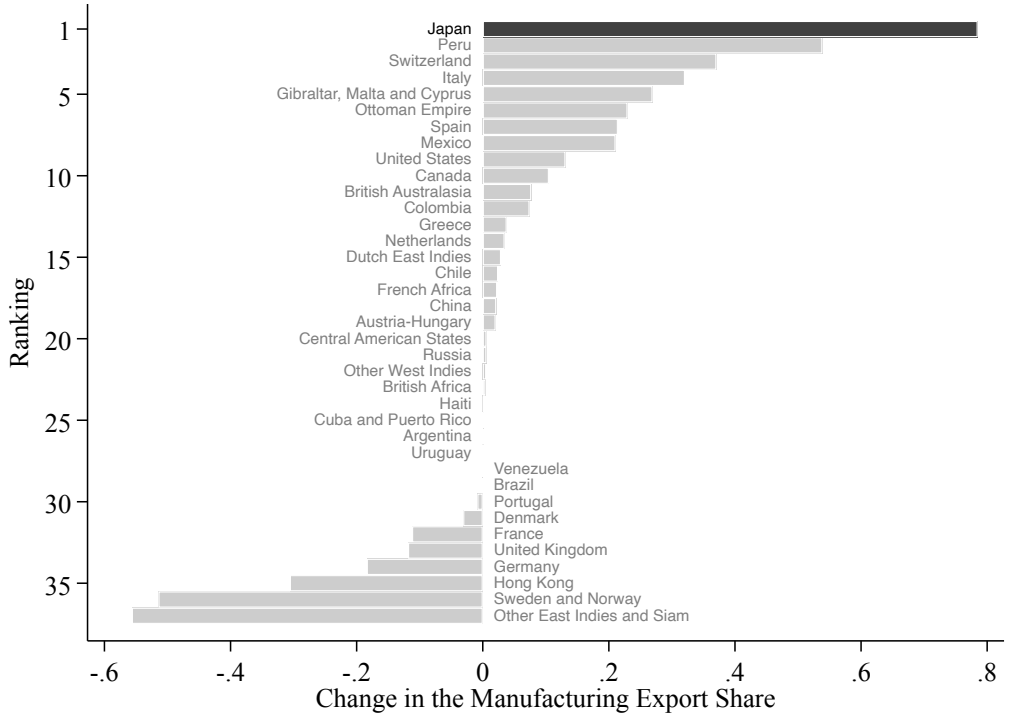
Our digitization of global trade data allows us to uncover another novel fact about the global spread of the Industrial Revolution: Japan experienced unique and explosive growth in its share of manufactured goods in exports over this period. Figure 9 plots the change in the manufactured exports expressed as a share of total exports for each region in our dataset. The Japanese manufacturing share of exports rose from close to zero to almost 80 percent over this period, indicating the Japanese economy was transformed from being principally an exporter of primary products to being a major exporter of manufactured goods.

These data paint a very different picture of Japanese industrialization than what one obtains using aggregate Maddison data. In the Maddison data, Japan's economic development is almost absent. Between 1850 and 1870—a period bracketing Japan's opening to trade with the West—per capita income growth was 0.5 percent per year. This number is hardly higher than the 0.3 percent per year per capita income growth between 1820 and 1850, a time when Japan was closed to the West. Similarly, the Meiji creation of institutions is also not associated with a growth boom. Japanese per capita income growth between 1870 and 1885 was only 0.6 percent per year. Even more puzzling is that the Maddison data indicates that Japanese per capita income growth, while faster, remained unremarkable between 1885 and 1910, averaging only 1.2 percent per year. There are two reasons why the "Meiji Miracle" does not appear in the Maddison data. First, as Fukao et al. (2015) (the source for the Maddison data) documents, labor productivity growth between 1885 and 1899 averaged 2.7 percent per year: more than double per capita income growth. Second, the aggregate Maddison data obscures the rapid structural change happening in Japan as it shifted from an agricultural economy to a manufacturing one.

To understand the timing of the rapid growth of Japanese exports, we switch data sources to the *Foreign Trade of Japan*, which provides Japanese industry-level export data. These data are not directly comparable with the bilateral data we use in Figure 9 because they report total Japanese exports to all countries, not just the countries we use to estimate productivity. However, the data let us build a longer time series and provide a more accurate picture of the shift in Japanese

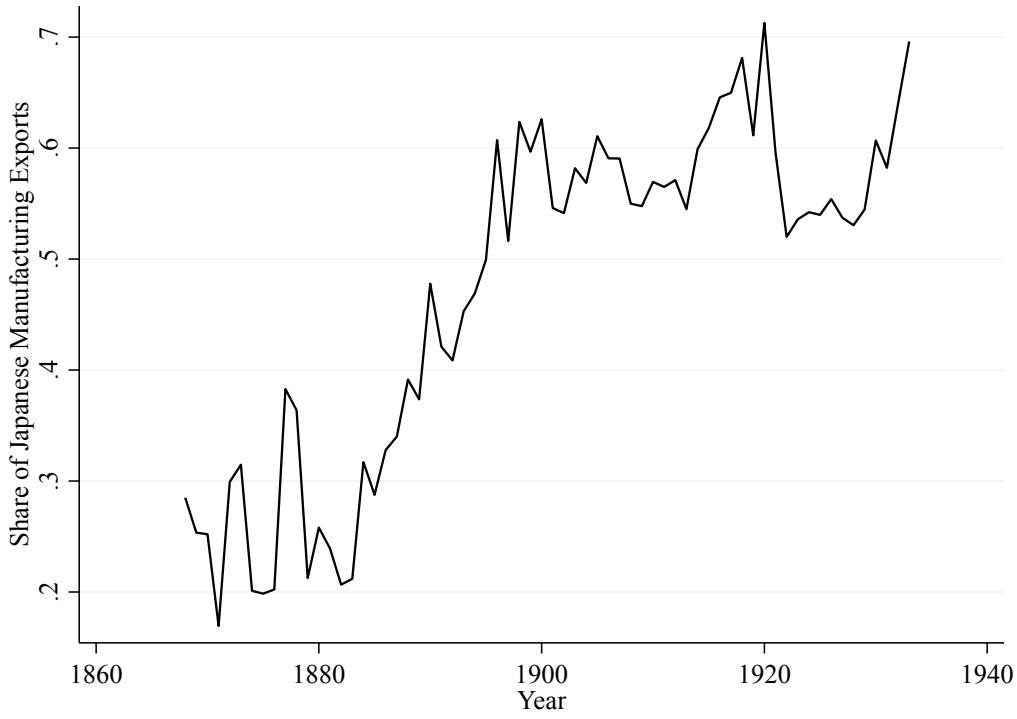
years.

Figure 9: Change in the Manufacturing Share of Exports Around the World (1880-1910)



Note: Manufacturing export share computed using historical bilateral trade data reported by Belgium, Italy, Japan, and the United States. 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 text for details on variable construction.

Figure 10: Share of Japanese Manufacturing Exports



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

exports. The Japanese manufacturing export share rose from around twenty percent in the early 1880s to seventy percent: an increase of fifty percentage points, which is smaller than what we saw in Figure 9, but quite large nonetheless. These data reveal no upward trend in the share of manufactured exports until the early 1880s. This result is surprising since the shift came more than two decades after Japan opened to trade and fifteen years after the Meiji Restoration. The fact that the manufacturing takeoff occurred shortly after the Japanese codified about as many books as Germany had in 1870 motivates our empirical analysis exploring whether the unique ability of Japanese people to read technical books affected their ability to use the knowledge contained in these books.

5 Estimating Productivity Growth

In this section, we show how to use trade data to build a global database to measure productivity growth at the region-industry level. Here, we explain how we estimate productivity growth for our set of regions. In appendix C, we explain how we convert our estimates into annual growth rates.

5.1 Estimating Productivity Growth

Our starting point is the framework of Costinot et al. (2012) who build a multisector Eaton and Kortum (2002) model featuring an economy with multiple countries, multiple industries, and one factor of production, labor. They show that one can write the value of exports from i to j in

industry k at time t (x_{ijkt}) as

$$\ln x_{ijkt} = \gamma'_{ijt} + \gamma'_{jkt} + \theta \ln z'_{ikt} + \epsilon'_{ijkt}, \tag{2}$$

where γ'_{ijt} is an importer-exporter that captures bilateral trade frictions and exporter-importer aggregate supply and demand forces (e.g., country size and distance) that matter for exports; γ'_{jkt} is an importer-industry fixed effect that captures deviations in importer demand in industry k; θ is the Fréchet scale parameter, z'_{ikt} captures comparative advantage, i.e., factors that shift productivity in a given exporter and industry; and ϵ'_{ijkt} is an error term that captures how trade costs deviate at the industry-exporter-importer level from the exporter-importer average.

Our interest is to estimate $\gamma_{ikt} \equiv \theta \Delta \ln z'_{ikt}$ using trade data. We will estimate it by noting that we can first-difference equation 2 and rewrite it in terms of fixed effects:

$$\Delta \ln x_{ijk} = \gamma_{ij} + \gamma_{jk} + \gamma_{ik} + \epsilon_{ijk}, \tag{3}$$

where we have suppressed the time subscripts 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.[15] This equation can be rewritten to yield

$$\Delta \ln x_{ijk} = \gamma_{jk} + \gamma_{ik} + \tilde{\epsilon}_{ijk}, \tag{5}$$

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

Estimation of equation (5) requires us to drop observations in which the initial bilateral export flow in an industry 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 and industry.[16]

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

[15] One can see this by noting that equation 3 can be rewritten as

$$\Delta \ln x_{ijk} = (\gamma_{ij} + \gamma_i + \gamma_j) + (\gamma_{jk} + \gamma_k - \gamma_j) + (\gamma_{ik} - \gamma_i - \gamma_k) + \epsilon_{ijk}, \tag{4}$$

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

[16] This property of the AW estimator is not shared by the Poisson pseudo-maximum likelihood (PPML) estimator. In particular, PPML estimates do not yield a simple mapping between the estimated parameters and trade growth. PPML fitted values match the observed aggregate export flows *in levels* by country, but not the bilateral flows. Therefore, the relative changes predicted by PPML, based on the coefficients

of moment conditions. First, the estimates aggregate to match total exports in every exporter-industry observation i:

$$\frac{\sum_j x_{ijk,t} - \sum_j x_{ijk,t-1}}{\sum_j x_{ijk,t-1}} = \gamma_{ik} + \sum_j \frac{x_{ijk,t-1}}{\sum_\ell x_{i\ell k,t-1}} \gamma_{jk}, \tag{6}$$

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:

$$\frac{\sum_i x_{ijk,t} - \sum_i x_{ijk,t-1}}{\sum_i x_{ijk,t-1}} = \gamma_{jk} + \sum_i \frac{x_{ijk,t-1}}{\sum_\ell x_{\ell jk,t-1}} \gamma_{ik}. \tag{7}$$

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 every region's export and import growth.

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

$$\gamma_{ik} = \gamma_i + \gamma_{1k} + \tilde{\gamma}_{ik}, \tag{8}$$

and

$$\gamma_{jk} = \gamma_j + \gamma_{2k} + \tilde{\gamma}_{jk}, \tag{9}$$

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 due to shifts in exporter characteristics (e.g., productivity or size). Second, $\tilde{\gamma}_{ik}$, the "comparative-advantage" component of productivity 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_k).[17] Finally, recalling that $\gamma_{ik} \equiv \theta \Delta \ln z'_{ikt}$, we can define $\Gamma_{ik} \equiv \tilde{\gamma}_{ik}/\theta$ as the change in exporters i's productivity in industry k that cannot be explained by changes in industry factors or general conditions in the exporting country.

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. All results reported below refer to annualized estimates.

differenced over time, average to the observed country-level relative change in exports if weighted by the initial PPML *fitted* bilateral flows. AW estimates have the property that they average to the observed aggregate relative changes in trade flows when weighted by the initial *observed* trade flows. We thank Kirill Borusyak for pointing this out.

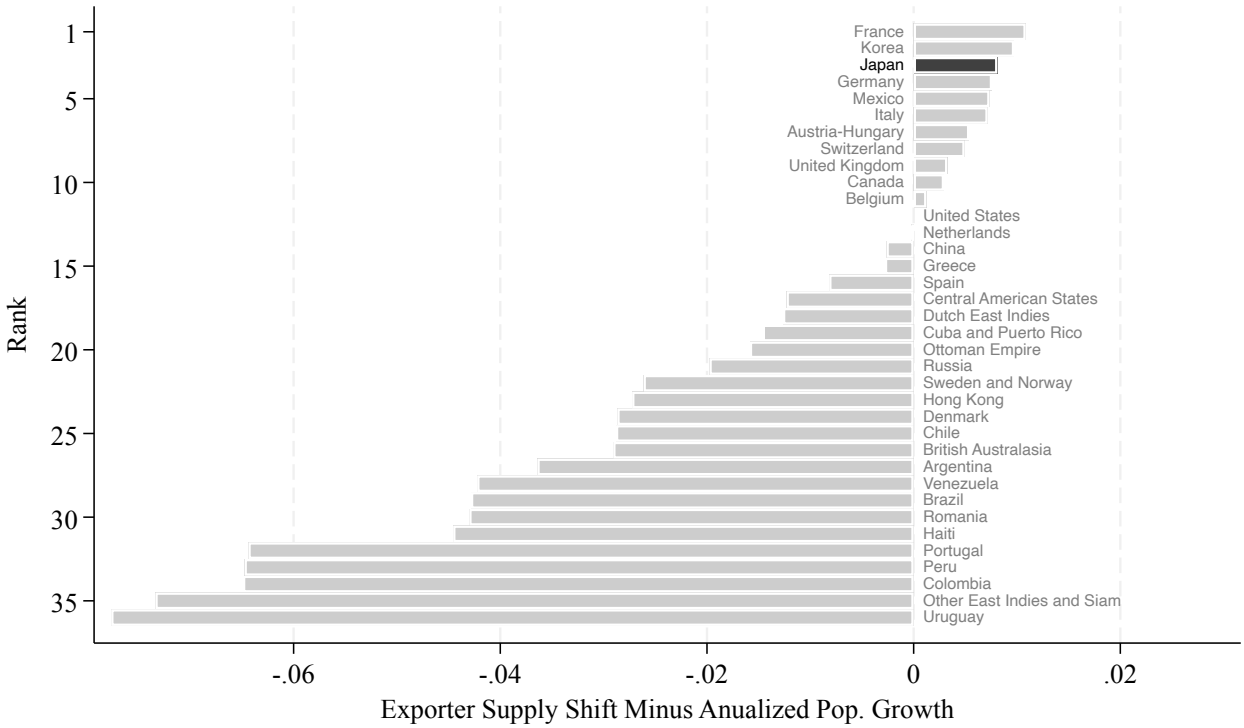
[17] Although we do not use the other normalization constants, we can recover them. $\hat{\gamma}_k \equiv \hat{\gamma}_{1k} + \hat{\gamma}_{2k}$ is the shift in exports that can be attributed to movements in industry k's characteristics (e.g., global productivity growth in k or global demand for k). 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.

6 The Meiji Miracle in Comparative Perspective

The methodology developed in the previous section allows us to provide the first systematic estimates of productivity growth for many regions in the late nineteenth and early twentieth century. 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.5 plots the annualized per capita shift in export supply net of population growth relative to the value for the US, i.e., $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$.

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 Portugal, Peru, Colombia, and Uruguay show weak performance. Notably, the export-supply shifter for Japan 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.

Figure 11: Relative Annualized Per Capita 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 $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$. Annual population growth is computed between {1870,1880} and 1913 depending on data availability in the Maddison data. See text for details on variable construction.

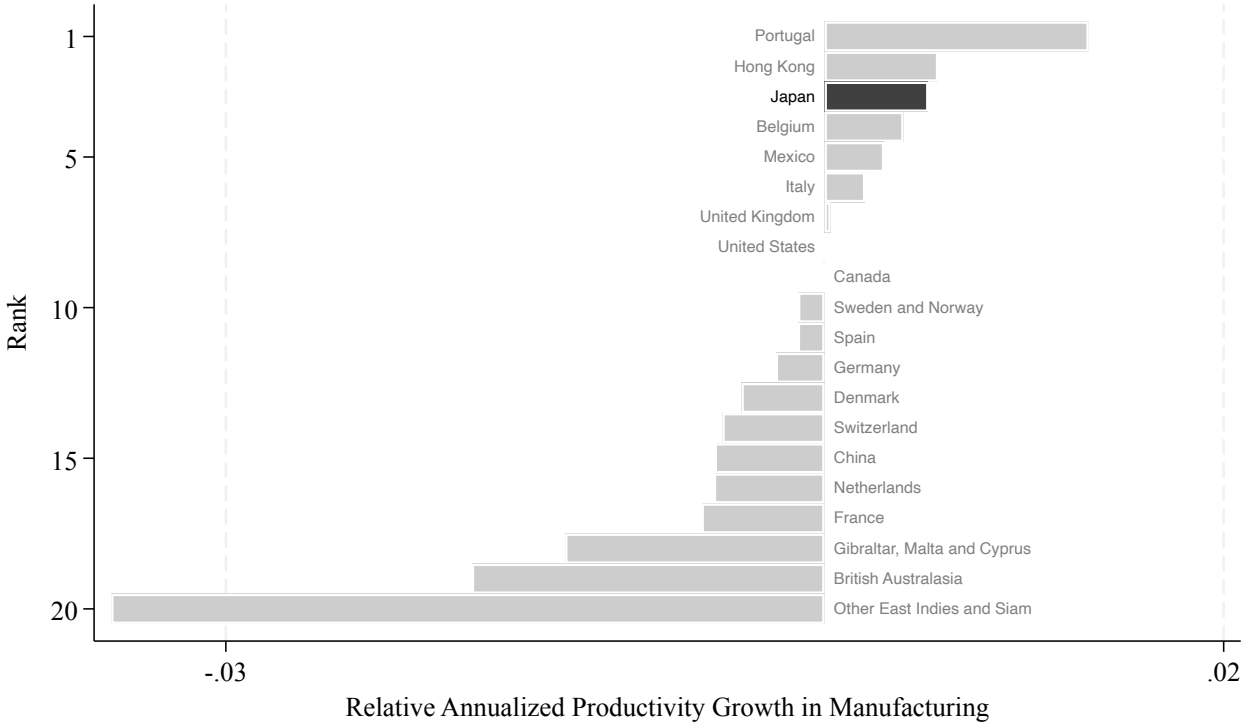
Next, we examine the extent to which productivity growth was biased towards manufacturing. We regress the comparative-advantage component of productivity growth, Γ_{ik} on broad industry dummies:

$$\Gamma_{ik} = \beta_i^{\text{Agg}} \times I_k^{\text{Agg}} + \beta_i^{\text{Mfg}} \times I_k^{\text{Mfg}} + \beta_i^{\text{Min}} \times I_k^{\text{Min}} + \epsilon_{ik},$$

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 12 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.5 indicate that these economies had low overall rates of productivity growth, which implies that while they did relatively well in manufacturing, their overall productivity growth was poor. The next seven countries (Japan, Belgium, Mexico, Italy, the UK, the US, and Canada) are all examples of regions that were industrializing over this period by exhibiting rapid productivity growth (as measured by δ_i) and having 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

Figure 12: 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 6 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.

that Japan's unparalleled shift towards specialization in manufacturing (Figure 9) was driven by productivity growth biased towards manufacturing—that is, shifting Ricardian comparative advantage. In the next section, we explore whether the growth in technical knowledge in Japanese allowed Japanese entrepreneurs to harness Industrial Revolution technologies.

7 Codification and Development

The previous sections established that i) Japan experienced strong productivity growth between 1880 and 1910, mainly driven by its manufacturing sectors, and ii) 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 could increase productivity across sectors. Intuitively, would-be entrepreneurs of textile yarn, which had undergone enormous changes in production methods during the Industrial Revolution, 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 Industrial Revolution technologies, had far fewer productivity benefits to reap from being able to read technical knowledge. We operationalize how much an

industry could benefit from access to technical knowledge with the British Patent Relevance (BPR) measure introduced in Section 3.1.

We test this relationship by estimating regressions of the form

$$g_{ik} = \alpha_i + \beta_J * BPR_k \times I_{iJ} + \beta_r * BPR_k \times I_{ir} + \epsilon_{ik}, \tag{10}$$

where g_{ik} is either annual export growth (raw data) or the growth in comparative advantage ($\tilde{\gamma}_{ik}$) 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.

We show outcomes for export growth and our structural estimates of growth in comparative advantage. 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.

Table 4: Annualized Export Growth and British Patent Relevance

	Export Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
British Patent Relevance × Japan	3.051***	3.051***	3.051***	3.051***	3.051***	3.051***	3.051***
	(0.793)	(0.793)	(0.794)	(0.794)	(0.794)	(0.794)	(0.794)
British Patent Relevance × Not Japan		-0.714***	-0.888***	-0.852***	-0.491*		
		(0.211)	(0.240)	(0.242)	(0.269)		
British Patent Relevance × English-Speaking			1.021**				
			(0.482)				
British Patent Relevance × British Colony				0.663			
				(0.488)			
British Patent Relevance × Language Similar to English					-0.458		
					(0.426)		
British Patent Relevance × High-Income						-0.322	-0.322
						(0.235)	(0.235)
British Patent Relevance × Medium-Income						-0.870	-0.777
						(0.543)	(0.556)
British Patent Relevance × Low-Income						-1.582***	-1.126*
						(0.514)	(0.655)
British Patent Relevance × Asia							-0.977
							(0.819)
Observations	71	1394	1394	1394	1394	1394	1394
R^2	0.125	0.234	0.235	0.234	0.234	0.236	0.237
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}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry *k*. 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. "British Colony" is a dummy for whether a region was a British colony in the sample period. "Language Similar to English" is a dummy equal to 1 if an English native speaker can attain professional proficiency in the region's plurality language in six months or less. {High, Medium, Low}-Income are dummies which use 1870 GDP per capita from the Maddison Project to flag 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 one if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors in parentheses: $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.001$.

We hypothesize that $\beta_J > 0$; that is, Japanese industries that benefitted more from the codification of knowledge witnessed faster productivity growth in Japan. Column (1) in Tables 4 and 5 show the results from estimating this regression using export growth and productivity growth as outcomes, respectively. Appendix figures A.3 and A.4 plot the corresponding scatterplots. Consistent with our hypothesis, industries with a higher BPR_k experienced faster export and productivity growth during the sample period. The coefficient is both economically meaningful and highly statistically significant. Our estimates imply that a Japanese industry whose British Patent Relevance was in the 75th percentile had export growth that was 15 percentage points per year faster than an industry in the 25th percentile and productivity growth that was 1.4 percentage points per year faster. These large effects help account for the sudden shift of Japanese exports from primary products to manufactures.

A causal interpretation of the parameter of interest, β, requires that BPR_k is uncorrelated with the error term ϵ_{Jk}. The main concern in this context is omitted variable bias, namely that

unobserved factors correlated with BPR_k drive the pattern of productivity growth in Japan. For example, it is conceivable that BPR_k is correlated with distance to the technology frontier. Or, it could be that some other Japan-specific factors, such as fundamental comparative advantage or institutional reforms implemented during the Meiji Restoration, are correlated with BPR_k.

If β_J is picking up general trends in the data, such as distance to the frontier, we would not expect the coefficient for Japan to differ from that estimated for other countries. If, on the other hand, β_J is driven by the codification of technical knowledge in Japan, we would expect the coefficient estimated for Japan to be different from most countries, on average.

Table 5: Annualized Productivity Growth Γ and British Patent Relevance

	Γ						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
British Patent Relevance × Japan	0.277***	0.277***	0.277***	0.277***	0.277***	0.277***	0.277***
	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)
British Patent Relevance × Not Japan		-0.010	-0.021	-0.025	-0.021		
		(0.024)	(0.028)	(0.028)	(0.033)		
British Patent Relevance × English-Speaking			0.063				
			(0.053)				
British Patent Relevance × British Colony				0.071			
				(0.053)			
British Patent Relevance × Language Similar to English					0.023		
					(0.049)		
British Patent Relevance × High-Income						-0.006	-0.006
						(0.026)	(0.026)
British Patent Relevance × Medium-Income						0.048	0.067
						(0.059)	(0.061)
British Patent Relevance × Low-Income						-0.078	0.014
						(0.062)	(0.075)
British Patent Relevance × Asia							-0.193**
							(0.096)
Observations	56	1243	1243	1243	1243	1243	1243
R^2	0.067	0.010	0.010	0.011	0.010	0.012	0.015
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable, Γ_{ik}, 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 how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry *k*. 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. "British Colony" is a dummy for whether a region was a British colony in the 1880-1910 window. "Language Similar to English" is a dummy equal to 1 if an English native speaker can attain professional proficiency in the region's plurality language in six months or less. High, Medium, and Low are dummies that are one 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 one if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors are in parentheses: $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.001$.

Tables 4 and 5 report the estimated coefficients for these pooled specifications. To begin, we estimate the effect of BPR on our growth outcomes using only the Japanese sample (column 1). The estimated coefficient is economically large and statistically highly significant. All other columns

estimate pooled specifications across the full region-industry sample. Column 2 shows that, on average, other countries in our sample did not exhibit the same pattern of export and productivity growth. Export growth in this period was *negatively* and statistically significantly correlated with BPR_k in countries outside of Japan, while there is essentially no correlation with productivity growth.

Of course, English-speaking countries had access to technical knowledge in their vernaculars. Based on our hypothesis, we may thus expect a positive relationship for this group of countries. Indeed, in column (3), we find a positive and statistically significant effect in the export growth specifications. However, the relationship is noisily estimated and not statistically significant at conventional levels in the productivity growth specifications. Note, however, that the mere existence of codified technical knowledge in Britain or the U.S. did not imply that an entrepreneur in, for example, the colony of New South Wales would have had easy access to this knowledge if technical books were not in wide local circulation. Thus, being a majority English-speaking region implied a lower cost of accessing technical knowledge, though not necessarily equivalent to what was available to Japanese entrepreneurs. Consistent with this, in columns (4) and (5), we estimate the effects for British colonies and regions speaking a plurality language similar to English. While the estimated coefficients are generally positive, they are not statistically distinguishable from zero.

In columns (6) - (7), we delve deeper into understanding whether certain region groups besides English-speaking countries displayed a similar productivity growth pattern. 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 Asia outside of Japan, show a negative correlation, though the patterns are never consistently statistically different from zero across export and productivity outcomes. In summary, the pooled specifications suggest that Japan's productivity growth pattern was unusual. Regional trends or structural factors, such as distance to the technology frontier, are unlikely to explain the relationship.

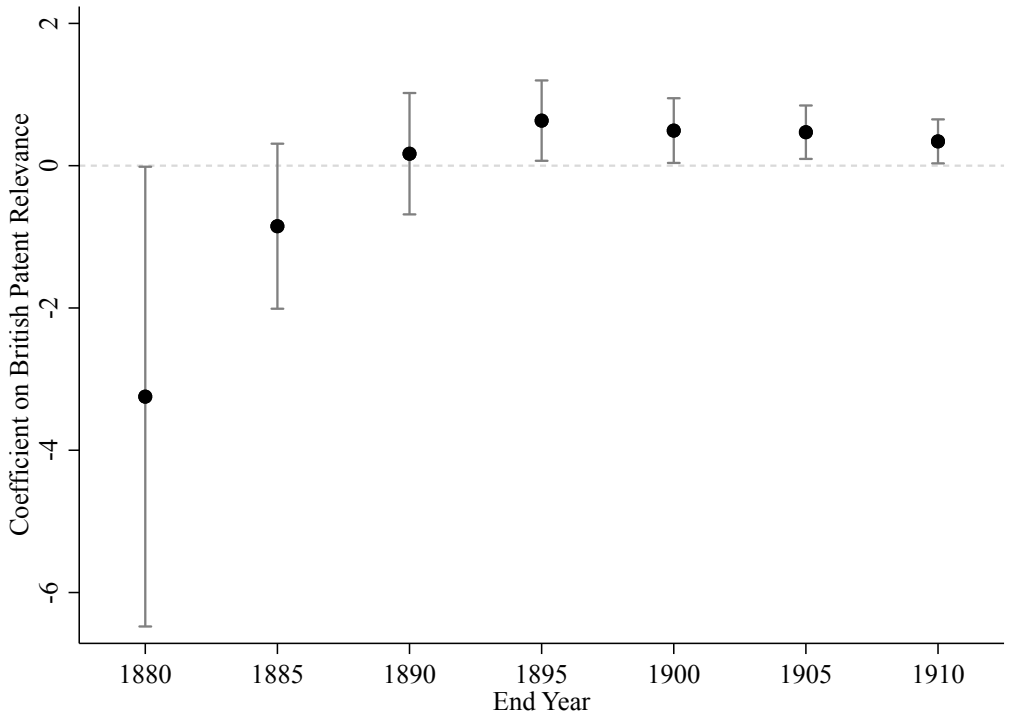
Second, we utilize the sharp timing of Japan's codification of technical knowledge to examine whether there is a "Japan-specific" confounder. In particular, while Japan was undergoing major economic and political changes in the second half of the 19th century, the previous sections have established that 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 the timing is suggestive, we now test this more formally with industry-level variation. In particular, we estimate the placebo treatment effect of BPR_k on Japanese industry growth *before* Japan became technically literate, which, following the discussion in Section 4.1, we define to be 1887.

If technical literacy mattered for Japan, we should expect to see British patent relevance only matter for Japanese exports after the Japanese could read Western technical books. We estimate regressions of annualized Japanese export growth rates by industry between 1875 (a year in which Japan had less than half as many technical books as Germany had in 1870) and an end year that varies in five-year increments starting in 1880 on British Patent Relevance.[18] Figure 13 plots the estimated coefficients from these regressions, along with the 95% confidence intervals.[19]

[18]Data limitations preclude us from estimating these regressions using productivity growth as the outcome variable. In particular, we only have Japanese trade data in 1875, not data from other countries.

[19]Appendix Table A.1 reports the estimated coefficients from the same specifications.

Figure 13: Japanese Export Growth and BPR: Estimated coefficients across different time intervals



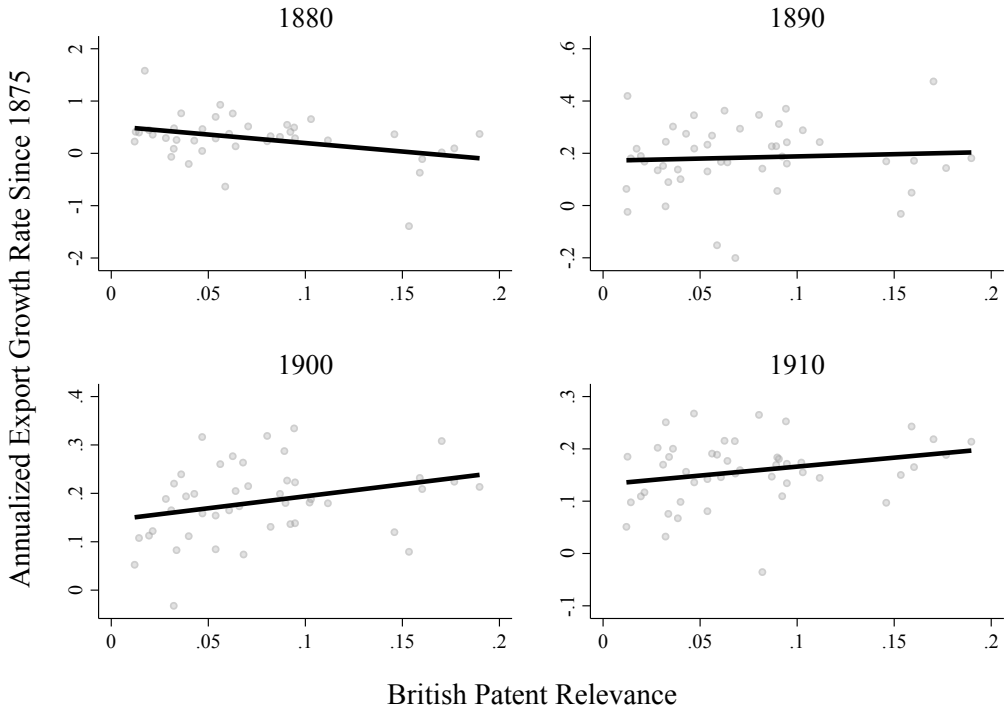
Note: The dependent variable is Japanese export growth for the year reported relative to 1875. We plot the estimated coefficient on BPR_k, as well as the 95% confidence interval.

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 Industrial Revolution. This result is similar to what we saw in Table 4 for countries other than Japan and other Asian economies in particular. In other words, before Japan became technically literate, its export growth patterns looked similar to those of other countries in the global periphery: losing comparative advantage in sectors where the potential to learn from the West was the highest.

However, shortly after Japan became technically literate in 1887, we see that the pattern flips significantly. By 1895, the coefficient of British patent relevance is positive, indicating that Japanese industries that stood the most to learn from British technologies grew faster. This coefficient remains positive throughout the rest of our sample period, suggesting a persistent effect. The timing of this effect is hard to explain with conventional stories about Japan. We do not detect an impact until 37 years after Japan opened to trade, so it is hard to see why trade openness is the sole driving variable. Figure 14 shows the scatterplots for 1880, 1890, 1900, and 1910 to show that outliers are not driving these results. As one can see from the plots, there is a clear negative relationship in 1880 and a clear positive relationship in later years, and these results are not driven by outliers.

In summary, the cross-region evidence and the timing of when technical knowledge became predictive of industry growth in Japan provides a strong case for a causal interpretation of codification on Japanese productivity growth. By making technical knowledge widely available in the vernacular, the Meiji government relaxed a critical bottleneck for industrialization. Any alterna-

Figure 14: 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.

tive explanation of Japanese productivity growth needs to account for both its distinctive pattern in cross-regional comparison and its timing in Japan.

8 Conclusion

This paper shows evidence in support of the argument that the public provision of technical knowledge in the vernacular eliminated an important friction impeding the absorption of Western technology in Meiji Japan. Our results show an empirical pattern unique to Japan: industries that had more to benefit from Western technology experienced faster growth, but only after Japan became technically literate. This suggests that regions hoping to emulate European industrialization in the nineteenth-century context, particularly those linguistically or physically distant from Western Europe, needed to provide complex public goods, such as access to technical knowledge, to emulate Britain successfully. While these public goods were unlikely sufficient to foster modern industrial development, our results suggest they may have been necessary. Other ethnically and linguistically distinct countries that received Japanese institutions and were forced to learn Japanese through annexation or colonization, such as the Ryukyu Kingdom (now Okinawa), Ezo (now Hokkaido), Taiwan, and Korea, also have per capita incomes that are now similar to Japan. We leave it to future researchers to examine whether Japanese colonial institutions, like British ones, had any salutary effect on their growth.

The obvious question is why the Japanese government was unique among regions in the periphery in 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,

which aligned the elite in support of a strategy of aggressive defensive modernization. Importantly, Japan did not discover the policy tools themselves. State support of technology adoption, 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.

References

Allen, R. (2012). Technology and the great divergence: Global economic development since 1820. *Explorations in Economic History 49*(1), 1–16.

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).

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.

Bible (1885). *The Holy Bible: Revised Version*. Oxford and Cambridge: Oxford University Press and Cambridge University Press.

Bo, S., C. Liu, and Y. Zhou (2023). Military investment and the rise of industrial clusters: Evidence from China's self-strengthening movement. *Journal of Development Economics 161*.

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.

Bolt, J. and L. van Zanden (2020). The Maddison Project: Maddison style estimates of the evolution of the world economy. A new 2020 update.

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.

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*.

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.

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

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.

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

Fukao, K., J.-P. Bassino, T. Makino, R. Paprzycki, S. Tokihiko, M. Takashima, and J. Tokui (2015, March). *Regional Inequality and Industrial Structure in Japan: 1874-2008*. Tokyo: Maruzen Publishing.

Gerschenkron, A. (2015). *Economic backwardness in historical perspective (1962)*. Cambridge MA.

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

Gillispie, C. C. (2009). *Science and polity in France: The end of the old regime*. Princeton University Press.

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.

Horn, J. (2006). *The Path Not Taken: French industrialization in the age of revolution*. Cambridge, MA: MIT Press.

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.

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

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

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

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.

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

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.

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

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.

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.

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

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

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).

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*.

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.

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.

Online Appendix

A Additional Tables

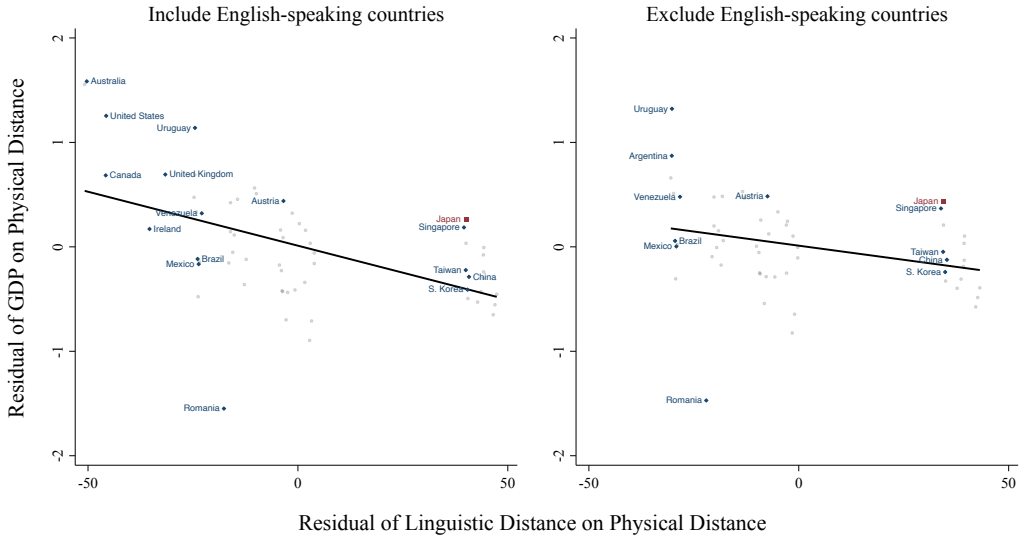
Table A.1: Japanese export growth and British Patent Relevance 1875-1910

	Annualized Export Growth Between 1875 and						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1880	1885	1890	1895	1900	1905	1910
British Patent Relevance	-3.246**	-0.851	0.168	0.633**	0.493**	0.471**	0.342**
	(1.596)	(0.575)	(0.423)	(0.280)	(0.226)	(0.186)	(0.153)
Observations	40	45	46	47	45	46	47
Constant	✓	✓	✓	✓	✓	✓	✓

Note: The dependent variable is 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: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.001$.

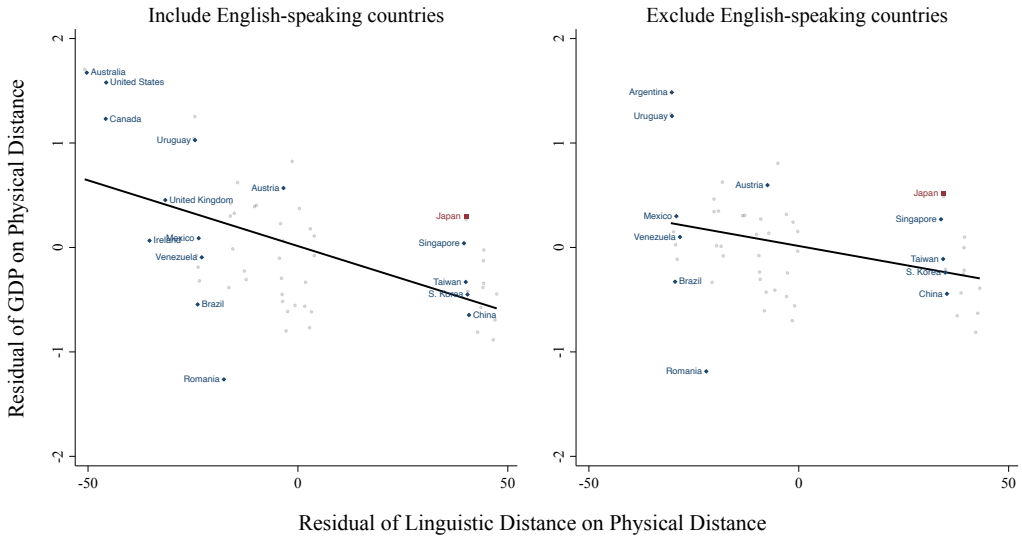
B Additional Figures

Figure A.1: Linguistic Distance Partial Regression Plot for 1870



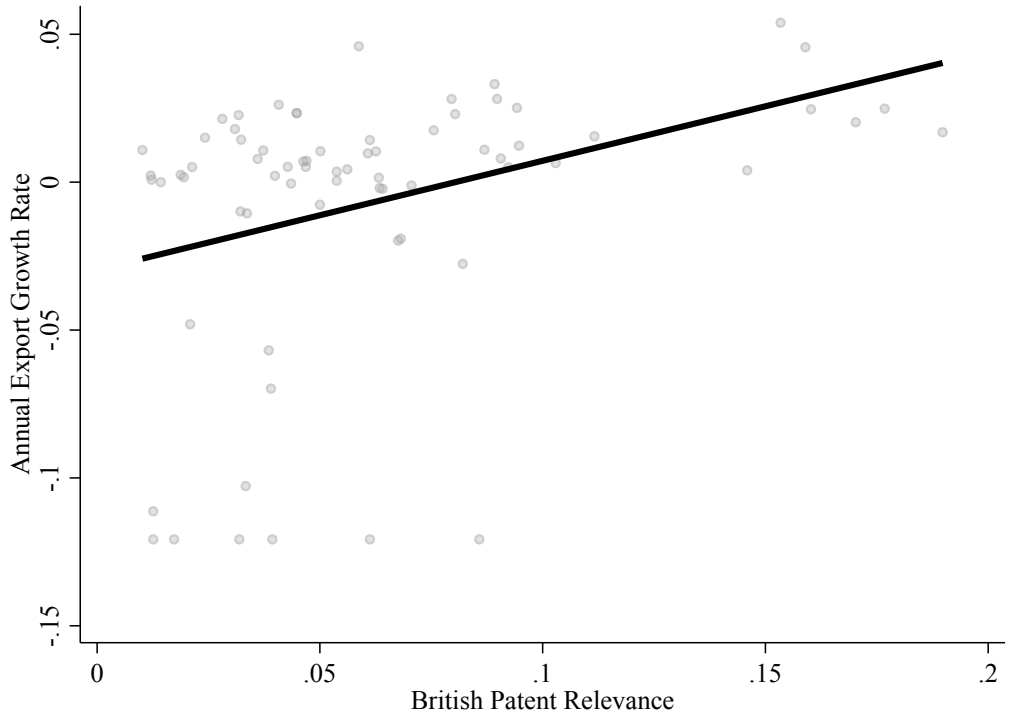
Note: This figure plots the relationship between log GDP per capita in 1870 and linguistic distance controlling for physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.2: Linguistic Distance Partial Regression Plot for 1913



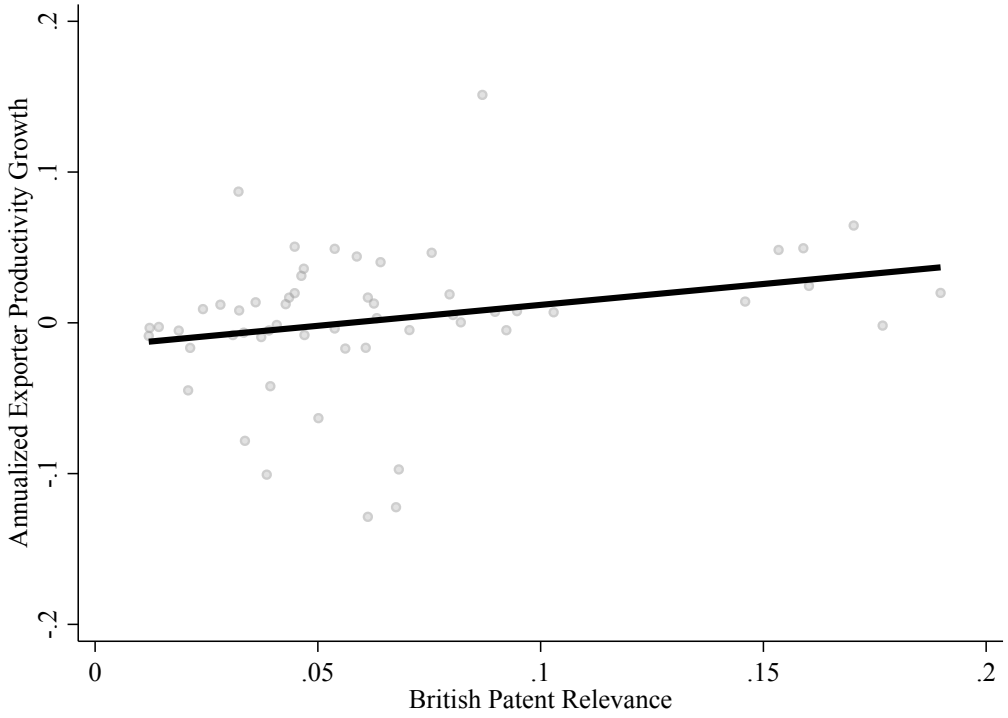
Note: This figure plots the relationship between log GDP per capita in 1913 and linguistic distance controlling for physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.3: Annualized Export Growth and British Patent Relevance for Japan



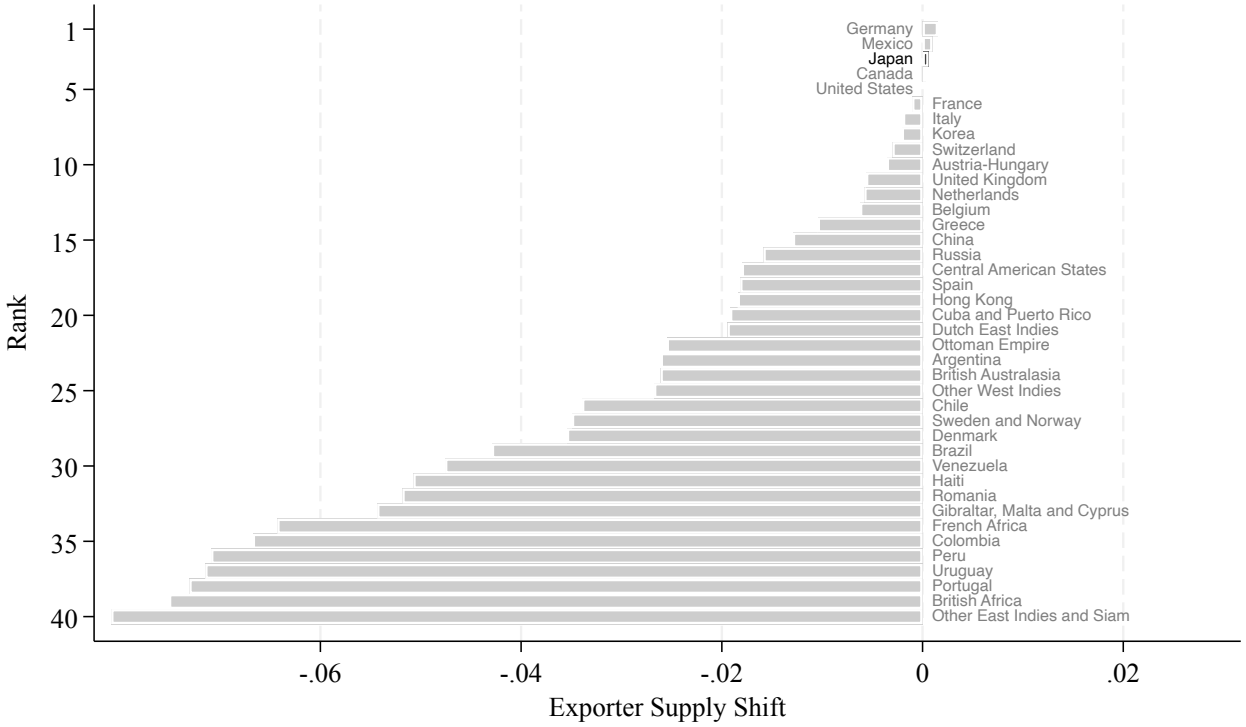
Note: British Patent Relevance is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry *k*. See text for details on variable construction.

Figure A.4: 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 a region *i*'s industry *k* between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry *k*. See text for details on variable construction.

Figure A.5: Relative Annualized Per Capita 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.

C Constructing Annual Growth Rates

One issue in the data is that we form 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).

To be flexible about start and end years when computing annual export growth rates, 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. ijk We annualize the data in two steps. If the reporting region exports the product in 1880 or 1885 (i.e., $\sum_j x_{ijks} > 0$, we set s equal to the first year that satisfies $\sum_j x_{ijks} > 0$. 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

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

Given this annual growth rate, we define the implied level of exports in 1881 as $x_{ijk,1881} \equiv \left(1 + g_{ijk}^C\right) x_{ijk,1880}$.

We face a different problem if the data source region reports positive exports in the start year ($\sum_j x_{ijks} > 0$), but exports to the region j are zero in the start year ($x_{ijks} = 0$). To deal with this problem, we define the average growth rate in exports of new varieties as

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

where $\mathcal{N}_i$ is the set of new importers j such that $x_{ijks} = 0$ and $x_{ijkf} \geq 0$. In this case, we compute the implied level of new export varieties 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 in year $s+1$ equal to the amount that we would have observed if the growth in exports for a country with $x_{ijkf} > 0$ was the same as that of new varieties overall. With these annualized values for exports in hand, we can write the left-hand side of equation 6 as

$$\frac{\sum_j x_{ijkf} - \sum_j x_{ijks}}{\sum_j x_{ijks}} = \frac{\sum_j x_{ijk,s+1} - \sum_j x_{ijks}}{\sum_j x_{ijks}}, \tag{A.2}$$

and the left-hand side of equation 7 as

$$\frac{\sum_i x_{ijkf} - \sum_i x_{ijks}}{\sum_i x_{ijks}} = \frac{\sum_i x_{ijk,s+1} - \sum_i x_{ijks}}{\sum_i x_{ijks}}. \tag{A.3}$$

We then can apply the AW estimation procedure in equations 6 and 7 to estimate the γ_{ik}.

D 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 section.

D.1 Ethnologue (2023): Plurarity Language by Country

Reference Ethnologue, https://www.ethnologue.com/.

From this website we obtain the modern (2023) plurality language spoken in each country and whether at least 10% of the population of the country speaks a Germanic or Italic language.

D.2 Federico et al. (2011): Historical Italian Trade

We obtain this historical Italian trade data for 1880, 1885, 1905, and 1910 from this source. This dataset harmonizes several historical trade records from the Italian customs between 1862 and 1950 by concording the different product lines to SITC codes. This source reports the bilateral trade between Italy and its major ten commercial partners, as well as the import and export series of the most important categories of products.

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

Reference "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 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.

D.4 High-Income Countries

We reference 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 than the US GDP per-capita, based on data from the World Bank (2024), specifically, variable "GDP per capita, PPP (current international $)". To know what countries comply this criterion, we divide each country i's GDP per-capita in 2022 by the GDP per-capita of the US. The code that generates the countries that comply the criterion can be found in the replication file.

D.5 Fouquin and Hugot (2016): Historical Exchange Rates

Fouquin and Jules created the Historical Bilateral Trade and Gravity Dataset (TRADHIST) from which we obtain the yearly exchange rates for the 1870-1915 window from the Great British Pound to the Belgian Franc, the German Mark, the Japanese Yen and the U.S. Dollar. Specifically, they provide us the value of one unit of the local currency in Pounds.

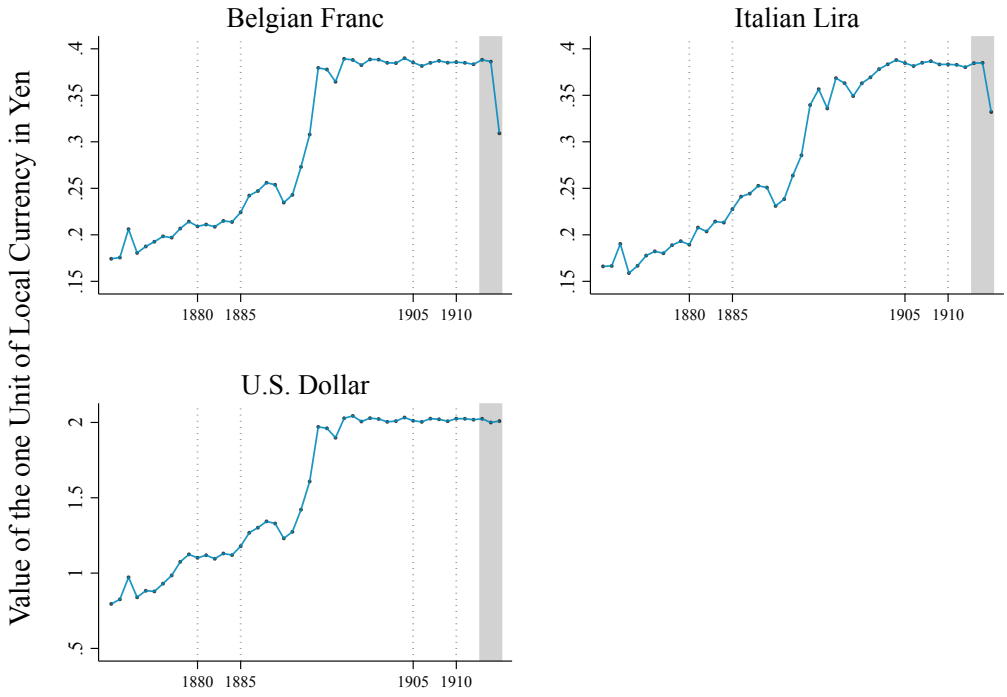
Yen Exchange Rates

Calculate the exchange rate from Yen to Belgian Francs, Italian Lira and Dollars:

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

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.

Figure A.6: Exchange Rates



Note: Exchange rate calculated as described in the equation using Great British Pound Exchange Rates from Fouquin and Hugot (2016) . World War I shaded in gray.

D.6 Mayer and Zignago (2011): GeoDist Database (2011) - Distance to U.K.

The *Centre d'Etudes Prospectives et d'Informations Internationales* (CEEPI) estimated 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 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.

Distance from United Kingdom

We restrict the dataset to the United Kingdom cross-section. The **`dist`** variable corresponds to the great-circle distance from the UK to other countries.

D.7 Huberman et al. (2017) - Historical Belgian Trade

We obtain the Belgian bilateral trade data for 1880, 1885, 1905 and 1910 from this source. Huberman et al. (2017) use the *Tableau générale du commerce extérieur* published by the Belgian government as their primary source. The authors record trade *in manufacturing* at five-year intervals between 1870 and 1910. In 1900, 50% of Belgian exports and 20% of its imports were in manufacturing. Belgian non-manufacturing exports to other reporting countries are computed using the bilateral import data of the reporting countries.

Import and export prices The *Tableau* recorded official prices (not declared prices) of products. This shortcoming is minimized because prices were adjusted annually. Although reliability of prices varies across commodity. The sources of official prices are unclear. For details see *Unit Values* section on page 80 of Huberman et al. (2017).

D.8 Long Term Economic Statistics of Japan - Historical Nominal Production of Japan

The Long Term Economic Statistics (LTES) of Japan database is a thirteen-volume collection of estimated and processed historical statistics of early modern Japan on economic activities in various fields based on the System of National Accounts, edited by Ohkawa et al. (1965).

Real and Nominal Production of Japan 1874-1910

We use Volumes 7 (*Government Expenditure*), 9 (*Agriculture*) and 10 (*Mining and Manufacturing*) to obtain nominal and real production value data for several sectors between 1874 and 1910. The data was hand-entered from select pages of volumes 9 and 10 onto an Excel spreadsheet that details which pages we used from these volumes.

D.9 Maddison Project Database - Historical GDP and Population by Country

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 (Bolt and van Zanden, 2020), which covers 169 countries up until 2018.

High-, Medium- and Low-Income

We classify countries 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 is the arithmetic mean of the GDP per capita of its corresponding modern state. We do not use any weights when taking this average.

3. We rank historic countries by GDP 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.

Annualized Population Growth

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

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

2. The population of historic country 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 have population 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).

D.10 Meissner and Tang (2018) - Historical Japanese Trade

We obtained bilateral Japanese export data at five-year intervals between 1880 and 1910 from Meissner and Tang (2018). This dataset was constructed from the trade statistics volumes published by the Japanese Ministry of Finance.

E Bilateral Trade Dataset

Our master bilateral trade dataset is made up of four main sections:

- American exports and imports in 1880, 1885, 1905 and 1910
- Belgian manufacturing exports and imports in 1880, 1885, 1905 and 1910
- Italian exports to and imports from top trading partners in 1880, 1885, 1905 and 1910
- Japanese exports and imports in 1875, 1880, 1885, 1905 and 1910

As noted in the previous section, we obtain the Belgian (Huberman et al., 2017) and Italy (Federico et al., 2011) trade data as well as Japanese 1880-1910 exports (Meissner and Tang (2018)) from external sources. We built the American bilateral trade, Japanese 1875-1910 import, and Japanese 1875 export datasets.

The Japanese trade data was 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 only use the tables from the "Quantity and Value of Commodities Imported/Exported from Various Countries" sections.

The American data was extracted from the yearly volumes of *Foreign Commerce and Navigation, Immigration and Tonnage of the United States* published by the Treasury Department's Bureau of Statistics (1900).

Both data sources are very rich, as both Japan and the U.S. kept very detailed records of their trade with foreign countries between 1880 and 1910. Each entry tells us the name of the product, its quantity, units, transaction value, year as well as the exporting and importing countries. Before we could use this dataset, we had to first digitize the trade book, harmonize its products, country names and currencies and deal with double reporting issues. The protocols we adopted are described in detail in the subsections below.

The US reports trade values in dollars (not in hundreds or thousands of Dollars). Japan reports trade values in yen and sen (100 Sen equals 1 Yen). Some of the Japanese books have two separately titled columns for yen and sen; in other years, only yen are reported, and in some years, values are reported (for example) as "Horse ... 2,350|000".

E.1 Harmonization of Product Lines

Historical trade statistics are not standardized. Compared with 1875, by 1910, the reporting countries logged a wider variety of transactions as they traded more, and their record records became much more detailed. For example, in 1875, Japan reported "iron." by 1910, there were dozens of iron categories (e.g., pig, bar, rod, etc...).

Even within one year, there are variations in the level of detail of trade records. Japan's trade with the United Kingdom is far greater than its trade with Portugal, so when Japan reports trade in iron with the U.K., it will report all the iron categories in detail as the volume of trade in each category is quite large. The total trade in iron between Portugal and Japan might not even reach the value of the trade of one category of iron between Japan and the U.K., so the trade records aggregate Japan's iron trade with Portugal.

As a result, it is very difficult to draw a direct comparison of trade of a particular sector across years and trade partners with a reporter, and much less when we want to compare the trade flows of two different reporters. We, therefore, standardized product names. Appealing to the literature, we use SITC-3 to standardize product lines.

We use the Meissner and Tang (2018) product-SITC mapping wherever possible for Japan and the U.S. to ensure consistency.

E.2 Harmonization of Countries

Country names are not standardized across reporters (Belgium, Italy, Japan, and the U.S.) and years because:

- Between 1875 and 1910, country and imperial borders changed substantially. Naturally, these changes are reflected in the way that countries report their trade statistics across time.

- Aggregation of regions is different across reporters. For example, The U.S. details its trade with Caribbean states than Japan, which trades less with this region.

- As countries trade more, they will be more specific in how they record countries in their trade statistics as the trade volume with each country is likely to increase. For example, in 1875, Japan grouped Singapore and British India into one small section; by 1910, each region had its own separate section.

In order to make comparisons across years and countries, we had to standardize country names first. To do we adopt the following method:

1. Make a list of all the country names that appear in all of the trade books from the four reporters, making note of the year and book where each name shows up.

2. Group names that clearly refer to the same country: Vietnam and French Indo-China both refer to the same political entity at the time, which was French Indochina.

3. We keep the group as is if each group is mentioned in the books of at least three reporters, and it appears in the 1880/5 *and* 1905/1910 books for each reporter.

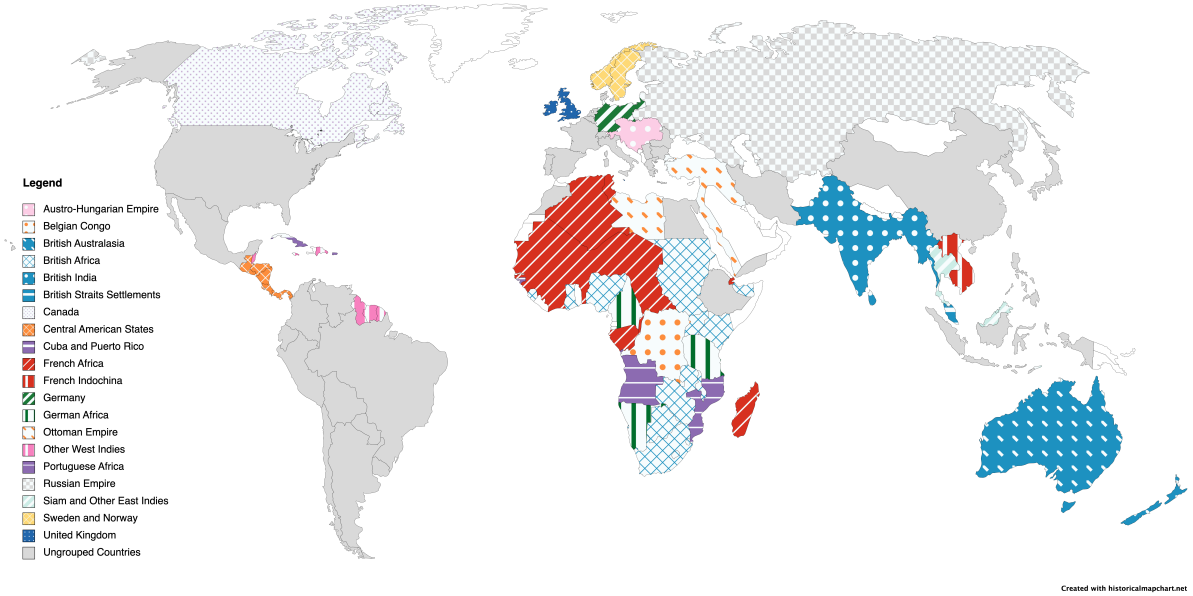
4. If the country group does not meet the previous requirement, then we try to build a regional group that does. For example, Honduras, Nicaragua, and Costa Rica do not have three reporters in the all the required years. If we group all Central American States together, this larger regional group meets our requirements.

5. If a country cannot be grouped and does not meet the reporter-year requirement, then we drop it.

6. If a region is too disaggregated, we drop it. For example, Singapore and Hong Kong are their own separate categories, each with substantial volumes of trade in our dataset. If one country, in one year, reports “Hong Kong & Singapore," we drop this observation. Since we cannot decompose the trade from aggregated observations and we do not wish to group Hong Kong and Singapore, we have no choice but to drop it.

The map illustrates how we grouped countries. Countries in grey were left as they were. We use the map of the world on the eve of World Word I (1914) as a baseline for our country groups.

Figure A.7: 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, countries in grey are left as they are. 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 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.

E.3 Harmonization of Currencies

The Belgian trade data is denominated in Belgian francs; the German, in marks, the Japanese, in yen; and the American, in Dollars. In order to compare trade values across reporters, we convert all currencies to yen using the exchange rate for the year in question. The historic exchange rate data is from Fouquin and Hugot (2016).

E.3.1 Double Reporting

Once the product lines, country groups, and currencies are harmonized, we can append the trade data from Belgium, Germany, Japan, and the U.S. and collapse the transaction value by exporting country, importing country, reporting country, SITC-3, and year.

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 just use their export data for their exports to reporting and non-reporting regions. Because Belgium does 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.

F British Patent Relevance in the Late 1800s

F.1 Intuition

We want to measure how relevant the ideas of British patents were to an industry in the late 19th century. To make this measure, we *assume* that the similarity between the word frequencies in books describing historical production techniques in an industry between the word frequencies in patent data tells us the relevance of patents to that industry. In practice, we start with unigrams (i.e., single words such as "steam") and bigrams (i.e., two-word combinations such as "steam engine"). We convert these into "terms" by stemming them and converting them into the terms "steam" and "steam engin". We also make use of two types of corpora. The first is the set of books describing production techniques in industry k, and the second is the words used in the book containing British patent synopses. Thus, we have a corpus for each industry and a separate corpus for the patents. To measure the relevance of the British patent corpus to industry k's corpus, we weight each term's frequency in a book by the total number of books divided by the number of books containing the term, i.e., we compute the Term Frequency-Inverse Document Frequency (TF-IDF). For each industry, we build a TF-IDF vector that characterizes its vocabulary (where each element is the TF-IDF of a term); we also build a TF-IDF vector for patents. Finally, we compute British patent relevance of industry k as the cosine similarity between the vector of TF-IDFs for industry k and that for the set of British patent synopses. We explain each of these steps in detail below.

F.2 Building the Terms

To build a term, we start with n-grams, we implement the following steps:

1. Split the raw text into sentences
2. Convert the words in the sentence to lower case, stem the words, replace UK spelling with US spelling
3. Turn each processed sentence in a sentence word list (where the position of a word on a list is the position it has in the sentence)
4. For each sentence word list, split it into n-grams
5. Count the number of times an n-gram appears in the sentence
6. Drop n-grams that include at least one stop word, (i.e., "a", "the", etc.)
7. Output a dataset with all the n-grams in the document and their count in the corpus

Example

1. **Start** "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

F.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 got the unigram data for the Bible from Hathi already processed as described above. We define Biblical words as the 1000 words with the highest frequency in the Bible. However, if one of these words is used in the definition of an SITC keyword, we do not count it as a Biblical word. For example, the word "brea" is a top 1000 word in the Bible, but it also happens to be a keyword in the SITC for cereal products, so we can't remove this word from the corpus since it is a critical word in characterizing the vocabulary of the cereal products industry.

F.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

$$\text{TF}(\tau, c) \equiv \frac{F_{\tau,c}}{\sum_{\tau' \in c} F_{\tau',c}} \tag{A.4}$$

where $F_{\tau,c}$ is the raw count of τ in 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 books) as

$$\text{IDF}(\tau, C) = \log\left(\frac{N}{N_\tau + 1}\right) \tag{A.5}$$

where N is the total number of books in C; N_τ is number of books in the corpus where the term τ appears.

The TF-IDF is then

$$\text{TF-IDF}(\tau, c, C) = \text{TF}(\tau, c) \cdot \text{IDF}(\tau, c) \tag{A.6}$$

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 British Patent Relevance of industry k as the similarity between the industry k and patent vocabulary characterization vectors. We use cosine similarity to measure the similarity between two vectors. If an industry uses the same words at the same frequency as the patent book, then the vectors are the same, and we conclude that British patents are very relevant in the industry. If there is no overlap in words, then the similarity score is low, and we conclude that British Patents are not relevant.

F.4.1 Cosine Similarities

Let $\vec{I_k}$ be the vector of TF-IDF scores for industry k and $\vec{P}$ be the vector of TF-IDF scores for British patents. The cosine similarity between the industry k corpus and the British patent corpus is given by

$$\text{cosine similarity}(\vec{I_k}, \vec{P}) = \frac{\vec{I_k} \cdot \vec{P}}{|\vec{I_k}||\vec{P}|}$$

F.5 Data Sources

Our text data (unless otherwise specified) was accessed through HathiTrust.

British Patents All patent text 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).

Industry For each industry (as defined by SITC-3) we constructed a list of books and sections of eighteenth century books that we believed to be relevant in describing the production process of the goods in the industry. The full list of books is given in "full_book_list.xlsx".

Bible (1885) English Revised Version of the Bible.[1]

UK-US Spellings We got *uk-us_spellings.csv* online from GitHub.[2]

G New Japanese Words in the Meiji Period

We utilize the etymology of Japanese words based on the revised edition of Nihon Kokugo Daijiten, 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 [3]. The number of new words by year can be seen on Figure 8.

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

H.1 Overview

We report the source libraries for our data on technical books in Table A.2. We scraped data for 33 languages, which include all of the 20 most spoken native languages on earth[4]. We define the set

[1]Wikepedia article Revised Version of the Bible: https://en.wikipedia.org/wiki/Revised_Version

[2]https://gist.github.com/heiswayi/12ca9081ae1f18f6438b

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

[4]We assume that if someone speaks Yue or Wu Chinese, they will be able to read Mandarin Chinese given that these languages all use the same set of characters.

of books comprising technical knowledge as those with a subject that can be 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.

For many major European and Asian languages, we were able to scrape the national libraries of countries where the language is the native tongue of a substantial fraction of the population. For many other languages (such as Arabic and Russian), we were not able to find a scrapable national library. Instead, we scraped WorldCat, which is an online catalog of thousands of libraries around the world covering dozens of languages. If we can scrape a language from a national library and WorldCat, we scrape this language from both sources and pick the source that yields the most books.

H.2 Issues with WorldCat

WorldCat is very reliable for European languages, but it can be very unreliable for non-Western languages. For example, while WorldCat has most of the catalog of the National Diet Library, the National Diet Library did not upload all the bibliographic information for each book. Many technical books had their subject field black so they would not show up in the subject search. In cases like this, we either drop the language or use the National Library as the source if possible.

WorldCat also pools data from several different libraries and sources. Often the same book is uploaded by different libraries with unstandardized bibliographic information, resulting many duplicate entries. We, therefore, prefer to use a national library when possible. Fortunately, we were able to scrape national libraries for the languages most likely to contain large numbers of codified books.

H.3 Search Filters

1. **Format:** We only search for books. No images, periodicals, articles, or newspapers.
2. **Language:** We always specify the language of the text. For example, when searching the National Diet Library, we only look for Japanese books.
3. **Publication Year:** 1500-1930
4. **Subject:** We always search by subject.
   - We search by subject code, if possible, because it is more precise than searching by subject keyword.
   - 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.

H.4 Handling Duplicated Books

When we make the book plots for each language, we drop duplicated books based on book ID. Although this rule does not guarantee that there won't be any duplicates, we think it best to rely on the library's own system of defining different books rather second guessing them.

Table A.2: 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	1500-1930	Nippon Decimal Classification	Technology (600) Industry (700)
Korean National Library	Link	Korean	0022-1980	Dewey Decimal Classes	Technology and Engineering (600)
Library of Congress	Link	English	1500-1930	Keyword Search	Made our own list
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	1500-1980	Chinese Library Classification System	Agriculture (S) Industry (T) Transportation (U)
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	Made our own list