Technology Diffusion through Cultural Links: Evidence from Industrial Firms in Late Imperial Russia*

Tamar Matiashvili† Timur Natkhov‡

November 22, 2025

Abstract

This paper examines the impact of ethnic connections on the diffusion of technology in a late-industrializing economy. Using machine learning we assigned probable ethnicities to nearly 16,000 industrial firm owners in the Russian Empire in 1894. We then compared firms owned by entrepreneurs with cultural connections to Europe (e.g., Germans and Poles) to Russian-owned firms. "Connected" firms exhibited higher productivity, employed modern machinery and advanced management practices. We established the knowledge diffusion channel by showing, among others, that "connected" firms traded more with Europe, and that German-origin entrepreneurs had the largest productivity advantage in chemicals – Germany's leading industry at the time. Our findings highlight the importance of cultural links in the diffusion of the Industrial Enlightenment from the industrial core of Europe to its periphery.

Keywords: technology diffusion, culture, Industrial Enlightenment, Russian Empire

JEL codes: N33, N43, N53, J47, O43

*We are grateful to Ran Abramitzky, Liran Einav, Oded Galor, Amanda Gregg, Pete Klenow, and Steve Nafziger for their invaluable feedback. We also thank the participants of various Stanford Economics Department workshops, 2023 Summer Workshop in the Economic History and Historical Political Economy of Eurasia, Asia-Pacific Economic and Business History Conference 2024, ASREC Chapman 2024, Cliometrics Conference 2024, Society for Institutional and Organizational Economics Annual Conference 2024, New Economic School, Asian Economic History Conference 2024.

†Stanford University, Department of Economics, tamrim@stanford.edu – corresponding author.

‡Higher School of Economics, Department of Applied Economics, timur.natkhov@hse.ru

1 Introduction

What factors facilitate the adoption of industrial technologies in developing economies? The existing empirical research has primarily examined macro-level cross-country diffusion patterns[1] or studied the adoption of specific technologies in individual industries.[2] In this paper, we take an economy-wide micro-level approach exploring the universe of manufacturing firms in the Russian Empire, which is a typical example of a late-industrializing economy. We show that firm owners' cultural connections with Western Europe, manifested by their ethnic origin, facilitate the adoption of advanced technologies, both at extensive and intensive margins, and increase firm productivity. These "connected" firms then serve as role models for local entrepreneurs inducing spillovers of useful knowledge and productivity.

We employ the uniquely detailed and comprehensive manufacturing census of about 16,000 private industrial firms conducted by the Ministry of Finance of the Russian Empire in 1894.[3] The census contains information on a firm's name, location, industry, employees, machinery, revenue, and other variables. Crucially, it includes the owners' surname, which allows us to identify the likely ethnicity of a firm owner. We employ a combination of automated approaches and a supervised machine learning algorithm to classify them into one of five different groups, namely Russians, Germans, Jews, Other Westerners (Poles, Lithuanians, etc.), and Eastern minorities (Tatars, Armenians, etc.). Of these five groups, we consider Germans, Jews, and Other Westerners culturally "connected" to Europe, while Russians and Eastern minorities as "non-connected".

We find that "connected" firms played a major role in the adoption of industrial machinery and advanced management practices in the early stages of Russian industrialization. In particular, firms owned by "connected" entrepreneurs were more productive in terms of revenue per worker and TFP, more likely to employ machinery instead of draft animals or human labor alone, and more likely to employ machinery with greater horsepower that ran on modern types of fuel such as oil or gas. They also adopted advanced management practices by sharing ownership with partners, employing women, and operating more regularly throughout the year. Focusing on private firms instead of corporations ensures that these advantages did not originate from access to stock market, protection of limited liability, or other exclusive legal rights.

We conjecture that the main reason for the higher productivity of "connected" entrepreneurs was their knowledge of modern production techniques and management practices acquired through cultural links with European countries. Unfortunately, the manufacturing census does not contain information on a firm's trade or investment activities, which would allow us to directly test this conjecture. Instead, we develop several alternative approaches using additional data sources.

[1] See Comin and Mestieri (2014) for the literature review.

[2] See Caselli and Coleman (2001) on the diffusion of personal computers, Manuelli and Seshadri (2014) on the diffusion of tractors, and Foster and Rosenzweig (2010) for the review of micro-level evidence; Griliches (1957) is the seminal paper on the topic.

[3] See Gregg (2020) for the first paper based on this census and for a detailed description of the data. We thank Amanda Gregg for kindly sharing the digitized version of the census.

First, we digitize data for all Russian firms that participated in the 1893 Chicago World Fair, which was an international exhibition of the technological advances of leading industrial nations in the late 19th century. We find that, even among these positively selected firms, the owner's ethnic origin predicts the extent of their commercial ties to Western Europe. "Connected" firms were significantly more likely to trade with European countries.

Second, we identify the industries in which Germany, one of the leading industrial nations at the time, had a comparative advantage, and see whether entrepreneurs of German origin in the Russian Empire also had an advantage in the same industries. We find that firms with owners of German origin had the largest productivity advantage in the chemical industry with about 80% higher revenue per worker than Russian-owned firms. This directly corresponded to Germany's world ascendance in the chemical industry in the late 19th century. Other connected groups had no such advantage. Moreover, we do not observe any "German premium" for industries in which Germany had no comparative advantage at the time, such as cotton textiles, wool- and silk-making.

Third, we examine the role of local spillovers. Presumably, a significant share of technology diffusion occurs at the regional level when lagging firms emulate a local "role model" in their industry. We then examine if higher exposure to "connected" firms facilitates technology adoption for "non-connected" firms operating in the same market. We find that "non-connected" firms exposed to a higher share of "connected" firms in a given industry and district were indeed more productive. Of all the productivity measures, the largest coefficient is for revenue TFP, implying that spillovers of know-how and unobservable management practices were probably the most consequential. This result highlights the importance of tacit knowledge in the process of technology adoption.

Fourth, we digitize a separate data set on patents issued in the Russian Empire in 1892-1896. We find that of all patents issued to the citizens of the Empire, about 75% had been granted to individuals with western surnames and only 24% to individuals of Russian or Eastern minorities origin. Thus, even domestic innovation in the Russian Empire appears to have been largely driven by the "connected" groups.

Finally, ease of communication and transportation to Europe could play a role in facilitating cultural connections. We find that the "connected" groups benefited more than the "non-connected" ones from locating in one of the 17 largest cities with population exceeding 100,000. The cities provided communication channels necessary to maintain connections with Europe. Furthermore, we find that "connected" firms closer to the European border benefited from the shortened distance, but these results are less precise.

Since our exercises are descriptive in nature, we provide additional evidence to rule out salient competing explanations. An important threat to the proposed mechanism is that firm owners with European surnames may have started their businesses with significantly larger capital. We address this concern showing that the productivity premium for "connected" firms existed even within the sub-sample of firms without any machinery, namely without significant initial capital investments.

Further, we try to rule out the human capital effect by focusing on a small subset of highly literate districts, and again obtain similar results. We also show that the productivity premium for "connected" firms could not be explained by ethnic network effects: being surrounded by upstream and downstream firms owned by individuals of the same ethnicity does not increase firm productivity and, if anything, is associated with lower performance. Finally, we rule out in-group effects for the "connected" group: the productivity premium of "connected" firms was just as strong when located in districts with ethnic Russian majority as in districts with non-Russian majority.

This paper contributes to two strands of literature on economic growth. First, we contribute to the literature on cultural determinants of technology diffusion. Ashraf and Galor (2011) empirically establish that low cultural diversity determined by geographical isolation in the pre-industrial period impedes the adoption of new technologies, delaying the onset of industrialization. Spolaore and Wacziarg (2009) show that cultural barriers, measured by the genetic distance to the technological frontier, explain a large portion of the productivity gap in a cross-section of countries. Historically, the diffusion of the Industrial Enlightenment (Mokyr, 2016) across Europe was associated with higher levels of upper-tail human capital (Squicciarini and Voigtländer, 2015), presence of scientific societies (Cinnirella et al., 2022), migration of skilled minorities (Hornung, 2014) and creative individuals (Serafinelli and Tabellini, 2022). Our paper shows that, in addition to the aforementioned factors, cultural connections via ethnic links facilitate technology diffusion even without migrations, and may compensate, to some extent, for rigid institutions and insufficient human capital in a receiving country.

Second, we contribute to the literature on the determinants of firm productivity (Syverson, 2011). One of the main factors of productivity variation across firms is management quality. Bloom et al. (2019) estimate that differences in management practices account for 20% of differences in productivity in US plants, while Bloom et al. (2020) find that management interventions had long-term productivity-boosting effects on Indian weaving firms. Giorcelli (2019) uses a quasi-experiment of the Marshall Plan when some Italian firms trained their management staff and received technological transfers from the US. She finds that the performance of firms that had the management training was higher for at least fifteen years after the intervention, while receiving new machines had only a short-run benefit. We complement this result showing that, first, "connected" firms displayed more advanced management practices on observable dimensions and, second, that productivity spillovers to "non-connected" firms likely occurred through the adoption of unobserved management practices.

In addition, we contribute to the growing empirical literature on the industrialization of the Russian Empire viewed from a micro-level perspective. Gregg (2020) demonstrates a strong association between corporate form and productivity; and Gregg and Nafziger (2019, 2024) show that Russian corporations demonstrated dynamism and flexible financial strategies despite operating in a rigid institutional environment; Gregg and Matiashvili (2022) show the role of factory operating time with respect to industrialization and growth. We show that cultural and ethnic origin of a firm

owner played an important role in operating decisions, such as capital investments, working days and organizational form, which contributed to industrialization prior to 1914.

In the following section, we briefly overview the historical setting in which these firms operated. Next, we introduce the dataset and describe the measurement of our key variable, i.e. the ethnic origin of a firm owner. We then evaluate how differences in productivity between firms reflect owners' cultural connections with the technological frontier, provide evidence for possible mechanisms of this correlation, and rule out alternative explanations for the observed patterns.

2 Historical background

The Russian Empire at the end of the 19th century was a multicultural and multiethnic society. According to the 1897 population census, those who listed Russian as their first language did not constitute a majority of the empire (44.3%). Over 40% were those who we classify as "connected" to Western Europe, Polish, Jewish (Hebrew, Yiddish, other), and German language speakers. Of these, Jewish and German language speakers constituted 4.03% and 1.43% each.

At the same time, the Russian Empire in the late 19th century was a sprawling and institutionally rigid polity undergoing a slow and uneven transition toward industrialization. While industrial output grew rapidly during this period, especially in textiles, mining, and metallurgy, the Empire remained behind Western Europe in terms of productivity, technological complexity, and infrastructure (Gerschenkron, 1962). The state played an active role in promoting industrialization, culminating in the ambitious economic policies of Prime-Minister Sergey Witte in the 1890s. The Ministry of Finance led efforts to systematically collect data on the country's manufacturing performance, resulting in the 1894 census of industrial establishments, which was one of the most detailed firm-level industrial surveys conducted in a relatively lagging country during the 19th century (Ministry of Finance, 1897).

In the late 19th and early 20th centuries, the Russian Empire also experienced a significant influx of foreign entrepreneurship. McKay (1970) shows that foreign investment in Russian industrial corporations surged from 17% in 1880 to 47% by 1914. In the critical industries of steel production and coal mining, foreign entrepreneurs introduced advanced technologies that significantly outpaced those used by local firms. For example, Belgian and French coal mining firms and German metallurgic companies adopted continuous production processes, vertical integration, and the latest coking and smelting technologies. These innovations allowed them to achieve greater efficiency and scale compared to domestic competitors. These firms did not merely replicate Western techniques; they adapted them to local conditions, facilitating a crucial technological transfer in a late-industrializing economy.

In addition to technological innovations, foreign entrepreneurs in the Russian Empire also introduced advanced forms of industrial management. Foreign-owned corporations, particularly those

from Belgium and France, tended to adopt more modern managerial structures than Russian-owned enterprises, including the delegation of responsibility to professional managers rather than relying solely on owner supervision (McKay, 1970, pp. 158–161). These firms often prioritized technical expertise in hiring decisions and created specialized supervisory roles facilitating more efficient labor organization and oversight. Furthermore, the introduction of Western-style management helped standardize production practices and improve administrative efficiency.

This paper seeks to extend insights drawn primarily from observations of foreign-owned corporations to the broader universe of private, locally owned firms with more limited access to foreign capital. We also focus beyond "business immigrants" by including ethnic minorities who maintained cultural and commercial ties to Western Europe. By adopting a cross-industry perspective, we assess the pervasiveness of knowledge transfer mechanisms across the economy, rather than confining the analysis to a handful of capital-intensive sectors.

3 Data and measurement

3.1 Data sources

Manufacturing census. Our primary source of data is the 1894 Russian Manufacturing Census, conducted by the Imperial Ministry of Finance and published in 1897 (Ministry of Finance, 1897) originally digitized by Gregg (2020). The census only includes manufacturing firms and excludes mining, farming, fishing and other non-manufacturing industries. Its purpose was to inform the central government about the industrialization of the Empire and the spatial distribution of industrial activities. The census did not inform tax collectors, so the usual concern of under-reporting in such surveys does not apply to these data. A factory was included in the census if at least one of the following conditions was met: it had at least 15 employees or had a steam engine, a steam boiler, or other mechanical device. The census includes a large number of variables describing a factory's operation: owners' first and last names, location, types of products the firm produced, number of employees by sex and age groups, number and types of machine, horsepower per machine, total revenue, and others. The total number of firms covered by the census was about 16,000.

Jewish surnames. We use "A Dictionary of Jewish Surnames from the Russian Empire" by Beider (2008), which includes all last names used by Jews of the Russian Empire, for a total of about 74,000 last names.[4] We use this source to distinguish Jewish last names on the one hand, and Russian or German last names on the other, since many Jewish last names had either Russian or German last name structure (see the following subsection for the ethnicity assigning algorithm).

Chicago World's Columbian Exposition. The Russian Empire presented 1,022 exhibits at the 1893 Chicago World's Columbian Exposition, which included art, inventions, and, notably, factory output and machines. The Ministry of Finance issued a catalog of all Russian exhibits, government- or

[4]We thank the author, Alexander Beider, who kindly shared the data.

privately owned (Ministry of Finance, 1893). If an exhibitor represented a factory, a short blurb describing the factory was attached. The blurb usually included the owner's name, founding year, number, type, and machine power of any engines in the factory, number of employees, and export-import operations. We digitize this information for all 262 firms in the catalog. We were able to match 122 of them to the 1894 manufacturing census using the firm's name, location, and industry. The rest of the firms would not have met the census inclusion criteria due to their size or production technology.

Patents. We use patent data to explore the differences in inventive activities between "connected" and "non-connected" ethnic groups. The Russian Empire had been issuing patents since 1812. Similarly to Western Europe and the US, inventors would apply for patents for their inventions, which would receive protection if deemed eligible. The duration of the patent could last from one to ten years, depending on the fee paid by the inventor. We digitize a comprehensive list of all patents issued in the years 1892-1896 from the official source published by the patent office of the Department of Trade and Manufacturing (1897). The list includes the assignee's name, citizenship, the nature of the patent, the year of the grant, and the patent term. While there are very few firm owners from the census that we can match to the patent database, the data provide additional evidence on innovative activity by various ethnic groups.

Population census. We augment the manufacturing census with data from the 1897 population census. The population census, among others, provides information on the share of each ethnic group in a district, allowing us to differentiate the districts by their ethnic labor and customer bases. In addition, we use data on literacy rates to isolate the effect of human capital on firm productivity.

3.2 Measurement: assigning ethnicities to firms

The primary measurement challenge of our study is to correctly assign 9,705 unique firm owners' surnames to a particular ethnic group. The assignment would enable us to classify owners into two broad categories: 1) "Connected" – ethnic groups with a large number of their representatives in European countries, such as Germans, Jews, and Poles; and 2) "Non-connected" – the rest of the sample, which include ethnic Russians and ethnic minorities of the Empire (Tatars, indigenous people of the Caucasus, etc.). Recognizing that the assignment can never be perfect, we combine several complementary approaches to construct the measure.

First, we use the comprehensive dictionary of all Jewish surnames in the Russian Empire, compiled by Beider (2008), to identify Jewish-owned firms. Separating Jewish surnames from Russian ones is important, since we classify Jews as a "connected" group (having a large diaspora in Europe), and Russians as "non-connected". Further, we would like to separate Jewish and German entrepreneurs, since these groups enjoyed different legal rights in the Russian Empire. This separation is never perfect, since Beider's dictionary contains a number of mixed-use surnames. For example, "Levin" could be a Jewish or a Russian surname, and "Schwartz" could be Jewish or German. However, we

believe this does not pose a serious threat to our estimation. By mistakenly classifying some Russian surnames as Jewish, we bias our coefficients of interest downward against our hypothesis. By doing the same with Germans, we are shrinking the German sub-group within the "connected" group. In either case, we are not inflating the coefficient on the "connected" group in our regressions. The matching between the dictionary and the manufacturing census shows that 27% of the unique surnames in the census were Jewish, corresponding to 35% of all firms.[5]

Having separated Jewish surnames, we are left with 7,119 unique non-Jewish surnames. We then use the fact that several ethnic groups in the Empire had characteristic last name endings, e.g., *"-ko"* for Ukrainian, *"-ok"* for Belarussian, *"-shvili"* for Georgian, *"-yan"* for Armenian, etc. We sort them out manually. Next, we sort out Russian-sounding surnames using common surname endings like *-ov(ova), -ev(eva)*, and *-in(ina)*. Within the Russian-sounding category there are at least two subgroups: ancestral Russian surnames and "Russified" surnames, namely the surnames of ethnic minorities, like Tatars, who changed their surname endings to Russian sounding ones in order to assimilate with the Russian majority. To distinguish between these two groups, two human trainers manually classified a part of the sub-sample. We then fitted a machine-learning algorithm (logistic regression), which produced high-accuracy results (94%).

The rest of the sample consists of non-Russian-sounding surnames, including German, French, Polish, Baltic peoples', and other European surnames. The goal was then to separate German-sounding surnames. Again, a part of this subsample was manually classified, and the machine-learning algorithm was then applied to the remaining list. Accuracy was realistically high - more than 73%. Having sorted out German surnames, we code the rest of the non-Russian-sounding sub-sample as "Other Connected" groups.

Thus, the whole sample of firms was divided into five mutually exclusive groups: Jewish firms (35%), Russian firms (34%), Other Connected firms (19%), German firms (12%), and Eastern minorities' firms (1%). The remaining 1% (109 firms) had mixed connected ownership (e.g. German and Other Connected). We assign "Other Connected" classification to this group.[6] In the empirical analysis that follows, we first compare the "connected" firms (German, Jewish and Other Connected firms combined) with "non-connected" firms (Russians and Eastern minorities combined). We usually repeat the analysis splitting the "connected" group into German, Jewish, and Other connected firms, and leaving Russians and Eastern minorities combined as a baseline comparison group.

[5] In addition to Beider (2008), we use the Jewish Name Index from Abramitzky et al. (2020) as a second classification method. Results obtained with this source are very similar.

[6] An additional challenge in our ethnicity assignment algorithm was the fact that some firms belonged to several individuals with different surnames. In rare cases, one owner would be "connected" while another would be Russian. We code these firms against our main hypothesis, such that all firms with at least one "connected" owner are coded as "connected". Similarly, every firm with no "connected" owner and at least one Russian owner is coded as Russian.

3.3 Descriptive statistics

Table 1 presents summary statistics for the whole sample. Connected firms accounted for about 65% of all firms in the Census, while non-Connected accounted for the remaining 35%. An average firm employed about 40 workers, generated about 1,700 rubles of revenue per year, 46% of firms had at least one machine, and 33% had a steam engine. About 44% were located in urban areas. Machine power was heterogeneous across firms, with a mean (20) and median (0) far apart, particularly due to many firms with no machine power.[7]

Table 1: Summary Statistics

	Mean	Median	Std. dev.	N
Firm ownership				
Connected firms	0.65	1	0.48	10,380
German	0.12	0	0.32	1,890
Jewish	0.35	0	0.48	5,492
Other Connected	0.19	0	0.39	2,998
Non-Connected firms	0.35	0	0.48	5,500
Russian	0.34	0	0.47	5,327
Eastern minorities	0.01	0	0.11	173
Productivity measures				
Revenue per worker, rubles	1743	973	2968	13,645
Revenue TFP	-0.013	0	1.1	13,645
Machine power, horsepower	20	0	76.14	15,880
Has at least one machine, %	46	0	50	15,880
Has at least steam engine %	33	0	47	15,880
Other firm's characteristics				
Number of employees	40.4	14	115.8	15,880
Women employees, %	7.5	0	17	15,880
Age of a firm, years	16.7	11	19.8	15,880
Urban location, %	43.8	0	49.6	15,880
Number of working days	215	240	79.9	15, 609

Notes: Data from the 1894 Manufactures Census of the Russian Empire (Ministry of Finance, 1897). Ethnicity of firm owners assigned according to our algorithm described in Section 3.2. Revenue TFP is derived from a Cobb-Douglas production function by regressing log revenue on log total employment and log total horsepower generated by machines and horses (+1). Machine power and presence of machine power exclude any power generated by animals.

Exploring one firm in detail allows us to understand the data better. A firm called "Solitermann" was a machine-intensive, high revenue-per-worker firm in the census. First, we focus on the firm owner's last name: Solitermann. This suggests that the firm was (1) a single proprietorship, since we do not have multiple last names or a plural of the same last name and (2) was most likely owned by a Jewish entrepreneur. Thus, the firm's name reveals both its legal status and the likely ethnicity of the owner.

[7] Given the prevalence of zeros, we always use Poisson Pseudo-Maximum-Likelihood estimation in the specifications where machine power is the dependent variable.

"Solitermann" was located in the city of Bratslav in the Podolsk province and operated in the flour production sub-industry within the food processing industry. It had one water wheel, which generated 75 horsepower in energy. It employed 14 workers (no women or children), was founded in 1888, operated for 240 days in 1894, and generated 65,202 rubles in revenue, resulting in 4,657 rubles per worker in revenue. A median firm in the census was somewhat less successful than "Solitermann". It employed 14 people (usually only men) and was slightly more likely to be located in a rural area. It operated in a relatively simple industry, like textiles or food processing, and had no horsepower units from machines. It was open for 240 days a year, suggesting it was most likely closed for the summer agricultural season, and generated 973 rubles per worker in revenue.

In addition to revenue per worker and machine power, we use revenue TFP as our productivity measure. We define revenue TFP as the residual of the simple Cobb-Douglas production function, with log capital (+1) defined as the log of a factory's horsepower[8] and log labor defined as the log of the total number of employees. Here we face a common challenge in industrial literature, where we observe factory-level revenues but not the quantities produced. Thus, our definition of TFP represents revenue (TFPR) rather than output (TFPQ) total factor productivity. While this measure may have flaws due to the underlying heterogeneity in prices, our data only allow us to measure TFPR. Crucially, Hsieh and Klenow (2014) show that TFPR rises with TFPQ, although by less than one-for-one.[9] That is, inasmuch as we find between-factory TFPR differences, these differences may understate true productivity differences based on TFPQ.

Table 2: Firms by industries and owner' ethnic origin

	Connected			Not Connected		
	Jewish	German	Other Connected	Russian	Eastern	Total
All industries	0.35	0.12	0.19	0.34	0.01	15880
Food processing	0.32	0.13	0.25	0.30	0.01	6843
Animal products	0.35	0.05	0.08	0.51	0.01	1668
Metals and Machines	0.35	0.17	0.19	0.28	0.01	1387
Mineral extraction	0.35	0.11	0.16	0.37	0.01	1181
Wool	0.44	0.20	0.16	0.20	0.00	1006
Wood	0.37	0.13	0.18	0.30	0.01	947
Paper	0.39	0.13	0.21	0.26	0.02	756
Chemical	0.37	0.10	0.14	0.38	0.02	627
Cotton	0.32	0.07	0.07	0.52	0.03	479
Mixed Materials	0.38	0.13	0.14	0.34	0.01	402
Flax	0.36	0.01	0.05	0.57	0.01	328
Silk	0.26	0.04	0.05	0.53	0.12	256

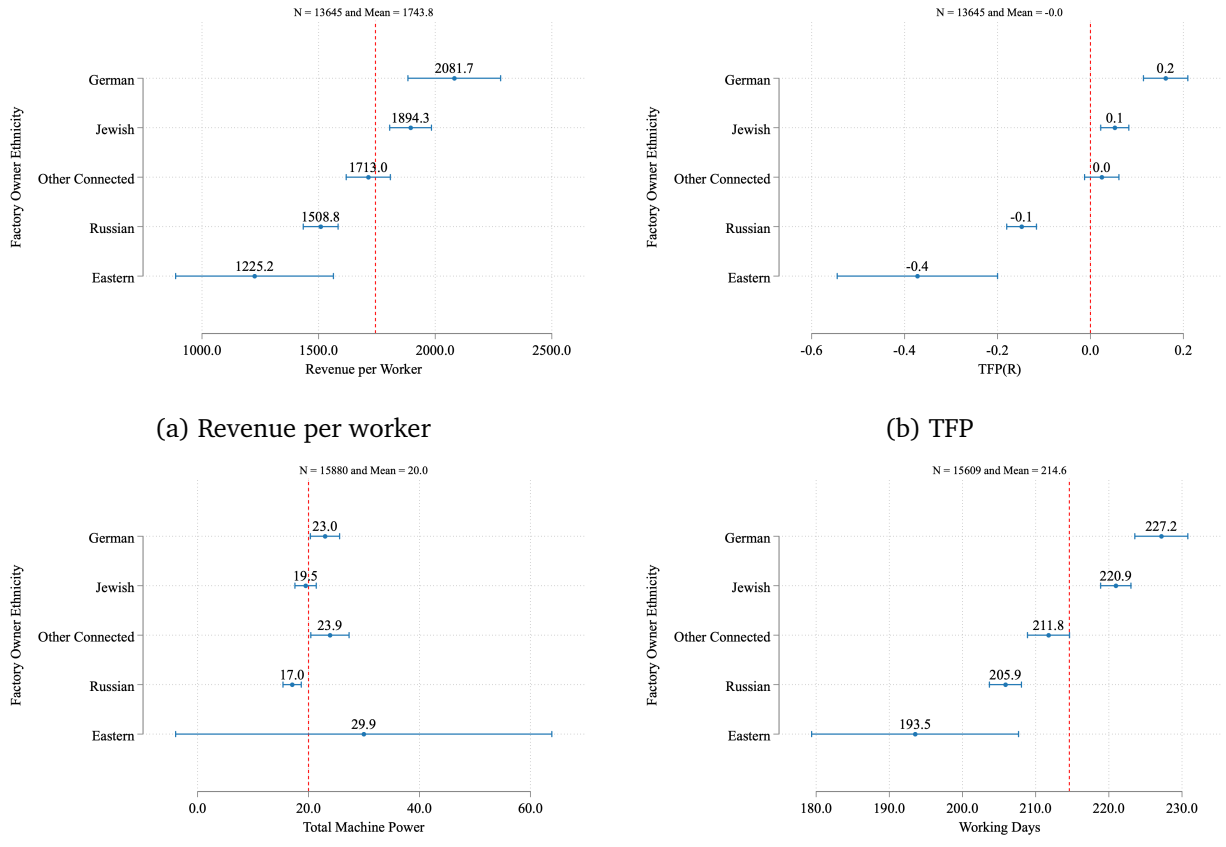
Notes: Data from the 1894 Manufactures Census of the Russian Empire (Ministry of Finance, 1897). Ethnicity of firm owners assigned according to our algorithm described in Section 3.2.

[8]Unlike in the rest of the paper, we use the horsepower generated by both machines and animals to account for all capital when defining TFP.

[9]Between-country variation in how much TFPR responds to rises in TFPQ varies from as low as 10 percent in the US to 50-60 percent in Mexico (Hsieh and Klenow, 2014).

Table 2 splits the sample by industry and ownership ethnicity. The upper row shows the share of each group in the whole sample. The following rows show the same shares separately for each industry. While Russian firms constituted 34% of all firms, they were over-represented in animal products processing, cotton and flax production, and mineral extraction. German firms accounted for 12% of all firms, they were over-represented in the production of machines and metals, and underrepresented in animal product processing and textiles (cotton, flax and silk production). The most striking observations concern Jewish entrepreneurs. First, they were heavily over-represented in the manufacturing census – while Jews constituted less than 5% of the population in the Russian Empire, they accounted for 35% of industrial firm owners. Second, they had an approximately even presence in all industries, with the share range between 25% and 44% of firms in each industry.

Next, we look at the differences in productivity measures across different ethnic groups. Figure 1 plots firms' characteristics for each of the five groups we identified with our ethnicity assignment algorithm.



(a) Revenue per worker

(b) TFP

(c) Machine power

(d) Working days

Figure 1: Outcomes by ethnic groups

Notes: The vertical red line represents mean values of each variable. TFP is calculated as a residual after regressing log revenue per worker on log number of workers and log machine power. Ethnic groups are defined according to the algorithm described in Section 3.2.

We observe that German firms were leaders in terms of revenue per worker, TFP, and the number of working days per year. Jewish firms were significantly higher than average in revenue per worker, TFP, and working days, but not in machine power. Other Connected firms (Polish, Baltic peoples) were higher than average in terms of machine power and close to the average in revenue per worker, TFP, and working days. Russian and Eastern minority firms were below average in all productivity measures and working days per year, although the Eastern group had wide confidence intervals for the machine power measure (see also Figure A1 for outcome distributions and Figure A2 for spatial distributions of connected firms and TFP). In the following section, we assess these patterns using regression analysis.

4 Cultural links and firm-level outcomes

Our baseline regression is an OLS model of the form:

$$Y_{idj} = \beta_0 + \beta_1 \times Connected_{idj} + \gamma_d + \delta_j + \nu X_{idj} + \epsilon_{idj} \tag{1}$$

where Y_{idj} are various outcomes, such as log revenue per worker, TFP, and machine power per worker, for firm i in district d and industry j. $Connected_{idj}$ is a dummy for a firm with owners connected to European countries. In an alternative specification, we subdivide connected firms into Germans, Jews and Other Connected to estimate separate coefficients for each group. In these cases, we use three dummies, namely Germans, Jews and Other Connected as explanatory variables. In both specifications the comparison group is the same – firms owned by ethnic Russians and Eastern minorities. All specifications include district-level fixed effects, γ_d, and industry and sub-industry fixed effects, δ_j, to ensure that we compare firms operating in the same market.[10]

In all specifications, we control for the logarithm of factory age, the number of employees, a dummy for urban location, and a dummy for a noble surname.[11] These are represented by vector X_{idj}. Controlling for factory age allows us to mitigate potential concern that younger factories would be more likely to be connected and to invest in a newer technology than incumbents, due to technological path dependence for older factories. Controlling for the number of employees ensures that the sheer differences in scale of employment are not responsible for "connected" firms' advantage; we do not have a prior on this coefficient. Controlling for urban location allows us to mitigate agglomeration effects, access to larger labor markets, and larger potential demand for a firm's products, all of which might be correlated with the "connected" status. Finally, controlling for surname nobility could partially mitigate differences in starting capital. The following subsections present the results for different outcome variables.

[10] We use industry classification as presented in the 1894 manufacturing census. The census also reports sub-industries (or primary activities) for each firm. On average, each of the 12 industries had about 40 sub-industries. For instance, 55% of the firms in the paper industry reported their primary activity as typography, and 11% as paper production. For this reason, in the regression analysis we introduce fixed effects for sub-industries in addition to industry fixed effects.

[11] Noble surnames identified from the General Armorial of the Noble Families of the Russian Empire by 1917 (Noble Assembly of the Russian Empire, 1917)

4.1 Productivity

Table 3 presents the main results for firms' productivity measures as dependent variables. In columns 1-3 we compare "connected" firms with the rest of firms in our sample. Connected firms had higher revenue per worker, higher total factor productivity, and more machine power. The economic effect of the owner's origins is substantial – "connected" firms had, on average, about 8% higher revenue per worker, about 7% higher total factor productivity, and about 8% more machine power. The effects are comparable to adding one more log point of labor or half the size of locating in a city (see Appendix Table B1 for coefficients on all control variables).

In columns 4-6, we decompose connected firms into three subgroups – Germans, Jews, and Other Connected, leaving the non-connected subgroups (Russians and Eastern minorities) as a baseline. German- and Jewish-owned firms exhibited larger productivity in terms of both log revenue per worker and TFP. While the coefficients do not attain 5% significance for any groups for machine power, the coefficient for the Jewish owners is sizably smaller than the other groups, whose coefficient is large but imprecisely estimated. Thus, their productivity premium may have stemmed from more efficient factor allocation or managerial practices. Other Connected groups (Poles, etc.) had a smaller advantage, but the coefficients are positive in all specifications.

Table 3: Connected firms and productivity

	log Revenue per worker (1)	TFP (2)	Machine Power (3)	log Revenue per worker (4)	TFP (5)	Machine Power (6)
Connected	0.084***	0.072***	0.079*			
	(0.023)	(0.022)	(0.043)			
German				0.122***	0.085***	0.091
				(0.032)	(0.031)	(0.083)
Jewish				0.085***	0.079***	0.055
				(0.024)	(0.024)	(0.047)
Other Connected				0.063**	0.045*	0.123*
				(0.027)	(0.026)	(0.069)
Observations	13,451	13,451	14,808	13,451	13,451	14,808
Mean dep. var.	6.851	-0.016	21.253	6.851	-0.016	21.253
R-squared	0.492	0.446		0.492	0.446	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at district level. The outcome variable in columns 1 and 4 is log revenue per worker; columns 2 and 5 is residual TFP from a Cobb-Douglas production function of regressing log revenue on log total employment and log total machine power (+1). Columns 3 and 6 estimate Poisson pseudo-maximum-likelihood model, with the total machine power as the outcome, due to prevalence of zero values in the outcome variable. Note that the number of observations increases in columns 3 and 6. This is due to missing revenue data for 2,235 firms in the original source net dropped 1,066 singleton observations, with overlap between the two. Controls are log factory age, a dummy for whether a factory is located in a city, log number of employees, and whether a last name is included in the Noble Assembly of the Russian Empire (1917).

The magnitude of the coefficients on revenue per worker and TFP is highest for the German- and Jewish-owned firms, suggesting that the advantage of the connected firms, documented in columns 1-3, is mostly driven by these two subgroups. Overall, the results provide strong statistical evidence on the importance of an owner's cultural background for a firm's productivity.

We further test the robustness of these estimates to several specification and sample choices: 1) only looking at firms with more than 15 employees (Table B2); 2) dropping Jewish-owned firms (Table C4); 3) separating Russian-sounding and non-Russian sounding Jewish owners (Table 9). In the latter exercise, we find that the Jewish advantage stems entirely from non-Russian sounding Jewish surnames, who would be more likely to have connections with Western European Jewish diasporas, strengthening our cultural connection argument.[12]

4.2 Machinery type

Next, we explore the proximate sources of the productivity advantage of connected firms, testing if they were more likely to possess advanced machinery. According to the manufacturing census, 41% of the firms did not have machines and employed only manual labor. Another 13% used animal power, usually horses, in addition to human labor. The rest of the sample had at least one mechanical device: steam engine (33%), water or wind wheel (14%), or a more advanced type of machinery such as locomotive, gas, kerosene, or benzine engine (11%). Only three firms used an electric motor, one of the most advanced industrial machines at the time. Based on these data, we generate three dependent dummy variables: 1) Has at least one machine; 2) Has at least steam engine; 3) Has above median machine power per engine, conditional on possessing a steam engine. The first takes the value 1 if a firm has any mechanical device (46% of the sample). The second takes the value 1 if a firm has a steam engine, locomobile, or other type of advanced machinery (44%). The difference between the two is that the second does not include firms with water or wind wheels as power sources. It is notable that these two groups are almost identical, meaning that by this time, firms that had any machine-powered energy source were likely to have a steam engine or a more advanced machine. The third takes the value 1 if a firm's machine has a horsepower above the median level in the whole sample. This variable captures the variation in intensive margin, which was substantial. In our sample, the average horsepower per machine ranged from 1 to 873.

Table 4 presents the estimation results. Columns 1-3 show that connected firms were 1.8% more likely to employ mechanical power sources in their production process and 2.6% more likely to have advanced machinery. They were also 3% more likely to have above median horsepower per machine. In columns 4-6, we again split the connected sample into three groups and repeat the analysis. It appears that the "machine advantage" of connected firms was largely driven by entrepreneurs of German origin. German and Other connected firms were more likely to have machine power, to use steam power, and to have more modern machines, as captured by higher power per machine, although the coefficient for German-owned firms is more than two times higher than for the Other Connected firms.

[12] In Appendix C, we explore the spatial distribution of Jewish-owned firms in the Pale of Settlement.

Table 4: Connected firms and machine types

	Has at least one machine (1)	Has at least steam engine (2)	Above median power per machine (3)	Has at least one machine (4)	Has at least steam engine (5)	Above median power per machine (6)
Connected	0.018**	0.026***	0.030*			
	(0.008)	(0.009)	(0.017)			
German				0.063***	0.069***	0.059***
				(0.013)	(0.013)	(0.020)
Jewish				0.009	0.019**	0.015
				(0.008)	(0.009)	(0.019)
Other Connected				0.024**	0.028**	0.047**
				(0.011)	(0.011)	(0.019)
N	15,673	15,673	6,726	15,673	15,673	6,726
Mean dep. var.	0.463	0.436	0.619	0.463	0.436	0.619
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: The outcome variable of columns 1 and 3 is whether a firm owned any engine type. Columns 2 & 4 are limited to only those that used at least steam (but could not have steam, but have more advanced engine types, such as gas, petrol, electricity). Controls are as in Table 3.

The results for Jewish firms are intriguing. They suggest that their productivity advantage stemmed not from capital investments but rather from non-tangible assets or management practices, which we explore in the next subsection. In general, Table 4 confirms our hypothesis that "connected" owners adopted modern technology at higher rates than non-connected firms.

4.3 Management practices

Differences in productivity could also stem from variations in management practices. Management practices is a blanket term that refers to decisions not directly related to the amount of capital and labor involved in production. Below, we investigate a few possible measures that can be attributed to management practices.

One facet of industrialization in the Russian Empire was the rising employment of women, which was previously largely limited to home production and agriculture.[13] The manufacturing census allows us to determine whether a firm hired female labor, and the share of women among the firm employees. The third measure of management practices is the number of days per year that a factory operated. Firms that stayed open more days per year left capital idle for less time and generally resembled the industrial factory-like operation more than seasonal manufacturing firms did.[14] They also seemingly succeeded in retaining labor through agricultural high seasons. Finally, we explore whether connected firms were more likely to form partnerships, namely have more than one owner with different surnames. Forming a partnership could introduce diverse expertise, thus creating more opportunities for innovation.[15]

[13] See Goldin and Sokoloff (1982) on the role of women labor in the industrialization of the US North-East states.

[14] See Gregg and Matiashvili (2022) for a discussion of the relationship between working days and factory-like operation in the late Russian Empire.

[15] Lamoreaux (1995), Lamoreaux and Rosenthal (2005), Artunç and Guinnane (2019), Gregg (2020), among many others, demonstrate the importance of ownership type on firm performance.

Although the number of partnerships in the data is small (only 690 firms) it is disproportionately concentrated among the connected firms as 89% of the partnerships had at least one connected owner.

Table 5: Connected firms and management practices

	Employs Women (1)	Proportion Women (2)	log Working Days in 1894 (3)	Partnership (4)	Employs Women (5)	Proportion Women (6)	log Working Days in 1894 (7)	Partnership (8)
Connected	0.014*	0.003	0.022**	0.041***				
	(0.008)	(0.003)	(0.010)	(0.004)				
German					0.014	0.007	0.036**	0.046***
					(0.011)	(0.005)	(0.015)	(0.008)
Jewish					0.012	0.001	0.025**	0.032***
					(0.009)	(0.003)	(0.011)	(0.004)
Other Connected					0.019*	0.007*	0.011	0.064***
					(0.010)	(0.004)	(0.012)	(0.008)
N	15,687	15,687	15,414	15,687	15,687	15,687	15,414	15,687
Mean dep. var.	0.242	0.075	5.254	0.043	0.242	0.075	5.254	0.043
R-squared	0.560	0.508	0.465	0.121	0.560	0.509	0.465	0.123
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Outcome variables: in columns 1 & 5 is whether a firm employs women; in columns 2 & 6 is what proportion of the workforce is women; in columns 3 & 7 is the log number of days the firm was open; in columns 4 & 8 is whether a firm was a partnership. Controls as in Table 3.

Table 5 presents the results. "Connected" firms were more likely to employ women (column 1), but did not employ a statistically significantly larger proportion of women (column 2). "Connected" firms operated more days per year (column 3) and were more likely to operate as a partnership (column 4). Next, we again look at the decomposition of the connected group into different subgroups. As discussed in baseline results, we found that Jewish-owned firms had higher TFP and revenue per worker than Russian and Eastern groups, but did not seem to have more capital. Here, we see a partial answer to the riddle. Jewish-owned firms were more likely to work more days during the year (column 7), which would have generated higher revenue per worker. They were also more likely to form a partnership (column 8), which would have offered more diversified expertise in management. German-owned firms exhibited similar or even stronger patterns. Finally, while the Other Connected firms did not operate longer, they were more likely to employ women and form partnerships.

5 Establishing the knowledge transmission mechanism

Having established baseline differences in productivity, capital, and management practices between connected and non-connected firms, we next explore the mechanisms behind this variation. We conjecture that "connected" firms enjoyed lower transaction costs in the adoption of Western technical and organizational know-how. However, testing this hypothesis is challenging, since the 1894 manufacturing census does not contain data on firms' foreign trade operations or any information on contacts with foreign companies. Hence, we employ additional data sources and combine them with the census to test the knowledge transmission mechanism to the extent possible. The following four subsections describe these empirical exercises.

5.1 1893 Chicago World's Fair

The Chicago World Fair (also known as World's Columbian Exposition) was held in Chicago in October 1893, representing technological and cultural advancements of industrial nations in the late 19th century. The Russian Empire had 1,022 exhibits from eleven different exhibition groups. While most of the exhibits were inventions or individual art pieces, approximately a fifth were private firm exhibits demonstrating firms' products. The Russian Ministry of Finance published a catalog of all exhibits, including those of private firms, with blurbs describing each firm in detail: industry, founding year, output, revenue, etc. (Ministry of Finance, 1893). We use this catalog to explore the connections of Russian industrial firms to the West.

We were able to match 122 firms from the exhibition catalog to the 1894 manufacturing census. This sub-sample of firms is heavily biased toward more productive and innovative ones since they were selected by the Russian government to demonstrate the industrial progress of the Empire.[16] On the other hand, the advantage of these data is that we observe whether firms: (1) export their products abroad to any foreign country, (2) export their products to a European country, (3) import materials from any foreign country, and (4) import materials from a European country. The data also include the founding year, number of workers and urban location indicator, similar to what we have in the manufacturing census. In addition, the fair firms show sufficient variation in the cultural origins of their owners, as assigned by our algorithm: 37% Russian, 38% Jewish, 10% German, and 10% Other Connected. Hence, despite a highly selected sample, we can perform a similar regression analysis as before, but with import and export variables on the left-hand side.

Table 6 presents these results. Columns (1)-(4) show the results the full sample of firms at the fair. Columns (1) and (2) present suggestive evidence that Other Connected firms were more likely to trade abroad. However, we are particularly interested in trade with Europe, shown in columns (3) and (4). We see that German and Other Connected firms were more likely to engage in both import and export with Europe. The same patterns hold for the sample of matched firms in columns (5)-(8), where we only look at firms that matched to the 1894 census, although the statistical significance of these estimates is lower because of a smaller sample.

The results suggest that even among the most successful firms in the Russian Empire, there were differences in their trade behavior that reflected owners' cultural origins. Connected owners were significantly more likely to trade with Europe and benefit not only directly in the form of higher revenue, but also indirectly in the form of know-how transfer from more advanced economies. Although we were unable to determine which technologies were imported from the West, we can state that the technology transfer was very likely given the economic ties we observed in the trade data.

[16] Table D6 shows that firms participated in the fair had 27% higher revenue per worker, 19% higher TFP, and 19% more machine power (conditional on industry and district fixed effects) than their counterparts.

Table 6: Foreign Trade: Data from 1893 Chicago World Fair

	Full sample				Matched sample			
	Abroad		To Europe		Abroad		To Europe	
	Import (1)	Export (2)	Import (3)	Export (4)	Import (5)	Export (6)	Import (7)	Export (8)
German	0.062	0.180	0.106*	0.315**	-0.125	0.141	-0.042	0.338**
	(0.051)	(0.161)	(0.044)	(0.119)	(0.095)	(0.125)	(0.099)	(0.120)
Jewish	0.019	0.007	0.016	0.039	0.011	0.011	0.076	0.068
	(0.059)	(0.101)	(0.087)	(0.033)	(0.051)	(0.133)	(0.062)	(0.035)
Other Connected	0.148***	0.161	0.126***	0.153	0.247*	0.207	0.202	0.210*
	(0.037)	(0.205)	(0.018)	(0.133)	(0.104)	(0.275)	(0.101)	(0.092)
N	227	249	191	207	107	114	89	92
Mean	0.357	0.305	0.162	0.135	0.383	0.360	0.180	0.152
R-squared	0.251	0.139	0.120	0.167	0.249	0.120	0.127	0.245
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at the Industry Section level in parentheses. Panel A includes all firms that participated in the 1893 Columbian Fair. Panel B includes firms we were able to match to the 1894 Manufacturing Census. Import abroad (column 1) is defined as an indicator of whether a firm reports imports from abroad. Exports abroad (column 2) is defined as an indicator for whether a firm reports sales abroad. European export (column 3) is limited to firms that specify what countries they are importing from. The same logic applies to exports in column (4)

5.2 Comparative advantage of German entrepreneurs in the Russian Empire

If the connected firms' productivity premium originates from the transfer of know-how, then the premium should be most pronounced in industries where the origin country had a comparative advantage at the time. Of ethnic groups of European origin, we are best able to identify Germans. Hence, in this subsection, we test whether firms owned by German-origin entrepreneurs in Russia had the productivity advantage in the same industries as the German economy did in the late 19th century. To test this, we estimate the same Equation (1) separately for each industry, and see if we observe any matches between the two sets of industries.

A large body of literature in German economic history documents that, after the unification in 1871, German economy experienced an industrial boom "especially in machinery, chemicals, and electro-technical goods" (Hungerland and Wolf, 2022). Germany's increasing industrial might in heavy machinery and chemicals came to rival Britain as Europe's primary industrial nation in the decades preceding World War I (Mokyr, 1990). By 1900 the German chemical industry dominated the world market for synthetic dyes with three major firms (BASF, Bayer, and Hoechst) producing about 60% of the world supply of dyestuffs and selling about 80% percent of their production abroad (Kindleberger, 1975).[17]

[17] See also Malein (2021) on the impact of the German minority on the industrialization across Russian regions, and Natkhov and Vasilenok (2021) on the diffusion of agricultural technologies from German settlers to Russian peasants in the Middle Volga region.

Based on these accounts and given industry classification in the 1894 Russian manufacturing census, we expect that firms owned by entrepreneurs of German origin exhibited higher productivity in such industries as Chemicals and Metals and Machines, and possibly other industries that employed heavy machinery to process raw materials into final products. Figure 2 shows the estimated coefficients on the German dummy for each industry in our sample. Indeed, we find a large and statistically significant productivity premium in chemicals: German-owned firms in this industry had, on average, 81% higher revenue per worker than Russian-owned firms.

Consistent with the proposed mechanism, we also observe a positive and 10% statistically significant coefficient for Metals and Machines. Similarly, the positive coefficients for Wood processing, Animal products and Food processing imply that German firms in these industries probably benefited from imported machinery allowing them to process raw materials more efficiently. However, other factors, such as quality of local inputs, might counterbalance the machine advantage. Interestingly, the estimated coefficient for textiles (Cotton and Wool) is zero, and for Silk production is even negative. Hence, we do not observe any "German premium" for industries in which Germany had no comparative advantage at the time.

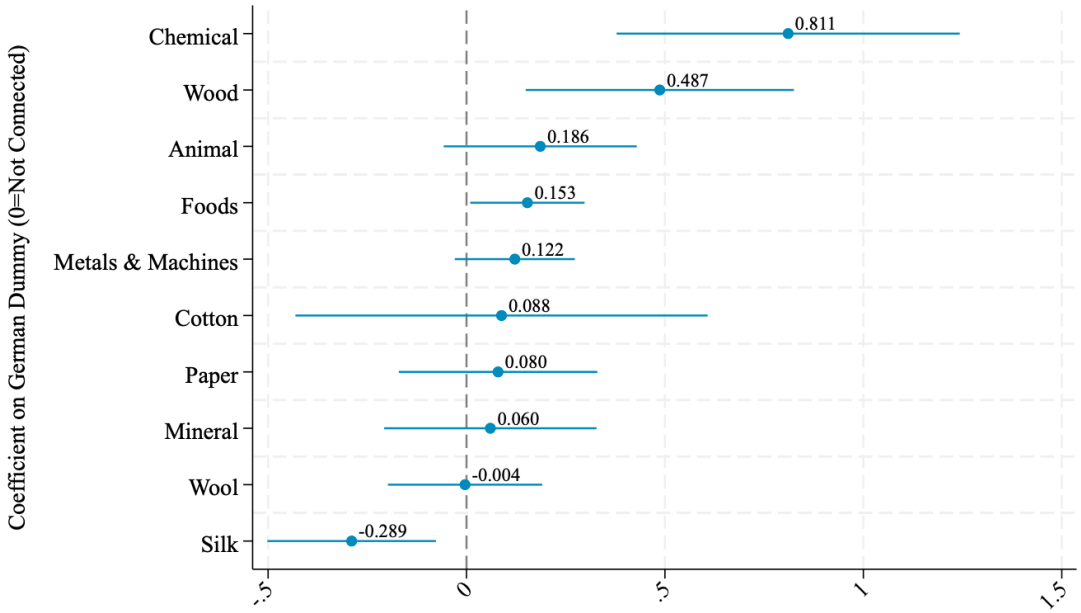


Figure 2: German productivity premium by industry (log Revenue per Worker)

Notes: The outcome variable in all specifications is log revenue per worker. The coefficients on the German dummy show the average difference in log revenue per worker between German-owned firms and non-connected firms (Russian and Eastern minorities) conditional on firm age, number of employees, urban location, and district and industry-primary activity fixed effects. Bars represent the 95% confidence interval. Firms classified by industry as in the 1894 manufacturing census. We omit Mixed Materials industry for being a nonspecific, catch-all group, with results difficult to interpret.

However, it is possible that the German advantage in chemicals was mainly due to Russian inferiority in it. In Appendix E.1, we look at all connected groups *except for* Germans. The coefficient on the chemical industry drops to zero, suggesting that the difference between connected and non-connected firms stemmed exclusively from the German comparative advantage.[18] Appendix E.2 presents qualitative evidence on the role of German-born entrepreneurs in the Russian chemical industry, showing that they indeed had direct connections to Germany, which enabled knowledge transmission.

5.3 Local spillover effects

In the previous subsections, we presented indirect evidence of technology diffusion from abroad. However, a significant share of innovation spillovers take place locally when less productive firms try to emulate a domestic "role model" in their industry. We then pose a question, did non-connected firms with higher exposure to "connected" firms operating in the same market adopt innovations at a higher rate than those with lower exposure? To answer this question, we calculate for each non-connected firm a continuous measure of "exposure to connected firms" as the share of labor employed by connected firms in a given district and industry, or more formally:

$$Exposure_{idj} = \frac{\text{Number of employees of connected firms }_{idj}}{\text{Total number of employees of all firms}_{idj}} \tag{2}$$

We then include this variable in the regression equation to compare the productivity measures of a non-connected firm in a given district-industry to other non-connected firms in a different district but the same province and industry. Hence, we estimate the following equation:

$$Y_{idj} = \alpha_0 + \alpha_1 \times Exposure_{idj} + \delta X_{idj} + \theta_j + \psi_p + \varepsilon_{idj}, \tag{3}$$

where Y_{idj} is an outcome for a non-connected firm i in district d in industry j; $Exposure_{idj}$ is an exposure measure for the same firm calculated as in Equation (2); θ_j is industry fixed effect, and ψ_p is province fixed effect.

Figure 3 presents the coefficient on the $Exposure$ measure for different outcomes. It appears that firms surrounded by more connected firms exhibited higher productivity measured by TFP and a higher likelihood of using machines.[19] Interestingly, we do not observe spillover effects for other outcomes such as working days, women employment or partnership status. Of all productivity measures, the largest coefficient is for revenue TFP. This probably implies that the spillovers of tacit knowledge and unobservable management practices were the most important part of the diffusion "package". In general, the results present compelling evidence of local productivity spillovers from firms with Western connections to firms without such connections.

[18] Appendix E.1 also presents Figure 2 in a table format.

[19] The coefficients from this exercise are an order of magnitude larger than in our baseline regression because the $Exposure$ measure is a continuous variable in contrast to the $Connected$ dummy.

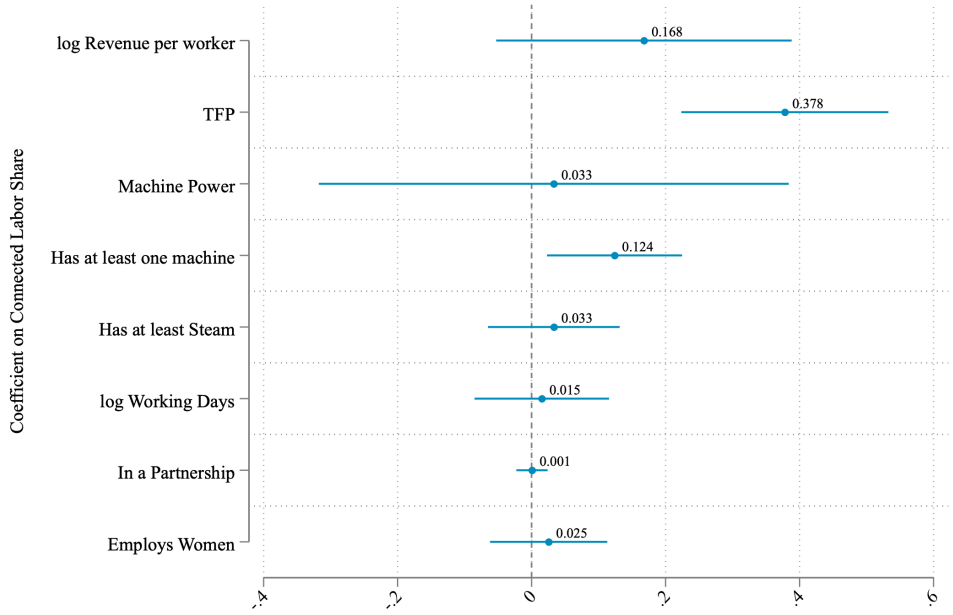


Figure 3: Spillovers from connected to non-connected firms

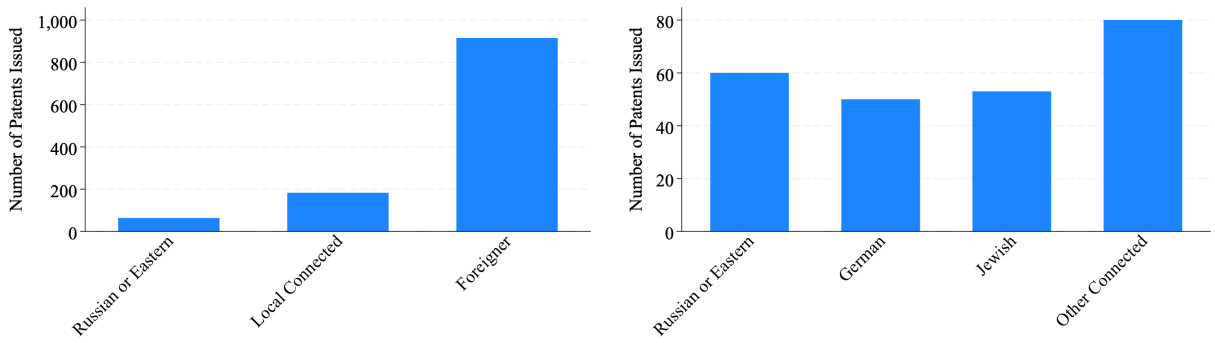
Notes: The coefficients on the share of connected firms show the average differences in productivity measures between non-connected firms with different exposure to connected firms conditional on firm age, number of employees, urban location, and industry-primary activity fixed effects. Bars represent 95% confidence interval.

5.4 Patents

Patents are often used as a proxy for innovative activity. We digitized the universe of patents issued in the Russian Empire in 1892-1896, around the time the manufacturing census was conducted. There were a total of 1,161 patents issued to individuals over this period.[20] Most of them, 915, were issued to foreign nationals (see Figure 4a), part of it due to cross-patenting from other nations. This in itself is a very telling fact.[21] However, even within the citizens of the Russian Empire, there remain stark differences in patenting by ethnicity. Of the remaining 246 patents, 50 had been assigned to Germans, 53 to Jews, 80 to other Western-sounding last names, 58 to Russians and only 2 to Eastern-sounding last names (see Figure 4b). The fact that Russians patented at similar levels as Germans is particularly striking, given that Russians accounted for 44.3% of the Empire's population, and Germans for only 1.43%.

[20] We drop patents issued to corporations.

[21] However, many patents assigned to foreign nationals may stem from international patenting, where the inventor resided in a different country but applied for a patent in many countries, as is common today.



(a) Number of patents assigned to foreigners and locals

(b) Number of patents assigned to locals, by ethnicity

Figure 4: Patents by citizenship and ethnicity, 1892-1896

Notes: The data come from the directory of all patents issued by the Russian Empire in the period 1892-1896 (Department of Trade and Manufacturing, 1897). (a) Foreigners are identified when referred to as "foreigner" or "foreign subject" or "subject of country X". Local Connected group is defined as having German, Jewish, or Other Connected last names. This classification, due to a small sample size, was conducted by the authors, and differences reconciled. Jewish last names were identified using Beider (2008)'s dictionary of Jewish last names in the Russian Empire.

In the Appendix, we show two additional pieces of evidence supporting the knowledge diffusion mechanism. In Appendix F, we show that the connected entrepreneurs benefited the most from locating in large cities, likely due to lower transportation and communication costs with Europe. We also find an insignificant but negative interaction of connectedness with distance to Europe, suggesting that shorter distances enabled the use of connections.

6 Eliminating alternative mechanisms

To isolate the knowledge transmission mechanism, we must rule out salient competing explanations. In this section, we present evidence that other potential mechanisms, such as initial wealth, education, ethnic bias in demand for a firm's products, or ethnic network effects in the supply of inputs and access to output markets, do not explain the "connected" firms' productivity premium.

6.1 Starting capital

The most salient competing mechanism is entrepreneurs' initial wealth that might correlate with cultural origins, enabling "connected" owners to invest in more and newer machinery from the beginning. While we do not have data on a firm's starting capital, we can exploit a large variation in capital per worker in our data to isolate the initial capital mechanism. In doing so, we compare log revenue per worker and TFP of firms that used no machines. If the productivity premium of the "connected" firms came mainly from larger starting capital, we should see no difference in productivity among firms without machines, controlling for the number of workers, and operating

in the same industry and same district. If, on the other hand, the productivity advantage persists even among firms without capital, we can at least tenuously conclude that something beyond initial wealth is responsible for the advantage. Hence, we estimate Equation (1) on the sub-sample of firms without machines.

Table 7: Sub-sample of firms without machines

	log Revenue per worker (1)	TFP (2)	log Revenue per worker (3)	TFP (4)
Connected	0.089***	0.089***		
	(0.031)	(0.031)		
German			0.156**	0.156**
			(0.061)	(0.061)
Jewish			0.079**	0.079**
			(0.033)	(0.033)
Other Connected			0.101***	0.101***
			(0.037)	(0.037)
N	5,350	5,350	5,350	5,350
Mean dep. var.	6.580	0.026	6.580	0.026
R-squared	0.576	0.575	0.576	0.575
District FE	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: Robust standard errors clustered at the District level. The sample is limited to the firms without any machines. Refer to Table 3 for details on outcome variables and controls.

Table 7 presents the results. The coefficient on log revenue per worker and TFP are roughly of the same magnitude as those in Table 3. The pattern of the coefficients is also similar to the whole sample estimation for the German and Jewish firms. The coefficient for the Other Connected group is almost twice as large, suggesting that this group's overall productivity advantage may largely stem from management rather than the capital margin, which we also observed in Table 5. It is also notable that there are no differences in the average number of employees between the connected and non-connected firms with no machines, also suggesting that these firms were similar along this margin that may reflect differences in starting finances. We cannot rule out unobservable advantages such as savings and access to higher-quality inputs.

6.2 Human capital

Another concern with our interpretation of the results is that we may be capturing the effect of a stock of human capital. It is possible and likely that the "connected" entrepreneurs were better educated. Perhaps it was due to this superior education that they were able to run more productive firms. First, this explanation does not hold in light of our findings with the German owners and chemicals industry. However, we present another exercise – we limit the sample to districts with high literacy rates ($\geq$ 75%, 6% of the sample) as captured by the 1897 population census. Table

8 shows that even within such a small and positively selected sample, the productivity premia for connected firms remain and are even higher than the baseline results.

Table 8: Literacy and Productivity Measures; districts with more than 75% literacy

	log Revenue per worker (1)	TFP (2)	Machine Power (3)	log Revenue per worker (4)	TFP (5)	Machine Power (6)
Connected	0.342**	0.346**	0.119			
	(0.143)	(0.137)	(0.233)			
German				0.317**	0.336**	0.065
				(0.147)	(0.152)	(0.189)
Jewish				0.368**	0.372**	0.180
				(0.155)	(0.147)	(0.257)
Other Connected				0.299*	0.292**	-0.040
				(0.148)	(0.132)	(0.182)
Observations	757	757	862	757	757	862
Mean dep. var.	7.215	0.198	13.629	7.215	0.198	13.629
R-squared	0.611	0.575		0.611	0.576	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at the District level. The sample is limited to the firms in districts with $\geq$ 75% literacy. Refer to Table 3 for details on outcome variables and controls.

6.3 "Connected" ethnic groups without connections

If our findings are truly due to connections to Europe and not some common characteristics within ethnic groups, we should not observe the productivity premium for groups that, although satisfying our definition of connected groups, did not truly maintain their connections with Europe. In other words, the "connected" ethnic groups were not homogeneous in terms of the strength of their connections to Europe. In particular, Volga Germans had lived in the Russian Empire since the late 18^{th} century in remote regions along the Volga river. They emigrated from German lands escaping religious persecutions and devastation after the Seven Years War. The Russian government granted them land along the sparsely settled frontier in the hopes that Germans will help to populate the area and bring advanced agricultural techniques. Thus, this group had lived in the depths of the Empire for over a century, focused mostly on agriculture. They were less likely to maintain strong ties with Germany. We expect this group not to have had the same advantage in manufacturing as the other groups of German origin. We define Volga Germans as those German owners who owned firms in one of the Volga basin provinces in the Empire. It is telling that only 29 of 1,053 firms in the Volga basin were owned by Germans. Germans owned 11.9% of all firms in the Empire, but only 2.7% of the firms in the Volga region.

The Jewish population of the Russian Empire also varied in terms of connections to Europe. In particular, small groups of Jewish entrepreneurs converted to Orthodox Christianity and changed their surnames into Russian-sounding ones to obtain a legal right to live and work outside the

Pale of Settlement. Although this group would still have enjoyed some benefits of connections through Jewish diasporas, arguably, on average, they were less likely to have direct connections to Europe than those who maintained their religious identity and their surnames. We separated our sample by Russian-sounding (Russian last-name endings such as *-ov*) vs. non-Russian-sounding Jewish surnames. Finding a difference in these two groups would suggest that it is indeed the connections, and not ethnic or religious identity per se, that matter for the diffusion of know-how. However, caution should be exercised when interpreting these results – given the measurement error present in the data, the Russian Jewish group will include some non-Jewish Russians, biasing this coefficient down. Similarly, the non-Russian Jewish group will include some Germans and other connected groups, biasing this coefficient up.

Table 9 presents the productivity results with decomposed German and Jewish groups. First, we focus on German firms. We find that the Volga Germans have consistently negative coefficients, confirming our hypothesis. We also find that the coefficients on non-Russian Jewish owners are considerably higher than the Russian Jewish owners, although the difference is statistically insignificant. Taken together, these results suggest that it is indeed the connections of these groups, rather than inherited wealth or human capital, that mainly drive our results.

Table 9: Connected groups without connections

	log Revenue per worker	TFP	Machine power
	(1)	(2)	(3)
Non-Volga German	0.143***	0.099***	0.121
	(0.037)	(0.034)	(0.093)
Volga German	-0.157	-0.158	-0.199
	(0.122)	(0.123)	(0.335)
Non-Russian Jewish	0.112***	0.094***	0.092
	(0.036)	(0.034)	(0.070)
Russian Jewish	0.068**	0.069**	0.020
	(0.027)	(0.027)	(0.052)
Other Connected	0.076**	0.053*	0.143**
	(0.031)	(0.028)	(0.069)
Observations	13,451	13,451	14,806
Mean depend. variable	6.851	-0.016	21.254
R-squared	0.492	0.446	
District FE	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓
Controls	✓	✓	✓

Notes: Robust standard errors clustered at the district level. The Russian Jewish group defined as Jewish surnames from Beider (2008)'s dictionary with Russian last-name endings, such as "-ov", "-ev", and "-in". Volga Germans defined as located in Astrakhan, Kazan, Samara, Saratov and Simbirsk provinces.

6.4 In-group bias

A potential confounder of our results is the fact that a lot of connected entrepreneurs lived in majority-connected districts. For instance, districts with a larger share of the Jewish population within the Pale of Settlement also had a higher share of Jewish-owned firms. These firms might have performed better because locals were more willing to purchase goods from their compatriots rather than from out-group entrepreneurs. We address this concern in Table 10, where we replicate Table 3 on a restricted sample of districts with more than 50% of ethnic Russians according to the 1897 population census. In these districts, non-Russian entrepreneurs would have had a smaller in-group advantage than Russian entrepreneurs, which whould introduce a downward bias in our estimates on connected firms dummy. Instead, the results are almost identical to Table 3, suggesting that they are likely not driven by in-group bias in non-Russian districts.

Table 10: Connected firms' productivity in districts with more than 50% Russian population

	log Revenue per worker	TFP	Machine Power	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.082***	0.073***	0.068			
	(0.023)	(0.022)	(0.044)			
German				0.096**	0.056	0.051
				(0.038)	(0.036)	(0.099)
Jewish				0.081***	0.077***	0.048
				(0.025)	(0.024)	(0.048)
Other Connected				0.079***	0.065**	0.116
				(0.027)	(0.026)	(0.076)
Observations	11,858	11,858	12,768	11,858	11,858	12,768
Mean depend. variable	6.820	-0.034	22.148	6.820	-0.034	22.148
R-squared	0.500	0.453		0.500	0.453	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at the District level. The sample is limited to the districts with majority Russian population. Refer to Table 3 for details on outcome variables.

6.5 Network effects

Similar to the in-group bias, but a somewhat distinct possible confounder to the knowledge diffusion mechanism is the ethnic networks mechanism – firms with same-ethnicity owners may be able to access cheaper input and larger output markets through ethnic networks, both locally and through diasporas in other provinces.[22] The network effect could also enhance knowledge diffusion – individuals of the same ethnicity may transmit entrepreneurial know-how within ethnic networks. Both of these mechanisms suggest a positive correlation between a firm's productivity and a share of firms with same-ethnicity owners in the same industry (but not the same primary activity - that

[22] Mejia (2024), for instance, finds that social networks mattered greatly during the industrialization in Colombia in the late 19th-early 20th centuries. Spitzer (2021) shows that Jewish emigration from the Russian Empire to the United States in same period exhibited a strong spatial diffusion pattern within social networks of relatives and friends.

is, direct competition) and district. Our data, however, do not allow us to reconstruct networks of firms and entrepreneurs. Instead, we repeat the analysis presented in equation (1) with an additional interaction variable reflecting the number of firms in the same district and industry owned by the same ethnicity-names as the i firm. This way, we proxy for upstream-downstream firms owned by the same-ethnicity entrepreneurs in a given market.

Table 11: Benefits of nearby upstream/downstream factories owned by same ethnicity

	log Revenue per worker		TFP	
	(1)	(2)	(3)	(4)
Share of the same ethnic group	0.054	0.224**	0.106	0.199**
	(0.114)	(0.093)	(0.105)	(0.088)
Jewish	0.309***	0.228***	0.268***	0.204***
	(0.075)	(0.064)	(0.069)	(0.061)
German	0.229***	0.253***	0.164**	0.179***
	(0.082)	(0.071)	(0.071)	(0.063)
Other Connected	0.217***	0.197***	0.193***	0.168***
	(0.073)	(0.067)	(0.066)	(0.062)
Share of the same ethnic group × Jewish	-0.492***	-0.225*	-0.400***	-0.188
	(0.155)	(0.129)	(0.142)	(0.122)
Share of the same ethnic group × German	-0.280	-0.195	-0.201	-0.116
	(0.204)	(0.148)	(0.190)	(0.143)
Share of the same ethnic group × Other Connected	-0.387**	-0.238*	-0.386***	-0.225*
	(0.163)	(0.138)	(0.148)	(0.131)
N	13,592	13,592	13,592	13,592
Mean depend. variable	6.660	6.660	-0.149	-0.149
R-squared	0.215	0.350	0.217	0.324
District FE	✓	✓	✓	✓
Industry and sub-industry FE		✓		✓
Firm Controls	✓	✓	✓	✓

Notes: Robust standard errors clustered at the District level. Nearby factories are defined as operating in the same district and industry. Refer to Table 3 for details on outcome variables.

Table 11 presents the results. Columns 1 and 3 present baseline results without industry fixed effects, while Columns 2 and 4 add industry fixed effects. The results suggest that having more firms in a district owned by the same ethnicity as one's own does not improve a firm's outcomes. If anything, it seems to be detrimental to productivity. All coefficients for the interaction terms are negative and often significant. These results suggest that the cheaper inputs or local networks story does not hold in this setting. More likely, having a higher share of upstream/downstream firms owned by the same ethnicity also translates to a higher share of direct competitors of the same ethnicity. If owners prefer to trade with same-ethnicity owners, this would create even higher competition within the ethnic enclaves, leading to negative interaction terms.

7 Conclusion

We document substantial productivity advantages of industrial firms with cultural connections to the technological frontier in the late 19th century Russian Empire. We show that these productivity advantages came from both the adoption of advanced machinery and superior management practices. We use additional data sources to show that more patents were awarded to "connected" ethnic groups, and "connected" firms were more likely to trade with Europe. "Connected" firms also exhibited higher productivity in the same industries as their European counterparts, i.e., the chemical industry for German-owned firms. In addition, we show that "connected" firms may have served as role models for local "non-connected" firms, inducing spillovers of best practices.

Our results provide new insights into the development paths of late-industrializing economies. Consider, for example, the contrasting experiences of the Russian and Ottoman Empires in the late 19th-early 20th century. Both had rigid economic institutions and were governed by autocratic political regimes. The Russian Empire, however, demonstrated impressive industrial growth of about 4% per year for almost three decades before World War I (Markevich and Nafziger, 2017), while in the Ottoman Empire industrialization remained very limited (Pamuk, 2019). Both empires were multi-ethnic, but only in the Russian case several minorities had direct cultural connections with Western Europe. These groups played a decisive role in the adoption of industrial technologies and advanced management practices. Hence, the presence of "connected" minorities in Russia compensated, to some extent, for institutional deficiencies and allowed for relatively faster growth of the private manufacturing sector. The Ottomans were deprived of that advantage, and industrialization before 1914 was mostly confined to the government-assisted construction of railways, while the private sector remained largely agrarian.

Unpacking the "black box" of cultural connections is the next step in studying technology diffusion. How do technologies and management practices spread through individual and family ties? What explains the variation in the extent of technology adoption among connected groups? How did best practices spread from high- to low-performing groups? These questions are worth exploring to advance our understanding of the sources of economic inequality around the world.

References

Abramitzky, Ran, Leah Platt Boustan, and Dylan Connor, "Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office," *National Bureau of Economic Research*, June 2020.

Artunç, Cihan and Timothy W Guinnane, "Partnership as Experimentation," *The Journal of Law, Economics, and Organization*, November 2019, *35* (3), 455–488.

Ashraf, Quamrul and Oded Galor, "Cultural Diversity, Geographical Isolation, and the Origin of the Wealth of Nations," *NBER Working Paper No. 17640*, 2011.

Beider, Alexander, *A Dictionary of Jewish Surnames from the Russian Empire*, Avotaynu, 2008.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts, "Do Management Interventions Last? Evidence from India," *American Economic Journal: Applied Economics*, April 2020, *12* (2), 198–219.

_ , Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen, "What Drives Differences in Management Practices?," *American Economic Review*, May 2019, *109* (5), 1648–1683.

Caselli, Francesco and Wilbur John Coleman, "Cross-Country Technology Diffusion: The Case of Computers," *American Economic Review*, 2001, *91* (2), 328–335.

Cinnirella, Francesco, Erik Hornung, and Julius Koschnick, "Flow of Ideas: Economic Societies and the Rise of Useful Knowledge," *CESifo Working Paper*, 2022.

Comin, Diego and Marti Mestieri, "Technology Diffusion: Measurement, Causes, and Consequences," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 2, Elsevier, 2014, pp. 565–622.

Department of Trade and Manufacturing, *Directory Chronological, by Subject and Alphabetical of Privileges issued in Russia from Jan 1, 1892 to July 1, 1896 (Ukazatel' Khronicheskogo, Predmetnogo i Alfavitnogo Vyddanykh v Rossii Privilegiy s 1 yanvarya 1892 g. po 1 iyulya 1896 g.)*, Saint Petersburg: Typography of V. Kirschbaum, 1897.

Foster, Andrew D. and Mark R. Rosenzweig, "Microeconomics of Technology Adoption," *Annual Review of Economics*, 2010, *2* (1), 395–424.

Gerschenkron, Alexander, *Economic Backwardness in Historical Perspective a Book of Essays*, Harvard University Press, 1962.

Giorcelli, Michela, "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 2019, *109* (1), 121–152.

Goldin, Claudia and Kenneth Sokoloff, "Women, Children, and Industrialization in the Early Republic: Evidence from the Manufacturing Censuses," *The Journal of Economic History*, 1982, *42* (4), 741–774. Publisher: Cambridge University Press.

Gregg, Amanda, "Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894–1908," *American Economic Review*, 2020, *110* (2), 401–427.

_ **and Steven Nafziger**, "Capital Structure and Corporate Performance in late Imperial Russia," *European Review of Economic History*, 2019, *23* (4), 446–481.

_ **and** _ , "The Births, Lives and Deaths of Corporations in Late Imperial Russia," *The Economic Journal*, 2024, *forthcoming*.

_ **and Tamar Matiashvili**, "Modernization in Progress: Part-Year Operation, Mechanization, and Labor Force Composition in Late Imperial Russia," *The Journal of Economic History*, 2022, *82* (4), 1143–1182.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 1957, pp. 501–522.

Hornung, Erik, "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia," *American Economic Review*, 2014, *104* (1), 84–122.

Hsieh, Chang-Tai and Peter J. Klenow, "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 2014, *129* (3), 1035–1084.

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

Kindleberger, Charles P., "Germany's Overtaking of England, 1806 – 1914 Part II," *Weltwirtschaftliches Archiv*, 1975, *111* (3), 477–504. Publisher: Springer.

Lamoreaux, Naomi R., "Constructing Firms: Partnerships and Alternative Contractual Arrangements in Early Nineteenth-Century American Business," *Business and Economic History*, 1995, *24* (2), 43–71. Publisher: Cambridge University Press.

_ **and Jean-Laurent Rosenthal**, "Legal Regime and Contractual Flexibility: A Comparison of Business's Organizational Choices in France and the United States during the Era of Industrialization," *American Law and Economics Review*, March 2005, *7* (1), 28–61.

Malein, Viktor, "Human Capital and Industrialization: German Settlers in late Imperial Russia," Technical Report, EHES Working Paper 2021.

Manuelli, Rodolfo E. and Ananth Seshadri, "Frictionless Technology Diffusion: The Case of Tractors," *American Economic Review*, 2014, *104* (4), 1368–91.

Markevich, Andrei and Steven Nafziger, "State and Market in Russian Industrialization, 1870–2010," in Kevin O'Rourke and Jeffrey Williamson, eds., *The Spread of Modern Industry to the Periphery Since 1871*, 2017, pp. 33–62.

McKay, John P., *Pioneers for Profit: Foreign Entrepreneurship and Russian Industrialization, 1885-1913*, University of Chicago Press, 1970.

Mejia, Javier, "Social Networks and Elite Entrepreneurship in Latin America: Evidence from the Industrialization of Antioquia," *The Journal of Economic History*, 2024, *84* (4), 1067 – 1106.

Ministry of Finance, *World's Columbian Exposition 1893, Chicago. Catalogue of the Russian Section.*, Imperial Russian Commission, Ministry of Finance, 1893.

_ , *Russian Factory Production: List of Factories and Plants. (Fabrichno-zavodskaya promyshlennost' Rossii: Perechen' fabrik i zavodov)*, Tipografia E.A. Efrona, 1897.

Mokyr, Joel, *The Lever of Riches: Technological Creativity and Economic Progress*, Oxford University Press, 1990.

_ , *A Culture of Growth: The Origins of the Modern Economy*, Princeton University Press, 2016.

Natkhov, Timur and Natalia Vasilenok, "Skilled Immigrants and Technology Adoption: Evidence from the German Settlements in the Russian Empire," *Explorations in Economic History*, 2021, *81*, 1–18.

Noble Assembly of the Russian Empire, *Obshchiy gerbovnik dvoryanskikh rodov Vserossiyskoy imperii (General Armorial of the Noble Families of the Russian Empire*, St. Petersburg, 1917.

Pamuk, Şevket, "Uneven Centuries: Turkey's Experience with Economic Development since 1820," *The Economic History Review*, 2019, *72* (4), 1129–1151.

Serafinelli, Michel and Guido Tabellini, "Creativity over Time and Space: A Historical Analysis of European cities," *Journal of Economic Growth*, 2022, *27* (1), 1–43.

Spitzer, Yannay, "Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States 1881–1914 *," *Maurice Falk Institute for Economic Research in Israel. Discussion paper series.*, 2021, pp. 0_1,1–41,43–83.

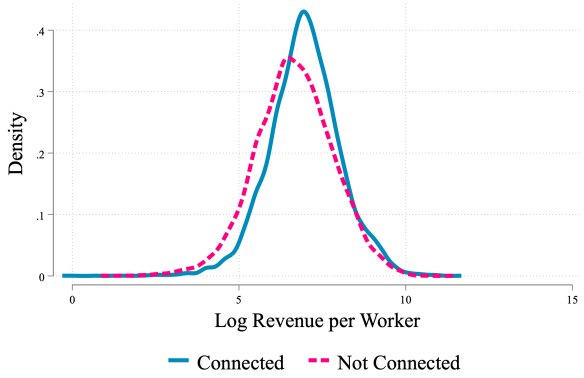
Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," *The Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

Squicciarini, Mara P and Nico Voigtländer, "Human Capital and Industrialization: Evidence from the Age of Enlightenment," *The Quarterly Journal of Economics*, 2015, *130* (4), 1825–1883.

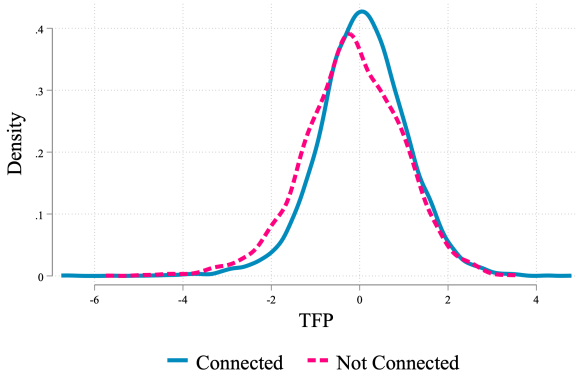
Syverson, Chad, "What Determines Productivity?," *Journal of Economic literature*, 2011, *49* (2), 326–365.

On-line Appendix

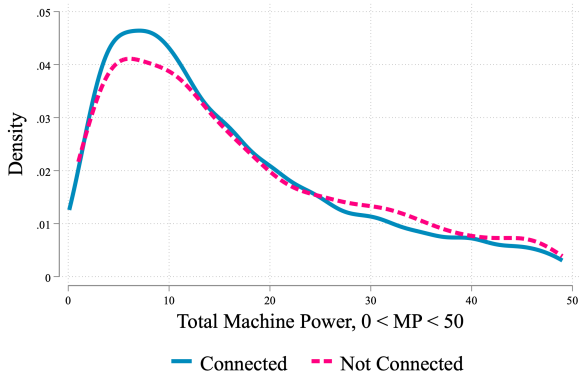
A Descriptive Statistics



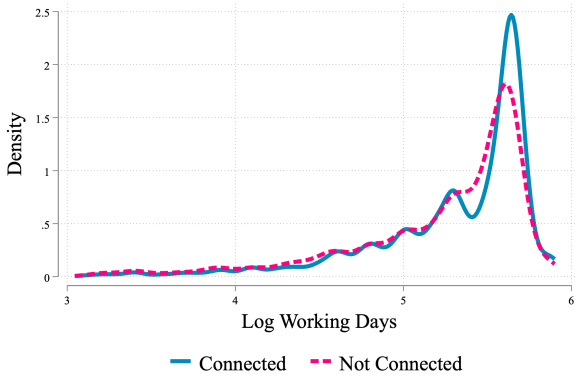
(a) Log Revenue Per Worker



(b) TFP



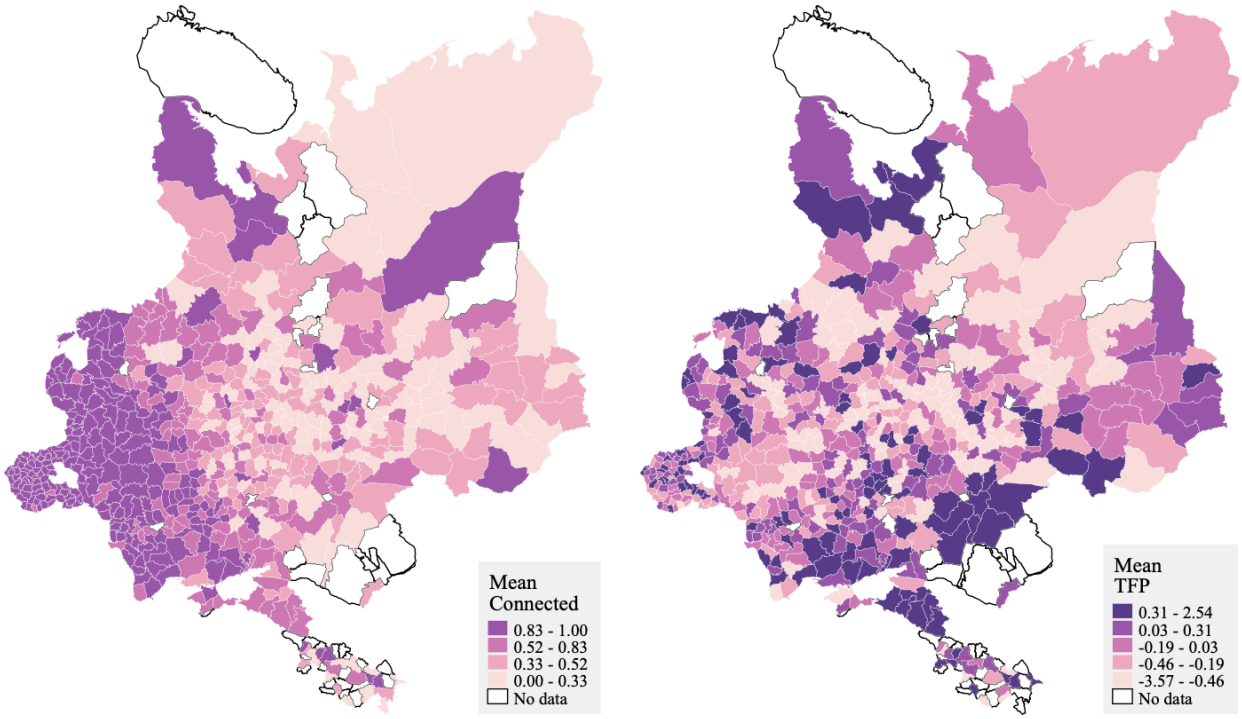
(c) Log Machine Power Per Worker



(d) Log Working Days

Figure A1: Density of select outcomes by group

Notes: Kernel densities use Epanechnikov kernel. TFP is calculated as a residual after regressing log Revenue per worker on log Number of workers and log Machine power (+1). Ethnic groups are defined as probable ethnic identity of the firm's owner according to the algorithm described in Section 3.2.



(a) Share Connected firms

(b) Mean TFP

Figure A2: Spatial Distribution of Connected Firms and TFP

Notes: The shapefile is from Kessler and Markevich (2021). TFP is calculated as a residual after regressing log Revenue per worker on log Number of workers and log Machine power (+1). Ethnic groups are defined as probable ethnic identity of the firm's owner according to the algorithm described in Section 3.2.

B Results, continued

B.1 Showing controls

Table B1: Connected firms and productivity (with control variables shown)

	log Revenue per worker	TFP	Machine Power	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.084***	0.072***	0.079*			
	(0.023)	(0.022)	(0.043)			
German				0.122***	0.085***	0.091
				(0.032)	(0.031)	(0.083)
Jewish				0.085***	0.079***	0.055
				(0.024)	(0.024)	(0.047)
Other Connected				0.063**	0.045*	0.123*
				(0.027)	(0.026)	(0.069)
log Age	0.046***	0.046***	-0.004	0.046***	0.046***	-0.004
	(0.007)	(0.007)	(0.013)	(0.007)	(0.007)	(0.013)
log Number of employees	0.068***	0.042**	0.931***	0.068***	0.043**	0.931***
	(0.017)	(0.018)	(0.030)	(0.017)	(0.018)	(0.029)
Factory is located in a city	0.194***	0.196***	-0.036	0.194***	0.196***	-0.032
	(0.034)	(0.032)	(0.048)	(0.034)	(0.032)	(0.047)
Noble last name	-0.021	-0.017	0.017	-0.021	-0.021	0.026
	(0.020)	(0.020)	(0.041)	(0.020)	(0.020)	(0.040)
Observations	13,451	13,451	14,808	13,451	13,451	14,808
Mean Dependent Variable	6.851	-0.016	21.253	6.851	-0.016	21.253
R-squared	0.492	0.446		0.492	0.446	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓

Notes: See notes to Table 3.

B.2 Larger firms

The inclusion criterion for this census was rather loose: a factory was included in the manufacturing census if one of the following were true: it had 15 employees, a steam boiler, a steam engine, or other factory devices. It is therefore possible that, by including many smaller factories possibly more likely to be operated by locals, we bias our results upwards. Here, we repeat the main productivity results, limiting our sample to factories with 15 or more employees.

Table B2: Productivity of larger connected firms
(more than 15 employees)

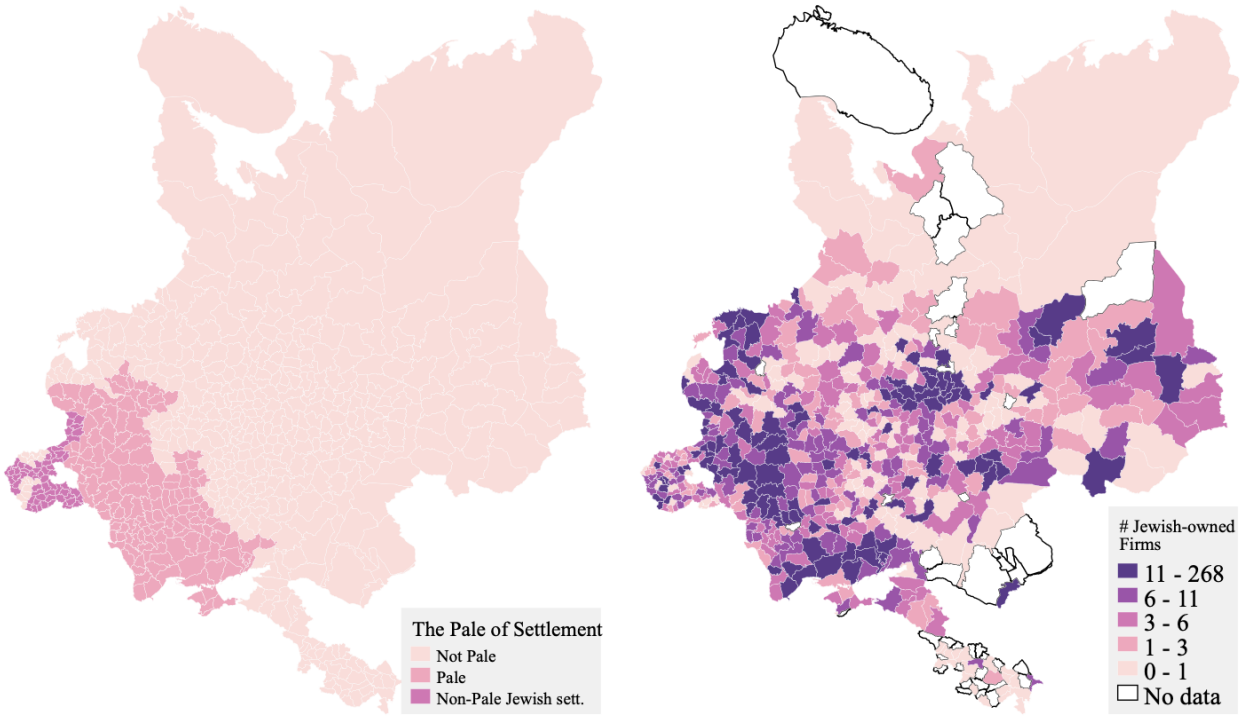
	log Revenue per worker (1)	TFP (2)	Machine Power (3)
Connected	0.084***	0.066***	0.106**
	(0.023)	(0.023)	(0.050)
log Age	0.046***	0.046***	-0.004
	(0.007)	(0.008)	(0.018)
log Number of employees	0.039**	-0.043**	0.990***
	(0.016)	(0.018)	(0.040)
Factory is located in a city	0.120***	0.145***	-0.031
	(0.033)	(0.030)	(0.053)
N	6,974	6,974	7,094
Mean	6.828	0.002	35.679
R-squared	0.606	0.541	
District FE	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓

Notes: Sample limited to firms with 15 or more employees. TFP is calculated as a residual after regressing log Revenue per worker on log Number of workers and log Machine power (+1). Connected ethnic groups are defined according to the algorithm described in Section 3.2.

C Jewish entrepreneurs

C.1 Jews in the Pale

We explore the spatial distribution of Jewish owners with respect to the pale of settlement. Large number of the Jewish-owned factories fell in the Pale of Settlement, but the distribution was indeed wider, as many Jews could live outside of the pale, especially if they were richer.



(a) Pale of Settlement (b) # factories owned by Jews

Figure C3: Pale of Settlement and Jewish Factories

Notes: Source of the shapefile: Kessler and Markevich (2021).

In Table C3 we show that Jews inside and outside of the Pale (here defined as the official Pale + the Congress Poland Provinces) ran very different factories indeed. Considering factories in the same activities, Jews in the pale were more likely to have higher log revenue per worker and TFP compared to the outside-of-the-pale Jews, which suggests that Jews in the pale perhaps faced less discrimination from outsiders from living in the pale. Jews outside of the pale, however, had significantly more machine power in their factories, consistent with the fact that Jews allowed to live outside of the pale were richer and thus would have been able to make larger capital investments.

Table C3: Jewish-owned firms within and outside of the Pale of Settlement

	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)
Jews in Pale	0.093***	0.080***	-0.156***
	(0.029)	(0.029)	(0.051)
log Age	0.032***	0.030***	-0.010
	(0.010)	(0.010)	(0.025)
log Number of employees	0.109***	0.090***	0.878***
	(0.014)	(0.013)	(0.030)
Factory is located in a city	0.325***	0.308***	0.268***
	(0.033)	(0.033)	(0.069)
N	4,635	4,635	5,010
R-squared	0.415	0.362	
Industry and sub-industry FE	✓	✓	✓

Notes: Robust standard errors clustered at the district level. "Jews in Pale" is a binary variable equal to one if a firm was owned by a Jewish entrepreneur and was located within the Pale of Settlement. TFP is residual in a regression of log revenue per worker on log number of workers and log total machine power(+1). Jewish ethnic origin is defined according to the algorithm described in Section 3.2.

C.2 Robustness to excluding Jewish owners

Jewish entrepreneurs were crucial in the economy of the Russian Empire. They operated 5,492 (34.6%) of all private enterprises. However, using the firms we classify as Jewish-owned could be problematic for two reasons. First is that the Russian Empire had a history of anti-semitic policies and outright violence, such as Pogroms. Therefore, Jewish entrepreneurs may have faced more administrative and social challenges while running their firms than other entrepreneurs. Second, our classification counts as Jews all Russian last names that were used as both Jewish ans Russian. Therefore, we may be incorrectly assigning some of the Russian Entrepreneurs to our Connected definition, thus biasing our results down. We replicate Table 3, dropping all observations we classify as Jewish-owned. Indeed, all the results hold up in this sample.

Table C4: Connected ownership and productivity (subsample without Jewish firms owners)

	log Revenue per worker	TFP	Machine Power	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.105***	0.078***	0.117*			
	(0.031)	(0.028)	(0.061)			
German				0.151***	0.109***	0.096
				(0.036)	(0.033)	(0.109)
Other Connected				0.086***	0.064**	0.126*
				(0.031)	(0.029)	(0.069)
log Age	0.044***	0.045***	-0.008	0.044***	0.045***	-0.008
	(0.008)	(0.008)	(0.017)	(0.008)	(0.008)	(0.017)
log Number of employees	0.083***	0.056***	0.953***	0.083***	0.056***	0.952***
	(0.021)	(0.021)	(0.032)	(0.021)	(0.021)	(0.032)
Factory is located in a city	0.191***	0.208***	-0.040	0.192***	0.209***	-0.041
	(0.043)	(0.041)	(0.054)	(0.043)	(0.041)	(0.055)
Noble last name	-0.012	-0.017	-0.012	-0.011	-0.017	-0.012
	(0.029)	(0.028)	(0.057)	(0.029)	(0.028)	(0.057)
Observations	8,725	8,725	9,489	8,725	8,725	9,489
Mean depend. variable	6.820	-0.051	21.893	6.820	-0.051	21.893
R-squared	0.500	0.459		0.500	0.459	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at the district level. The sample is limited to the firms with non-Jewish-classified owners. Refer to Table 3 for details on outcome variables and controls.

C.3 Jewish Name Index by Abramitzky et al. (2020)

Abramitzky et al. (2020) create a Jewish Name Index for various US Censuses, based on the percentage of individuals speaking Yiddish with a certain combination of first and last names. We use their US data of last names and assigned "Jewishness" probabilities to train our data on Jewishness of the last names in the 1894 Manufacturing Census. Because we were looking for a lower bound on Jewish counts, we used the most conservative approach. We use authors' suggested cutoff of 0.8 Jewishness probability of a last name to categorize their US data as Jewish versus not. We then use a machine learning algorithm (based on Logistic Regression), using their data as a training sample, and applying to our last names. Out of the 9,634 total last names, only 701 (7.3%) are considered Jewish by this algorithm. This is much lower than the yield of our main classification method, the Dictionary of Jewish last names, at 27%. This algorithm, however, clearly undercounts Jewish last names: while last names such as "Abramov" are correctly identified as Jewish, last names "Abramovskii" and "Abramovich" are not, as, most likely, these Eastern European last name structures were not as common in the US in the first half of the 20th century. Further, 194 last names this algorithm considers Jewish do not appear in the Dictionary, so we treat these as non-Jewish.

Thus, we are left with 507 unique Jewish last names in our data, corresponding to 1,340 (8.25%) firms.

We then re-ran our ethnicity-assignment algorithm described in the main body, on the non-Jewish portion of our sample. First, we explore what happened to the last names that were listed in the Jewish Dictionary but were not identified as Jewish by this algorithm. There were 4,299 such firms. 1,826 were now considered Other Connected; 1,524 were now considered Russian; 918 were considered German; only 31 were considered Eastern.

Next, we reproduce our main findings (Table 3) in Table C5. Reassuringly, while smaller, all the coefficients on Connected are positive (Columns 1-3). In addition, all the coefficients on German are also positive, al and significant (Columns 4-6). The coefficients on Jewish remain large and significant for log revenue per worker and TFP. The coefficients for Other Connected are positive but lose statistical significance in columns 4 and 5. These results reassure us that our exact classification of Jewish last names does not alter the takeaways of our findings, and validate using the Dictionary of Jewish Last Names as our main source.

Table C5: Jewish-owned firms within and outside of the Pale of Settlement

	log Revenue per worker	TFP	Machine Power	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)	(4)	(5)	(6)
Connected	0.063***	0.050**	0.092**			
	(0.022)	(0.022)	(0.046)			
German				0.101***	0.071**	0.086
				(0.034)	(0.033)	(0.076)
Jewish				0.061**	0.066**	-0.063
				(0.031)	(0.031)	(0.055)
Other Connected				0.039	0.022	0.141**
				(0.027)	(0.026)	(0.058)
log Age	0.046***	0.047***	-0.003	0.046***	0.047***	-0.004
	(0.007)	(0.007)	(0.013)	(0.007)	(0.007)	(0.013)
log Number of employees	0.067***	0.041**	0.931***	0.069***	0.042**	0.933***
	(0.018)	(0.018)	(0.030)	(0.018)	(0.019)	(0.030)
Factory is located in a city	0.199***	0.202***	-0.041	0.202***	0.204***	-0.036
	(0.035)	(0.032)	(0.048)	(0.035)	(0.033)	(0.050)
Noble last name	-0.016	-0.013	0.025	-0.015	-0.015	0.040
	(0.020)	(0.020)	(0.042)	(0.020)	(0.020)	(0.041)
Observations	13,297	13,297	14,639	12,923	12,923	14,196
Mean Depent Variable	6.850	-0.016	21.182	6.842	-0.020	21.463
R-squared	0.493	0.446		0.496	0.450	
District FE	✓	✓	✓	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at the district level. Jewish classification altered to use Abramitzky et al. (2020)'s Index. See Table 3 for details on outcome variables and controls.

Further, we again map the Jewish-owned factories by district, using this new classification method. We did this to confirm that even using this conservative estimate, that is based on more "definitely-Jewish" and fewer mixed-use last names, we still observe the geographic dispersion of Jewish-owned factories we saw in C3. In Figure C4, we see that this is in fact the case, and the concentration in the Pale of Settlement is even lower. This once again, gives us confidence in using our original method.

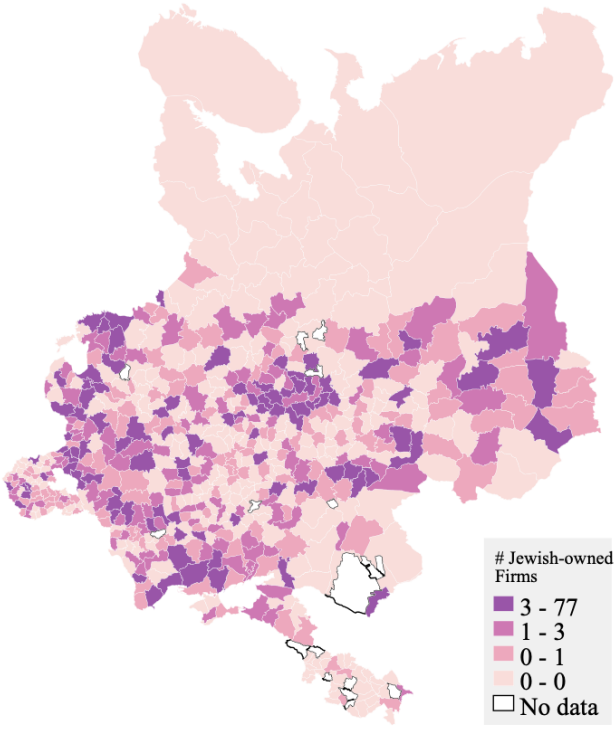


Figure C4: Number of factories owned by Jewish entrepreneurs using Jewish Name Index method from Abramitzky et al. (2020).

D Chicago World's Fair

Table D6 shows that firms at the Chicago World's Fair had higher baseline productivity measures than the firms that did not participate.

Table D6: Productivity of Russian firms at the 1893 Chicago World Fair

	log Revenue per worker (1)	TFP (2)	Machine Power (3)
Participated in Fair	0.270***	0.192**	0.192*
	(0.079)	(0.078)	(0.103)
log Age	0.046***	0.046***	-0.007
	(0.006)	(0.006)	(0.012)
log avg. total num. of workers	0.065***	0.040***	0.935***
	(0.009)	(0.008)	(0.018)
Factory is located in a city	0.199***	0.202***	-0.048
	(0.023)	(0.023)	(0.040)
Noble last name	-0.019	-0.015	0.022
	(0.019)	(0.019)	(0.038)
N	13,426	13,426	14,788
Mean Y_0	6.845	-0.022	20.731
R-squared	0.492	0.445	
District FE	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓

Notes: Standard errors clustered at the district level. Participated in the Fair is a binary variable equal to one if a firm was matched to the Chicago World's Fair directory. All other components of the specification same as in Table 3.

E German advantage by industry: additional evidence

E.1 German and non-German productivity premium by industry

	Log Revenue per Worker										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Animal	Chemical	Cotton	Flax	Foods	Metals & Machines	Mineral	Paper	Silk	Wood	Wool
German	0.186	0.811***	0.088	0.000	0.153**	0.122	0.060	0.080	-0.289**	0.487***	-0.004
	(0.123)	(0.214)	(0.254)	(.)	(0.073)	(0.076)	(0.135)	(0.123)	(0.094)	(0.169)	(0.096)
N	826	203	191	133	2,177	512	483	199	159	291	255
Mean	6.974	6.866	6.899	6.561	7.127	6.510	5.737	6.534	6.287	6.670	6.668
R-squared	0.611	0.805	0.491	0.495	0.530	0.560	0.611	0.489	0.445	0.657	0.598
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sub-industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

	Log Revenue per Worker										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Animal	Chemical	Cotton	Flax	Foods	Metals & Machines	Mineral	Paper	Silk	Wood	Wool
Connected (not German)	0.043	-0.016	0.302	0.487	0.043	0.127	0.043	0.066	-0.250*	0.189	0.294**
	(0.111)	(0.147)	(0.195)	(0.340)	(0.053)	(0.098)	(0.067)	(0.090)	(0.116)	(0.137)	(0.144)
N	861	220	198	140	2,853	507	551	253	161	336	252
Mean	6.981	6.812	6.897	6.592	7.056	6.511	5.740	6.495	6.280	6.747	6.683
R-squared	0.607	0.795	0.501	0.523	0.513	0.563	0.600	0.391	0.436	0.645	0.584
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sub-industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

E.2 Stories of German entrepreneurs in the chemicals industry

Below, we recount stories of some of the German chemicals entrepreneurs who appear in the Census and whose stories we have been able to find. These examples illustrate the lived experiences behind the "connections" we describe throughout the paper.

Three Russian-born German brothers, Edward, Robert, and Wilhelm Bremme, whose father was himself an immigrant entrepreneur, were sent to Göttingen University to study Chemistry.[23] All three returned to the Russian Empire and started an essential oils and essences firm called "Brothers Bremme". In the Census, the firm appears to have been successful - it had opened in 1886, operated 290 days a year, had a steam engine, employed 25 men, and generated 45,000 rubles in revenue. This story illustrates a likely channel through which cultural connections operated - immigrants who had studied in Germany or descendants of immigrants who had left the Empire to study in Germany due to their knowledge of the language, culture, and family ties came back and used their knowledge to start companies.

Another story is of direct immigrant transplants who started a chemicals firm - Sigismund and Emil Erlenbachs. We find Sigismund later in history involved in trade with Germany during World War I, where he is identified as a Bavarian national and treated with suspense.[24] In our data, Sigismund and Emil Erlenbachs own a Paints and Lacquer company founded in 1888, which operated 240 days a year, had a steam engine, employed 8 men and 2 boys, and generated 26,800 rubles in revenue. Thus, direct German immigrants are certainly a part of our data and explain the productivity differences.

However, Germans not only operated more productive companies but also innovated within the industry. In 1892, Max Geflinger, Pavel Emilev, and Rudolph Schmitz were the first to open a superphosphate - first chemical fertilizer - factory in the Russian Empire (Gusev, 2020). Two of the three founders had obviously German first and last names. Thus, German owners were pioneers within at least one important chemical subindustry in the Russian Empire.

[23] source: https://www.dp.ru/a/2019/08/29/Fabrichnij_aromat

[24] http://www.raruss.ru/excellent/2737-leman-gravure.html

F Distance to Europe and ease of communication

Establishing and maintaining ethnic connections, especially at a time of high transportation costs, would have been positively correlated with physical distance to Europe. Indeed, many ethnic groups we study, such as Poles, originally had connections with Western Europe due to their geographic location. To evaluate how distance and connectedness interact in our setting, we calculate Euclidean distances from county centroids to the closest point on the border of the Empire with Europe. We then interact this distance measure with our connected dummy. Table F7 presents these results. While negative, as expected, the effect of distance was small and statistically insignificant.

Table F7: Complementarities of connectedness and distance to Europe

	(1) log Revenue per worker	(2) TFP	(3) Machine Power
Connected	0.102**	0.083**	0.139***
	(0.040)	(0.038)	(0.045)
Connected * Log Distance	-0.010	-0.006	-0.035
	(0.017)	(0.016)	(0.022)
log Age	0.045***	0.046***	-0.005
	(0.007)	(0.007)	(0.014)
log Number of employees	0.069***	0.043**	0.931***
	(0.017)	(0.018)	(0.030)
Factory is located in a city	0.194***	0.197***	-0.036
	(0.034)	(0.032)	(0.048)
Noble last name	-0.019	-0.016	0.021
	(0.019)	(0.019)	(0.042)
Observations	13,427	13,427	14,789
Mean Depent Variable	6.850	-0.017	21.223
R-squared	0.491	0.445	
District FE	✓	✓	✓
Industry and sub-industry FE	✓	✓	✓

Notes: Robust standard errors clustered at the District level. Distance is calculated as the euclidean distance of the district in which the factory is located and the closest point on the Russian Empire's border with Europe. All other definitions as in Table 3.

However, physical distance may not capture the ease of "connection", especially in a country as vast as the Russian Empire. Connections, instead, would have been facilitated by telegraph lines, railroads, and postal systems. All these existed in large cities.

Were connected firms more likely to enjoy the benefits of cities compared to the non-connected firms? It is well-known that locational choices have a big effect on outcomes – for instance, Abramitzky et al. (2020) find that the upward mobility of immigrants' children in the United States is largely driven by the immigrants choosing wealthier locations as their destinations. To make sure we were not picking up this urban versus rural choice of factories, we have included dummies for urban location throughout. However, a deeper dimension to this choice in our case may be that the Connected entrepreneurs benefit *more* from locating in major cities, as big cities would provide easier connection to and trade with other countries. While we cannot directly test the latter part of the argument, we can see whether connected entrepreneurs benefitted more from locating in cities, providing at least suggestive evidence.

We identify seventeen cities from the 1897 Russian Empire Population Census with a population greater than one hundred thousand. These were the largest cities of the Empire. First, we find that Connected entrepreneurs were more likely to locate in a big city (21.7% vs 14.8%). We then repeat the analysis found in Table 3 but include the interaction terms with the binary variable of whether a factory was located in one of these major cities. Indeed, we find that the interaction term on Connected and locating in a major city is large and significant for all outcomes (Columns 1 - 3 of Table F8). The interaction was positive for all ethnicities, and statistically significant for the Jewish and Other Connected groups (Columns 4-6 of Table F8). These results suggest that cities were more beneficial to the connected than the non-connected group, potentially due to higher access to communication with and transport to Europe.

Table F8: Connected Entrepreneurs and Location Choices

	log Revenue per worker	TFP	Machine Power	log Revenue per worker	TFP	Machine Power
	(1)	(2)	(3)	(4)	(5)	(6)
Connected × City >100k pop.	0.127***	0.097***	0.237**			
	(0.039)	(0.037)	(0.095)			
German × City >100k pop.				0.135*	0.093	0.120
				(0.069)	(0.067)	(0.186)
Jewish × City >100k pop.				0.117**	0.087*	0.254***
				(0.046)	(0.047)	(0.096)
Other Connected × City >100k pop.				0.151***	0.124***	0.277
				(0.044)	(0.039)	(0.177)
City >100k pop.	0.058	0.092	-0.171	-0.089	0.091	-0.161
	(0.079)	(0.065)	(0.129)	(0.091)	(0.065)	(0.130)
Connected	0.064***	0.057**	0.034			
	(0.024)	(0.024)	(0.040)			
German				0.096***	0.064*	0.086
				(0.036)	(0.036)	(0.069)
Jewish				0.064**	0.068***	0.007
				(0.026)	(0.026)	(0.044)
Other Connected				0.036	0.023	0.067
				(0.029)	(0.029)	(0.048)
N	13,451	13,451	14,808	13,451	13,451	14,808
Mean	6.851	-0.016	21.253	6.851	-0.016	21.253
R-squared	0.490	0.443		0.492	0.443	
District FE	✓	✓	✓	✓	✓	✓
Indsutry and sub-industry FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors clustered at District level. "City > 100k pop" is a binary variable for cities with more than 100,000 population in the 1897 Population Census. Please refer to Table 3 notes for details on outcome variables.

References

Abramitzky, Ran, Leah Platt Boustan, and Dylan Connor, "Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office," *National Bureau of Economic Research*, June 2020.

Gusev, Igor, *History of Riga from Founding to the Beginning of the 20-th century*, Zorkis, 2020.

Kessler, Gijs and Andrei Markevich, "Russian Empire Historical GIS Maps (1897)," December 2021.