Codification and the Spread of the Industrial Revolution*

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

Shogo Sakabe
LMU and Hitotsubashi‡

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

December 15, 2023

Abstract

PRELIMINARY AND INCOMPLETE, PLEASE DO NOT CITE WITHOUT PERMISSION

JEL CLASSIFICATION:
KEYWORDS:

*We thank Benjamin Eyal, Isaac Loomis, Zachary Marcone, Ojaswee Rajbhandari, Roshan Setlur, Alex Zhang, and especially Michael Duarte, Veronic Perez, Angela Wu, and Dongcheng Yang for excellent research assistance. We also want to thank Treb Allen, Andrew Bernard, Davin Chor, John Fernald, Shizuka Inoue, Robert Staiger, and Dan Trefler for their great 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.

1 Introduction

It is widely believed that Meiji Japan was the only non-Western country to successfully industrialize during the first wave of globalization. In the space of a few decades between 1880-1910, Japan transformed from a relatively poor, predominantly agricultural economy specialized in the exports of primary materials, such as silk and tea, to an economy specialized in the export of light manufactures, such as cotton yarn and textiles, which were in turn produced in large, factory-based mechanized mills. How and why did Meiji Japan succeed in starting structural transformation, while so many other, similar countries failed in this period?

We answer this question by bringing several novel datasets to bear on the question that lets us test one of the main theories proposed by Mokyr (2002): namely, that an essential component of the Industrial Revolution (IR) was the development of "technical literacy," i.e., the *codification* of engineering, commercial, and industrial practices. Following Kuznets, Mokyr (2002) calls this knowledge "useful knowledge," a term we will adopt as well. While there is extensive evidence about the codification of knowledge in English and French, there is little understanding of the extent of the cross-sectional or time-series patterns of codification more broadly. But how important was the codification of useful knowledge for industrialization in the nineteenth century? An ideal experiment would require both time-series and cross-sectional variation in useful knowledge so that a researcher could examine whether productivity growth rose in industries where more useful knowledge existed only after entrepreneurs became technically literate. We argue that the late nineteenth century provides precisely this empirical setting.

Our analysis proceeds by first establishing four new stylized facts about the spread of the Industrial Revolution globally and then testing the hypothesis that technical literacy explains the sudden change in comparative advantage observed in Japan but not in other countries. The first stylized fact is that those who were physically close to England or spoke languages that were linguistically close to English had significantly higher per capita incomes in 1870. This stylized fact establishes the plausibility of the theory that interactions with England, either through distance-related activities like trade or through the ability to communicate in English, seem to have mattered for development in the nineteenth century. [See Table 1 and Figures 4 and 5].

The second stylized fact is that the codification of useful knowledge was almost entirely done in three languages: English, French, and German. We document this fact by scraping major libraries containing books for dozens of languages and find that people who could only read other languages were technically illiterate in the sense that they had almost no books containing useful knowledge. For example, Japan's National Diet Libary contains twelve million books, but only fourteen "technical books (i.e., books with useful knowledge) were published before 1870. Libraries for other major non-European languages, such as Chinese, Arabic, and Turkish, have even larger collections of books but contain similarly small numbers of technical books. By contrast, speakers of English and French had access to over 10,000 technical books. [see Figure 11]

The third stylized fact is that Japan is a major outlier in the rate at which technical books are published. While the growth in the number of technical books was slow (typically five percent per year) in most languages, Japan had a growth rate of technical books

of 40 percent per year from 1871 to 1883. This explosive growth rate in technical translations was the result of a major government effort to codify Western knowledge by creating technical English-Japanese dictionaries, paying for translators, and investing broadly in education. As a result, there were were over 10,000 technical books in Japanese in 1910, which meant that Japanese were as technically literate as Germans.

The last stylized fact is that the rise of Japanese manufacturing happened immediately after Japan became technically literate. 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. However, ten years after Japan had published a thousand technical books and had a high degree of codification of useful knowledge, the share of manufactured exports grew to 60 percent and stayed at this level for the next 40 years [Figure 13]. We show that this increase in manufacturing specialization was unique in the 19th century [see Figure 1]. Thus, Meiji development didn't create a gradual increase in growth rates as institutions improved, but rather a very rapid increase in manufacturing that happened only after Japan succeeded in codifying a large amount of Western knowledge *in Japanese*. [see Figures 6-10]

In the second part of the paper, we turn to understanding the determinants of Japan's industrial development. 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, it requires no data from individual reporting countries in the sample. This allows us to examine industry-level productivity trends globally for the first time. We find that late nineteenth-century Japan had high country-level productivity growth rates in international comparison [Figure 2], which were concentrated in manufacturing sectors [Figure 3]. We develop a method to study whether industry productivity growth was concentrated in sectors that had more to benefit from the absorption of Western technologies. We measure the relevance of Western useful knowledge for each sector by using a text-based approach that closely follows how codified 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 compare the similarity of text from these historical technical manuals to the text of British patents using cosine similarity, the standard metric for comparing the similarity of two documents in natural language processing (NLP). We call this *British Patent Relevance*. Intuitively, industries such as textiles that 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, which also featured prominently in patent texts. As such, we say that the contents of patent texts are relevant for manufacturing textiles. On the other hand, for an industry such as musical instruments, which benefited relatively little from Industrial Revolution technologies, the cosine similarity to patent descriptions is smaller, which suggests that the makers of musical instruments benefitted little from the Industrial Revolution [Figure 14].

We argue that the Japanese government overcame a difficult public goods problem that facilitated technical literacy by around 1880. The problem that publishers, translators, and entrepreneurs faced was that jargon for new technologies simply did not exist. The Japanese government solved the problem by creating a large dictionary that contained jargon for many technical words. Indeed, we find that new word coinage in the Japanese language took off after the dictionary was published, as did the translations of technical books.

How much did these public good investments in knowledge acquisition matter for the absorption of Western technology? To probe these questions, we examine the pattern of Japanese productivity growth before and after these public good investments. Consistent with our mechanism, we find that Japanese productivity growth is associated with the British Patent Relevance measure but only after these investments were made [see Figure 12]. In particular, the Meiji government made substantial investments in acquiring knowledge from the West by sending Japanese on foreign study missions, through the hiring of foreign workers (Jones, 1980), and through the large-scale translation of technical knowledge from the West (Montgomery, 2000). This latter effort may have been particularly important, as by translating the technical manuals into Japanese, they could be widely disseminated and accessed by potential industrial entrepreneurs. This created two sub-periods in Meiji Japan: a period before 1880, in which Japan was largely technically illiterate, and a period afterward in which Japan had the capacity to read, write, and translate useful Western knowledge.

We find a strong, positive association between industry productivity growth in Japan and British Patent Relevance after Japan had codified Western useful knowledge but not before. Indeed, until 1885, Japan looked remarkably similar to the rest of the developing world in which there was no relationship between the relevance of British patents for an industry and the industry's productivity growth rate. Indeed for Japan before it became technically literate, and for other low and middle-income countries that remained technically illiterate, we find evidence of divergence, suggestive of the international division of labor, which was emerging during this period. These results suggest that Meiji Japan was uniquely effective at absorbing Western technologies. The obvious questions are why Japan succeeded and what prevented others from doing the same.

2 Related Literature

To be added

3 Historical context [Incomplete]

Late 19th century Japan presents an interesting study of late industrialization. There is wide-ranging consensus amongst economic historians that Meiji Japan was the only non-Western economy which managed to industrialize in this time period. After centuries of isolation, Japan was forcibly opened up to trade by Western powers in the middle of the 19th century. At this juncture, Japan was a poor, feudal and agricultural society. Due to its limited contact with the West in preceding centuries, mid-19th century Japan had little knowledge of Western science and technology. In fact, it was virtually untouched by Industrial Revolution technologies. As an illustration, there were no words in Japanese

to describe canonical Industrial Revolution products such as the railroad, steam engine or telegraph. Consistent with these facts, following the rapid move from near-autarky to quasi-free trade led Japan to become specialized in the export of primary commodities such as coal and silk, in exchange for manufactured products.

The arrival of Westerners, and the perceived economic weakness of Japan instigated a national crisis which led to the fall of the Tokugawa shogunate which had ruled Japan for the past two and half centuries. The Meiji Restoration of 1868 restored imperial rule under Emperor Meiji and the new government undertook a rapid modernization effort. This effort entailed a broad overhaul of society and the economy. The feudal system, along with the hereditary status of the samurai class was abolished and the government undertook a large centralization effort. Similar to state-led development efforts elsewhere, they built railroads and telegraph lines, reformed the banking and education system, and set up model factories. The defining and perhaps distinguishing feature of the Meiji modernization effort however, was its single-minded focus on acquiring Western science and technology. The government sent students on study-abroad missions, hired foreign skilled workers and translated books on Western science and technology. While there was precedent for state-led technology absorption of a similar nature elsewhere, the scale of Japanese efforts was unmatched at the time.

Starting in about the 1880s, the historical record points to a marked shift in the Japanese economy. Pockets of modern, factory-based manufacturing begin to emerge, predominantly in textiles. These textile mills used British machinery, inanimate power-sources and a modern industrial labor force. Indeed, consistent with this, Figure 13 shows a remarkably swift and dramatic shift in export composition away from agricultural products and commodities towards manufactured goods.

There is widespread consensus in the literature surrounding these basic facts. What seems more contested are the causes of Japan's swift and unique industrialization among non-Western countries. Moreover, quantitively establishing the growth miracle and studying its sources has been a challenge. In fact, based on state of the art measures of historical per capita income levels from the Maddison project, Japanese income growth is by no means remarkable in this period.

4 Estimating Productivity and Productivity Growth

4.1 Estimating Productivity

Our starting point for modeling comparative advantage in a multisector Ricardian setup is the framework or Costinot, Donaldson, and Komunjer (2013) [Henceforth CDK]. CDK develop a multisector Eaton and Kortum (2002) model in which the value of exports from i to j in industry k (x_{kij}) is given by

$$x_{kij} = \frac{(w_i d_{kij}/z_{ki})^{-\theta}}{\sum_{i'}(w_{i'} d_{ki'j}/z_{ki'})^{-\theta}} \cdot \alpha_{kj} w_j L_j, \tag{1}$$

where w_i is the wage in in country i; $\theta > 0$ is the Fréchet shape parameter; $z_i^k > 0$ is industry i's scale parameter, which CDK call "fundamental productivity"; d_{kij} is an iceberg trade cost; L_j is the amount of labor in country j; and α_{kj} is the share of expenditures on industry k in country j.

In order to simplify terms, we follow CDK to write

$$\begin{aligned}
\ln\left(\frac{x_{kij}x_{ki'j}}{x_{k'ij}x_{ki'j}}\right) &= \ln(w_i d_{kij}/z_{ki})^{-\theta} + \ln(w_i d_{k'i'j}/z_{k'i'})^{-\theta} - \ln(w_i d_{k'ij}/z_{k'i})^{-\theta} - \ln(w_i d_{ki'j}/z_{ki'})^{-\theta} \\
&+ \ln\left(\alpha_{kj} w_j L_j\right) + \ln\left(\alpha_{k'j} w_j L_j\right) - \ln\left(\alpha_{k'j} w_j L_j\right) - \ln\left(\alpha_{kj} w_j L_j\right) \\
&- \ln\left(\sum_{i'}(w_{i'} d_{ki'j}/z_{ki'})^{-\theta}\right) - \ln\left(\sum_{i'}(w_{i'} d_{k'i'j}/z_{k'i'})^{-\theta}\right) \\
&+ \ln\left(\sum_{i'}(w_{i'} d_{k'i'j}/z_{k'i'})^{-\theta}\right) + \ln\left(\sum_{i'}(w_{i'} d_{ki'j}/z_{ki'})^{-\theta}\right) \\
&= \theta \ln\left(\frac{z_{ki} z_{k'i'}}{z_{k'i} z_{ki'}}\right) - \theta \ln\left(\frac{d_{kij} d_{k'i'j}}{d_{k'ij} d_{ki'j}}\right),
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
\ln x_{kij} = &\left[\theta \ln \frac{z_{k'i'}}{z_{k'i}} - \ln \frac{1}{x_{k'ij}} - \theta \ln \frac{1}{d_{kij}}\right] \\
&+ \left[\theta \ln \frac{1}{z_{ki'}} - \ln \frac{x_{k'i'j}}{x_{ki'j}} - \theta \ln\left(\frac{d_{k'i'j}}{d_{ki'j}}\right)\right] + \theta \ln z_{ki} - \theta \ln d_{kij}
\end{aligned} \tag{2}$$

or

$$\ln x_{kij} = \delta_{ij} + \delta_{kj} + \theta \ln z_{ki} + \epsilon_{kij}, \tag{3}$$

where δ_{ij} is a importer-exporter fixed effect; δ_{kj} is an importer-industry fixed effect; and $\epsilon_{kij} \equiv -\theta \ln d_{kij}$. If we first difference this equation, we obtain

$$= \Delta\gamma_{ij} + \Delta\gamma_{kj} + \theta\left(\Delta \ln z_{ki}\right) + \Delta\epsilon_{kij} \tag{4}$$

Since we do not observe $\Delta \ln z_{ki}$, we introduce a method to back it out from the trade data. Consider running a regression specification of the form

$$\Delta \ln x_{kij} = \gamma'_{ij} + \gamma'_{kj} + \gamma'_{ki} + \epsilon'_{kij}, \tag{5}$$

where γ'_{ij}, γ'_{kj}, and γ'_{ki} are fixed effects. Estimating this equation allows us to obtain identification of γ'_{ij}, γ'_{kj}, and γ'_{ki} up to a choice of normalization. In particular, any set of parameter estimates $(\gamma'_{ij}, \gamma'_{kj}, \gamma'_{ki})$ and normalization choices $(\kappa_i, \kappa_j, \kappa_k)$ that satisfy

$$\gamma'_{ij} = \Delta\gamma_{ij} - \kappa_i - \kappa_j,\ \gamma'_{kj} = \Delta\gamma_{kj} + \kappa_j - \kappa_k,\ \text{and}\ \gamma'_{ki} = \theta\left(\Delta \ln z_{ki}\right) + \kappa_i + \kappa_k \tag{6}$$

will also satisfy equation 5.

One of the problems with directly estimating equation (5) is that the left-hand side cannot be estimated whenever exports are zero. Amiti and Weinstein (2018) [AW] propose an alternative estimation approach that have a number of properties that will prove useful for our application. Noting that we can approximate log change in exports between a start (s) and final (f) period with percentage changes, we can rewrite equation (5) as

$$\frac{x_{kijf} - x_{kijs}}{x_{kijs}} = \gamma'_{kj} + \gamma'_{ki} + \left(\gamma'_{ij} + \epsilon'_{kij}\right) = \gamma^{AW}_{kj} + \gamma^{AW}_{ki} + \epsilon^{AW}_{kij}, \tag{7}$$

where γ_{kj}^{AW} and γ_{ki}^{AW} are the estimates of γ'_{kj} and γ'_{ki}, respectively; and ϵ_{kij}^{AW} is the Amiti and Weinstein estimate of the term in parentheses. Importantly, AW show that moving γ'_{ij} to the error term only and estimating γ_{kj}^{AW} and γ_{ki}^{AW} only affects the choice of normalization constants $(\kappa_i, \kappa_j, \kappa_k)$.[1] In particular, in the AW procedure all variation at the exporter-industry level will be contained in γ_{ki}^{AW}, and all variation at the importer-industry level will be contained in γ_{kj}^{AW}. As a result, γ'_{ij} only captures country-pair variation that is orthogonal to importer-industry and exporter-industry characteristics (e.g., bilateral trade costs). A second useful property of the AW procedure is that the γ_{kj}^{AW} and γ_{ki}^{AW} exactly aggregate to the growth rate of total exports by country-industry for all sectors in which the industry's *aggregate* growth rate is well defined: i.e., the exporting country has positive exports in an industry in the initial period. Similarly, the estimates aggregate to match country-import levels as long as a country has *some* positive imports in the initial period. In particular, the estimates will satisfy two moment conditions. First, the estimates will aggregate to match total exports in every country-industry observation:

$$\frac{\sum_j x_{kijf} - \sum_j x_{kijs}}{\sum_j x_{kijs}} = \gamma_{ki}^{AW} + \sum_j \frac{x_{kijs}}{\sum_\ell x_{ki\ell s}} \gamma_{kj}^{AW}, \tag{8}$$

where the left-hand side equals the growth rate of total exports in sector k from origin i, and the right-hand side is the sum of the exporter fixed effect and a bilateral export weighted average of the importer fixed effects. Secondly the estimates will aggregate to match total imports in every country-industry observation:

$$\frac{\sum_i x_{kijf} - \sum_i x_{kijs}}{\sum_i x_{kijs}} = \gamma_{kj}^{AW} + \sum_i \frac{x_{kijs}}{\sum_\ell x_{k\ell js}} \gamma_{ki}^{AW}. \tag{9}$$

Thus, the AW estimates aggregate to match every country's export and import growth.

Once we have estimates of γ_{ki}^{AW}, we can obtain estimates of industry productivity growth by using equation (6) and running the following regression:

$$\gamma_{ki}^{AW} = \kappa_i + \kappa_k + \epsilon_{ki}^{\gamma}, \tag{10}$$

where ϵ_{ki}^{γ} is a regression residual; and κ_i and κ_k are exporter and industry fixed effects. In this case, κ_i is the exporter normalization that captures the fact that comparative advantage is unaffected if all sectors in a country increase in productivity by the same amount. In this sense, we can think of κ_i as an export-supply shifter (e.g., a Hicks-neutral technical change or an increase in the size of the exporter). κ_k is an industry normalization that captures the fact that any common shift in the productivity of the industry or demand for that industry that is common across all countries will not affect a country's comparative advantage in producing that good. Finally, changes in country-industry comparative advantage, which correspond to $\Delta \ln z_{ki}$ are captured by $\epsilon_{ki}^{\gamma} = \theta \Delta \ln z_{ki}$. In this case, we can write the deviations in an exporter-industry's productivity from global trends as well as from average productivity growth in that country as $\Gamma_{ki} \equiv \epsilon_{ki}^{\gamma}/\theta$.

[1] One can always recover an estimate of γ'_{ij} by regressing the error term from the regression on a exporter-importer fixed effect.

5 Data (Still in Note Form)

We begin by defining some notation. Let the value of exports from region i to region j in industry k in year t be denoted by x_{ijt}^k. Define the value of exports from region i to the world in industry k in year t as $x_{it}^k \equiv \sum_j x_{ijt}^k$.

We next describe how to compute the average annual export growth rates. For each exporter-sector pair ik, we set the start year as the first non-zero observation in the years 1880 or 1885. We denote exports in this year by x_{is}^k, where s denotes the year chosen. Similarly, we set the final year as the last non-zero observation in the years 1905 or 1910 and denote it by x_{if}^k, where f denotes the year chosen. We set the average annual growth rate of exports for a country-sector as

$$g_{iksf} = \frac{\ln(x_{if}^k / x_{is}^k)}{f - s}. \tag{11}$$

Denote the cosine similarity between British patents and industry k's productivity technology by Φ_k. 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. Use the Maddison data (mpd2020.xls or mpd2020.dta) to divide the set of *non-Japanese* exporters into three terciles—High (H), Medium (M), and low (L)—according to per capita GDP (gdppc) in 1870. For regions that do not correspond to modern countries (e.g., Korea) use the average gdppc for the countries we have in our sample of exporters. Let JPN denote Japan. Thus, any exporter i belongs in one or more sets of countries given by Ω^C, where $C = \{\text{JPN, ASIA}, H, M, L, \text{Missing from Maddison}\}$.

5.1 Conversion to Annual Growth Rates

One concern with applying the AW procedure to the problem is that f and s may vary across countries and sectors. In order to deal with this problem, we first annualize the data. We do this in two steps,

1. If $x_{ijs}^k > 0$, the annual growth rate can be computed as

   $$g_{kij}^C \equiv \left(\frac{x_{ijf}^k}{x_{ijs}^k} \right)^{\frac{1}{f-s}} - 1$$

   Given this growth rate we can define the implied level of exports in period $s+1$ as $x_{ij,s+1}^k \equiv \left(1 + g_{kij}^C\right) x_{ijs}^k$.

2. If $x_{ijs}^k = 0$, $x_{ijf}^k > 0$ and $\sum_j x_{ijs}^k > 0$, we can define the average growth rate in exports of new varieties as

   $$g_{ki}^N \equiv \left(\frac{\sum_{j \in \mathcal{C}_i} x_{ijf}^k}{\sum_j x_{ijs}^k} \right)^{\frac{1}{f-s}}, \tag{12}$$

   where $\mathcal{C}_i$ is the set of j such that $x_{ijs}^k = 0$ and $x_{ijf}^k > 0$. Given an observed x_{ijf}^k, we can compute the implied level of new exports in period $s+1$ as $x_{ij,s+1}^k \equiv \left(1 + g_{ki}^N\right)^{-(f-s-1)} x_{ijf}^k$.

Once you have constructed these values for $x_{ij,s+1}^k$, you can use $x_{ij,s+1}^k$ instead of x_{ijf}^k and rerun the AW shocks on the annualized data.

6 Data and Measurement

In this section, we describe the main datasets used in the empirical analysis. Appendix XXX contains a complete discussion of all data used, including all data construction steps and sources.

6.1 Cross-country, bilateral, industry level trade flows

We construct what is, to the best of our knowledge, the first cross-country 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).[2] We build on existing, country-specific data sources, and add two, additional sources: US trade flows, and Japanese imports. We use 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). An observation in this dataset, x_{kij}, refers to an export flow from i to j in sector k. Using the fact the 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 countries' export flows.

Product lines were harmonized to 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 to ensure that similar product lines were consistently concorded to the same SITC category across all datasets. Country names (and boundaries) were harmonized within and across datasets. All trade values were converted to yen (at current exchange rates) using historical exchange rates from Fouquin and Hugot (2016). Our final dataset consists of export values for XX countries in YY sectors. Table X contains the summary statistics.

6.2 Constructing the British Patent Relevance measure

A key challenge for this paper is to quantify new technical knowledge by industry available to Japan and other technological follower countries. Our approach uses natural language processing to quantify one of the main channels through which codified technical knowledge diffused during this period: through 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) were a source of knowledge transmission for technical follower countries. We then describe how we use natural language processing to quantify new, Industrial Revolution technologies contained in each industry .

6.2.1 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

[2]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.

created through an 'industrial public sphere' which included scientific societies, and *the coding, storing and transmission* of useful knowledge" (Berg, 2007, p. 125, own emphasis).

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).

These manuals gave detailed, practical descriptions of the technological and organizational aspects of an industry. Their audience was the practitioner; the entreprenuer setting up a plant or the manager overseeing production. For example, take the "The American Cotton Spinner, and Managers' and Carders' Guide", published in 1851. This manual provides guidance on the layout of the factory building, on 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 knowledge entrepreneurs would need to familiarize themselves with for the setting up of modern, factory-based manufacturing and their day-to-day operation.

We use industry specific technical manuals to capture how intensively these technologies benefited each industry. Technical manuals were a standard way of codifying best-practice industry knowledge at the time. 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.

6.2.2 Quantifying new 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 steps. First, we use the text of technical manuals as a measure of frontier production methods. Second, we take patent text as a measure of new 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, for each industry we compute the similarity of text in industry technical manuals and patent text. This similarity score captures how much production processes by industry benefit from Industrial Revolution technologies. 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 thus 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* using a combination of automated keyword search and manual curation within Hathi's full repository of English language books published between XXXX and XXXX. We take the full text of these technical manuals to represent frontier knowledge of codifiable production techniques.

To proxy the full body of new technical knowledge created during the Industrial Revolution, we digitize the text of British patent titles (1617-1851) from Bennet Woodcroft's "Subject Matter 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 also harder to diffuse and absorb, and will be absent from technical manuals. For these reasons, patent text should proxy new technologies of the Industrial Revolution that could most easily diffuse.

With these two sources of data, we have measures of both 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 text. Each document (technical manual, patent text) i is represented by a vector of size N, where N is the size of the vocabulary across the entire corpus. The vocabulary includes unigrams and bigrams, and employ 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 new technology content of technical manuals. We want to know how similar British patent text is to the technical manual for an 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 vectors. Intuitively, if an industry uses the technologies of the Industrial Revolution intensively, the text of its technical manuals will feature many words that are also frequently found in the text of British patents. 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;

$$BPR_i = \frac{\boldsymbol{BW} \cdot \boldsymbol{TM_i}}{\|\boldsymbol{BW}\|\|\boldsymbol{TM_i}\|} = \frac{\sum_{j=1}^{n} BW_j TM_{ij}}{\sqrt{\sum_{j=1}^{n} BW_j^2}\sqrt{\sum_{j=1}^{n} TM_{ij}^2}} \tag{13}$$

which we call the *British Patent Relevance (BPR)* measure.

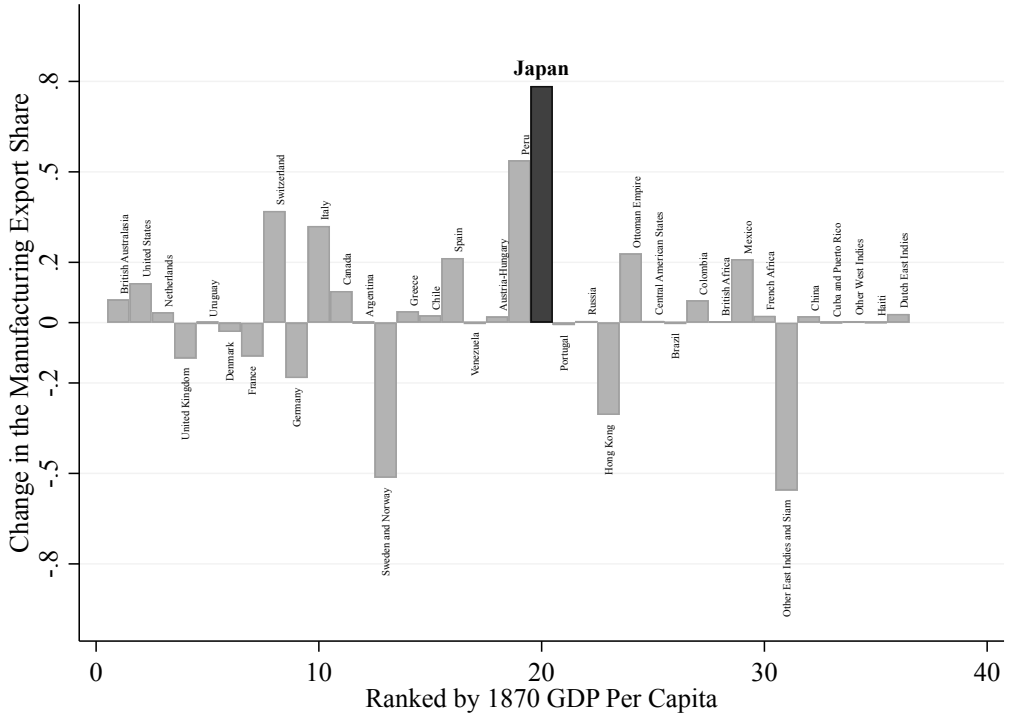
This measure has several important advantages for our setting. Most importantly, by focusing on knowledge codified in technical manuals, this measure captures one of the key channels through which Meiji Japan acquired Western knowledge: the translation of

these documents. Moreover, this measure naturally accounts for how a given technology benefits different industries through input-output linkages. For example, this measure will naturally quantify which industries benefited more from macro-inventions such as the steam engine – a distinct advantage relative to using the industry classification of patents.

7 Tables and Figures

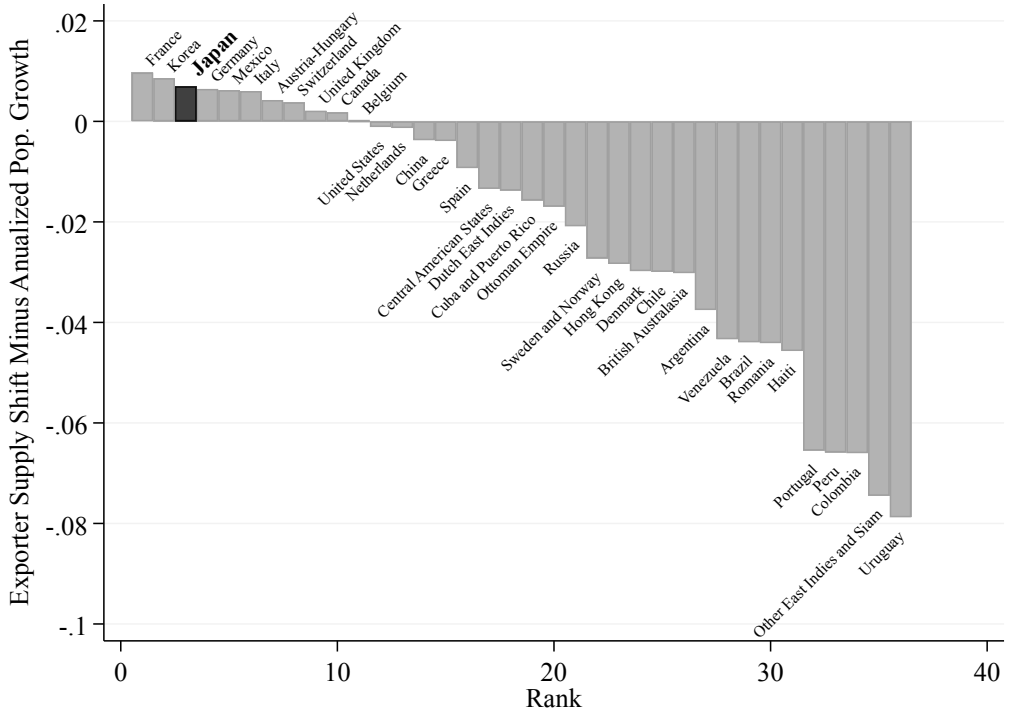
7.1 The World in 1880

Figure 1: Change in the Manufacturing Share of Exports Around the World (1880-1910)



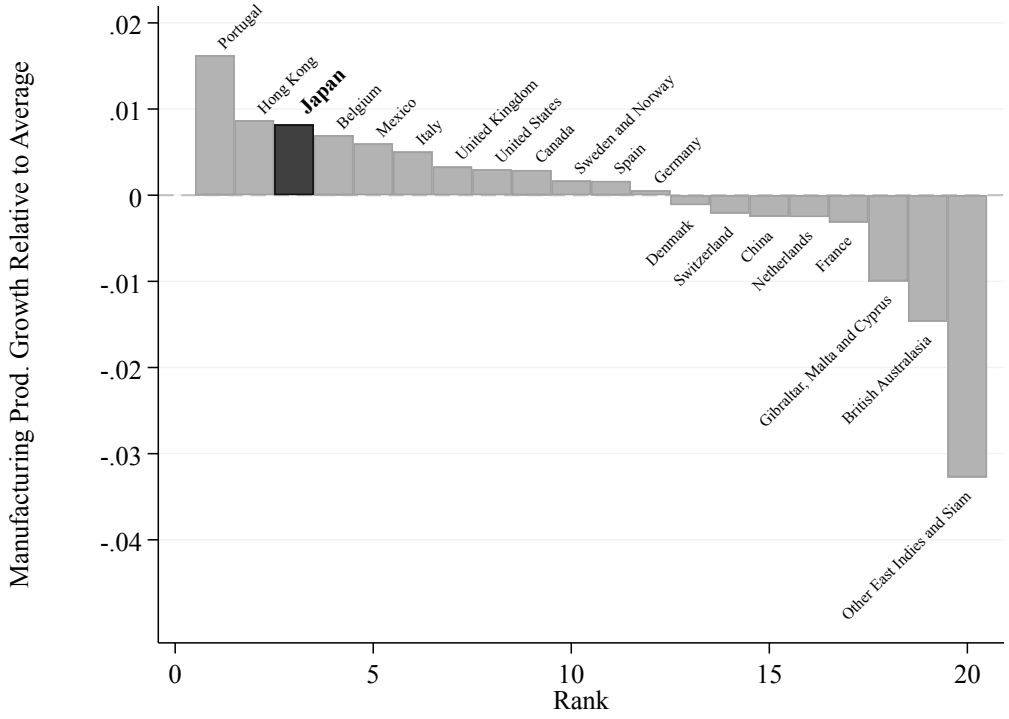
Note: 1870 GDP per Capita from the Maddison project. Primary export share computed using historical bilateral trade date reported by Belgium, Italy, Japan and the United States. An industry is considered non-primary if it has SITC code 6 (Manufactured goods classified chiefly by material), 7 (Machinery 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 2: Exporter Supply Shift κ_i Minus Annualized Population Growth



Note: Exporter Supply Shift refers to κ_i from the following regression: $\gamma_{ik}^{AW} = \kappa_i + \kappa_k$ where the κ terms refer to a dummy for each country i and industry k. See text for details on variable construction. Annualized population growth refers to the annualized population growth of each country between {1870,1880} and 1913, the data comes from the Maddison Project. See text for details on variable construction.

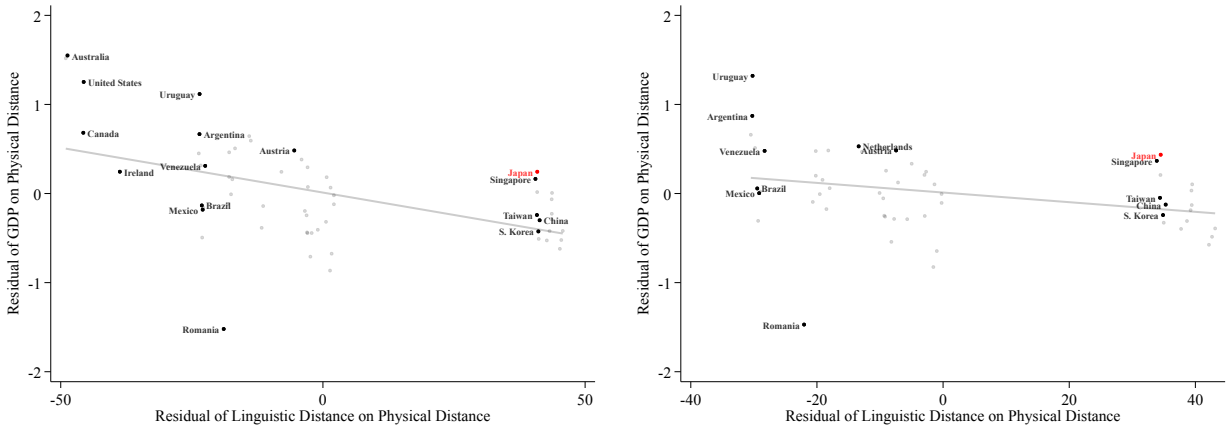
Figure 3: Relative Annualized Productivity Growth Γ in Manufacturing



Note: We begin by calculating the productivity growth of country $i$in industry k (Γ_{ik}) according equation 22 of the Export Regression notes (version 11). Using Γ_{ik}, we run the following regression:
$\Gamma_{ik} = \sum_{j \in \text{Set of Countries}} \left(\beta_{Aj} A_k \times I_i^j + \beta_{MAi} MA_k \times I_i^j + \beta_{MIj} MI_k \times I_i^j \right)$
where I_i^j is a dummy equal 1 if the country of the observation is country j, A_k is a dummy equal 1 if sector k is an agricultural or fishing industry, MA_k equals 1 if sector k is a manufacturing industry and MI_k is equal 1 if the sector k is in mining. We interpret the coefficient β_{MAj} as country j productivity growth in manufacturing relative to the average growth of all countries in manufacturing. The plot presents β_{MAj} coefficients for countries in which Manufacturing sector's export share in 1880 is at least 0.5% and countries with productivity growth estimates for least five non-primary sectors and five primary sectors. See text for details on variable construction.

7.2 Language Regressions

Figure 4: Linguistic Distance Partial Regression Plot for 1870

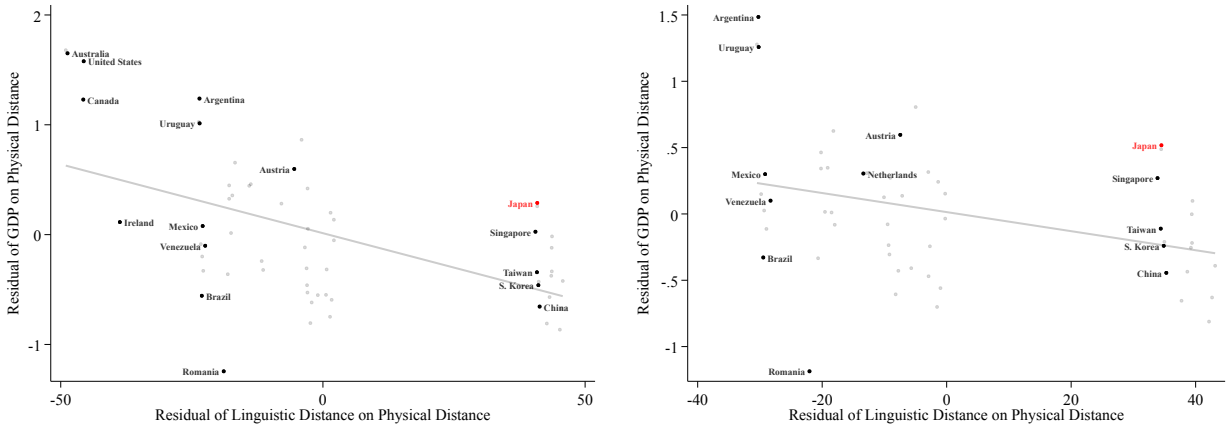


(a) Include English-speaking countries

(b) Exclude English-speaking Countries

Note: GDP refers to log GDP per Capita from the Maddison Project. Physical Distance refers to log Physical Distance between Country and the UK is from the CEPII databasedatabase using the Great Circle Formula which uses a weighted average of latitude and longitude of two countries using population. Number of Weeks Required to Learn the Plurality Language is the number of weeks an English native speaker will take to obtain "Professional Working Proficiency" the majority language in the country as estimated by the U.S. Department of State's Foreign Service Institute. See text for details on variable construction. Since we take the log of weeks to learn the language in the regressions and the log of zero is undefined, we drop English speaking countries from the plots.

Figure 5: Linguistic Distance Partial Regression Plot for 1913



(a) Include English-speaking countries (b) Exclude English-speaking Countries

Note: GDP refers to log GDP per Capita from the Maddison Project. Physical Distance refers to log Physical Distance between Country and the UK is from the CEPII databasedatabase using the Great Circle Formula which uses a weighted average of latitude and longitude of two countries using population. Number of Weeks Required to Learn the Plurality Language is the number of weeks an English native speaker will take to obtain "Professional Working Proficiency" the majority language in the country as estimated by the U.S. Department of State's Foreign Service Institute. See text for details on variable construction. Since we take the log of weeks to learn the language in the regressions and the log of zero is undefined, we drop English speaking countries from the plots.

Table 1: Linguistic Distance from English and GDP - OLS Specification

	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	✓	✓	✓			

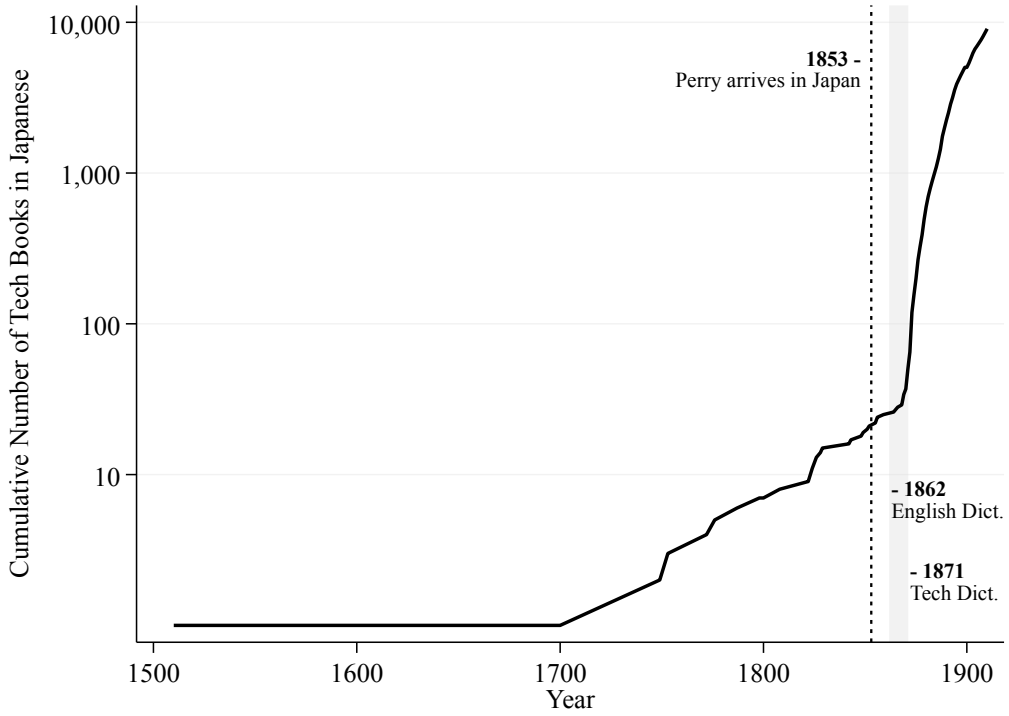
Standard errors in parentheses

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

Note: The log GDP per Capita variables are from the Maddison Project. Log Physical Distance between Country and the UK is from the CEPII database using the Great Circle Formula which uses a weighted average of latitude and longitude of two countries using population. This means that the distance of the UK to itself is not zero (it is a measure of how far apart its urban centers area), therefore the UK is included in this regression as the observation with the shortest distance to the UK. Number of weeks an English native speaker will take to obtain "Professional Working Proficiency" the majority language in the country as estimated by the U.S. Department of State's Foreign Service Institute. See text for details on variable construction.

7.3 Technical Books in the 1800s and Early 1900s

Figure 6: The Number of Technical Books in Japanese Takes Off c. 1870



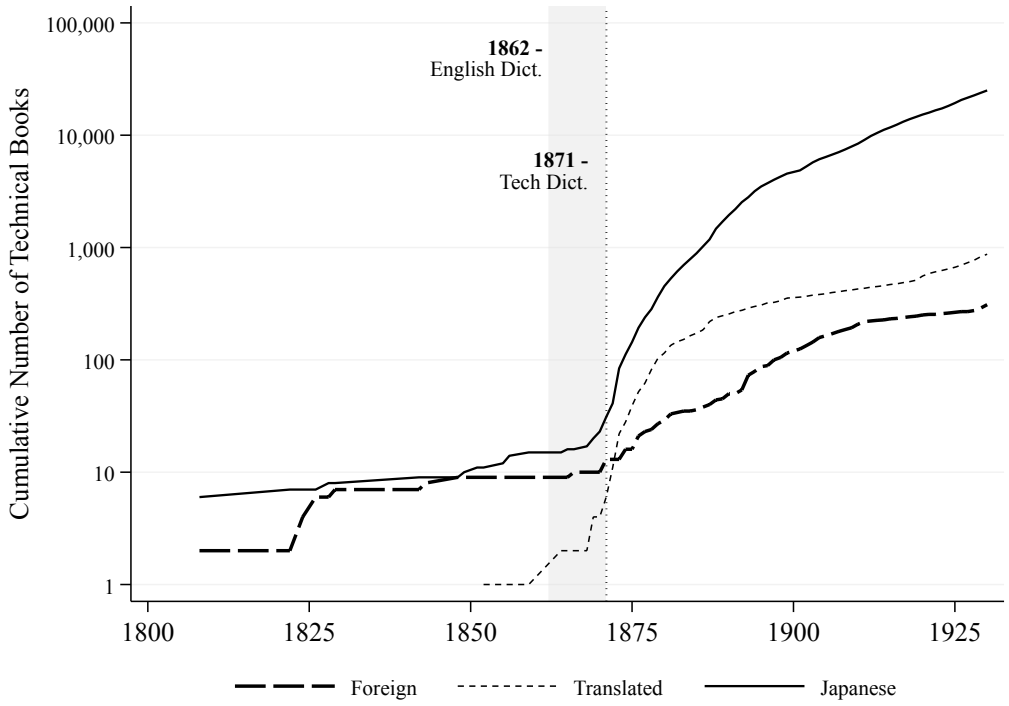
Note: The Cumulative Number of Technical Books variable refers to all books in the National Diet Library catalog published or translated by year t the classified under divisions “500 - Technology and Engineering" and “600 - Industry and Commerce" from the Nippon Decimal Classification System. See text for details on variable construction.

Figure 7: Total Number of Technical Books Translated into Japanese in Log Scale



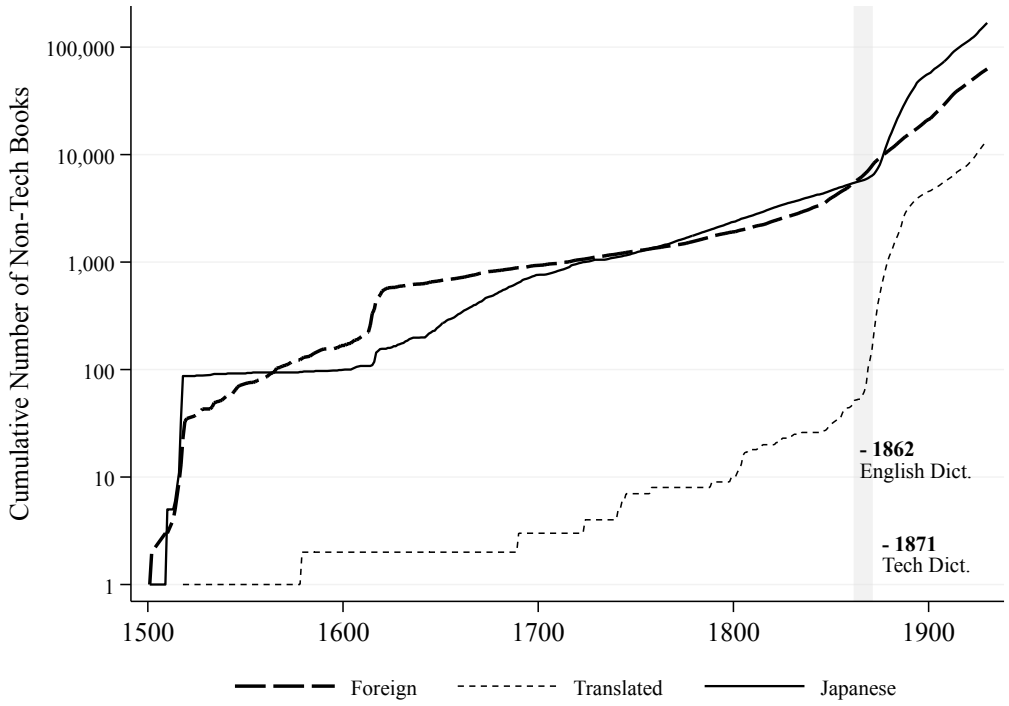
Note: The Cumulative Number of Technical Books variable refers to all books in the National Diet Library catalog published or translated by year t classified under divisions "500 - Technology and Engineering" and "600 - Industry and Commerce" from the Nippon Decimal Classification System. A book is foreign if its title is not in Japanese and book is considered a translation if the National Diet Library flags the book as a translation. Since up until 1798 there were no translated books, the log scale plot with 0 will look very strange since the 0-1 portion of the y-axis will take up the majority of the axis. See text for details on variable construction.

Figure 8: Total Number of Technical Books in Log Scale in Japan



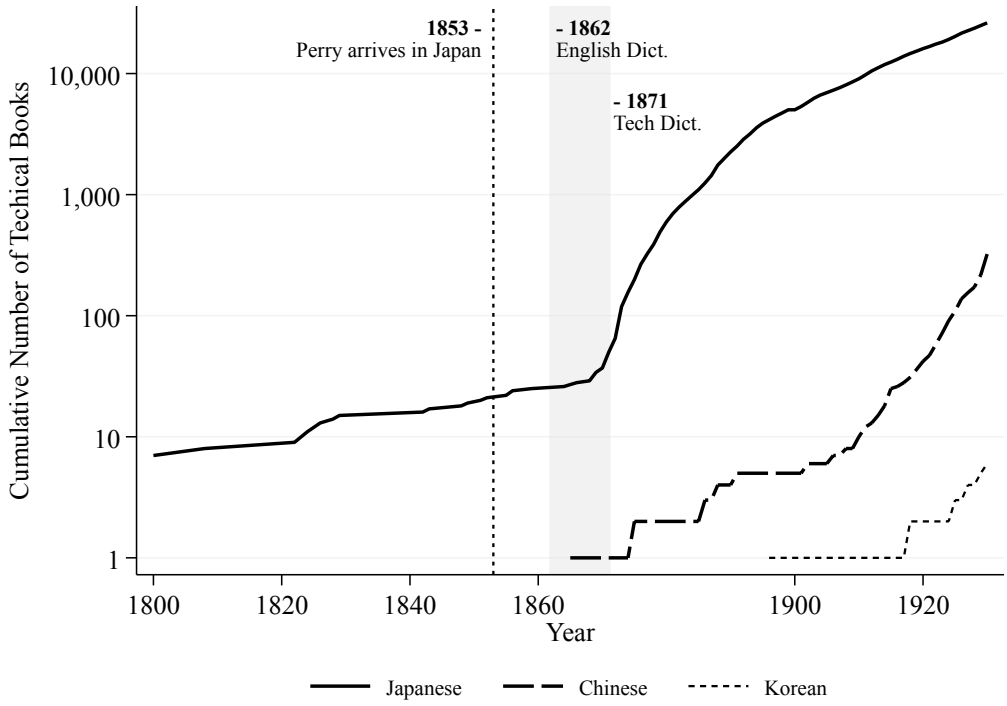
Note: The Cumulative Number of Technical Books variable refers to all books in the National Diet Library catalog published or translated by year t classified under divisions "500 - Technology and Engineering" and "600 - Industry and Commerce" from the Nippon Decimal Classification System. A book is foreign if its title is not in Japanese and book is considered a translation if the National Diet Library flags the book as a translation. Since up until 1798 there were no translated books, the log scale plot with 0 will look very strange since the 0-1 portion of the y-axis will take up the majority of the axis. See text for details on variable construction.

Figure 9: Total Number of Non-Technical Books Translated into Japanese



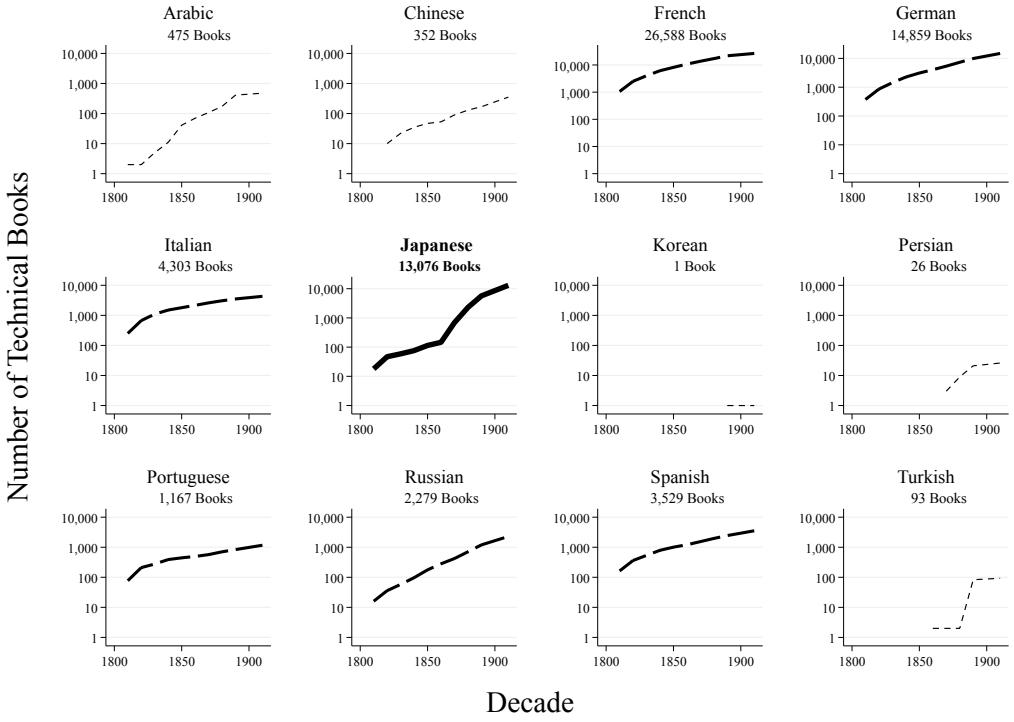
Note: The Cumulative Number of Non-Technical Books variable refers to all books in the National Diet Library catalog published or translated by year t the not classified under divisions "500 Technology and Engineering" and "600 Industry and Commerce" from the Nippon Decimal Classification System. A book is foreign if its title is not in Japanese and book is considered a translation if the National Diet Library flags the book as a translation. A book is considered Japanese it if it has a Japanese title and is not a translation. See text for details on variable construction.

Figure 10: By 1910 Japanese has 1000 Times More Tech Books than Korean and Chinese



Note: The list of Japanese technical books was scraped from the National Diet Library's catalogue of books classified under divisions "500 - Technology and Engineering" and "600 - Industry and Commerce" from the Nippon Decimal Classification System. Korean books were scraped from the Korean National Library. In this library technical book as classified under category "600 Technology and Engineering" from the Dewey Decimal System. Lastly, the Chinese books were scraped from the Shanghai library. A Chinese book is technical if it has subject Agriculture (S), Industry (T), or Transportation (U) in the Chinese Library Classification System. See text for details on variable construction.

Figure 11: Technical Books Available in the World's Major Languages



Note: European languages represented by thick dashed lines. Non-European languages represented by think dashed lines. Japanese, Chinese and Korean books scraped from the National Diet Library, the Shanghai Library and the National Library of Korea respectively. All other books were scraped from WorldCat.

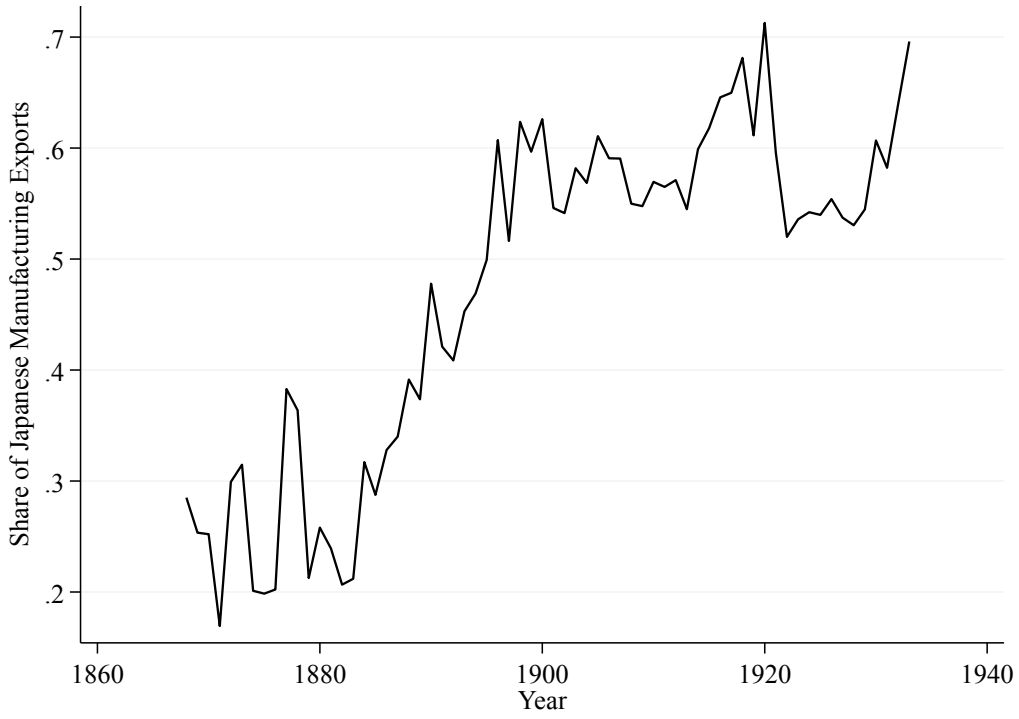
7.4 Meiji Japan

Figure 12: Japanese Government Expenditure



Note: Government Educational Expenditure from *Meiji Taisho Zaisei Shoran* (1926). See text for details on variable construction.

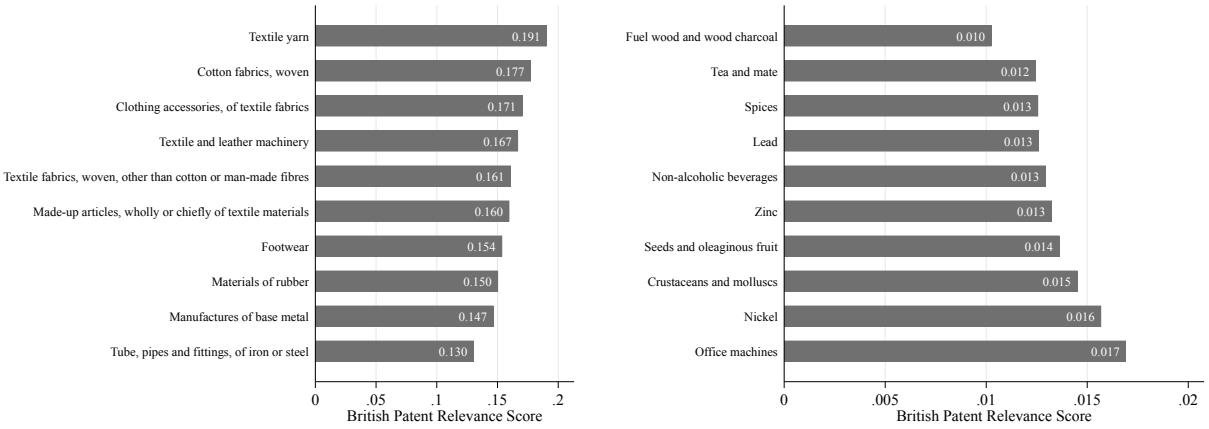
Figure 13: Share of Japanese Manufacturing Exports



Note: **Missing information.** See text for details on variable construction.

7.5 British Patent Relevance Scores

Figure 14: Industries Ranked by British Patent Relevance Score



(a) Top 10 (b) Bottom 10

Table 2: Top 10 Industries Based on British Patent Relevance

SITC-3	SITC-3 Description	BPR
651	Textile yarn	0.191
652	Cotton fabrics, woven (not including narrow or ...	0.177
847	Clothing accessories, of textile fabrics, nes	0.171
724	Textile and leather machinery, and parts thereo...	0.167
654	Textile fabrics, woven, other than cotton or ma...	0.161
658	Made-up articles, wholly or chiefly of textile ...	0.160
851	Footwear	0.154
621	Materials of rubber (e.g., pastes, plates, shee...	0.150
699	Manufactures of base metal, nes	0.147
678	Tube, pipes and fittings, of iron or steel	0.130

Note: See text for details on variable construction.

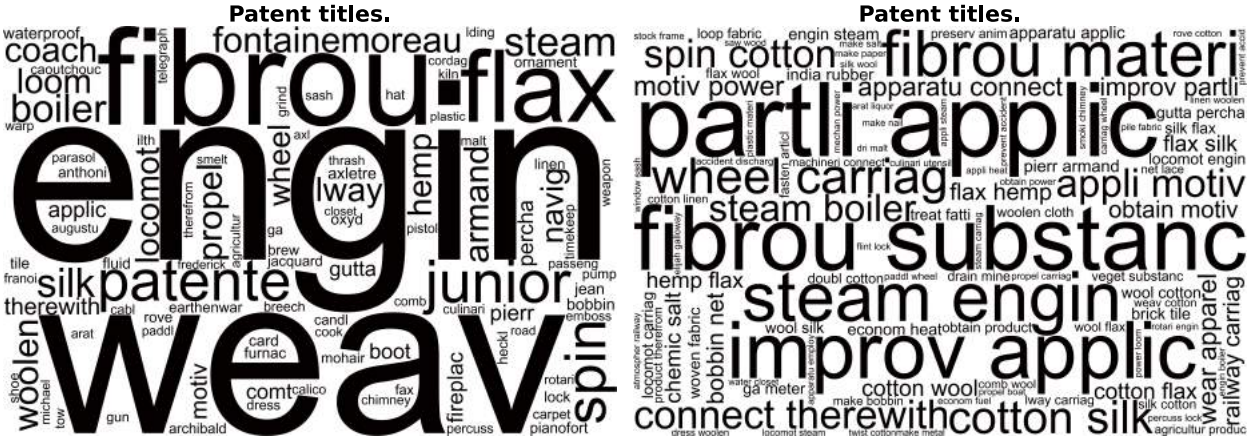
Table 3: Bottom 10 Industries Based on British Patent Relevance

SITC-3	SITC-3 Description	BPR
245	Fuel wood and wood charcoal	0.010
074	Tea and mate	0.012
075	Spices	0.013
685	Lead	0.013
111	Non-alcoholic beverages, nes	0.013
686	Zinc	0.013
222	Seeds and oleaginous fruit, whole or broken, fo...	0.014
036	Crustaceans and molluscs, fresh, chilled, froze...	0.015
683	Nickel	0.016
751	Office machines	0.017

Note: See text for details on variable construction.

7.6 Visualizing Uni- and Bi-grams

Figure 15: Patent Titles



(a) Unigrams (b) Bigrams

Figure 16: Industry with Highest BRP



(a) Unigrams (b) Bigrams

Figure 17: Industry with Lowest BRP

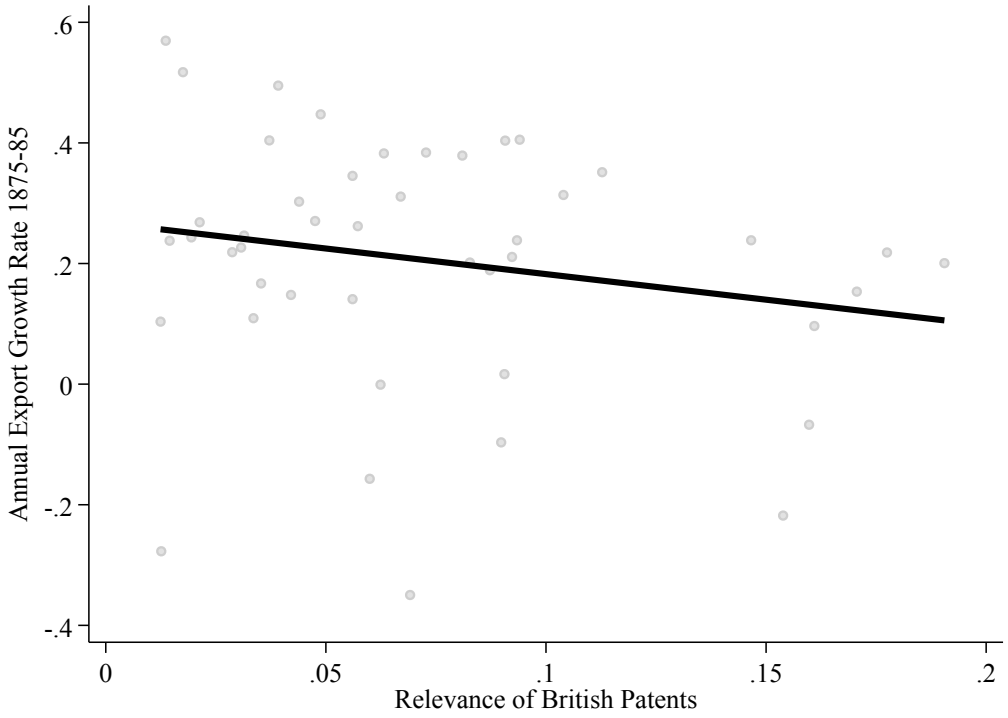


(a) Unigrams

(b) Bigrams

7.7 Main Results

Figure 18: Placebo Test



Note: The dependent variable Export Growth is the annualized export growth rate for a country *i*'s industry *k* between {1880,1885} and {1905,1910}. See text for details on variable construction.

Table 4: Placebo Test

	(1) Export Growth
British Patent Relevance	-0.848
	(0.577)
Constant	0.267***
	(0.056)
Observations	45

Standard errors in parentheses

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

Note: The dependent variable Export Growth is the annualized export growth rate for a country *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.

Table 5: Placebo Differences-in-Differences

	Annualized Export Growth Rate	
	(1)	(2)
Annualized Export Growth Rate 1875-1885	-0.234***	-0.174*
	(0.078)	(0.089)
British Patent Relevance	3.833***	3.792**
	(1.366)	(1.411)
Observations	41	41
R^2	0.372	0.253
Sample	Japan	Japan
Start year	1885	1880 or 1885
Finish year	1910	1905 or 1910

Note: Annualized export growth rate was computed using the following formula $(value_f/value_s)^{\frac{1}{f-s}} - 1$ where s refers to the start-year and f refers to the finish year, $value$ refers to the sum of exports in an industry for year $\{s, f\}$. In the first version of the regression, we require that the start and end years are 1885 and 1910, in the second version (depending on data availability) we set the start and end year to $\min(1880, 1885)$ and $\max(1905, 1910)$ respectively. 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.

Table 6: Annualized Export Growth and British Patent Relevance

	Export Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
British Patent Relevance × Japan	3.052***	3.052***	3.052***	3.052***	3.052***	3.052***	3.052***
	(0.799)	(0.799)	(0.799)	(0.799)	(0.799)	(0.799)	(0.800)
British Patent Relevance × Not Japan		-0.709***	-0.882***	-0.847***	-0.481*		
		(0.210)	(0.239)	(0.241)	(0.268)		
British Patent Relevance × English-Speaking			1.016**				
			(0.479)				
British Patent Relevance × British Colony				0.664			
				(0.486)			
British Patent Relevance × Language Similar to English					-0.469		
					(0.424)		
British Patent Relevance × High-Income						-0.323	-0.323
						(0.234)	(0.234)
British Patent Relevance × Medium-Income						-0.873	-0.781
						(0.543)	(0.556)
British Patent Relevance × Low-Income						-1.550***	-1.100*
						(0.511)	(0.650)
British Patent Relevance × Asia							-0.964
							(0.814)
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

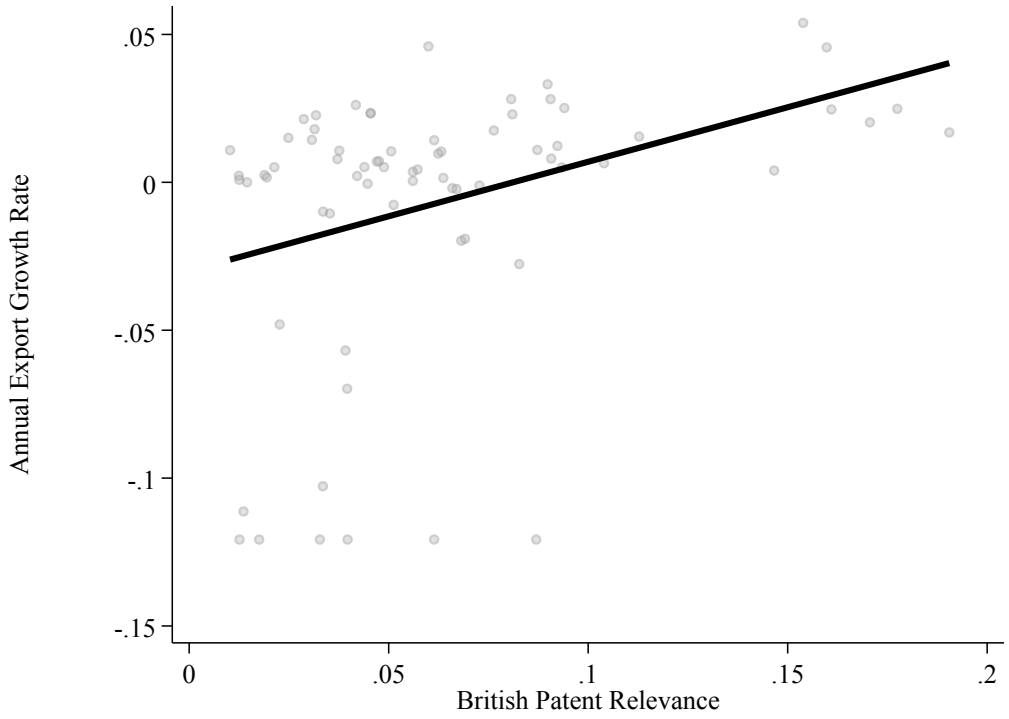
Note: Dataset restricted to countries with a GPD per capita value and industries with a British Patent Relevance estimate. The dependent variable Export Growth is the annualized export growth rate for a country *i*'s industry *k* between {1880,1885} and {1905,1910}. British Patent Relevance is a variable 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 country is Japan and zero otherwise, Not Japan is analogously defined. English-Speaking is an indicator equal to 1 if the country's plurality language is English. British Colony is dummy for whether a country was a British colony in the 1880-1910 window. Language Similar to English is an indicator equal to 1 if an English native speaker can attain professional proficiency in the country'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 country is in the top third of the income distribution (high), middle third (medium) or in the bottom third (bottom); we set these dummies to 0 for Japan. Asia dummy equals 1 if the country is in Asia and 0 if it is Japan or not in Asia. See text for details on variable construction.

Table 7: Annualized Productivity Growth Γ and British Patent Relevance

	Γ						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
British Patent Relevance × Japan	0.279***	0.279***	0.279***	0.279***	0.279***	0.279***	0.279***
	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)	(0.086)
British Patent Relevance × Not Japan		-0.010	-0.020	-0.025	-0.020		
		(0.024)	(0.028)	(0.028)	(0.033)		
British Patent Relevance × English-Speaking			0.062				
			(0.053)				
British Patent Relevance × British Colony				0.070			
				(0.053)			
British Patent Relevance × Language Similar to English					0.021		
					(0.049)		
British Patent Relevance × High-Income						-0.006	-0.006
						(0.026)	(0.026)
British Patent Relevance × Medium-Income						0.047	0.067
						(0.059)	(0.061)
British Patent Relevance × Low-Income						-0.077	0.014
						(0.062)	(0.075)
British Patent Relevance × Asia							-0.191**
							(0.096)
Observations	56	1243	1243	1243	1243	1243	1243
R^2	0.068	0.010	0.010	0.011	0.010	0.012	0.015
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

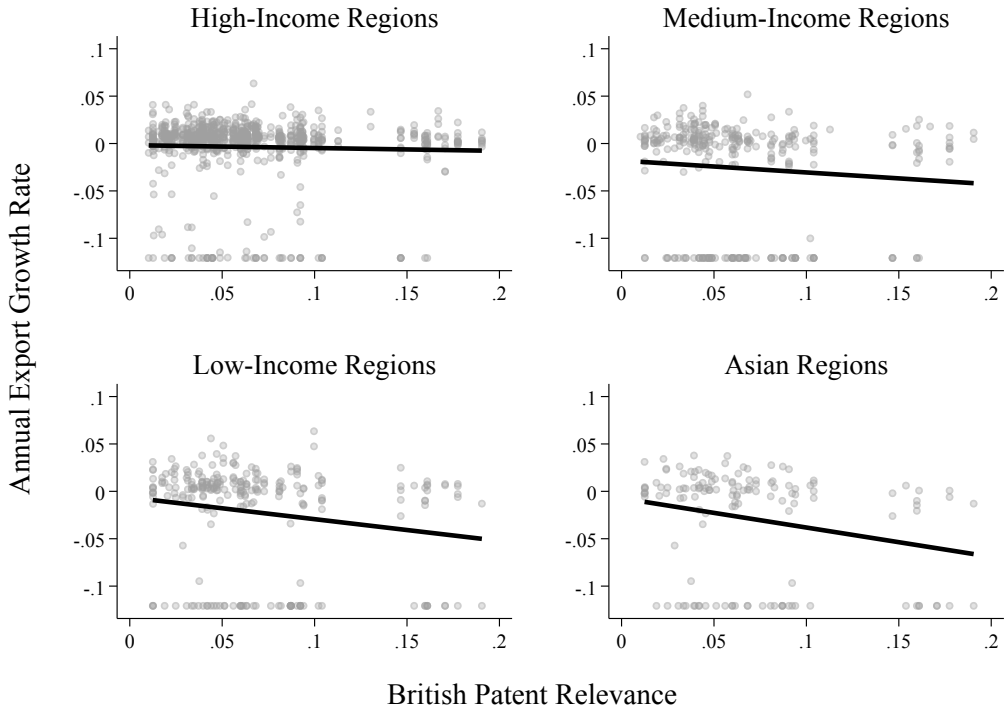
Note: Dataset restricted to countries with a GPD per capita value and industries with a British Patent Relevance estimate. The dependent variable Γ_{ik} is the annualized productivity growth rate for a country *i*'s industry *k* between {1880,1885} and {1905,1910}. Γ_{ik} was constructed using γ_{ik}^{AW} which was estimated using the Amiti-Weinstein (JEP 2018) methodology and calibrated by $\theta = 8.28$ from Eaton-Kortum (2002). 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 country is Japan and zero otherwise, Not Japan is analogously defined. English-Speaking is an indicator equal to 1 if the country's plurality language is English. British Colony is dummy for whether a country was a British colony in the 1880-1910 window. Language Similar to English is an indicator equal to 1 if an English native speaker can attain professional proficiency in the country'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 country is in the top third of the income distribution (high), middle third (medium) or in the bottom third (bottom); we set these dummies to 0 for Japan. Asia dummy equals 1 if the country is in Asia and 0 if it is Japan or not in Asia. See text for details on variable construction.

Figure 19: Annualized Export Growth and British Patent Relevance for Japan



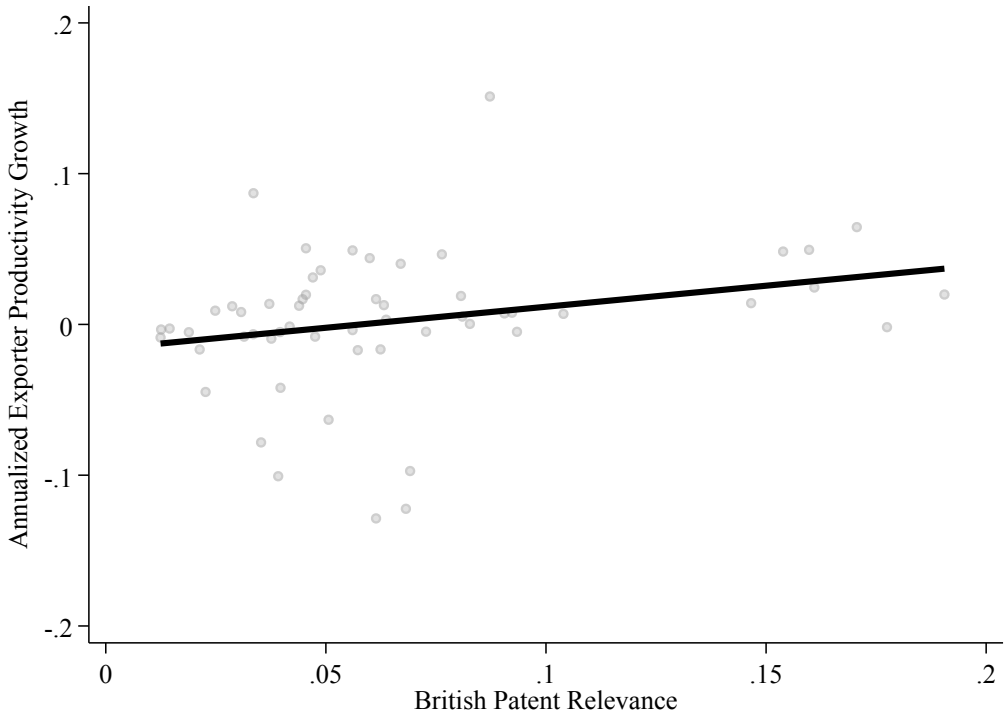
Note: The dependent variable Γ_{ik} is the annualized productivity growth rate for a country *i*'s industry *k* between {1880,1885} and {1905,1910}. Γ_{ik} was constructed using γ_{ik}^{AW} which was estimated using the Amiti-Weinstein (JEP 2018) methodology and calibrated by $\theta = 8.28$ from Eaton-Kortum (2002). 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 20: Annualized Export Growth and British Patent Relevance by Income Group



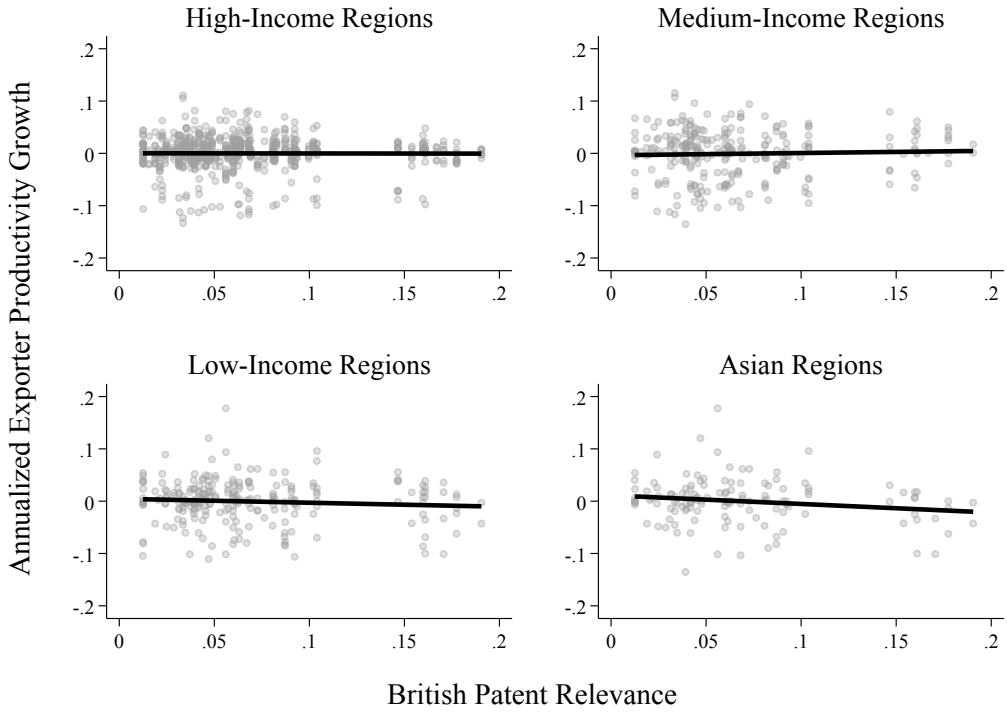
Note: The dependent variable Γ_{ik} is the annualized productivity growth rate for a country i's industry k between {1880,1885} and {1905,1910}. Γ_{ik} was constructed using γ_{ik}^{AW} which was estimated using the Amiti-Weinstein (JEP 2018) methodology and calibrated by $\theta = 8.28$ from Eaton-Kortum (2002). 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. {High, Medium, Low}-Income Regions refer to dummies which use 1870 GDP per capita from the Maddison Project to flag if a country 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. Asian dummy equals 1 if the country is in Asia and 0 if it is Japan or not in Asia. See text for details on variable construction.

Figure 21: Annualized Prod. Growth Γ and British Patent Relevance for Japan



Note: The dependent variable Γ_{ik} is the annualized productivity growth rate for a country *i*'s industry *k* between {1880,1885} and {1905,1910}. Γ_{ik} was constructed using γ_{ik}^{AW} which was estimated using the Amiti-Weinstein (JEP 2018) methodology and calibrated by $\theta = 8.28$ from Eaton-Kortum (2002). 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 22: Annualized Prod. Growth Γ and British Patent Relevance by Income Group



Note: The dependent variable Γ_{ik} is the annualized productivity growth rate for a country i's industry k between {1880,1885} and {1905,1910}. Γ_{ik} was constructed using γ_{ik}^{AW} which was estimated using the Amiti-Weinstein (JEP 2018) methodology and calibrated by $\theta = 8.28$ from Eaton-Kortum (2002). 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. {High, Medium, Low}-Income Regions refer to dummies which use 1870 GDP per capita from the Maddison Project to flag if a country 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. Asian dummy equals 1 if the country is in Asia and 0 if it is Japan or not in Asia. See text for details on variable construction.

References

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

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*.

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.

Horn, J. (2006). The path not taken: French industrialization in the age of revolution. *Cambridge, MA*.

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.

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

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.

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

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

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

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

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