Transportation Networks and the Rise of the Knowledge Economy in 19th Century France*

Georgios Tsiachtsiras†

JOB MARKET PAPER

Latest version here

November 8, 2021

Abstract

This paper exploits an episode of French history to study the relationship between the roll-out of railroads and the rise of the knowledge economy. I take advantage of the exogenous variation in railway access arising from a straight line time variant instrument, to document that access to rail network increases the innovation activity at the canton level. I explore two underlying mechanisms behind the main results. First, I introduce a mechanism of access to knowledge to study how patent activity is affected by the reduction of transportation costs, due to the expansion of rail and canal network, to the cantons where inventors reside. Second, using text analysis techniques, I am able to determine the technology of each patent application in the historical database of the National Institute of Industrial Property of France and to explore how connectivity with the global city of Paris is associated with the diffusion of novel technologies. Finally, I introduce a back of the envelope exercise based on canals to show that in the absence of railroads the invention rate of the French cantons would have been, on average, 21.3% lower.

Keywords: innovation, patent data, railroads, access to knowledge

JEL classification: L92, N73, O31, O33, P25, R12

---

*I would like to thank from the bottom of my heart, Rosina Moreno and Ernest Miguelez for their guidance and support at all the stages of this paper. Special thanks to Anastasia Litina and Sergio Petralia for helpful discussions and comments. I am grateful to Steeve Gallizia, Christophe Mimeur, Teresa Sanchís and Nicolas Verdier for sharing with me the historical database of INPI, the railroad data, the GDP data and the postal offices of France. I would like to express my gratitude to Richard Hornbeck for sharing with me his replication files. I have benefited from the comments of Thomas Piketty, Sascha O. Becker, Mara Squicciarini, Diogo Britto, Carlo Schwarz, Jan Bakker, Francesco Quatraro, Stefano Breschi, Denis Cogneau, Laurent Bonnaud, Tommaso Ciarli, Cédric Chambru, Zelda Brutti, Ester Manna, Xavier Fageda, Emilie Bonhoure, Abel Lucena, Mercedes Teruel, Dolores Añon Higon, András Jagadits, Vera Rocha, Alexander Klein, Dirk Martignoni, Bjoern Brey, Giuseppe Cappellari, Filippo Tassinari, Eric Melander, Michel Philippe Lioussis, Ivan Hajdukovic, Emanuele Caggiano, Maria Teresa Gonzalez and the participants of the DRUID 2021 (Copenhagen Business School), the 15th North American Meeting of the Urban Economics Association (Arizona University), the PhD - Economics Virtual Seminar, the F4T Seminar Series (Bocconi University), the 6th Summer School on Data & Algorithms for ST&I studies (organized by Aristotle University and KU Leuven, 2021), 11th ifo Dresden Workshop on Regional Economics (2021), the EPIP Conference (2021), the Young Economist Symposium 2021 (sponsored by Columbia, NYU, UPenn, Princeton and Yale), the XXIII Applied Economics Meeting (Universitat de València, 2021), the Economic History Seminar at Paris School of Economics (2021), the PhD in Economics Workshop at University of Barcelona (2020), the Economic Student Seminar at Universitat of Autònoma de Barcelona, Virtual Meeting of the Urban Economics Association (2020), 13th RGS Doctoral Conference in Economics on Regional Disparities and Economic Policy in Europe (Technische Universität of Dortmund, 2020), the 5th Geography of Innovation Conference (University of Stavanger, 2020), the 7th International PhD Workshop in Economics of Innovation, Complexity and Knowledge (University of Turin, 2020) and the 4th Summer School in Economics (University of Ioannina, 2019). Financial support from Agaur (FI-DGR). Any remaining errors are mine.

†Tsiachtsiras: Ph.D. candidate, University of Barcelona School of Economics; Department of Statistics, Econometrics, and Applied Economics, Barcelona 08034, Spain. email: gtsiachtsiras@ub.edu

"There is no country in the world where so small a proportion of the capital invested within the last forty years in canals and railroads has been wasted or where traveling is safer, or in which travel and trade are accommodated at more reasonable rates than in France" (Moncure Robinson, American Philosophical Society, 1880 source: Smith (1990)).

1 Introduction

Productivity growth rely in two elements: new ideas bringing new technologies, and the diffusion of these ideas across people and places. Connecting places via transportation networks is essential in order to promote knowledge diffusion across the space, and hence long run economic growth. Over the last decades, the World Bank dedicated a high proportion of money to support infrastructure investments (World Bank, 2007, 2013, 2017).[1] Despite an emphasis on reducing transportation costs, we lack empirical evidence on understanding how transportation infrastructure projects could actually create meaningful connections to places that matter for boosting the innovation performance and to facilitate channels related to the diffusion of knowledge.

In this paper, I exploit an episode of French history to study the impact of the largest-scale construction of French railroads on innovation performance. Over the second half of 19th century, the establishment of railroads transformed the French economy. Railways triggered economic relations and cultural environments, stimulated commerce and created new economic opportunities. The creation of the rail network changed the perception of time (Schwartz et al., 2011). According to Thévenin et al. (2013) a striking example of this transformation is the travel time from Paris to Marseille. In 1814, the duration of this trip was four to five days while in 1857, this trip needed about 13 hours.

Remarkably, the roll-out of the network coincided with a rise of innovative activity. Until 1850, the number of patent applications owned by inventors in the historical French patent database was 22,978 while from 1850 until 1902 increased to 285,597. Some of the greatest inventions in history took place in France during the second half of 19th century. Louis Pasteur in 1865 and 1871 developed the modern method for wine and beer pasteurization. Another example is the invention in 1853 by the French chemist, Charles Frederic Gerhardt, who was the first to prepare acetylsalicylic acid (aspirin). Finally, the very first patented film camera introduced by Louis Le Prince in 1888.

[1]Based on the reports of 2007, 2013 and 2017 WB committed approximately $105 billion in transport-related projects.

I use rich historical data to construct a panel dataset at the canton level for France (2,925 cantons) over the period 1850-1890.[2] I combine a very recent dataset of communes with access to rail station over the 19th century (Mimeur et al., 2018) with the historical patent applications database of the National Institute of Industrial Property office (INPI from now on) (INPI, 2019) over the time period 1850-1890 to construct a unique dataset.[3] The historical patent database of INPI is still quite unexplored. Previous papers, have used the INPI database during the first half of 19th century to study the effect of technology transfer from Britain to France (Nuvolari et al., 2020), the influence of the patent system on the economic performance (Galvez-Behar, 2019) and the role of women in enterprise and invention in France (Khan, 2016).

Equipped with this dataset, this paper relies on the expansion of the French railroad network which provides a unique setting to identify the causal impact of transport infrastructure on innovation outcomes. I use the three different French railway plans over the 19th century to construct a time variant straight line instrument. I document that access to a rail station increases the innovation activity of a canton. As a robustness analysis to complement the instrumental variance approach, I apply an inconsequential units approach in which I rely on the randomly chosen subset of municipalities that received railway access because they lie on the most direct route between the nodal destinations that were used for the creation of the straight line instrument. In addition, I exploit the richness of the patent data to create a balance test in which I regress the distance to rail stations on invention rate of the cantons over the period 1800-1840 one decade prior to the actual arrival of railroads. I report no empirical evidence on pre-trends related to the railroads and innovation performance. Also, I provide a conventional difference-in-differences regression model with staggered treatment where I use as independent variable a binary indicator switching to 1 if the canton has access to a rail station. This design takes place from 1800 to 1890 allowing me to control for a long period of time, 1800-1840, before the arrival of railroads in which all the cantons do not have access to a rail station. Finally, I explore cross section variation for every decade over the period 1850 to 1890 to identify the year that railroads started to have an impact on patenting rate and also to test the validity of the instrument by decade.

In a second step I explore the mechanism behind the main results. This paper adds to the literature by using the rationality of "market access" framework, developed by Donaldson and Hornbeck (2016), to provide a supply side mechanism related to potential interactions (Akcigit et al., 2018) and knowledge spillovers (Jaffe et al., 1993) among the inventors residing in different cantons. Instead of using the population in the market access index, the paper uses the number of inventors. The potential interactions and spillovers among inventors residing in different can-

[2] I include only the cantons from the mainland of France.

[3] The patent database is available after request.

tons could occur by establishing less costly routes, due to infrastructure improvements, within mainland France. I call this mechanism access to knowledge. A previous attempt in the literature relies on a market access framework (Perlman, 2015) to identify the mechanism that boosts innovation performance. Other papers in the literature (Agrawal et al., 2017; Andersson et al., 2021; Inoue and Nakajima, 2017; Wong, 2019; Gao and Zheng, 2020; Dong et al., 2020; Sun et al., 2021; Hanley et al., 2021; Komikado et al., 2021; Cui et al., 2020; Huang and Wang, 2020) provide evidence on how transportation networks facilitate the diffusion of knowledge but without connecting directly the knowledge diffusion mechanism to innovation performance. This the first paper to create a direct link between a supply side mechanism and the invention rate.

Finally, this paper uses the unique setting of France to introduce as a special case the importance for a given canton i to have a less costly connection to a global city such as Paris. Prior literature argues that Paris can work as a gatekeeper of knowledge which connects the national innovation system to global innovation networks (Miguelez et al., 2019). This is confirmed by the fact that inventions in Paris are more diverse and spread across the different technologies (Kogler et al., 2018). However, for only a small sample of patent applications, INPI database contains information about their technological class. In order to explore this special case, I assign technological classes to the remaining patent applications in the historical database of INPI. I assemble a vocabulary which is based on key words from the titles of the patent applications for which I have their technological class and thanks to this vocabulary, I am able to assign technological classes to all patent applications in the INPI database.

This paper directly speaks to the literature about transportation networks and innovation activity. Closer to my research are the papers of Perlman (2015), Agrawal et al. (2017) and Andersson et al. (2021). Perlman (2015) establishes a relationship between network access and innovative activity in the USA over the nineteenth century. Agrawal et al. (2017) find that the stock of highways has a positive effect on patenting in metropolitan statistical areas of the USA. Finally, Andersson et al. (2021) explore the effect of Swedish railroad on patent activity over the time period 1830-1910. They contend that network access fosters innovation activity. They find solid evidence that independent inventors start to specialize in specific technologies when they enter the market. General railroad access facilitates the inventors to develop ideas beyond the local economy and to invent in new technological classes.

There is a growing literature exploring the interplay between infrastructure and economic outcomes from a historical perspective. Railroads in the USA, Sweden, England, Wales and Switzerland transformed all towns into new places. They manage to lure banks and to increase urbanization (Atack et al., 2008, 2009; Atack and Margo, 2009; Berger and Enflo, 2017; Büchel and Kyburz, 2020; Bogart et al., 2022). They are responsible for the growth and the economic development of US and Nigeria cities (Nagy, 2016; Okoye et al., 2019). Railroads are meaningful for the development of the agricultural sector (Donaldson and Hornbeck, 2016; Donaldson, 2018) and have an impact on fertility and human capital (Katz, 2018). Transportation linkage has a adverse effect on health in rural US (Zimran, 2019). Railroads manage to boost manufacturing productivity in US (Hornbeck and Rotemberg, 2019; Pontarollo and Ricciuti, 2020). Finally, the steam railway led to the first large-scale separation of workplace and residence (Heblich et al., 2020).

In general, infrastructure networks could affect economy through different channels. The reallocation of road investments after the division of Germany creates regional income inequalities in terms of GDP per capita (Santamaria, 2020). Highways affect population and employment (Baum-Snow, 2007; Duranton and Turner, 2012). Finally, the adoption of the steamship between 1850 and 1900 boosts globalization of trade (Pascali, 2017).

In addition, this paper proposes as a direct policy that may smooth the persistent inequalities among urban and rural areas the targeted investments to infrastructure projects which facilitate access to big cities. This policy can increase the probability of smaller cities to innovate in new patent classes. Recent papers rise concerns about the correct allocation of infrastructure funds (Flyvbjerg, 2009) and which are the proper circumstances for these projects to be efficient (Crescenzi et al., 2016). At the same time, complex economic activities concentrate disproportionately in a few large cities, compared to less complex activities (Balland et al., 2020). Large cities are also the places with more interactions between higher-ability participants (Davis and Dingel, 2019).

Finally, this paper speaks to an extended literature on the economic history of France around the turn of the 19th and 20th century. Other papers in the literature explore the effect of knowledge elites in France or religion on economic development (Squicciarini and Voigtländer, 2015; Squicciarini, 2020). Apart from economic development it was less likely for the more religious cantons to give birth to scientists (Lecce et al., 2021). Furthermore, regions in the French Empire which became better protected from trade from the British for exogenous reasons during the Napoleonic Wars increased the capacity in mechanized cotton spinning more than regions which remained more exposed to trade (Juhász, 2018). Yet, the major technological breakthroughs like mechanized cotton spinning tend to be adopted slowly (Juhász et al., 2020). Other factors that affected the economic development of France are the large income shock of the phylloxera (Banerjee et al., 2010), emigration intensity (Franck and Michalopoulos, 2017), population swifts (Talandier et al., 2016) and the early industrialization (Franck and Galor, 2019). Daudin et al. (2019) establishes a relationship between fertility rates and the diffusion of cultural and economic information. One previous attempt related to railroads focus on the effect of railroads on population distribution for France, Spain, and Portugal from 1870 to 2000 (Mojica and Martí-Henneberg, 2011). Finally, the literature has explored the wealth concentration in Paris and France (Piketty et al., 2006) and regional inequality in France (Díez-Minguela and Sanchis Llopis, 2019).

The rest of the paper is organized as follows: Section 2 contains the historical background, Section 3 presents the data, Section 4 shows the empirical strategy, Section 5 presents the main results, Section 6 explores the mechanisms and Section 7 concludes.

2 Historical Background

2.1 Railroads

The first railway line in France established in 1828 from mining companies to connect St. Etienne to the Loire River. They used this line to transfer coal. The first line for passengers opened in 1837. It was a short line from Paris to Le Pecq (Dunham, 1941). At that period France was already lagged behind comparing to Britain and Belgium. In 1842 Britain had 1,900 miles of railways in operation while France only 300 (Lefranc, 1930).

After the success of the first line, the government of France understood the importance of a national railway network. However, the next attempt for a rail line between Paris and Versailles was unsuccessful. Two different companies created one line each to connect Paris with Versailles but they failed financially. The government recognized the problem of the rivalry of local interests (Dunham, 1941). Even though, the government was aware of the importance of a national railroad did not provide funding until 1842 (Ratcliffe, 1976).

According to Dunham (1941) the planning of rail network was given to $Corpsdespontsetchaussees$, an organization of highly trained engineers which was in charge for major infrastructure projects in France. The expansion of railroads during the period 1840-1860 took place based on Legrand Star, a design of $Corps\ des\ ponts\ et\ chaussees$. According to Legrand plan the purpose of the government was to connect major communes, borders, and coastlines with Paris. In 1865, based on the Migneret law, the government established direct connections lines among all prefectures. The final expansion of rail network constructed according to Freycinet plan which introduced in 1878 (Thévenin et al., 2013).

The most important railroad expansion occurred from 1860 to 1900 when 45,000 km of rail lines established. The final phase was until 1930 and by the end of this phase 56,000 km of railroad were operational (Thévenin et al., 2013). The introduction of motor vehicles in the end of 1930 limited the expansion of railroad network.

2.2 Patent System

Great Britain, France and the United States were the first countries that adopted a legislation on patents in 1791 (MacLeod, 1991; Galvez-Behar, 2019). Prior to the establishment of patent legislation and more specifically, during the 18th century the name of the patent applications were "les privileges" and they were granted by the king and registered by the parliament in Paris (MacLeod, 1991). In addition the inventors were receiving remarkable awards for their inventions. These awards included things like national or local production, pension for a life or exception from the taxes. Finally, a major difference with the English patent system is that the application procedure was much more severe in case of French patent system (MacLeod, 1991).

In 1791, Boufflers' bill became the first Patent Act in France. The 1791 law established three different categories for patenting: 1) Patents for invention, 2) Patents for improvement and 3) Patents for importation. The patents were granted from five until fifteen years. The name of the inventions changed to "brevets de invention". However, their cost was prohibitive. Apart from the initial cost, several taxes were added during the examination process. Patent system did not facilitate the spread of patents, because a lot of inventors could not afford the high cost (Galvez-Behar, 2019).

The rise of patenting activity was underpinned by the introduction of a patent law in 1844. This legislation allowed the inventors to export their patents in foreign countries and also changed the method of payment for the inventors. Now, patentees could spread the payment of the tax over the entire duration of their patent (Galvez-Behar, 2019).

3 Data

3.1 Setting

I start the collection of data by first extracting shapefiles of 19th century France. I make use of the recent constructed shapefiles of Gay (2020) over the period 1870-1940. Over my sample period, France does not experience large changes in the national administrative system except for the departments that become part of Germany. I rely on the shapefile of 1870 to conduct my analysis. The shapefile of 1870 contains 2,925 cantons for the mainland of France.

Next, I am able to determine the majority of the city areas in order to use the centroids of the cities. The first data source is Perret et al. (2015) and contains the polygons of the cities, towns, domains and forts during the 18th century. For a given canton i, I first use the cities within the polygon then the towns, the domains and the forts. If one canton has more than one city, I keep as a reference point the centroid of the largest (in terms of area) city. The same rule applies in case of a town, domain or fort. I determine the location of 1,961 cities, 69 towns, 112 domains and 3 forts which corresponds to 2145 cantons, 73,33% of the sample. I complement my analysis using the population raster file of 1840 from HYDE (2020). I use the year 1840 because it is the closest available year before the arrival of railroads. Based on the second source of data, I transform in QGIS the raster file to pixels and I choose for the remaining cantons, without a reference point, the pixel with the highest population value. Furthermore, I assign a reference point in 738 cantons, which corresponds to 25.23% of the sample. For the remaining 42 cantons without a reference point, I use the centroid of the polygon, 1.44% of the sample. Figure 1 illustrates an example of the empirical setting. The brown areas are the cantons and the pink areas are the city polygons. The black dots, within the pink areas, are the city centroids. The white dots are the rail stations while the blue line is the navigable waterway. Finally, concerning the black dots without a pink area, they are the points with the highest population value within the canton. I believe that using these black dots as reference points makes the analysis more realistic.

3.2 Transportation Data

3.2.1 Railroad Data

The data about the communes that have a rail station over the period 1860-1920 come from Mimeur et al. (2018). It is a very detailed database which contains five variables of interest for all the communes.[4]

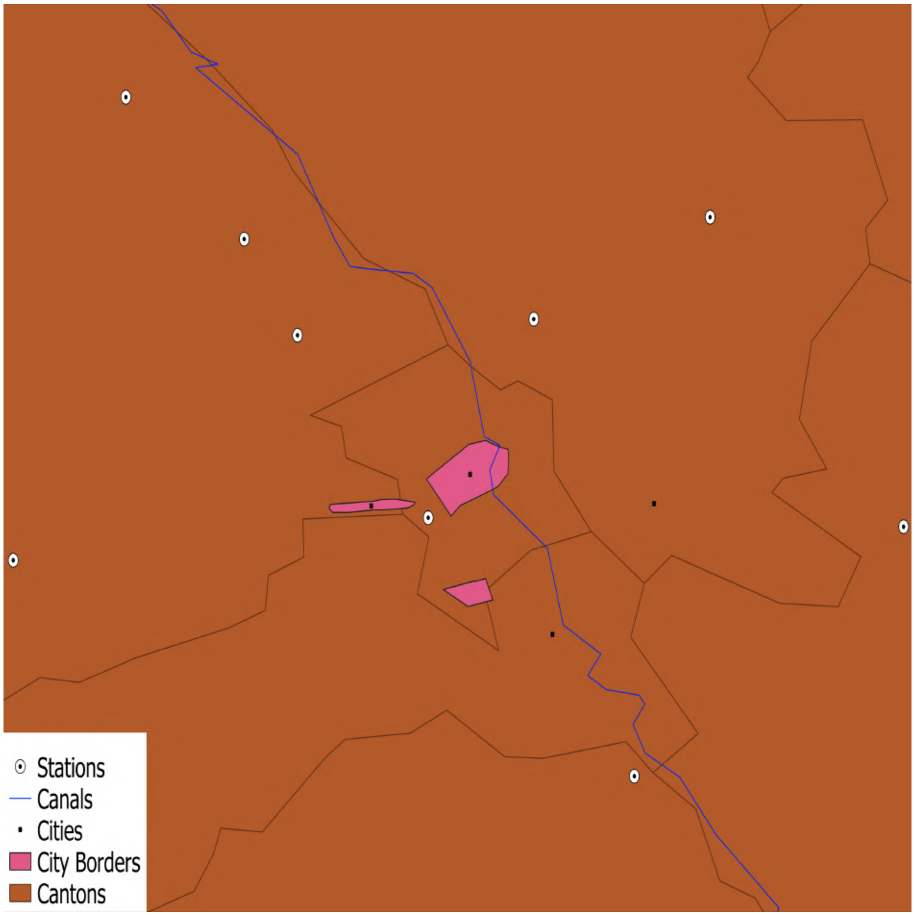
I make use of an historical image[5] with rail lines until 1860 to add one additional decade to my sample. I extract the lines from the image that were already established until 1850 to identify the communes that had a rail station in 1850. In order to do that, I rely on the historical rail station data of 1860 from Mimeur et al. (2018) and I preserve only the communes which are crossed by a rail line of 1850. This exercise allows me to identify the communes with a rail station in 1850.

Distance to Rail Station: It is the straight line distance in km of any canton city point to the closest commune centroid with access to a rail station.

[4]It contains a dummy variable of access (=1 if the commune has a rail station), a variable of the type of line, and three variables with travel time to reach from any commune to Paris, regional centers and departmental centers.

[5]For more details see Figure C1 in the Appendix C.

Figure 1: Setting



Notes: This figure illustrates the research setting of the paper.

Furthermore, I attempt to create the historical rail line network of France. In the absence of high resolution historical maps, I use the communes with rail stations over the period 1850-1890[6] and modern rail lines of France.[7]Baum-Snow et al. (2017, 2018) argue that historical infrastructure networks are good predictors and facilitate the construction of a lower cost and more modern infrastructure network. In this paper, I use a modern infrastructure network to re-create an old one. I use a shapefile with the rail lines over the period 2000-2020 (Jeansoulin, 2019) and a shapefile with the lines until 1992 (DIVA-GIS, 2020). Figure 2 presents the results of this exercise. According to the findings a significant development of rail network occurred in France over the 19th century.

[6]For more details see Figure C2 in the Appendix C.
[7]For more details see Figure C3 in the Appendix C.

3.2.2 Canal Data and Ports

In addition to the railroads, France developed a canal network. I make use two sources of data in order to identify the canal network of France.

The first source of navigable waterways comes from Ryavec and Henderson (2017) and their exercise 2 about Cities and Water Transportation in 19th Century France. It is a shapefile developed by Jordi Marti Henneberg (http://europa.udl.cat/projects/inland-waterways/). This shapefile includes the canal network of France in 1850. The second source of data is the collection of historical maps of University of Chicago (2020). It contains very detailed maps of navigable waterways of France over the second half of 19th century.[8]I digitized these historical maps in QGIS. Figure 2 shows the expansion of canals. Next, I construct the indicator of access to the canal network.

Access to Navigable Waterways: It is a binary indicator which switches to 1 for a canton if there is a canal within 3 km distance from the city centroid. In the Appendix A, I apply robustness check with different thresholds.

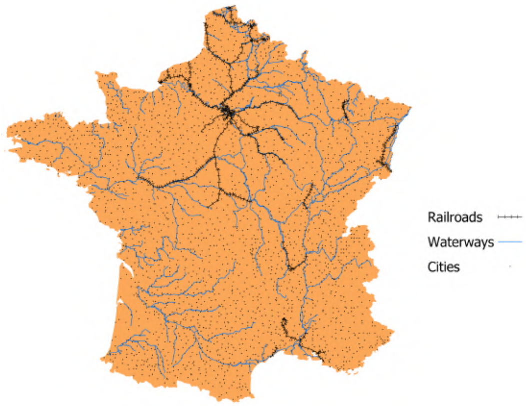
In addition, I make use of a historical map from David Rumsey (2020) collection to identify the ports of France in 1877.[9]I complement this data with active ports over the period 1662-1855 from García-Herrera et al. (2006). Their database is about the reconstruction of oceanic wind field patterns for this period that precedes the time in which anthropogenic influences on climate became evident. Nevertheless, they include the variables "voyagefrom" and "voyageto" which contain the names of the places where the ship departed from or sailed to. I apply geocoding on Stata (Zeigermann, 2018) and I identify 13 active ports[10] for France. I use the ports to construct the instrument.

3.3 Patent Data

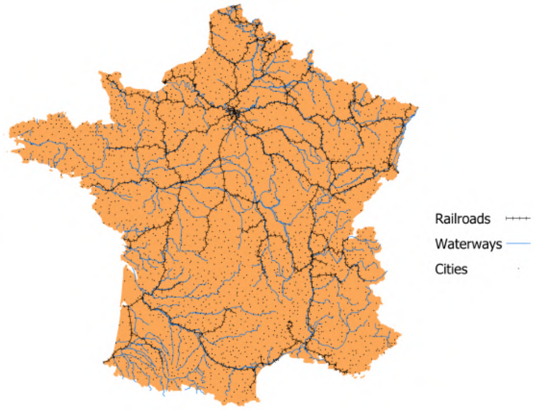
The historical database of INPI (National Institute of Industrial Property) contains all the patent applications (409,324 applications with their additions) covering the period 1791-1902. The database includes the application number of the patents, the filing year, the name of the applicant(s), the commune, the street and number of the applicant(s), the title of the patents and the expiration date of the patent applications. Additional details are included in the database like if an application is a patent that was imported from abroad, the number of additions of

[8]For an example of the maps see Figure C4 in the Appendix C.
[9]For more details see Figure C4, Appendix C.
[10]At least one ship departed or sailed to these ports.

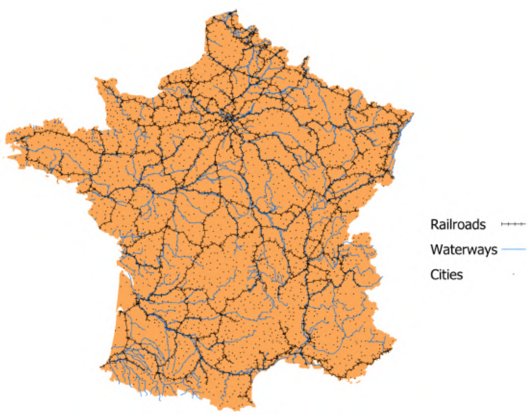
Figure 2: Expansion of Railroads and Navigable Waterways



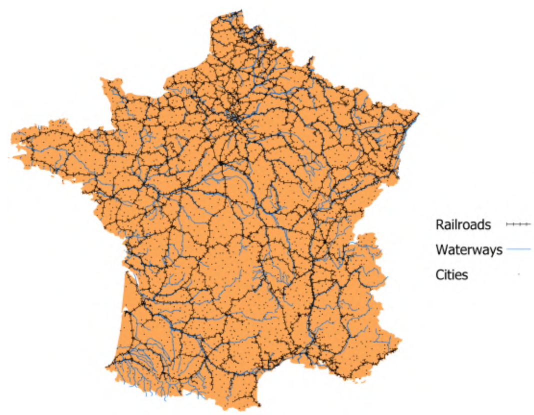
(**a**) 1850



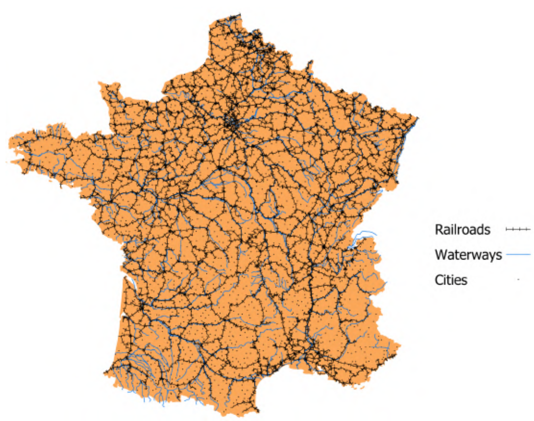
(**b**) 1860



(**c**) 1870



(**d**) 1880



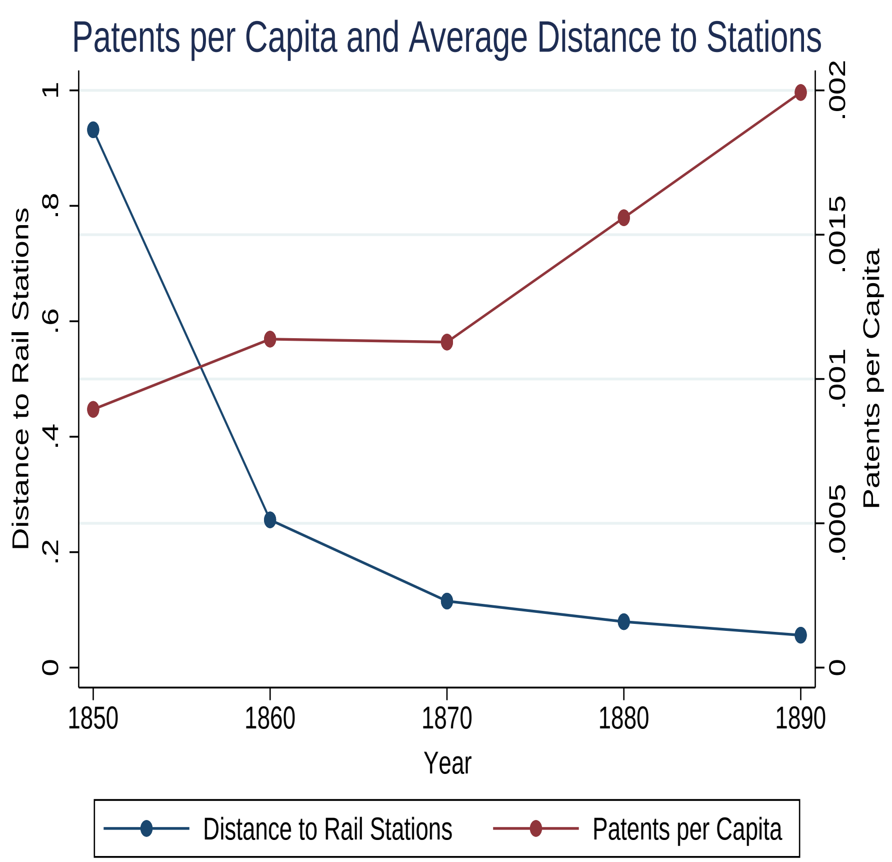
(**e**) 1890

Notes: This figure shows the expansion of navigable waterways for France over the time period 1850-1890. Source: the shapefile of canal network for 1850 comes from Ryavec and Henderson (2017). For the period 1860-1890, it is based on author's digitization of the historical maps from University of Chicago (2020). The lines of railroads are constructed by the author based on historical rail stations of Mimeur et al. (2018) and recent shapefiles of rail lines from DIVA-GIS (2020) and Jeansoulin (2019).

every patent application and the profession of the applicant.

I restrict my sample only to applications with at least one inventor residing in France. I remove applications that belong only to firms using key words.[11] Next, I exclude 116 applications without a filing date. When one application has several improvements, I do not take into account the improvements, but only the initial design. Finally, I exclude from the sample the patent applications that are importations since it is more likely the inventor to be resident of another country and only the patent agent to be in France. I end up with 308,926 patent applications over the entire period. Figure 3 presents the evolution of patent applications per capita over the sample of period of the paper as well the time trend for the average distance to rail stations.

Figure 3: Innovation per Capita and Average Distance to Rail Stations



Notes: Summary graph showing the time trend of patent applications per capita and average distance to rail station. Source: author's computations based on the patent applications from INPI (2019) database, population from HYDE (2020) and rail stations from Mimeur et al. (2018).

Number of patent applications: I calculate the sum of the patents[12] for each decade from 1850

[11]I exclude all the applications that have the words "COMPAN" or "COMPAGNI" or "SOCIETA" or "SOCIET" or "GESELLSCHAFT" or "SYNDICAT" or "GESELLSCHAF" or "MANUFACTUR" in the applicant name.

[12]It is the collapse sum command in Stata.

until 1890 like in Andersson et al. (2021).
Number of patent applications that received an addition: INPI database contains the number of additions of each patent application. The patent law of 1844 gave the applicants the possibility to apply for certificates of addition (Galvez-Behar, 2019). These certificates allowed the applicants to protect a minor improvement to the initial patent during its term. An addition of the initial design has an extra cost of 24 francs (Nuvolari et al., 2020). As a quality measure of innovation, I restrict the sample of patents to the ones that received an addition.

Apart from the main database, INPI includes a sample of 38,527 patent applications with their technological field. These patents are divided in 20 main classes.[13] For only this sample of patents, I have their main class and also their title. I use the "lsemantica" command in Stata (Schwarz, 2019) to extract key words from the titles of the patent applications. I complement these key words with additional words from the names of the main and sub classes of the patent applications. Then, I keep the unique words among classes to create a vocabulary. Based on this vocabulary, I manage to assign classes to the rest of patent applications in INPI database. More technical details regarding this method and the vocabulary can be found in the Appendix B. Figure 4 illustrates the time trend for every class. As a result of this analysis, I construct two more indicators of innovation.

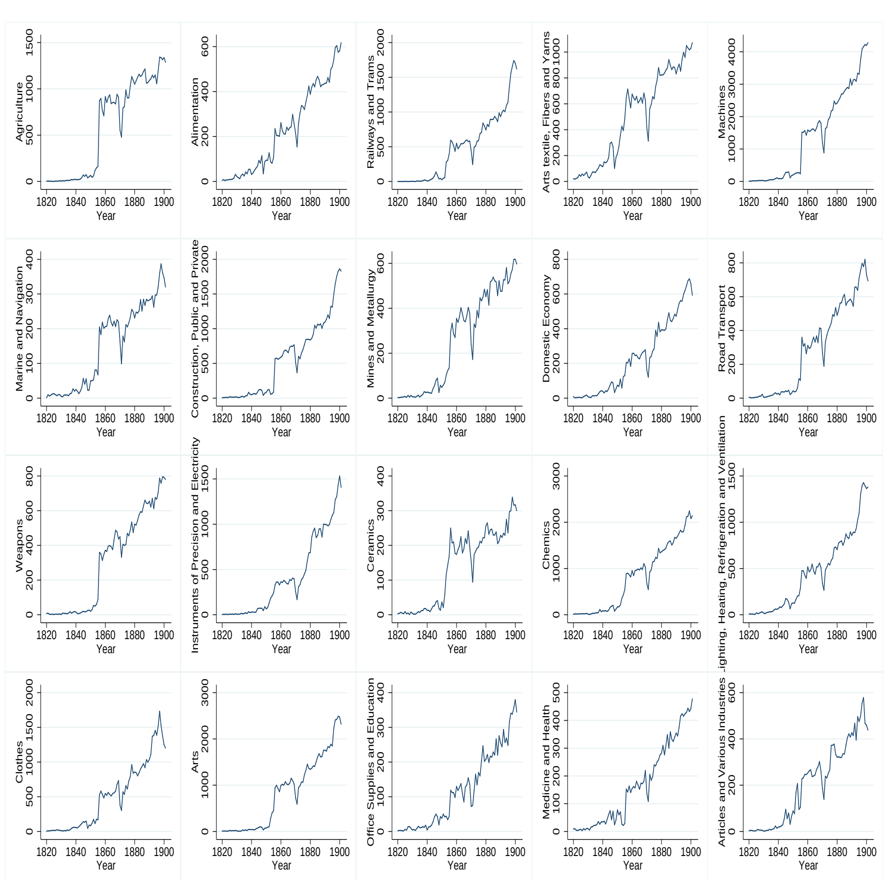
Number of Novel Classes: Number of Novel Classes it is the number of new technology fields that a canton has at least one invention in a given year
Stock of Classes: Stock of classes is for every year the number of technological classes that a canton already has at least one invention.

Next, in order to be sure that I do not double count inventors residing in the same canton, I apply a simple applicants' disambiguation based on the name surname and their address. This exercise allows me to assign a unique identifier to the inventors. I apply fuzzy matching (Raffo, 2020) on the name and surname of applicants residing in the same street. Then, I consider the applicants with a matching score more than 73% to be the same person. I identify 334,504 unique applicants from 353,017 patent applications. Even though the disambiguation is simple and "kills" the mobility of the inventors, since one of the criteria is applicant's address, it serves its purpose. In the sample every period is a decade and this simple disambiguation allows me to build an indicator of stock of inventors. One additional advantage of this method is that the stock of inventors is not be driven by outliers in the data. (Bahar et al., 2020)[14]

[13]The 20 classes are: Agriculture, Alimentation Railways and Trams, Arts textile, Fibers and Yarns, Machines, Marine and Navigation Construction, Public and Private, Mines and Metallurgy, Domestic Economy, Weapons, Road Transport, Instruments of Precision and Electricity, Ceramics, Chemics, Lighting Heating Refrigeration and Ventilation, Clothes, Arts, Office Supplies and Education, Medicine and Health, Articles and Various Industries.

[14]As already explained in Bahar et al. (2020) there could be fluctuations in the stock of inventors if one inventor

Figure 4: Applications by classes



Notes: Summary graph presenting the evolution of patent applications by classes. If an application has several classes, I allocate one application to all of them. The classes are: Agriculture, Alimentation Railways and Trams, Arts textile, Fibers and Yarns, Machines, Marine and Navigation Construction, Public and Private, Mines and Metallurgy, Domestic Economy, Weapons, Road Transport, Instruments of Precision and Electricity, Ceramics, Chemics, Lighting Heating Refrigeration and Ventilation, Clothes, Arts, Office Supplies and Education, Medicine and Health, Articles and Various Industries. Source: author's computations based on patent applications from INPI (2019).

3.4 Other Data

I rely on HYDE (2020) database to export (gridded) time series of population and land use. I rely on raster files to form the indicators in QGIS at the canton level. All the indicators have time variation. I extract data about population (defined as the population counts, in inhabitants/gridcell), cropland (defined as the total cropland area, in km2 per grid cell) and grazing (defined as the total land used for grazing, in km2 per grid cell). I use the baseline estimates of the database.[15]

In addition, I construct a binary indicator which switches to one if the canton has a university

residing in a commune has a patent in year $t-1$ and $t+1$ but not in year t. By taking the average of ten years these fluctuations are not an issue any more.

[15] I use the version 3.2 which was released in 04-08-2020.

or a grande ecole like in Lecce et al. (2021). I manually collect the locations of the universities from Ruegg (2004) and check if the universities were abolished during the French Revolution in 1793. Next, I extract manually the addresses of the grand ecole from the Conference des Grandes Ecoles (www.cge.asso.fr). I geolocalize in STATA the addresses of the universities and grand ecole using the command developed by Zeigermann (2018). Finally, I control for the number of post offices a canton has, using the database of Verdier and Chalonge (2018), to be in line with the literature about state capacity and innovation (Acemoglu et al., 2016).

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Innovation Variables					
INPI Applications	17.5745	776.0184	0	68635	14625
Probability to Innovate	0.4755	0.4994	0	1	14625
Additions	2.8867	107.7245	0	7232	14625
Number of Novel Classes	1.1625	1.9716	0	20	14625
Stock of Classes	3.9177	5.1607	0	20	14625
INPI Applications by technological class	1.9523	107.4823	0	25804	292500
Transportation Variables					
Distance to Rail Station	41.1569	75.6703	0.0608	594.5292	14625
Access to Waterways	0.0003	0.0185	0	1	14625
Travel Cost to Paris	7790.6959	7213.0369	0	92328.3125	14625
Travel Cost to Lyon	7498.2269	6672.2881	0	87256.3359	14625
Travel Cost to Marseille	9951.6185	6981.023	0	84630.1094	14625
Distance to Straight Line (Instrument)	28.6276	39.1383	0	339.3349	14625
Access to Knowledge	0.0002	0.0002	5.64e-06	0.0032	14625
Access to Knowledge by technological class	0.0001	0.0000453	1.00e-06	0.0013	292500
Control Variables					
Population	13244.6819	35836.6914	0	2524117	14625
Average Cropland Area	29.4377	16.1494	0	61.8082	14625
Average Grazing Area	18.3853	11.993	0	55.9528	14625
University	0.0116	0.1072	0	1	14625
Post Offices	2.1933	1.8605	0	40	14625
Robustness Analysis - Variables					
Travel Time to Paris	931.1978	625.1727	0	7817.8740	11700
Distance to a Discrete Station	183.4004	183.5866	0.1922	759.4374	14625
Distance to International Collaboration	100.2293	69.7108	0	409.8712	14625
Access to Knowledge based on Elevation	0.0002	0.0002	8.80e-07	0.0029	14625
Differences in Differences Model - Variables					
INPI Applications	9.3656	550.2264	0	68635	29250
Access to rail within 3km distance	0.1115	0.3148	0	1	29250
Access to rail within 5km distance	0.1532	0.3602	0	1	29250
Access to rail within 7km distance	0.1817	0.3856	0	1	29250
Population	12090.7981	27842.4077	0	2524117	29250
Average Grazing Area	18.302	11.9084	0	55.9528	29250
Average Cropland Area	28.6418	16.2708	0	61.9237	29250

Notes: Summary statistics for all the main variables (2,925 cantons, every 10 years is 1 period in the sample, 5 time periods in total.). Innovation variables: INPI Applications is the total number of patent applications, Probability to innovate is a binary indicator which switches to one if the canton has a patent application, Additions is the total number of patent applications that received an addition, Number of Novel Classes contains the number of novel technological classes that a canton has in a given year and Stock of classes the number of technological classes that a canton already has. Transportation variables: Distance to Rail Station is the distance (in kilometers) from any city centroid to the closest commune centroid with access to a rail station, Access to Waterways is defined as a dummy variable which takes the value 1 if the centroid of a canton is within 3 kilometers distance from the closest canal, travel cost to Paris, Lyon and Marseille is the computed travel cost based on rail lines and canals of every city centroid to Paris, Lyon and Marseille and distance to straight line is the instrument (in kilometers). Access to knowledge variables: Access to knowledge is the computed access of every canton based on the share of inventors over population and accessibility in rail lines and canals and market access is the computed market access of every canton based on the number of citizens and accessibility in rail lines and canals. Control variables: population is the total number of inhabitants, cropland is the average cropland area, grazing is the average land used for grazing, university is a binary variable which switches to 1 if the canton has a university and post offices is the number of post offices in a canton. The dimension of the variables INPI Applications by technological class and Access to Knowledge by technological class is 2,925 cantons, 20 technological classes every 10 years, 5 time periods in total. The time period for the differences in differences model is from 1800 to 1890.

4 Empirical Strategy

I start the analysis by estimating the main model through OLS regressions (Correia, 2015). The estimation equation is:

$$Pat_{it} = \alpha_0 + \beta DistStat_{it-1} + \gamma_i + \delta_t + \zeta X_{it} + \epsilon_{it} \tag{1}$$

where, Pat_{it} is the number of patents in the canton i, in time period t. The main variable of interest is $DistStat_{it-1}$, which contains the distance in km of any canton city centroid to the closest commune centroid with a rail station as computed in the previous period.[16] I include canton fixed effects, γ_i and year fixed effects, δ_t. X_{it} contains all the controls at the canton level such as the population, the average cropland area, the average grazing area, the number of post offices and the existence of a university.

The variable Pat_{it} and the population have been transformed using the log transformation[17] to reduce the effect of the extreme values in the sample (Squicciarini and Voigtländer, 2015). As a robustness test, I apply also a Poisson pseudo-maximum likelihood regressions model in the Appendix A without any transformation in the data.

However, I have to take into consideration the potential endogeneity problem due to omitted variable bias. The placement of the actual network may be endogenous and affected by unobservable local economic conditions (Andersson et al., 2021). To mitigate any concerns that could naturally arise, I complement the analysis with an instrumental variable approach and several robustness tests.

4.1 Endogeneity and Instrumental Variables

Following a similar strategy as in Katz (2018), I propose the use of a time variant instrument for the rail network of France. The identification strategy builds on straight lines (Perlman, 2015; Katz, 2018; Banerjee et al., 2020) as they represent the Euclidean (least cost) distance between two places. In addition, Dunham (1941) mentions about French rail network that "*Corps des ponts et chaussees* believed in making railroads as straight as possible, no matter what important centres of trade or industry they might by pass on the way". This team of highly trained engineers was not interested in trade or industry, nor in the problems of economics. The

[16] A one-lag structure for the effect of railway variable on innovation outcomes is intuitive in this framework because the railway stations constructed near the end of the calendar year are likely to affect innovation outcomes only in the following year (Melander, 2020).

[17] I add 1 in the variables: number of patent applications and population before I apply the log transformation.

above theoretical framework justifies the use of straight lines as an instrument in case of France.

In order to create time variation for the instrument, I rely on the 3 different French rail plans.
Legrand plan (1840-1860): The purpose of the government was to connect the major cities, borders, and coastlines to the capital of Paris (Thévenin et al., 2013). For the year 1850, I connect with straight lines Paris and with the rest of regional centers (major cities) as they are closer than the ports, Figure 5 (a). Next, for the year 1860, I make use of the data about French ports and I draw straight lines from Paris to the communes with access to a port (coastlines) Figure 5 (b).
Migneret law (1865): According to Thévenin et al. (2013) the second phase of railroad expansion involved connections among all departmental seats (prefectures). Based on this plan, I use a spanning tree to connect all the departmental centers with regional centers and ports, Figure 5 (c).
Freycinet plan (1878): The Freycinet plan has no effect on travel time to reach in Paris. Its purpose was to facilitate access to departmental centers (Mimeur et al., 2018) and to eliminate the regional disparities among rural and urban areas (Schwartz et al., 2011). According to Le Bris (2012) all the sub-prefectures should be connected to the railway network. For the year 1880, I connect all the sub-prefectures to the closest straight line. Finally, for the year 1890, I create an updated spanning tree among all nodal destinations. Then, I estimate the following equations:

$$DistStat_{it-1} = (Distance\ to\ Straight\ Line)_{it-1} + \alpha_0 + \gamma_i + \delta_t + \zeta X_{it} + \epsilon_{it} \tag{2}$$

and

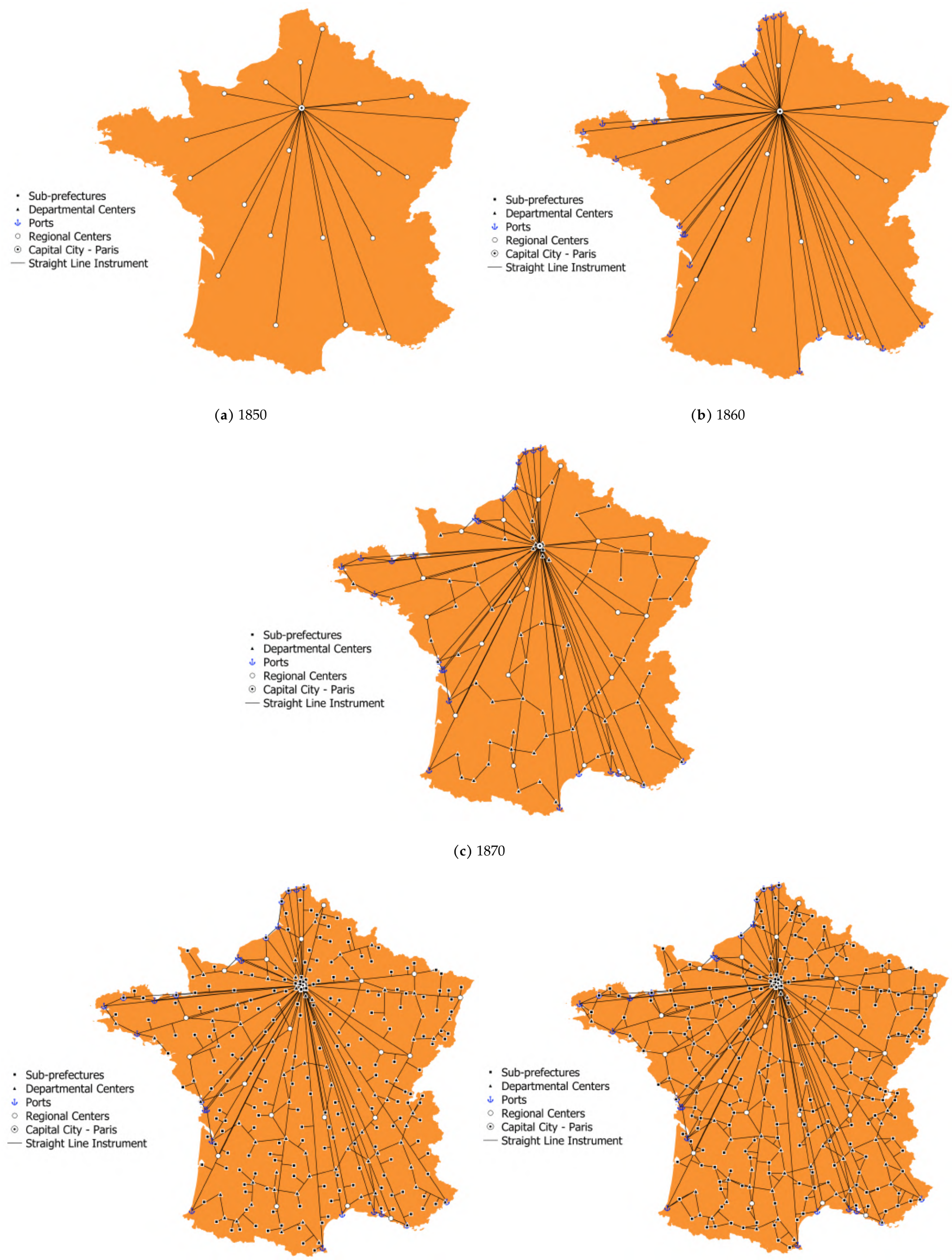
$$Pat_{it} = \alpha_0 + \beta Dist\hat{S}tat_{it-1} + \gamma_i + \delta_t + \zeta X_{it} + \epsilon_{it} \tag{3}$$

$Distance\ to\ Straight\ Line_{it-1}$ in equation 2 is the railway instrument. It includes for every time period the distance from any canton city centroid to the closest straight line. I use the "ivreghdfe" command of Correia (2018) in Stata. In all the regressions, I cluster the standard errors at the canton level.

5 Results

Table 2 provides the first-stage results. Distance to straight line is associated positively with the distance to a rail station. The positive relationship means that for a canton city centroid which is far away from the straight line instrument, it is also far away from an actual rail station. Straight line distance and distance to station are both expressed in kilometers. This means that a one standard deviation increase in the distance to the straight line corresponds to a 0.41% increase

Figure 5: Evolution of the Straight Line Instrument



(**a**) 1850 (**b**) 1860

(**c**) 1870

(**d**) 1880 (**e**) 1890

Notes: This figure illustrates the expansion of straight line instrument based on the regional centers in 1850, regional centers and ports in 1860, a spanning tree among regional centers, departmental centers and ports in 1870, subprefecture centers in 1880 and a spanning tree among all the nodal destinations in 1890. Source: author's computations.

in the distance to the actual rail stations. Finally, the instrument is highly significant.

Table 2: First stage: distance to straight line instrument and railway stations

Dep. var. =	Distance to Station
	(1)
Distance to Straight Line	0.413***
	[0.016]
Sample Size	14625
Canton FE	Yes
Year FE	Yes
Canton Controls	Yes

Notes: First stage regressions. The dependent variable is the distance to the nearest constructed railway station. Distance to straight line is the distance of any centroid to the closest straight line. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The population is transformed using a log transformation. The results are based on the equation 2. Clustered standard errors at the canton level are reported in the parenthesis.

Moving now to the main results, Table 3 presents the OLS and the IV results. The dependent variable in the first three columns is the number of patent applications. Both OLS and IV estimates are highly significant and negative, column 1 and 2 OLS and column 3 the IV. I rely on the IV estimates, column 3, to interpret the results. Since all the variables are standardized the interpretation is that for a given canton a one standard deviation decrease in the distance to a rail station is associated with a 0.049% increase in the number of patent applications. A one standard deviation increase in the population is associated with approximately 0.22% increase in the number of patent applications, column 3. In addition, in line with the literature, the number of post offices has a positive and significant effect meaning that a one standard deviation increase in the number of post offices in a canton corresponds to a 0.025% increase in the number of patent applications. Average cropland and grazing land have no significant effect on patenting rate. Finally, access to waterways does not affect innovation activity. In the Appendix, I use alternative indexes of accessibility to canals based on different thresholds. Finally, Kleibergen-Paap Wald rk F statistic is high meaning that the instrument performs well. In column 4 and 5, I use as dependent variable a binary indicator which switches to 1 if the canton has at least one patent application. Based on the findings of the IV regressions, column 6, a one standard deviation decrease in the distance to a rail station is associated with a 0.07% increase in the probability of a canton to innovate. Even though population has a strong impact on the number of patent applications it appears to have no effect on the probability of a canton to innovate. A possible explanation is that the population affects innovation performance through the theoretical channel of agglomeration and does not boost the innovation performance across cantons in contrast with a connection to a network.

Given that literature considers number of patents as a crude measure of innovation, at least in the recent years, (Aghion et al., 2019) I complement my analysis by introducing as a quality index of innovation the patents that received an addition. I present the results in the Appendix A. In addition, in the Appendix A, I include Poisson pseudo-maximum likelihood regressions models where I do not transform the dependent variable.

Table 3: Baseline results: distance to rail stations and innovation

Dep. var. =	Log (Applications+1)			Patent =1	
	(1)	(2)	(3)	(4)	(5)
	OLS	**OLS**	**IV**	**OLS**	**IV**
Distance to Rail Station	-0.069***	-0.074***	-0.049***	-0.075***	-0.071***
	[0.007]	[0.007]	[0.018]	[0.010]	[0.024]
Log (Population+1)		0.241***	0.223***	0.084	0.081
		[0.068]	[0.071]	[0.075]	[0.077]
Post Offices		0.026***	0.025**	0.041***	0.041***
		[0.010]	[0.010]	[0.013]	[0.013]
Average Cropland Area		-0.063	-0.065	-0.020	-0.020
		[0.095]	[0.096]	[0.127]	[0.128]
Average Grazing Area		0.025	0.028	0.024	0.024
		[0.051]	[0.051]	[0.072]	[0.072]
Access to Waterways		0.007	0.011	0.010	0.011
		[0.015]	[0.015]	[0.024]	[0.025]
University		-0.036	-0.038	-0.003	-0.003
		[0.040]	[0.040]	[0.006]	[0.006]
R-squared	0.80	0.80	-	0.50	-
Sample Size	14625	14625	14625	14625	14625
Canton FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap Wald rk F statistic	-	-	689.38	-	689.38
Kleibergen-Paap rk LM statistic Chi-sq(1)	-	-	273.595	-	273.595

Notes: Baseline results. Innovation is measured by the number of INPI patent applications in the first three columns and the probability for a given canton to innovate in the last two columns. Distance to rail station is the distance to the nearest constructed railway station. Population is the total number of inhabitants. Post offices is the number of post offices that a canton has. Average cropland area and average land used for grazing is the average cropland and grazing area in km2 per grid cell. Accessibility to waterways is a binary indicator which switches to 1 if there is a canal within 3 km distance from the city centroid. University is a binary indicator which switches to 1 if a canton has a university. The dependent variable INPI patent applications and the population are transformed using a log transformation. Column 1, 2 and 4 contain the OLS results based on the equation 1 and column 3 and 4 the IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

5.1 Robustness Analysis

5.1.1 Inconsequential Units Approach

To further mitigate any concerns, in this section, I adopt an inconsequential units approach as a robustness test. This method is widely used in the literature (Büchel and Kyburz, 2020; Möller and Zierer, 2018; Faber, 2014). The intuition is that in the early stages of transport infrastructure developments, major destinations are typically connected first. This IV approach relies on the randomly chosen subset of municipalities that received railway access because they lie on the most direct route between these nodal destinations. Even though, I follow the actual rail plans when I draw the straight lines the selection of the nodal destinations could be endogenous. To this end, I re-estimate equation 1 and equation 3, using the inconsequential units approach, to eliminate any concerns.

Based on this approach, I remove from the sample all the focal destinations that were used in the construction of the instrument. By doing that, I also exclude the capital city of Paris. This is crucial since it takes into consideration an another type of endogeneity. The only office of INPI was in Paris and as a result for an inventor it could be easier to use in the application file an address of a patent agent in Paris. This could introduce bias in the benchmark analysis, a bias that is possible not addressed by the instrument.

Moving to the results in Table 4, I re-estimate the equations 1 and 3 after I remove the nodal destinations. Again, I rely on the IV estimates to interpret the results. The magnitude is now lower comparing to the benchmark estimates in Table 3 but again significant at 5%. A possible explanation for the lower magnitude it could be the exclusion of the big urban centers from the sample.

5.1.2 Balance Test

In this section, I present the results of a balance test exercise. I re-estimate equation 3 but this time I use as dependent variable the number of patents from 1800 until 1840. The intuition behind this empirical exercise is that allows me to explore if there is a pre-trend effect of railroads on innovation performance before the actual sample period.

Table 5 presents the IV results. According to the findings distance to rail station has no effect on innovation performance. This result confirms that the distance to railroads variable is not associated with pre-trends regarding the patent activity.

5.1.3 Difference-in-Differences Model

Next, I apply as a robustness test a conventional difference-in-differences regression model with staggered treatment. I collect data about population, cropland and grazing area over the period 1800 until 1840 to complement the panel dataset of the benchmark analysis. I create three different variables of access to a rail station based on the distance to the closest commune with a rail station. The advantage of this method is that allows me to control for pre-trends since the

Table 4: Distance to rail stations and innovation - inconsequential units approach

Dep. var. =	Log (Applications+1)	
	(1)	(2)
	OLS	**IV**
Distance to Rail Station	-0.062***	-0.040**
	[0.007]	[0.016]
R-squared	0.64	-
Sample Size	13010	13010
Canton FE	Yes	Yes
Year FE	Yes	Yes
Canton Controls	Yes	Yes
Kleibergen-Paap Wald rk F statistic	-	620.68

Notes: I exclude from the sample all the destinations which I use to create the instrument. Innovation is measured by the number of INPI patent applications. Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables and the population are transformed using a log transformation. Column 1 contains the OLS results based on the equation 1 and column 2 the IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

arrival of railroads. I am able to do that by using the period 1800 to 1840 in which there is not any canton with access to a rail station. I estimate the following equation:

$$Pat_{it} = \alpha_0 + \beta Acc_{it-1} + \gamma_i + \delta_t + \zeta X_{it} + \epsilon_{it} \tag{4}$$

where Acc_{it-1} is the access variable switching to one if a canton has access to a rail station. I consider three different variables of access depending on the distance to the closest rail station. It can be 3, 5, or 7 km away. I include canton and year fixed effects and time variant controls like log of population, grazing area and cropland area. Table 6 summarizes the results. The sample period now is from 1800 until 1890. I find that access to a rail station has a positive and significant effect which confirms the results of the benchmark analysis.

5.1.4 Cross Sectional Analysis by decade

Finally, I apply cross sectional analysis to further explore the interplay between the distance to railroads and innovation performance. I include arrondissement fixed effects and cluster the standard errors at the department level. I estimate the following equation for each decade:

$$Pat_i = \alpha_0 + \beta Dis\hat{t}Stat_i + \gamma_i + \zeta X_i + \epsilon_i \tag{5}$$

Table 5: Distance to rail stations and innovation - Balance Test 1800-1840

Dep. var. =	Log (Applications+1)
	(1)
	IV
Distance to Rail Station	0.003
	[0.010]
Kleibergen-Paap Wald rk F statistic	689.38
Sample Size	14625
Canton FE	Yes
Year FE	Yes
Canton Controls	Yes

Notes: I repeat the same IV regression but this time the depedent variable is the number of patents from 1800 until 1840 one decade prior to the arrival of railroads. Innovation is measured by the number of INPI patent applications. Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variable and the population are transformed using a log transformation. IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

where γ_i is the arrondissement fixed effects. The rest of the variables are the same as in equation 3. Table 7 presents the results. According to the Panel A, the 2SLS estimates, the significant effect of the distance to railroads on innovation activity starts in 1870. In 1865 started the expansion of railroads based on the second plan, the Migneret law. It is the phase that the departmental seats joined also to the network through direct connections. In addition, the effect gets stronger in column 4 and 5. This means that the increase in innovation performance is not driven only by the large urban centers which got connected during the first decades. Finally, the instrument predicts quite well the expansion of railroads for every decade.

5.2 OLS vs IV estimates

This section attempts to shed more light in the gap between OLS and IV estimates in Table 3. IV estimates are lower than the corresponding OLS in every column. According to the literature the transportation improvements are not randomly assigned and this cause the difference in the OLS and IV coefficients (Redding and Turner, 2015). The OLS estimates capture the impact of transportation investments assigned through the existing political process while on the contrary the corresponding IV estimates the impact of transportation investments assigned through quasi-experimental variation. Baum-Snow (2007) explains in his paper that if the only source of endogeneity of actual lines to changes in the focal variables is through the effect of state and local governments, then a valid instrument should produce IV estimates smaller in magnitude than the corresponding OLS estimates.

Table 6: Distance to rail stations and innovation - Differences in Differences

Dep. var. =	Log (Applications+1)		
	(1)	(2)	(3)
Access to rail 3km distance	0.102***		
	[0.008]		
Access to rail 5km distance		0.103***	
		[0.008]	
Access to rail 7km distance			0.088***
			[0.008]
R-squared	0.67	0.67	0.67
Sample Size	29250	29250	29250
Canton FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes

Notes: Access to rail is a binary indicator switching to 1 if a canton has access to railroads within a given distance threshold. Innovation is measured by the number of INPI patent applications. Canton controls contain the population, the average cropland area and the average land used for grazing. The dependent variable and the population are transformed using a log transformation. Differences in differences model based on equation 4. Clustered standard errors at the canton level are reported in the parenthesis.

In case of France, indeed, the rail lines that were planned in the borders with Germany was an act of political interference. The reason was the Franco-Prussian war in 1870. According to Jordan W. Jonathan (2005) "the French government constructed long stretches of strategic railways in eastern France along the German border that served strategically crucial ends". The endogenous selection of these places as recipients of rail lines violates the assumption that transportation improvements are randomly assigned and introduces bias to the OLS estimates. Table 8 explores the effect of distance to a rail station on innovation activity after I remove from the sample areas that could gain access to a rail station for defensive reasons because they are close to the borders with Germany. The OLS coefficient is being reduced while the IV increases as we move from column 1 to 3. In column 3, after the exclusion from the sample the cantons that are within 35 km distance from the borders, the two coefficients have the same magnitude.

The war ended with the Treaty of Frankfurt on May of 1871. According to this Treaty the departments Bas-Rhin, Haut-Rhin, Moselle, one-third of the department of Meurthe, including the cities of Château-Salins and Sarrebourg, and the cantons Saales and Schirmeck in the department of Vosges became part of the German Empire. In the Appendix A, I re-estimate the equations 1 and 3 after I exclude the cantons that were affected from the war.

Table 7: Distance to rail stations and innovation by decade

Dep. var. =	Log (Applications+1)				
	(1)	(2)	(3)	(4)	(5)
Decade	**1850**	**1860**	**1870**	**1880**	**1890**
Panel A - Second Stage	**IV**	**IV**	**IV**	**IV**	**IV**
Distance to Rail Station	-0.304	-0.539	-2.475***	-6.417***	-7.014***
	[0.187]	[0.687]	[0.547]	[1.350]	[1.776]
Kleibergen-Paap Wald rk F statistic	38.55	14.91	38.05	26.25	22.82
Panel B - First Stage	**OLS**	**OLS**	**OLS**	**OLS**	**OLS**
Distance to Straight Line	0.207***	0.078***	0.115***	0.087***	0.069***
	[0.033]	[0.020]	[0.019]	[0.017]	[0.015]
Sample Size	2924	2924	2924	2924	2924
Canton FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes	Yes	Yes

Notes: Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Innovation is measured by the number of INPI patent applications. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variable and the population are transformed using a log transformation. Panel A contains the IV results based on the equation 5 and panel B the first stage results. Clustered standard errors at the department level are reported in the parenthesis.

6 Mechanisms

6.1 Access to Knowledge

Dittmar (2011) studies the adoption of printing from cities in 16th century. Printing was the major technological innovation of 16th century. The author finds that places with ports and cheap water transportation benefited more from this invention. The adoption of the printing press required face to face interactions and the cities close to water transportation receive higher benefits than comparing to the rest of the sample. Network connections could reduce the obstacles involved in knowledge diffusion (Breschi and Lissoni, 2009). Recent literature has shown that in contrast with physical stock of capital, human capital is not transferred easily and inventors interact with each other to combine their skills (Jones, 2009). In the absence of one of the collaborators, the remaining inventors lose in terms of patent production and earnings (Jaravel et al., 2018). An inventor acquires knowledge through interactions with others who are more knowledgeable than him (Akcigit et al., 2018) and this fact fosters innovation activity.

The intuition of the mechanism relies on a recently paper of Akcigit et al. (2018). The authors show that inventors built their knowledge and improve their skills by interacting with others and learning from them. The knowledge of the inventors could work as an input in the production function (like the R&D) and not just as an output (like a patent). I employ the share of inventors over population instead of population as a numerator in the traditional market access

Table 8: Distance to rail stations and cantons in the borders

Dep. var. =	Log (Applications+1)		
	(1)	(2)	(3)
Panel A	**OLS**	**OLS**	**OLS**
Distance to Rail Station	-0.069***	-0.068***	-0.067***
	[0.007]	[0.007]	[0.007]
R-squared	0.80	0.80	0.80
Panel B	**IV**	**IV**	**IV**
Distance to Rail Station	-0.059***	-0.061***	-0.065***
	[0.017]	[0.017]	[0.017]
Kleibergen-Paap Wald rk F statistic	708.68	712.28	716.22
Sample Size	14440	14420	14370
Canton FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes
Distance from the German Borders	25 km	30 km	35 km

Notes: Column 1 does not contain the cantons that are within 25 kilometers from the borders, column 2 30 kilometers and column 3 35 kilometers. Innovation is measured by the number of INPI patent applications. Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables and the population are transformed using a log transformation. Panel A contains the OLS results based on the equation 1 and panel B the IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

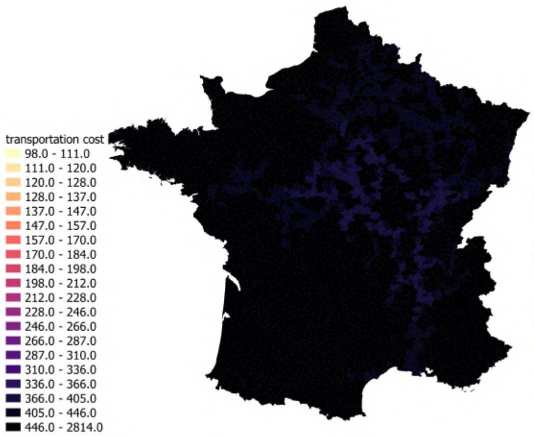
index of Donaldson and Hornbeck (2016). I call it access to knowledge since the purpose of this index is to explore how for a given canton *i* connectivity to other places where inventors reside affects the invention rate of the canton *i*. It is a supply side mechanism which relies on how important are the interactions and potential knowledge spillovers among inventors during the innovation process.

More details about the construction of the costs for the access to knowledge indicator can be found in the Appendix C. Figure 6 shows the accessibility of every canton based on railroads and canals over the period 1850-1890. The areas which are less accessible are the darker ones. As the network expands the areas become brighter.

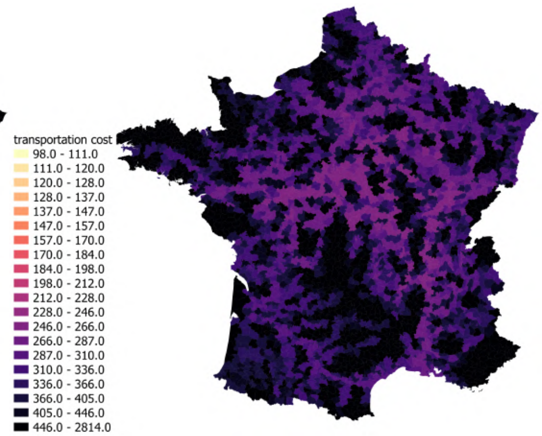
In order to disentangle the effect of knowledge spillovers from the market size, I divide the number of inventors of every canton with the population. I rely on the inventors' disambiguation as not to double count inventors who live in the same canton. Next, I form the access to knowledge index as:

$$Access\,to\,Knowledge_{it} = \sum_{h \neq i} \frac{Inventors_{ht}/Pop_{ht}}{Cost_{iht}} \tag{6}$$

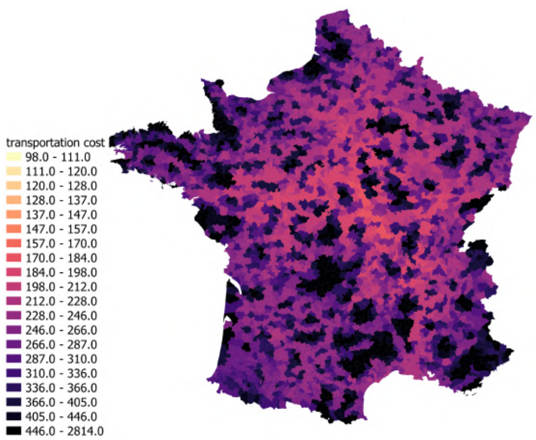
Figure 6: Transportation Cost and Accessibility of every Canton



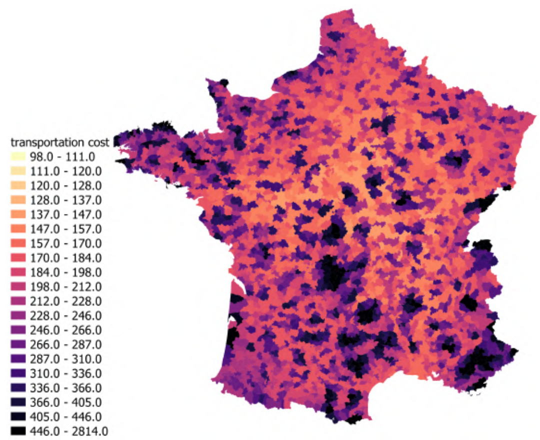
(**a**) 1850



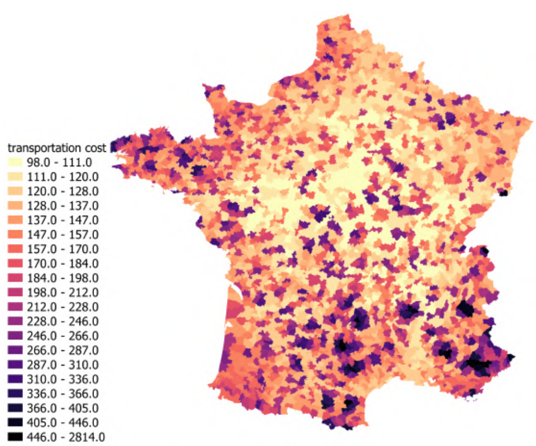
(**b**) 1860



(**c**) 1870



(**d**) 1880



(**e**) 1890

Notes: This figure shows the reduction in the transportation cost and the accessibility of every canton due to the expansion of railroads and canals over the time period 1850-1890. Source: author's computations based on the shapefiles of canal network and railroads.

where for a canton i access to knowledge is defined as the sum of the share of inventors over population residing in all the other cantons except i divided by the cost to reach in these cantons. To be in line with the rest of the analysis since the access to knowledge indicator contains inventors, it has been transformed using the log transformation.[18]

6.1.1 Estimation Method and Results

I estimate the following equation:

$$Pat_{it} = \alpha_0 + \beta AK_{it-1} + \gamma_i + \delta_t + \zeta X_{it}\epsilon_{it} \tag{7}$$

The independent variable, AK_{it}, is the index of access to knowledge. Second, I do not control for access to waterways of 19th century because I use the canals to compute the travel costs. The rest of the control variables are the population, the average cropland area, the average grazing area, the number of post offices and the existence of a university.

Table 9 summarizes the results. Column 1 explore the reduce form equation where I control only for canton and year fixed effects. A one standard deviation increase in the share of access to knowledge mechanism corresponds to a 0.060% increase in the number of patent applications. I include all the controls in column 2. The coefficient is now slightly higher. In the third column, I control also for market access, constructed in line with Donaldson and Hornbeck (2016) using the population of the cantons, in order to rule out alternative mechanisms. Again, the coefficient is positive and significant. Yet, the high increase in the coefficient of the variable could arise concerns related to multicollinearity. For this reason, in column 4, I re-estimate equation 7 but this time I add a third dimension in the model which is the technological class of the patent data. The updated equation is:

$$Pat_{ict} = \alpha_0 + \beta AK_{ict-1} + \kappa_{it} + \lambda_{ic} + \mu_{tc} + \epsilon_{ict} \tag{8}$$

This allows me to introduce three pairwise fixed effects: canton-year fixed effects, κ_{it}, canton-technological class fixed effects, λ_{ic}, and year-technological class fixed effects, μ_{tc}. These combinations of fixed effects deal with all possible variables that do not vary by canton, year and technology like market access. Given that, I find again in the Table 9 that access to knowledge has a positive and significant effect on the number of patent applications, column 4.

[18] I multiply both access to knowledge indicators with 10000000. These two indexes do not have zeros but according to Bellemare and Wichman (2020) the elasticities derived from log transformation hold for large enough average values of the variables. Their suggestions for applied econometricians is to use approximate elasticities for values of a variable no less than 10.

Table 9: Mechanism: access to knowledge and innovation

Dep. var. =	Log (Applications+1)			
	(1)	(2)	(3)	(4)
	OLS	**OLS**	**OLS**	**OLS**
Log Access to Knowledge	0.060***	0.064***	0.988***	0.174**
	[0.017]	[0.017]	[0.110]	[0.073]
R-squared	0.79	0.79	0.80	0.82
Sample Size	14625	14625	14625	292500
Canton FE	Yes	Yes	Yes	No
Year FE	Yes	Yes	Yes	No
Canton Controls	No	Yes	Yes	No
Log Population Market Access	No	No	Yes	No
Canton x Year FE	No	No	No	Yes
Canton x Technology FE	No	No	No	Yes
Year x Technology FE	No	No	No	Yes

Notes: Access to knowledge is the sum of the share of inventors over population residing in all the other cantons except i divided by the cost to reach in these cantons. Innovation is measured by the number of INPI patent applications. Canton controls contain the population, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variable, the index of access to knowledge, the index of market access and the population are transformed using a log transformation. OLS model based on the equations 7 and 8. Clustered standard errors at the canton level are reported in the parenthesis.

6.2 Diffusion of Novel Technologies and Access to a Global City

This section illustrates the special case of the connectivity to Paris. This mechanism is related to the arrival of new technologies in a canton. Perlman (2015) argues that local transportation access has no effect on how fast new technologies appear in a new county (the paper is about the USA). In this section, I build on her initial idea and I combine it with new literature based on the importance of global cities.

Larger cities are places with more idea exchanges between higher-ability participants (Davis and Dingel, 2019). According to Sassen (2001) a global city is a city generally considered to be an important node in the global economic system. Paris works as a gatekeeper that connects the national innovation system to global innovation networks (Miguelez et al., 2019). This is in line with recent findings that patents in Paris are more diverse and spread across the spectrum of IPC classes while, for instance, in Toulouse are more concentrated (Kogler et al., 2018). I hypothesize that in the absence of ICT technologies, due to the historical framework, the accessibility to a global city is the only crucial factor when it comes to the diffusion of novel technologies.[19] Finally, Paris was the first city in 1826 that had at least one invention in all the 20 technological classes. Second, instead of using words, like Perlman (2015), which are more difficult to capture similar inventions, I rely on patent classes.

6.2.1 Estimation Method and Results

Next, I describe the estimation method:

$$Number\,of\,Novel\,Technologies_{it} = \alpha_0 + \beta Travel\,Cost_{iut-1} + \gamma_i + \delta_t + \zeta X_{it} + \epsilon_{it} \tag{9}$$

$Number\,of\,Novel\,Technologies_{it}$ is a variable which counts the number of novel classes in a canton.[20] $Travel\,Cost_{itu}$ is the cost to reach from any canton centroid to a big urban center. u could be Paris, Lyon or Marseille. I do not control for the distance to waterways in 19th century, because I use them to compute the travel costs. I use also stock of classes fixed effects, a variable which takes into consideration the number of technological classes that a canton already has.[21] The rest of the control variables are the population, the average cropland area, the average grazing area the number of post offices and the existence of a university.

[19] Other types of communication could be the telegraphic communication or the telephone. Regarding, the telegraphic communication, according to the France Tél (2021) the French Post Office gradually absorbed the telegraph service and I control for the number post offices that a canton has. Furthermore, the diffusion of telephone in France occurred only in the end of 19th century. The year 1883 installed the first telephone exchange in Rheims while Société Générale du Telephones's telephone network was nationalized in September 1889. In 1911 there were only 0.6 telephones per 100 people in France.

[20] The minimum value is 0 while the maximum is 20, since 20 are all the possible technological classes.

[21] A canton with a high number of technological classes is less likely to adopt a novel technology.

Given the nature of the dependent variable[22], I estimate the new equation using the Poisson pseudo-maximum likelihood regressions model of Correia et al. (2019) which allows for a high dimension of fixed effects. I apply the log transformation only to population. I exclude from the sample the cantons that already have all the technological fields before the arrival of railroads. These are Paris, Lyon, Marseille, Lille, Rouen and Bordeaux since it is not possible to have a patent application in a novel field. Finally, I cluster the standard errors at the canton level.

Table 10 explores the effect of access in a global city on the diffusion of novel technologies. The main independent variable in column 1 is the travel cost from the city centroid of each canton to Paris. The coefficient of travel cost to Paris is highly significant and negative meaning that one standard deviation decrease in the cost to reach in Paris is associated with a 0.079% increase on the number of novel classes. Moving to column 2 and 3, I test the effect of travel cost to Lyon and Marseille on the number of novel classes respectively. According to the results, travel cost to other big urban centers do not have a significant effect on the number of novel classes. Finally, in column 4, I perform a regression using the travel costs to all nodal destinations. Again, travel cost to Paris is the only that is significant.

Table 10: Diffusion of novel technologies and access to a global city

Dep. var. =	Number of Novel Classes			
	(1)	(2)	(3)	(4)
	PPML	**PPML**	**PPML**	**PPML**
Travel Cost to Paris	-0.079***			-0.727***
	[0.028]			[0.145]
Travel Cost to Lyon		-0.035		0.424
		[0.026]		[0.291]
Travel Cost to Marseille			-0.017	0.230
			[0.028]	[0.213]
Sample Size	11484	11484	11484	11484
Canton FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes	Yes
Stock of Classes	Yes	Yes	Yes	Yes

Notes: I exclude from the sample the cantons of Paris, Lyon, Marseille, Lille, Rouen and Bordeaux because they already had all the technological classes before the arrival of railroads. The dependent variable is the number of novel technological classes in a canton. Travel cost to Paris, Lyon and Marseille is the computed travel cost to reach in Paris, Lyon and Marseille. Canton controls contain the population, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. Population transformed using a log transformation. PPML model based on the equation 9 with clustered standard errors at the canton level are reported in the parenthesis.

[22]The 63.25% of the values in the dependent variable are zeros. According to Bellemare and Wichman (2020) as a rule of thumb, they state that if the data has more than one third zero-valued observations, it is perhaps best to explicitly apply Tobit or zero-inflated Poisson or negative binomial model.

One possible concern could be that connectivity to Paris is associated with a lower cost for registration of the patent applications and has nothing to do with the diffusion of novel technologies. Since, the patent office of INPI was in Paris a faster connection, because of the railroads, could facilitate the registration of the patent applications by the inventors. According to Galvez-Behar (2019) there were patent agents in Paris that could fill in the patent applications on behalf of the inventors. Even after the construction of railroads and the reduction in transportation costs, it could be much easier for the inventors just to send a letter to the patent agents through a post office and to avoid to face bureaucratic procedures. Given that, controlling for the number of post offices should take into consideration this potential concern. Furthermore, I repeat the same regressions by excluding the Seine department and the region of Ile de France in the Appendix, Table A4, because for the inventors that were living close to Paris it could be possible the same cost to apply through a post office or to go to INPI's office. The effect of transportation to Paris remains strongly significant.

6.3 Limitations of the Access to Knowledge Approach

The use of freight rates makes comparable the transportation through canals with the transportation through railroads. However, according to Donaldson and Hornbeck (2016), there are several important limitations that someone has to take into consideration when uses a similar approach to the market access approach based on freight rates. This paper attempts to address several of these issues.

To begin with, Donaldson and Hornbeck (2016) use Fogel's average national rates and these rates remain constant over the entire sample period. According to the authors freight rates may vary with local demand and market power in the transportation sector. In this paper, I take into consideration the improvements in the transportation sector and I allow the freight rates to vary over time. As I document in the Appendix C, there is a significant reduction in the freight rates between years. The initial value for canals is 4.1, in 1850, and in the end of the sample period, 1890, is 1.8. The decrease of railway cost is even more intense. In 1850 the value of railroads is 14.5 and in 1890 is 6.8.

Second, the authors report that there are no congestion effects or economies of scale in transporting goods. They do not control for locations where trains can turn or switch tracks, so actual railroad transportation routes may be less direct. Third, they do not take into consideration the speed of the lines. In order to deal with these issues, at least partially, I make use of an additional variable computed in Mimeur et al. (2018). According to the authors, this variable is a discrete measure which combines two information: the number of axes converging in a station and the quality of the infrastructure. Next, I compute for every canton most populous point the straight line distance to the rail stations that combine the above two characteristics and I re-estimate the equation 6 included this variable.

Table 11: Access to Knowledge index and innovation - distance to discrete stations

Dep. var. =	Log (Applications+1)	
	(1)	(2)
	OLS	**OLS**
Log Access to Knowledge	0.056***	0.058***
	[0.017]	[0.017]
Distance to a Discrete Station	-0.079***	-0.085***
	[0.012]	[0.012]
R-squared	0.79	0.80
Sample Size	14625	14625
Canton FE	Yes	Yes
Year FE	Yes	Yes
Canton Controls	No	Yes

Notes: Innovation is measured by the number of INPI patent applications. Access to knowledge is the sum of the share of inventors over population residing in all the other cantons except *i* divided by the cost to reach in these cantons. Distance to a discrete station controls for the straight line distance in meters from the most populous point of a canton to a station that combines information based on the fastest lines and the number of axes converging in. Canton controls contain the population, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables, the index of access to knowledge and the population are transformed using a log transformation. OLS model based on the equation 7. Clustered standard errors at the canton level are reported in the parenthesis.

Table 11 presents the results. Distance to a discrete station is significant and negative meaning that cantons which are close to a discrete station witness an increase in the number of INPI applications. On the other hand, the magnitude of access to knowledge is very close to the baseline estimates reported in Table 9.

Next, I make use of the an additional variable in Mimeur et al. (2018) to confirm the effect of Paris as a global city on the diffusion of novel technologies. This variable contains the time duration of the fastest route for all the communes in France to reach in Paris over the period 1860-1890. It relies on the graph theory definition of shortest path using time rather than length as edge weight. It combines information based on the fastest lines and the number of axes converging in a station. However, the variable does not take into account the frequency of train at the time and the transition from the walking network to the train network is considered instantaneous. Since this study uses the cantons as level of analysis, I first associate every commune centroid with the closest canton most populous point in order to match the two different level of analysis and then, for every canton *i* I compute the average travel time based on the communes that this canton i contains. By definition this measure does not take into consideration the canal network.

Table 12: Diffusion of novel technologies and access to a global city based on travel time

Dep. var. =	Number of Novel Classes
	(1)
	PPML
Travel Time to Paris	-0.296***
	[0.063]
Sample Size	8702
Canton FE	Yes
Year FE	Yes
Canton Controls	Yes
Stock of Classes	Yes

Notes: I exclude from the sample the cantons of Paris, Lyon, Marseille, Lille, Rouen and Bordeaux because they already had all the technological classes before the arrival of railroads. The dependent variable is the number of novel technological classes in a canton. Travel time to Paris is the ability to reach in Paris based on the time duration of the fastest route. Canton controls contain the population, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. Population transformed using a log transformation. PPML model based on the equation 9 with clustered standard errors at the canton level are reported in the parenthesis.

In line with the baseline results in Table 10, I find that accessibility to the city of Paris has a strong effect on the number of novel technologies, Table 12. The difference with Table 10 is that the magnitude of the coefficient of travel cost to Paris is much larger. One explanation is that this cost is computed only based on the rail lines and not in canals and it could over estimate the impact of railroads.

Another one limitation is that the network is restricted to transportation linkages within France. The computation of transportation cost includes only the cantons of France and does not take into consideration the exposure of the cantons to international patent collaborations. I exploit the richness of the patent database and I am able to track and exclude from the sample all the patent applications that come as a result of an international collaborations. Then, I compute again the number of patent applications at the canton level. Furthermore, I include as a control the straight line distance in kilometers from any most populous canton point to the closest canton point with a patent application based on international collaboration. I control for the straight line distance because areas that are close to the cantons which benefited from exposure to international patent collaborations could also be benefited. Table 13 provides the results. I find that even though straight line distance to a canton with an international collaboration has a negative effect it is not significant. On the other hand, the variables of interest, distance to rail stations and access to knowledge, have a similar effect with the benchmark analysis.

Finally, as in Donaldson and Hornbeck (2016), the freight rates are not allowed to vary by direction. However, it could be more costly to cross areas that are in high altitude by train or boat. To this end, I compute in QGIS the least cost paths but this time I weight the cost values of the cantons that have an altitude value more than 200 with their elevation rate. Figure C5 in the Appendix C presents these cantons of France. I extract the elevation raster file from DIVA-GIS (2020). Table 14 summarizes the results with the weighted cost. The magnitude of the coefficient access to knowledge is a little bit lower comparing to the baseline estimates in Table 9 but again highly significant.

Table 13: Access to Knowledge and innovation - exposure to international collaborations

Dep. var. =	Log (Applications+1)		
	(1)	(2)	(3)
	OLS	**IV**	**OLS**
Distance to Rail Station	-0.074***	-0.048***	
	[0.007]	[0.018]	
Log Access to Knowledge			0.065***
			[0.017]
Distance to International Collaboration	-0.075	-0.070	-0.064
	[0.054]	[0.054]	[0.054]
R-squared	0.80	-	0.79
Sample Size	14625	14625	14625
Canton FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Kleibergen-Paap Wald rk F statistic	-	698.49	-
Kleibergen-Paap rk LM statistic Chi-sq(1)	-	274.314	-
Canton Controls	Yes	Yes	Yes

Notes: Innovation is measured by the number of INPI patent applications except the ones that are result of an international collaboration. Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Canton controls contain the population, the average cropland area, average land used for grazing, the existence of a university and the number of post offices. Additional control is the distance to a canton with an international patent application. The dependent variables, the indexes of access to knowledge and the population are transformed using a log transformation. OLS and IV model based on the equation 1 and equation 3 for column 1 and column 2 and on the equation 7 for columns 3 and 4. Clustered standard errors at the canton level are reported in the parenthesis.

6.4 Removal of Railroads - Back of the Envelope Exercise

This section provides addition evidence about importance of railroads on innovation performance. I assume that railroads have never been constructed and I rely on transportation by canals in order to create this back of the envelope exercise. Next, I compute in QGIS the least cost paths among cities in the absence of railroads using the same cost values for canals as in the baseline model. I find that the new access to knowledge measure based on canals has on average 26.37% lower values (with a standard deviation of 24.17) comparing to the access to knowledge measure produced by the equation 7.

Table 14: Access to Knowledge index based on elevation and innovation

Dep. var. =	Log (Applications+1)
	(1)
	OLS
Log Access to Knowledge	0.054***
	[0.017]
R-squared	0.79
Sample Size	14625
Canton FE	Yes
Year FE	Yes
Canton Controls	Yes

Notes: Innovation is measured by the number of INPI patent applications. I weight the cost of the access to knowledge indexes by elevation. The dependent variables, the indexes of access to knowledge and the population are transformed using a log transformation. OLS model based on the equation 1. Clustered standard errors at the canton level are reported in the parenthesis.

Following the same approach as in Donaldson and Hornbeck (2016), I store the coefficient of the baseline regression of access to knowledge on innovation, Table 9 column 1, and I use it together with the updated index of access to knowledge, keeping everything else constant, to predict which would be the number of patents in the absence of railroads. I find that in the absence of railroads the invention rate of the French cantons would have been, on average, 21.3% lower, Table 16.

Table 16: Back of the Envelope Exercise based on Canals

Time Period	**Percent Lower - Invention Rate**	**Standard Errors**	**Obs.**
1850-1890	21.3	(5.227993)	14,625

Notes: This Table presents the counterfactual impact on invention rate from the removal of railroads. Robust standard errors clustered by canton are reported in parentheses.

The purpose of this exercise is to promote the importance of railroads for innovation in 19th century France but it is not a counterfactual scenario. In a counterfactual scenario other things has to be taken into consideration like the re-allocation of population like in Donaldson and Hornbeck (2016).

7 Concluding Remarks

In this paper, I make use of two historical databases, a recent constructed database of rail stations (Mimeur et al., 2018) and the database of the French National Institute of Industrial Property to study the effect of access to railroads on patent activity. I exploit the enormous expansion of railroads in France over the second half of 19th century to document that access to rail network fosters the patent activity at the canton level. Controlling for navigable waterways and the number of postal offices, I am able to capture the net effect of rail network. Finally, the results are also confirmed by an IV strategy and several robustness tests regarding the validity of the instrument.

I explore access to knowledge as an underlying mechanism behind the results. I compute accessibility measures for the largest cities of the 2,925 French cantons. Findings suggest that for a given canton i, the reduction of the transportation costs to all the other cantons with high percentage of inventors has a positive impact on patenting activity of the canton i. I state, that this effect is mainly driven by potential interactions among inventors residing in different cantons. Back of the envelope exercise based on canals provides additional evidence that in the absence of railroads the invention rate of the French cantons would have been, on average, 21.3% lower.

In the last part of the paper, I use text analysis techniques to determine for the first time the technological class of each patent application in the historical database of INPI. Equipped with this dataset this paper documents that less costly access to the global city of Paris, compared to other large urban centers, is associated with a higher probability for a given canton i to innovate in a novel patent class. The intuition behind this effect is that Paris works as a gatekeeper of knowledge (Miguelez et al., 2019) that connects the national innovation system to global innovation networks.

This paper contributes to a growing literature about infrastructure and economic outcomes from a historical perspective and more specifically to the papers about railroads and innovation activity. It also adds to the literature about the industrial era in France. In terms of policy, the paper shows that general network contributes to the increase of innovation. Yet, in the case of diffusion of novel technologies what matters more is connections to global cities. This is crucial, especially in our days with the intense concentration of complex activities in big cities (Balland et al., 2020).

References

Abrams, David, Ufuk Akcigit, and Jillian Grennan, "Patent Value and Citations: Creative Destruction or Strategic Disruption?," *NBER Working Paper 19647*, 2018.

Acemoglu, Daron, Jacob Moscona, and James A. Robinson, "State capacity and American technology: Evidence from the nineteenth century," *American Economic Review*, 2016, *106* (5), 61–67.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Banque De France, Richard Blundell, and David Hemous, "Innovation and Top Income Inequality," *Review of Economic Studies*, 2019, *86* (1), 1–45.

Agrawal, Ajay, Alberto Galasso, and Alexander Oettl, "Roads and innovation," *Review of Economics and Statistics*, 2017, *99* (3), 417–434.

Akcigit, Ufuk, Santiago Caicedo Soler, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi, "Dancing with the Stars: Innovation Through Interactions," *NBER Working Paper 24466*, 2018.

Andersson, David, Thor Berger, and Erik Prawitz, "Making a Market: Infrastructure, Integration, and the Rise of Innovation," *The Review of Economics and Statistics*, 2021, pp. 1–44.

Atack, Jeremy and Robert Margo, "Agricultural Improvements and Access to Rail Transportation: The American Midwest as a Test Case, 1850-1860," *NBER Working Paper 15520*, 2009.

_, Fred Bateman, Michael Haines, and Robert Margo, "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60," *NBER Working Paper 14640*, 2009.

_, Michael Haines, and Robert Margo, "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70," *NBER Working Paper 14410*, 2008.

Bahar, Dany, Prithwiraj Choudhury, and Hillel Rapoport, "Migrant inventors and the technological advantage of nations," *Research Policy*, 2020, p. 103947.

Balland, Pierre Alexandre, Cristian Jara-Figueroa, Sergio G. Petralia, Mathieu P.A. Steijn, David L. Rigby, and César A. Hidalgo, "Complex economic activities concentrate in large cities," *Nature Human Behaviour*, 2020, *4* (3), 248–254.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 2020, p. 102442.

—, —, Gilles Postel-Vinay, and Tim Watts, "Long-run health impacts of income shocks: Wine and phylloxera in nineteenth-century France," *Review of Economics and Statistics*, 2010, *92* (4), 714–728.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *Quarterly Journal of Economics*, 2007, *122* (2), 775–805.

—, J. Vernon Henderson, Matthew A. Turner, Qinghua Zhang, and Loren Brandt, "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2018.

—, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang, "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 2017, *99* (3), 435–448.

Bellemare, Marc F. and Casey J. Wichman, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Berger, Thor and Kerstin Enflo, "Locomotives of local growth: The short- and long-term impact of railroads in Sweden," *Journal of Urban Economics*, 2017, *98*, 124–138.

Bogart, Dan, Xuesheng You, Eduard J. Alvarez-Palau, Max Satchell, and Leigh Shaw-Taylor, "Railways, divergence, and structural change in 19th century England and Wales," *Journal of Urban Economics*, 2022, *128*, 103390.

Breschi, Stefano and Francesco Lissoni, "Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows," *Journal of Economic Geography*, 2009, *9* (4), 439–468.

Büchel, Konstantin and Stephan Kyburz, "Fast track to growth? Railway access, population growth and local displacement in 19th century Switzerland," *Journal of Economic Geography*, 2020, *20* (1), 155–195.

Correia, Sergio, "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix *," 2015.

—, "IVREGHDFE: Stata module for extended instrumental variable regressions with multiple levels of fixed effects," *Statistical Software Components*, 2018.

—, Paulo Guimarães, and Thomas Zylkin, "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," *Papers 1903.01690, arXiv.org*, 2019.

Crescenzi, Riccardo, Marco Di Cataldo, and Andrés Rodríguez-Pose, "Government Quality and the Economic Returns of Transport Infrastructure Investment in European Regions," *Journal of Regional Science*, 2016, *56* (4), 555–582.

Cui, Jingbo, Tianqi Li, and Zhenxuan Wang, "High-Speed Railway and Collaborative Innovation: Evidence from University Patents in China," *SSRN Electronic Journal*, 2020.

Daudin, Guillaume, "Domestic trade and market size in late-eighteenth-century France," *Journal of Economic History*, 2010, *70* (3), 716–743.

_ , Raphaël Franck, and Hillel Rapoport, "Can Internal Migration Foster the Convergence in Regional Fertility Rates? Evidence from 19th Century France," *Economic Journal*, 2019, *129* (620), 1618–1692.

David Rumsey, "David Rumsey Map Collection," 2020.

Davis, Donald R. and Jonathan I. Dingel, "A spatial knowledge economy†," *American Economic Review*, 2019, *109* (1), 153–170.

Díez-Minguela, Alfonso and M. Teresa Sanchis Llopis, "Regional income inequality in France 1860–1954: Methods and findings," *Historical Methods*, 2019, *52* (1), 1–14.

Dijkstra, E. W., "A note on two problems in connexion with graphs," *Numerische Mathematik*, 1959, *1* (1), 269–271.

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

DIVA-GIS, "DIVA-GIS | free, simple effective," 2020.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Dong, Xiaofang, Siqi Zheng, and Matthew E. Kahn, "The role of transportation speed in facilitating high skilled teamwork across cities," *Journal of Urban Economics*, 2020, *115*, 103212.

Dunham, Arthur L., "How the First French Railways Were Planned," *Journal of Economic History*, 1941, *1* (1), 12–25.

Duranton, Gilles and Matthew A. Turner, "Urban Growth and Transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440.

Faber, Benjamin, “Trade integration, market size, and industrialization: Evidence from China’s national trunk highway system,” *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Flyvbjerg, Bent, “Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it,” *Oxford Review of Economic Policy*, 2009, *25* (3), 344–367.

France Tél, “France Tél | Encyclopedia.com,” 2021.

Franck, Raphaël and Oded Galor, “Flowers of evil? Industrialization and long run development,” *Journal of Monetary Economics*, 2019.

_ **and Stelios Michalopoulos**, “Emigration during the French Revolution: Consequences in the Short and Longue Durée,” *NBER Working Paper 23936*, 2017.

Galvez-Behar, Gabriel, “The patent system during the French industrial revolution: Institutional change and economic effects,” *Economic History Yearbook*, 2019, *60* (1), 31–56.

Gao, Yanyan and Jianghuai Zheng, “The impact of high-speed rail on innovation: An empirical test of the companion innovation hypothesis of transportation improvement with China’s manufacturing firms,” *World Development*, 2020, *127*, 104838.

García-Herrera, Ricardo, Gunther P. Können, Dennis A. Wheeler, Maria R. Prieto, Phips D. Jones, and Frits B. Koek, “Ship logbooks help analyze pre-instrumental climate,” *Eos, Transactions American Geophysical Union*, 2006, *87* (18), 173.

Gay, Victor, “Mapping the Third Republic. A Geographic Information System of France (1870–1940).,” *SSRN Electronic Journal*, 2020.

Hanley, Douglas, Jiancheng Li, and Mingqin Wu, “High-speed railways and collaborative innovation,” *Regional Science and Urban Economics*, 2021, p. 103717.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm, “The Making of the Modern Metropolis: Evidence from London*,” *Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Hornbeck, Richard and Martin Rotemberg, “Railroads, Reallocation, and the Rise of American Manufacturing,” *NBER Working Paper 26594*, 2019.

Huang, Yue and Yebin Wang, “How does high-speed railway affect green innovation efficiency? A perspective of innovation factor mobility,” *Journal of Cleaner Production*, 2020, *265*, 121623.

HYDE, “History Database of the Global Environment - the Netherlands Environmental Assessment Agency (PBL),” 2020.

Inoue, Hiroyasu and Kentaro Nakajima, “The Impact of the Opening of High-Speed Rail on Innovation,” *RIETI Discussion Paper Series 17-E-034*, 2017.

INPI, "INPI (Institut National De La Propriété Industruelle)," 2019.

Jaffe, A. B., M. Trajtenberg, and R. Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Jaravel, Xavier, Neviana Petkova, and Alex Bell, "Team-specific capital and innovation," *American Economic Review*, 2018, *108* (4-5), 1034–1073.

Jeansoulin, Robert, "rail data (railway lines, stations) - France Dataverse 2000-2020," *Harvard Dataverse, V1*, 2019.

Jones, Benjamin F., "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

Jordan W. Jonathan, "Reign of the Railroads," in Rod Paschall, ed., *The Quarterly Journal of Military History*, 4 ed., Vol. 17, Primedia Inc, 2005.

Juhász, Réka, "Temporary protection and technology adoption: Evidence from the Napoleonic Blockade," *American Economic Review*, 2018, *108* (11), 3339–3376.

_ , Mara Squicciarini, and Nico Voigtländer, "Technology Adoption and Productivity Growth: Evidence from Industrialization in France," *NBER Working Paper 27503*, 2020.

Katz, Ori, "Railroads, Economic Development, and the Demographic Transition in the United States," *MPRA Paper*, 2018, (88869).

Khan, B. Zorina, "Invisible Women: Entrepreneurship, Innovation, and Family Firms in Nineteenth-Century France," *Journal of Economic History*, 2016, *76* (1), 163–195.

Kogler, Dieter Franz, Gaston Heimeriks, and Loet Leydesdorff, "Patent portfolio analysis of cities: statistics and maps of technological inventiveness," *European Planning Studies*, 2018, *26* (11), 2256–2278.

Komikado, Hiroshi, So Morikawa, Ayushman Bhatt, and Hironori Kato, "High-speed rail, inter-regional accessibility, and regional innovation: Evidence from Japan," *Technological Forecasting and Social Change*, 2021, *167*, 120697.

Le Bris, David, "Les grands travaux du plan freycinet : De la subvention à la dépression ?," *Entreprises et Histoire*, 2012, *69* (4), 8–26.

Lecce, Giampaolo, Laura Ogliari, and Mara P. Squicciarini, "Birth and migration of scientists: Does religiosity matter? Evidence from 19th-century France," *Journal of Economic Behavior and Organization*, 2021, *187*, 274–289.

Lefranc, Georges, "The French Railroads, 1823-1842," *Journal of Business and Economic History*, 1930, *2*, 299–331.

MacLeod, Christine, "The Paradoxes of Patenting: Invention and Its Diffusion in 18th- and 19th-Century Britain, France, and North America," *Technology and Culture*, 1991, *32* (4), 885.

McCormick, Michael, Guoping Huang, Giovanni Zambotti, and Jessica Lavash, "Roman Road Network (version 2008)," *Harvard University, Cambridge MA 02138*, 2013, *DARMC Scho* (DARMC, Center for Geographic Analysis,).

Melander, Eric, "Transportation Technology, Individual Mobility and Social Mobilisation," *CAGE Online Working Paper Series*, 2020, (471).

Michaels, Guy and Ferdinand Rauch, "Resetting the Urban Network: 117-2012," *Economic Journal*, feb 2018, *128* (608), 378–412.

Miguelez, Ernest, Julio Raffo, Christian Chacua, Massimiliano Coda-Zabetta, Deyun Yin, Francesco Lissoni, and Gianluca Tarasconi, "Tied in: the Global Network of Local Innovation," *Cahiers du GREThA (2007-2019)*, 2019.

Mimeur, Christophe, François Queyroi, Arnaud Banos, and Thomas Thévenin, "Revisiting the structuring effect of transportation infrastructure: An empirical approach with the French railway network from 1860 to 1910," *Historical Methods*, 2018, *51* (2), 65–81.

Mojica, Laia and Jordi Martí-Henneberg, "Railways and population distribution: France, Spain, and Portugal, 1870-2000," *Journal of Interdisciplinary History*, 2011, *42* (1), 15–28.

Möller, J. and M. Zierer, "Autobahns and jobs: A regional study using historical instrumental variables," *Journal of Urban Economics*, 2018, *103*, 18–33.

Nagy, David, "City location and economic development," *2016 Meeting Papers*, 2016, (307).

Nicholas, David, *The growth of the medieval city : from late antiquity to the early fourteenth century*, Longman, 1997.

Nuvolari, Alessandro, Gaspare Tortorici, and Michelangelo Vasta, "British-French technology transfer from the Revolution to Louis Philippe (1791-1844): evidence from patent data," *CEPR Discussion Papers*, 2020.

Okoye, Dozie, Roland Pongou, and Tite Yokossi, "New technology, better economy? The heterogeneous impact of colonial railroads in Nigeria," *Journal of Development Economics*, 2019, *140*, 320–354.

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

Perlman, Elisabeth Ruth, "Dense Enough To Be Brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America," *CEH Discussion Papers*, 2015, (036).

Perret, Julien, Maurizio Gribaudi, and Marc Barthelemy, "Roads and cities of 18th century France," *Scientific Data*, 2015, *2* (1), 1–7.

Piketty, Thomas, Gilles Postel-Vinay, and Jean Laurent Rosenthal, "Wealth concentration in a developing economy: Paris and France, 1807-1994," *American Economic Review*, 2006, *96* (1), 236–256.

Pontarollo, Nicola and Roberto Ricciuti, "Railways and manufacturing productivity in Italy after unification," *Journal of Regional Science*, 2020, p. jors.12484.

Raffo, Julio, "MATCHIT: Stata module to match two datasets based on similar text patterns," *Statistical Software Components*, 2020.

Ratcliffe, Barrie, "Railway Imperialism: the example of the Pereires' Paris-Saint-Germain Company, 1835-1864," *Business History*, 1976, *18*, 66–84.

Redding, Stephen J. and Matthew A. Turner, "Transportation Costs and the Spatial Organization of Economic Activity," in "Handbook of Regional and Urban Economics," Vol. 5, Elsevier B.V., 2015, pp. 1339–1398.

Ruegg, Walter., *Universities in the nineteenth and early twentieth centuries (1800-1945)*, Cambridge University Press, 2004.

Ryavec, KE and M Henderson, "Mapping Political Economies over Time, GIS Exercise 2: Cities and Water Transportation in 19th Century France," *University of California, Merced*, 2017.

Santamaria, Marta, "Reshaping Infrastructure : Evidence from the division of Germany," *CAGE Online Working Paper Series*, 2020, (456).

Sassen, Saskia., *The global city : New York, London, Tokyo*, Princeton University Press, 2001.

Schwartz, Robert, Ian Gregory, and Thomas Thévenin, "Spatial history: Railways, uneven development, and population change in France and Great Britain, 1850-1914," *Journal of Interdisciplinary History*, 2011, *42* (1), 53–88.

Schwarz, Carlo, "lsemantica: A command for text similarity based on latent semantic analysis," *Stata Journal: Promoting communications on statistics and Stata*, 2019, *19* (1), 129–142.

Smith, Cecil O., "The Longest Run: Public Engineers and Planning in France," *American Historical Review*, 1990, *95* (3), 657–692.

Squicciarini, Mara P., "Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France," *American Economic Review*, 2020, *110* (11), 3454–3491.

_ **and Nico Voigtländer**, "Human capital and industrialization: Evidence from the age of enlightenment," *Quarterly Journal of Economics*, nov 2015, *130* (4), 1825–1883.

Sun, Daxin, Saixing Zeng, Hanyang Ma, and Jonathan J. Shi, "How Do High-Speed Railways Spur Innovationx003F;," *IEEE Transactions on Engineering Management*, 2021.

Talandier, Magali, Valérie Jousseaume, and Bernard-Henri Nicot, "Two centuries of economic territorial dynamics: the case of France," *Regional Studies, Regional Science*, 2016, *3* (1), 67–87.

Thévenin, Thomas, Robert Schwartz, and Loïc Sapet, "Mapping the distortions in time and space: The French railway network 1830-1930," *Historical Methods*, 2013, *46* (3), 134–143.

Toutain, J., "Les transports en France de 1830 à 1965," *Cahiers de l'ISEA*, 1967, *9* (8), 7–306.

University of Chicago, "University of Chicago Collections," 2020.

Verdier, Nicolas and Ludovic Chalonge, "The issue of scales in geohistory. Post offices from the 18th century to the present day," *Cybergeo : Revue européenne de géographie / European journal of geography, UMR 8504 Géographie-cités*, 2018.

Wong, Jason C Y, "Blue-sky Thinking: Connectivity Impacts on Regional Economies and Innovation in the United States *," 2019.

World Bank, "A Decade of Action in Transport : An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005," *Washington, DC: World Bank Publications*, 2007.

_ , "Improving Institutional Capability and Financial Viability to Sustain Transport," *Washington, DC: World Bank Publications*, 2013.

_ , "Mobile Metropolises: Urban Transport Matters," *Washington, DC: World Bank Publications*, 2017.

Zeigermann, Lars, "OPENCAGEGEO: Stata module for forward and reverse geocoding using the OpenCage Geocoder API," 2018.

Zimran, Ariell, "Transportation and Health in the Antebellum United States 1820-1847," *NBER Working Paper 24943*, 2019.

Zobl, Franz Xaver, "Regional economic development under trade liberalisation, technological change and market access: evidence from 19th century France and Belgium," *PhD thesis, The London School of Economics and Political Science (LSE)*, 2018.

Appendix A: Additional Findings

This section provides additional findings that confirm the benchmark analysis. Table A1 summarizes the results of the patents that have an addition as a quality index of innovation. The results hold for all the three models. Yet, it should be taken into consideration that this is a very crude measure based only on the additional cost for an inventor to have an addition. [23]

Next, I estimate as a robustness test a PPML model in Table A2 to take into account, a possible large number of zeros in the dependent variables. The dependent variables are now the INPI Applications without any transformation. The results are again highly significant at 1% column 1 and column 2 and 5% in column 2.

In Table A3, I re-estimate equation 1 and equation 3 after I exclude from the sample the canton that became part of Germany in the end of the Franco-Prussian War in 1870. I rely on the shapefile of 1880 from Gay (2020) to identify these cantons. Again the results, both OLS and IV are highly significant.

Road Network of 18th Century

Historical infrastructure networks may be correlate with the construction of modern infrastructure networks (Baum-Snow et al., 2017, 2018). This is true in case of France. According to Smith (1990), the same organization, the *Corps des ponts et chaussees*, was responsible for the construction of the Legrand Star rail plan of 1842 and the royal highways of 1770. They continued from the highway system in the eighteenth century through the waterways and railroads of the nineteenth century to electric power and economic planning in the twentieth. The purpose of the rail network was to imitate the improved highway system of the eighteenth and early nineteenth centuries (Schwartz et al., 2011). The road network of 18th century seems as a good potential instrument. However, there are two main drawbacks: the absence of the time variation, since the road network is static compared to the rail network and possible correlation direct or indirect with outcomes of interest (Baum-Snow et al., 2017, 2018). These reasons discourage the use it as an instrument. In contrast, I incorporate it into the analysis to control for pre-trends in infrastructure network since the arrival of railroads. I extract data about road network of France during the 18th century from Perret et al. (2015). I use only the main road network. Figure in the Appendix C presents the old network. Based on this data, I create an indicator of accessibility in the infrastructure network of 18th century.

[23] It is common in the recent time period the production of defensive patent applications from the firms which do not reflect valuable inventions and the way to deal with this issue is the use of patent citations (Abrams et al., 2018). A possible reason could be the high cost of patent applications.

Transportation of 18th century: I compute straight line distance of any canton city centroid to the closest infrastructure network line of 18th century.

Roman Road Network

Recent literature finds an association among Roman roads and modern urban network in case of France. Michaels and Rauch (2018) explore the resetting of the urban network for Britain and France. They find that France's urban network was largely shaped by its Roman origins. In their online Appendix, they argue that bishops had crucial roles in the evolution of the town life in France after the fall of the Western Roman Empire. According to Nicholas (1997) bishops were instrumental in the survival of towns in France after the end of the Roman Empire. I extract data for the main Roman network from McCormick et al. (2013). Figure in the Appendix C presents the Roman road network for France. Next, I define accessibility to Roman road network.

Roman Road Network: I compute straight line distance of any canton city centroid to the closest line of Roman road.

Table A4 summarizes the results. In the first three columns, I use different thresholds for the binary indicator access to navigable waterways. Comparing to the results in Table 3 they are similar in magnitude even though the results now are significant at 5%. In column 5, I control for the distance to canals, a measure similar to the main independent variable distance to rail stations. There is a reduction in the magnitude but the coefficient is still highly significant. Finally, in the last column, I add the additional variables, distance to road network of 18th century and distance to Roman road network multiplied with time dummies. The coefficient of the IV estimator is similar to the one in Table 3.

Table A5 presents the results for the diffusion of novel technologies model when I remove from the sample the department of Seine in column 1 which I use the transportation cost based on canals and rails and in column 3 which I use the travel time measure based on railroads. In columns 2 and 4, I remove the entire region of Ile de France. The results are highly significant. This robustness test confirms that the results are not driven by inventors who are close to Paris and are facilitated by the fact that can now travel to INPI's office faster for bureaucratic procedures which are not related to the diffusion of novel technologies.

Table A1: Distance to rail stations and quality indexes of innovation

Dep. var. =	Log (Additions+1)		Additions
	(1)	(2)	(3)
	OLS	**IV**	**PPML**
Distance to Rail Station	-0.051***	-0.032*	-0.135**
	[0.007]	[0.019]	[0.063]
R-squared	0.76	-	-
Sample Size	14625	14625	7550
Canton FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes
Kleibergen-Paap Wald rk F statistic	-	689.38	-

Notes: Innovation is measured by the number of INPI patent applications that have an addition. Distance to rail station is the distance from the most populous point of a canton to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables and the population are transformed using a log transformation. Column 1 contains the OLS results based on the equation 1 column 2 the IV results based on the equation 3 and column 3 the PPML results. Clustered standard errors at the canton level are reported in the parenthesis.

Table A2: Distance to rail station and innovation - PPML model

Dep. var. =	Applications
	(1)
	OLS
Distance to Rail Station	-0.143**
	[0.063]
Sample Size	11830
Canton FE	Yes
Year FE	Yes
Canton Controls	Yes

Notes: Innovation is measured by the number of INPI patent applications. Distance to rail station is the distance to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The population is transformed using a log transformation. PPML model. Clustered standard errors at the canton level are reported in the parenthesis.

Table A3: Distance to rail stations and innovation - German Empire

Dep. var. =	Log (Applications+1)	
	(1)	(2)
	OLS	**IV**
Distance to Rail Station	-0.061***	-0.073***
	[0.007]	[0.016]
R-squared	0.81	0.01
Sample Size	14145	14145
Canton FE	Yes	Yes
Year FE	Yes	Yes
Canton Controls	Yes	Yes
Kleibergen-Paap Wald rk F statistic	-	732.91

Notes: Innovation is measured by the number of INPI patent applications in the first column. Distance to rail station is the distance to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables and the population are transformed using a log transformation. Panel A contains the OLS results based on the equation 1 and panel B the IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

Table A4: Distance to rail stations and innovation - Other types of Networks

Dep. var. =	Log (Applications+1)				
	(1)	(2)	(3)	(4)	(5)
Panel A	**OLS**	**OLS**	**OLS**	**OLS**	**OLS**
Distance to Rail Station	-0.073***	-0.073***	-0.073***	-0.073***	-0.075***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
R-squared	0.80	0.80	0.80	0.80	0.80
Panel B	**IV**	**IV**	**IV**	**IV**	**IV**
Distance to Rail Station	-0.048***	-0.047**	-0.047**	-0.038**	-0.048**
	[0.018]	[0.018]	[0.019]	[0.018]	[0.019]
Kleibergen-Paap Wald rk F statistic	690.35	691.14	692.09	723.22	580.36
Sample Size	14625	14625	14625	14625	14625
Canton FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes	Yes	Yes
Access to Waterways	5km	7km	9km	No	No
Distance to Waterways	No	No	No	Yes	Yes
Distance to other Road Networks x Time Dummies	No	No	No	No	Yes

Notes: In the first column, I include a variable which switches to 1 if a city is within 5 km from a canal, in column 2 if it is within 7 km from a canal, in column 3 if it is within 9 km from a canal and in fourth column I control for the distance to navigable waterways. In the last column, I control for the distance to other road networks like the distance to roman road network and to the road network of 18th century. I have multiply the distances with time dummies. Innovation is measured by the number of INPI patent applications. Distance to rail station is the distance to the nearest constructed railway station. Canton controls contain the population, the accessibility to waterways, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. The dependent variables and the population are transformed using a log transformation. Panel A contains the OLS results based on the equation 1 and panel B the IV results based on the equation 3. Clustered standard errors at the canton level are reported in the parenthesis.

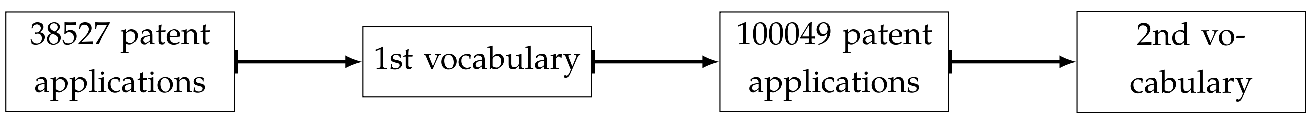
Table A5: Diffusion of novel technologies and access to a global city - Robustness

Dep. var. =	Number of Novel Classes			
	(1)	(2)	(3)	(4)
	PPML	**PPML**	**PPML**	**PPML**
Travel Cost to Paris	-0.080***	-0.069**		
	[0.028]	[0.028]		
Travel Time to Paris			-0.296***	-0.290***
			[0.063]	[0.064]
Sample Size	11472	11174	8702	8478
Canton FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Canton Controls	Yes	Yes	Yes	Yes
Stock of Classes	Yes	Yes	Yes	Yes

Notes: I exclude from the sample the cantons of Paris, Lyon, Marseille, Lille, Rouen and Bordeaux because they already had all the technological classes before the arrival of railroads. In addition, I exclude from the sample the department of Seine in columns 1 and 3 and the whole Region of Ile de France in columns 2 and 4. The dependent variable is the number of novel technological classes in a canton. Travel cost to Paris is the computed travel cost to reach in Paris. Travel time to Paris is the ability to reach in Paris based on the time duration of the fastest route. Canton controls contain the population, the average cropland area, the average land used for grazing, the existence of a university and the number of post offices. Population transformed using a log transformation. PPML model based on the equation 9 with clustered standard errors at the canton level are reported in the parenthesis.

Appendix B: Assignment of Classes

INPI provides me a sample of 38,527 (out of 308,513) patent applications that already have a patent class. I use these patents to create a vocabulary which helps me to assign classes to the rest of patent applications. Even though INPI contains a subclass, I use only the main class. I start my analysis by splitting the sample of patents based on their main class. The 35,663 patent applications have a unique class. Using the "lsemantica" command in Stata (Schwarz, 2019), for a given class i, I keep the words from the title of patent applications that appear at least 3 times[24] in the sample of patents of the class i. I exclude from the vocabulary the words with length less than 2 letters. Finally, I keep in the vocabulary the words that are unique among classes. Then, I use this vocabulary to assign classes to the rest of patent applications. Based on this methodology, I manage to delegate classes in 138,611 patent applications out of 290,761. From these 138,611 patent applications the 100,049 have a unique class. I repeat the same analysis but this time, I expand my initial vocabulary based on the 100,049 patent applications with a unique class. I complement the vocabulary with key words from the titles of the classes and sub classes of the patents.



I repeat the same procedure manually until there are in the sample 28,645 patent applications with no class. However, the allocation of patent classes could be bias from the number of words in each individual step. For instance, it could be possible agricultutae, to have more key words in the 1st vocabulary than weapons because it has a higher number of patent applications in the initial sample. This bias could be generate more patent applications for agriculture in the second step. A second problem regarding this methodology, is that for instance in the second step machines has more applications, which leads to more words but if the class machines had these words in the first step it may even had more patent applications. To this end, after I sum up the words from all the vocabularies and keep these which are unique among classes, I repeat the procedure from the beginning using only the final vocabulary. This again leaves me with the 28,645 patent applications with no class. For these patent applications, I apply fuzzy matching in Stata (Raffo, 2020) between the key words and the titles of the patent applications. I keep the key word which has the maximum similarity score with the patent title and then I assign a class according to this key word[25]. shows the number of patent applications in the initial and final sample by main class.

[24]Only for class 11 I keep the words that appear 2 times because it has a very limited number of patents

[25]If more key words has the same maximum similarity score I assign two classes in the patent application.

Table B1: Initial and Final Statistics by Main Class

main class	name	initial applications	initial percentage	final applications	final percentage
1	Agriculture	1206	3.13	44168	6.71
2	Alimentation	1751	4.54	17460	2.65
3	Railways and Trams	1693	4.38	36788	5.59
4	Arts textile Fibers and Yarns	5526	14.34	38411	5.84
5	Machines	4017	10.42	107912	16.4
6	Marine and Navigation	947	2.46	11732	1.78
7	Construction Public and Private	1783	4.63	42037	6.38
8	Mines and Metallurgy	1209	3.14	19356	2.94
9	Domestic Economy	1394	3.61	18107	2.75
10	Weapons	1013	2.63	22125	3.35
11	Road Transport	641	1.66	23211	3.53
12	Instrument of Precision and Electricity	1716	4.44	31152	4.73
13	Ceramics	932	2.42	10607	1.61
14	Chemics	3901	10.12	60939	9.26
15	Lighting Heating Refrigeration and Ventilation	2860	7.42	34452	5.23
16	Clothes	2279	5.91	38898	5.91
17	Arts	2492	6.47	62577	9.51
18	Office Supplies and Education	714	1.85	9292	1.41
19	Medicine and Health	1065	2.76	12721	1.93
20	Articles and Various Industries	1388	3.59	16038	2.44

Notes: This table presents the initial number of patent applications by main class, second column, the percentage of the patent applications by class in the initial sample, third column, the final number of patent applications by main class, fourth column and the final percentage of patent application by main class, fifth column. When a patent application has several classes, I allocate one patent application to each class.

Vocabulary

Agriculture: abri agricole agricoles agricultur animau araire aratoire aratoires arracher arrosoir avoine basée batteuse battoir beche bestiaux bineuse brabant brouette brouettes bétail bêche charancon charançons chariot charrue charrues colza cultiver culture dessèchement décanteur décortication défoncement défoncer défricher dépiquer ecuries engrai ensemencement etrille extirpateur extraire faire faucher faucheuse feve fleau fléau foin foins fourrage fourrages fourragère fourragères fruit fruitier fruitiers fumure fève fèves gelee gelée granuleuse granuleuses greffage greffe greffer greffoir grisou grêle hache haricot haricots harpon herse houe huitre huîtres jardinage l'herbe labour labourage labourer liege liège luzerne magnanerie maladie moissonner moissonneur moissonneuse monder nourriture oidium oïdium peche phosphorée phylloxera phylloxéra phylloxérée phylloxérées piège pièges plantation plante planter planteur plantoir pois prairie prairies printemp printemps pulverulent pulvérulent pyrale rateau ratière rats rural râteau râteaux récoltes secateur semence semences semer semoir siege siège soc socs soufrage soufre soufrer soufreuse souri souricière souris sulfatage sulfate sulfater sécateur taille taupe taupes traire trefle trèfle vegetal versoir viand viande viandes vigne végétal végétale végétative échala échalas écosser écurie écuries égrainer égrener étrille

Alimentation: aliment amandes anille arome arôme avarie avariés baratt baratte barattes baril barils barrique barriques beurr beurre bluter bluterie boulanger boulangers bruts brûloir cannelle cannelles claircage clairçage comestible comestibles confire confiture crible cribles cristallisable cristalliser crues depot douves défécation dépôt fecule ferment ferments froid froide fromage fromages fécale fécule fécules gazeificateur gazéificateur gland gruau insufflation levain levur levure limonade lors maceration macérateur macération marrons melasse menthe meuneri meunerie moka mousseuse mousseuses mousseux moutarde mélasses patisserie petrissage petrisseur produi prune pâtissier pétrin pétrir pétrissage pétrisseur raffinerie rafraîchissante sante santé semoule sucreries sucrée terrage tonneau tonnellerie tonnes vannage vermicelle vinasse vineuse vinificateur étuvage évitant

Railways and Trams: accrochage adherence adhérence aiguille aiguilleur amortir attraction automobile automoteu automoteur barrière bourrage chemi chocs circuler collisions conducteur conducteurs connexion convergent convergents convoi couplage croisement d'isolation d'isolement d'éclissage decrochage destination detacher décrochage dégagement déraillement déraillements désincrustation détacher enrayeur etanche ferrée ferrées filtration franchir frein freins funiculaire gare gomme gravir gravite gravité indéraillable installation isolant isolante isolateur isoler isolé l'impression l'isolation l'isolement locomobile locomobiles locomoteur locomoteurs locomotion locomotiv locomotive manoeuvre manoeuvrer mecanicien mécanicien nettoyeur parachoc parcour parcourir parcours passage pentes porte-crayon porte-mine porte-plume preservateur pression pressions protecteur préservateur pétard rail railway ralentir rarefie raréfié rayon refroidissement relier rencontre rencontres retarder reçoit roulement routier routière serre-frein station stations superstructure suppression tender tournant tournante tournantes traction train tramwa tramway travail travailler traverse trolley truck valve voie étanche

Arts textile, Fibers and Yarns: agissant aloès araignée armoire batiste blousse bobin bobine bobiner bobines bobineuse bobinoir bobinoirs broche brocheurs broché brochée brochées brode broder broderie broderies brodeur brodeuse brodé brodés calicot calicots canettes canevas canneler cannetilles cardeuse cardée cardées cardés caret cerf-volant chardons chaussettes chenille cheviller chiffon châle cirer coco cocon cocons coller collet contours cordonnet coupage coupé coupées coupés crapaudine crins d'écheveaux decoupage dentelle devidoir draperie défiler démêler déroulement dérouler détisser dévidage dévide dévider dévideur dévidoir dévidoirs effiler effilochage façonnées fibr fibre fibres fibreuse ficelles filage filament filamenteuse filaments filature filatures filer fileur fileuse filigraner filoir foulard foulards galons gaze grège grèges grége guipure guipures imprimée instar jacquard jaspés jenny jennys lainage laine lainer lainières largeur largeurs laveuse linge lingerie linon lins lisage lisse lisser lisses longueurs loquette loquettes lustrage maille mailles mailleuse maillon maillons malines marchure moirage moire moirer moiré moirés moquette moquettes moulinage mousse mousseline mousselines mâché noué ombrer ombrée ombrées ombrés ondulées papie papier pein peignage peigne peigner peigneur peigneuse peigneuses peignée peignées peignés peintes perrotine phormium picot piler pinceaux plisser presseur profiler rayer rayure rayures rayé rayée rayés retordre satinage satins satinés soierie soieries sylphide tissag tissage tissant tisse tisserand tisseur tisseurs tissé tissée tissées tissés tondeuse tondeuses tonte tordage tordoir tordre tricot tricoter tricoteur tulle tulles veloutées vergé vernie volant volante volantes volants vélin écheveau écheveaux épinglé épinglés

Machines: actionne actionner actionné actionnée agencement anse articulation articule articulee articulée ascenseur balancier bielle bouche bouchon broyeur bâti casse-fil charnière chaudiere chaudière chéneaux clape clapet cliquet condensateur couverture d'entraînement d'introduction d'ouverture d'élévateur d'émaillage d'étamage entretien equilibre essieu fardeau feux fluide freinage gouttière hydraulique l'ouverture lanière levag machine machines-outil manivelle manivelles manutention manèges molette monte-charge motocycle moufle mouvant multiplicateur multitubulair multitubulaire mâchoire mèche noria obtention obturateur obturateurs opaque opère organe orientant oscillant oscillante ouverture ouvertures ouvrier palan pedale perforateur perforer pesanteur physico pivot pneumato pompe porte porte-outil porter porteur poulie pouvant pressoirs primitif produire produit produites progressif progressive précieux puiser puissance puissances pulverisation pulveriser pulvérisateur pulvérisation pulvériser pédale pédalier quadruple rabot rabots raccords rectification reflux refoulante refouler regulateur relief reparation roche rogner rondelles rotateur rotule roulant roulante rouli râpe réchauffeur réfrigérant régulateur régulatrice réparation scier scierie scieries scies sieur sieurs siphon siphons sortie soufflante soufflantes soufflerie soufflet soufflets soulever tamiser tamiseur tampon tarare tarauder tare tarière teindre tombereau trancher transbordement transportable transports transvaser treuil trieur tubulaire tuyère vanne vannes vapeur varier vasista vendange verticalement viroles visser vitess vitesse vitesses wagon élévateur élévatoire épuisements équilibre équilibré étau

Marine and Navigation Construction: ailette ancre aviation aviron barque barrage boisseau bouée buee buée canalisée canaux canot ciseaux d'hélice dock embarcation fabriquer fleuve flot flottante flotteur gouverner hélic hélice latéraux marin marine maritime navigable navigation navire propulseur propulsive sonore sous-mari sous-marin toile torpille trame

Construction Public and Private: alluvion architecture armé assemblage aériann aérienne balayeuse boulangerie brosse canalisation caniveau citerne classement constructio construction courbe cuvier d'aronde d'assemblage d'attelage d'echelle d'ecluse d'indicateur d'échelle d'écluse dallage dalle destruction dispositif disposition echelle elastique excavateur façade ferme fermeture fermé fondation fosse frottement frotter hourdi incassable incendie indicateur interieur l'assemblage latte magasi mecanique mortier mécanique ondule ondulé palier passementerie pavage pelle pincette plafond planche plastique plate poutre poutrelle pratique quai scellement solaire substitution suspendu suspendus terrain terrassement terrassements terrasses terrassier toit toits tourbe tranchées treilli triangulaires trottoir trottoirs utilisable viaducs vieux vitrage vitre vitrerie vitrine voiri volcanique voûte voûtes échafaud échafaudage échafaudages échelles écluses élastique

Mines and Metallurgy: acier aciers affinage affinerie alliage alliages argenter artésiens braser camphre cannelure ciselés couche couches coudés coulé creuset cuivrage cuivre cémentation cémenté damasquinage décapage element emboutir emboutissage enclumes fonte fontes forage forge forgeage forger fourche galvanisation guimpier laminage laminoir laminoirs laminé laminés maillechort malléabl malléable martinets metallurgie mine miniere métallurgie métallurgique métallurgiques pilon platinage plombage proportions puddlage puddler raccord recuit revêtir réchauffage salants scories sondage soudage soudés souterrains spéciaux sucre treillage treillages trempe tréfilerie trépan tôles zincage élément étirés

Domestic Economy: affiche affiler affiloir affiloirs alcôves ameublement ameublements baignoire balcon banquette banquettes bassinoire berceau bibliothèque bouillante cache-entrée cadena cadenas cafetiere cafetière cage canapé cane casseroles cisaille coffre coffres coisées commode contrevent contrevents corbeille coucher couchette coutellerie cremon croisee croisée crémon crémone crémones d'affiche d'appel diffusion divan divans dossier dossiers espagnolette espagnolettes fauteuil ferrements ferronnerie feuillure fiches fusion galet gonds gorge gorges hachoir hermétiquement incrochetable incrochetables infusion infusions literie loquets manteau menage meuble mobilier mousqueton mécaniquement patères paumelle paumelles pene pentures persienne pincettes pliant poinconneuse poinçonneuse pène pêne pênes quincaillerie rallonges rappel rasoirs repliable rideaux roulette râpes secret serrure serrurerie serrures sonnettes sphériques table tabourets tapisserie tenture théière tournebroche tringle tringles valise verrou verrous vols vrilles

Weapons: arme artillerie baionnette barreau barreaux baïonnette bouclier calibre canon carneaux carreaux cible classeur d'alarm d'armure detendeur detente dynamite défensive détendeur détente embrasure emploi engin explosible faible flexible force fusee fusil fusée gaine hammerles l'art l'artillerie lance masque militaire noyau nuance pistolet piston poignées portatif portativ portative poudrière projectile projectiles rayées recharger revolve revolver revolvers rotative sabre sabres schakos silex tendeur tente tente-abri tir tonnerre troupes épée équipement

Road Transport: appelées bitumineux brancard brancards bridon brisés burette cabriolet cabriolets calcination capote carrosserie d'arçon diligences dételage ferrement ferrer flèche gourmette harnachement inversable inversables landau licou longe lubrifiante luxe marc marchepied marechalerie messageries mors moyen moyeu muselière oxyde oxydé porte-brancard portière publique rais renferment roulage selle selles separateur separation sulfuré séparateur séparation transport velocipedie voiturette écartement éperon étrier

Instrument of Precision and Electricity: abat-jour accumulateur acoustique aimant anéroïde bascul bascule binocle blutoir communication commutateur compensateur compteur consommation curseur d'arpentage d'horloge d'électricit demontage dosage dynamique dynamo dynamo-électrique electricite electrique electro-automatique electro-magnetique electro-moteur electrode electrolyse electrolytique enregistrer enregistreur enregistreuse galvanique horloge instrument l'electrolyse l'électricit l'électrolyse lentille lorgnette lorgnettes lorgnon lunette magnétiques magnétisme magnéto manomètre manomètres mathematique mesurage mesurant mesure mesureur microscope montage mètre mètres médecine méridien métrique métriques niveau niveaux nombres objectif obscure observation pesage peseur peson pesée piles pince-ne pince-nez planimètre polymètre pondérateur precision pyromètre pèse quadrature quantième rapporteur recepteur reconnaître remontage remontoir remontées rouage récepteur secondes sonnerie sonneries spectacle stadia surveillance telegraphie telephon tempes terrestres thermique thermo-électrique thermomètre thermomètres tracer transmetteur transmettre triangulaire télégraphe télégraphes télégraphie télégraphique télégraphiques téléphone téléphoni téléphonie téléphonique voltaïque échappements électricité électrique électro-automatique électro-chimique électro-magnétique électro-moteur électrode électrolyse électrolyte électrolytique

Ceramics: argile argileuse brique briqueterie carrelage creuse céramiqu céramique faience faïence faïences glaise goulots incrustés manchon manchons marqueterie plâtr plâtre porcelain porcelaine poterie poteries potier pouzzolane rebord rebords recuire refractaire réfractaire réfractaires soufflage tain tuile verrerie verreries vitrifiées

Chemics: absorbante acétate acétates acétique alambic albumine alcali alcalines alcalins alcalis alcool alcooliques alcools altération aluminate alumine alumineuses aluminium alun ambulant ammoniac ammoniaque amorphe amylacée antiseptique appret apprêt apprêter aqueuse aromatique asphaltes asphaltiques azote azotique azurer barium baryte baryum bases bichromate bioxyde bisulfate bisulfite borate borax borique bouillon brome brute cachou caoutchouc carbonatation carbonate carbonates carburation carburé caustique caustiques celluloid chaleu chaleur chimique chimiquement chlor chlorate chlore chlorhydrate chlorhydrique chlorure chlorures cidre cirage cire colle colorants composition compositions compound conservateur conservateurs copal coque corroyer couleur couperose creme cristallisé crème crèmes cuir cuve cyanogène cyanure cyanures cérus céruse d'aluminium d'ammonium d'appret d'apprêt d'aprè d'eau d'eaux d'encre d'encrier d'epuration d'ether d'hydrate d'oxydation d'oxyde d'épuration d'éther d'évaporatio d'évaporation decantation dentifrice dessécher dissou dissoudre distillation distillatoire distillatoires distille distiller distillée distillés dorer dorure doseur décantation déchet décreusage désagrégation désinfectante désulfuration encre encres encrier epuration essence evaporation evaporer explosive extracteur extrait extraite extraits fermentation fermentescible fermentés feutre filtr filtrante filtrantes filtre filtrer filtres fluorure foie fondants frigorifique galle gallique galvanoplastie garancine gelatine glucose glucoses glycérine gommes gommeuses gommo goudron goudrons graisseur gras hydrate hydraté hydratée hydrochlorique hydrofuge hydrofuges hypochlorite imperméabl imperméable imputrescible inattaquable incolore indicatrice industriel industrielle inflammable injecter injecteur injection inodor inodore instantane instantanee instantanée iode l'alcoo l'aluminate l'aluminium l'amiante l'amidon l'ammoniaqu l'anhydride l'apprêt l'eau l'epuration l'evaporation l'hydrate l'industrie l'oxyde l'ozon l'ozone l'évaporation laques lessives levûre lichen liquide lubrifiantes lubrifier luisant magnesium magnésie magnésium manganèse margarique maroquin melange melangeur methyle minérale minéraux mixtion morue mucilagineuse muriate muriatique mélange mélanger mélangeur mélangé mélangée métallisation méthyle nature naturel naturelle naturels neutralisation neutre neutres nitrate nitreux nitrique ocres olive oléine oléique onctueux ourdir oxalique oxydant oxygène parfum parfumerie parfums pendant periodique perméables peroxyde phenique phenol phosphate phosphore phénique phénol plaque potable potasse potassium protoxyde prussiate prussiates précipitation pulpe pulpes purifier pyrites périodique pétrol pétrole queue quinine reducteur resine rouille réactif réducteur réfrigération résine résines résineux saccharatus saccharification saccharine saline salins salpêtre sang saponification saponifier saturation savon savonneuse savonneuses savons schiste siccatif silicate silice siphoïd siphoïde sodium soi sorgho souder soudes soutache steriliser strontiane strontium stéarine stéarique stéariques stériliser sucré suif suifs sulfates sulfite sulfonique sulfures sulfureux sulfuriqu sulfurique surtout tableau taches tanin tanner tannerie tannin tanné tannés taquet tartre tartrique terreux tinctoriales térébenthine vaporisation vernir vernis volatil volatils végéto émailler évaporation évaporatoires évaporer

Lighting Heating Refrigeration and Ventilation: absorber active activer aeration agglomération agglomérer aggloméré allumage allume-fe allume-feu allumer allumette allumeur allumoir amadou appliquee appliquée artificiel artificielle aérage aéro bobèche bougeoir bougeoirs bouilloire brai braise briquet briquets bruler brûlant brûler brûleur bûches calorifèr calorifère caloripède candélabre carboniser carbure carbures carcel centra central chandelier chandeliers chapeliers charbonnières chaud chaude chauffage chauffages chauffant chaufferette chaufferettes cheminée combustible congélateur crémaillère cuisines cuisinière culinaire d'anthracite d'extincteur d'extinction d'hydrocarbure d'éclairage d'écran dirigeable eclairage ecran enflammer etincelle etui extincteur flambeau flambeaux flamm flamme forêts fournaises fumeurs fumi fumifuge fumivores galerie galeries gazo gazomètre gazomètres glacière houille hydrocarbure hydrocarbures hydrogene hydrogène hydrostatique ignition illuminations incandescence incandescente inflammables ininflammable intensité intermittente l'incandescence l'éclairage l'étincelle lampe lanterne lanternes lenticulaire lignite longues-vue lumineux phares phosphoriques pignon porte-allumette porte-bougie porte-meche porte-mèche poussier poêle pyro pyrogène pyrogènes pyrogénées pyrotechnique rayonnant refrigeration reverbere réchaud réchauds réflecteur réflecteurs réflexion réverbère réverbères salles sechoir supérieur séchoir séchoirs thermosiphon thermostat ustensile veilleuse veilleuses ventilation volume vues échauffer éclairant éclairer écran écrans éteignoir étincelle étui étuves

Clothes: anneau ardillon ardillons benzidine bijou blanche bleue botte bottine boutonnière boutonnières bretelle brodequin brodequins brun brune brunir buanderie buanderies caleçon caleçons cambre cambrée cambrées canne canne-paraplui caoutchoute caoutchouté casquette chapea chapeau chapellerie chaussette chausson chaussure coiffes coiffure col concentree concentrée corsages corset coulant coulants coupes cousu cousues couture couturiere couturière cravache cravat cravate crinoline culotte d'adaptation d'alizarine d'espadrille d'uniforme doublure drapé dressage embauchoirs empeigne epingle eventail ferrage fixe-cravat flanelle fleur forme fouet fourrure friser frisure ganse gant gibus gilet griffes guêtre guêtres gélatine habillements habit habits hiver invisible jambiere jambière jarretière jarretières jaune jupe jupon lacer lacet latanier lessive lessiver leur ligno manchette manchettes mannequin mannequins manteaux marquises musette ombrelle osier panier pantalon pantoufle paracrotte paraplui parapluie parasol parasols perruque plissage plume porte-canne poser propreté rebras repassage sandale semelle semelles socque socques soulier suspensoir tailleur talon teignant teinte toques toupets tournure tournures tournurière tresses tricoté uniform velocipede verte vetement violet violette visière visières vélocipède vélocipédiste vêtement épingle

Arts: accord accorder accords accordéon accordéons agrafe album amplificateur anime animé animée applique archet archets argentés arrondir artistique autographie autographique bague bagues baleine basses battant bijoux boiseries bonde bourrelet bouto bustes cahiers calorique carcasse carnet chant charge charger chargeur chemise chevalet chevalets cheveux chromatique chromatiques chromo cinematographe cinématographe ciselure clairons clarinette clarinettes clavi clavier claviers coloriées compositeur concertinas confection confectionner cor cornet cornets couronne couronnes coussin crin culot d'agrandissement d'album d'amiante d'image d'imitation d'instrument d'oreille d'ornement d'épreuve daguerréotype dessin dessinateur dessinateurs dessiner diamant dièses décors découpés déterminer déterminée développer ecremeuse emboitement embouti emboîtement emporte-piece emporte-pièce encadrement encrage estampes estampé estampés explosif expressif expressifs expressives faconne feuille feuillets fixateur fleurons flûte flûtes flûtinas fonction format graphique gravure gravures gravées guitare harmonie harmonique harmoniques harmonium harpe harpes hautbois image images imitation imprime imprimé joaillerie jonction l'image l'imitation l'imprimerie l'influence l'ivoire l'ornementation l'outillage lavis lithochromique lithographie lithographier lithographique lithographiques mate medaille medaillon miniature monnaie monnaies monétaire mouvoir murale musical musicale musicaux musique médaille médailles médaillon médaillons métronome nacre nappe notation note notes négatif négative octaves orchestre orfevrerie orfèvrerie orgue orgues ornement ornementé ouvrage paillette panoramique papeterie passe-partout peint peintre pelliculaire pellicule pendule perce perforation perfore perforé perle perles phonographe photographie photographique photographiques photogravure pian pianinos piano pianos pierre pierreries piquage pointure poinçons polychrome porte-monnaie portrait portraits pourvu presse-etoupe presse-étoupe production projeter précieuse précisio précision présentant recouvert reproduction rondelle s'appliquant sculpter sculpture sculptures sertissure sons stéréoscop stéréoscope stéréoscopiques tabletterie tabouret taille-douce teint ton tons touche toucher touches tourne-feuillet tourne-page transformable transformant transformer transparents travaux trompette trompettes types typo typographiques veloutage velouter vignettes violon violons vitraux vitrification zincographie écrémeuse

Office Supplies and Education: affichage albums almanach apprendre biblorhapte brochure calligraphie calligraphique copie copies copiste correspondance d'enveloppe document ecrire enseignement enseigner enveloppe envelopper facture géographie géographique journal journaux lecture lettre leçons manuscrit mettre orthographe paquets porte-journa postale postaux poste prendre public publication quotidien reliure reliures rendre reproduire sciences sphère sphères tendre uranographique écrire écrit écritoire

Medicine and Health: accouchement accouchements affections allaitement analeptique aromatiques aseptique baignoires balayage baume bibero biberon biberons blessé blessés cabinets cachet cadavre canules cautère cercueil cercueils chalumeau chauffe chauffe-bain chauffe-pied chauffer chirurgical chirurgicaux chirurgie clysoir clysopompe clysopompes crachoir cremation cuisses curage dentaire descente descentes difformités domestique douche douches délétères désinfecteur désinfecteurs déviations egout embaumement gymnastique hernie hernies humain hydrotherapie hygiene hygiène immondice jambes latrines lavabo linceul malade malades mamelon matrice mauvaise maux medecine mortuaire médical médicament médicamenteuse médicaments médicinales médicinaux méphitiques natation nettoyage orthopédique pansement pessaire pessaires pharmaceutique pharmacie plaies plumeau poitrine redressement remède salubrite sangsue sein stereotypie sterilisateur stérilisateur stéréotypie sulfureuses topique traitement urinoir urinoirs utérus varices ventouse ventouses vertébrale vidange égout émanations

Articles and Various Industries: banque bebe bidon bidons billard billards bille blouses bouquet bouquin bourse bourses bruyere bruyère bébé campement canonnière carnets casier cerceau cigare cigarette cigarettes concert coupon course d'échec domino dominos désigné emballage emballer embouchure embouchures empaqueter entonnoir estagnon fermoir fermoirs fragiles fumeur gainerie goulot guidon hochet illustre illustré jeux jouer jouet jouets joueurs loto marqueur nicotine oeillet ongles oranger panorama panoramas papier patin patins pipe pipes porte-papier poupée poupées serviette serviettes société tabac tabatières theatre théâtre tourniquet voyage échecs

Appendix C: Transportation Network and Least Cost Paths

The key element of the market access framework is the least cost paths among the city points of the cantons, the term $Cost_{iht}$. This cost term is allowed to vary by year, t. The source of variation relies on the expansion of railroads and waterways. Therefore, I compute the least cost paths in QGIS for each year based on the following steps:

1. Division France into $0.008^o \times 0.008^o$ grid[26].

2. Assignment of a cost to each grid cell using the same properties as in Zobl (2018), Toutain (1967) and Daudin (2010). The values, that the authors propose, are:

Table C1: French absolute freight rates by transport mode

Period	Road	Canal	Railway
1841-1844	25	4.1	14.5
1845-1854	25	3.8	10.6
1855-1865	25	3.2	8.7
1865-1874	25	2.4	7.5
1885-1894	25	1.8	6.8

Freight absolute rates are measured as centimes per ton and kilometre. The canals have a lower value in terms of cost than the railways. However, the reduction over the years is greater for the railways than the canals. Next, the roads have a fixed cost over the sample period. The effect of roads is then captured by the time fixed effect. I assign a cost of 231[27] for walking like in the paper of Donaldson and Hornbeck (2016).

3. Construction of cost raster file.

4. Calculation for each of the 2,925 cantons in the sample, the travel cost to every other canton. This is a minimisation problem based on Dijkstra (1959) algorithm over the least cost surface which selects the optimal route.

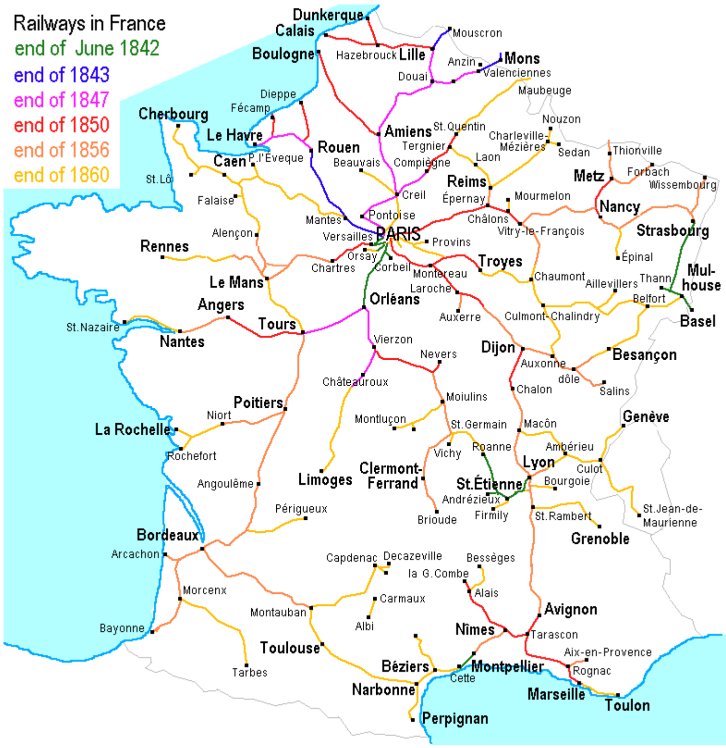
Finally, I compare the costs between canton pairs and between two different consecutive years (t and $t + 1$) and I assign always in year t the lower of the two costs (Perlman, 2015). This methodology produces 8,555,625 pairwise least cost terms for every year. These cost terms can be collected in a cost matrix:

[26] One grid is approximately 1.2 square kilometers.

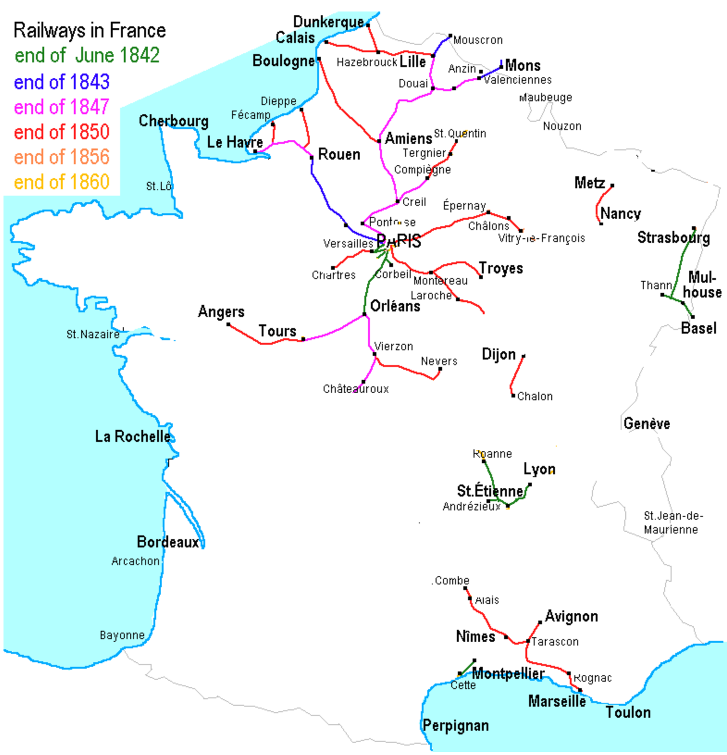
[27] The value of Donaldson and Hornbeck (2016) is 23.1. I have multiplied it by 10 in order to be in line with the other values.

$$\mathbf{C}_t = \begin{bmatrix} \mathrm{Cost}_{11}^{-\theta} & \mathrm{Cost}_{12}^{-\theta} & \dots & \mathrm{Cost}_{1n}^{-\theta} \\ \mathrm{Cost}_{21}^{-\theta} & \mathrm{Cost}_{22}^{-\theta} & \dots & \mathrm{Cost}_{2n}^{-\theta} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cost}_{n1}^{-\theta} & \mathrm{Cost}_{n2}^{-\theta} & \dots & \mathrm{Cost}_{nn}^{-\theta} \end{bmatrix}_t$$

Figure C1: Rail lines until 1860



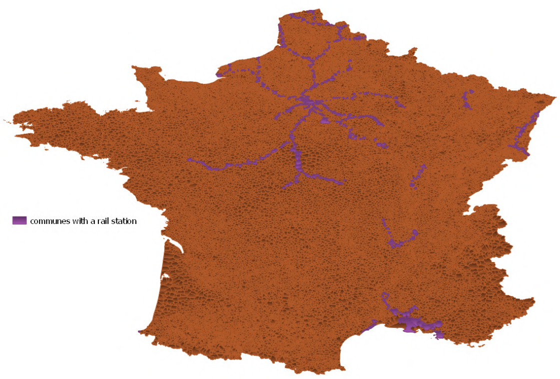
(**a**) initial image



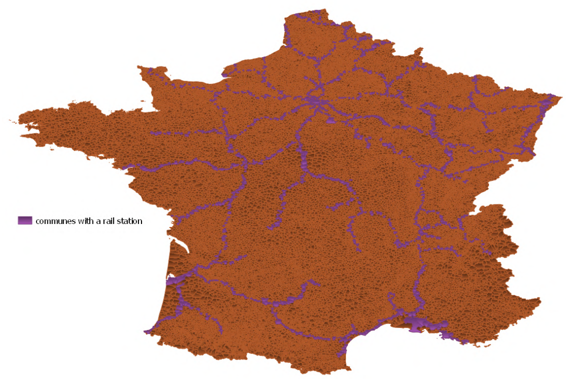
(**b**) 1850

Notes: Railroads until 1860. Source: https://www.wikiwand.com.

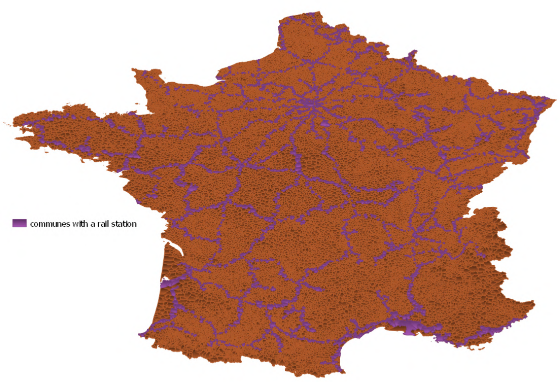
Figure C2: Communes with rail stations



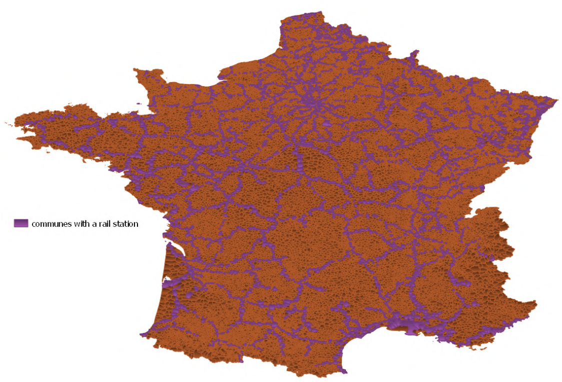
(**a**) 1850



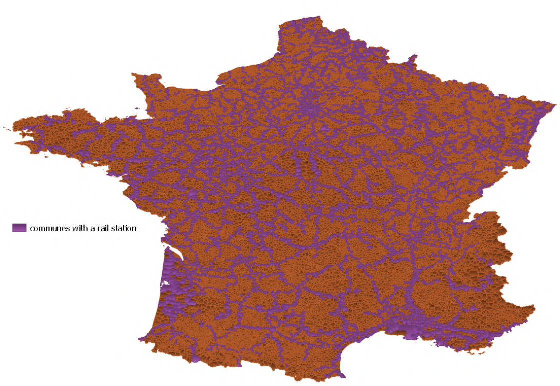
(**b**) 1860



(**c**) 1870



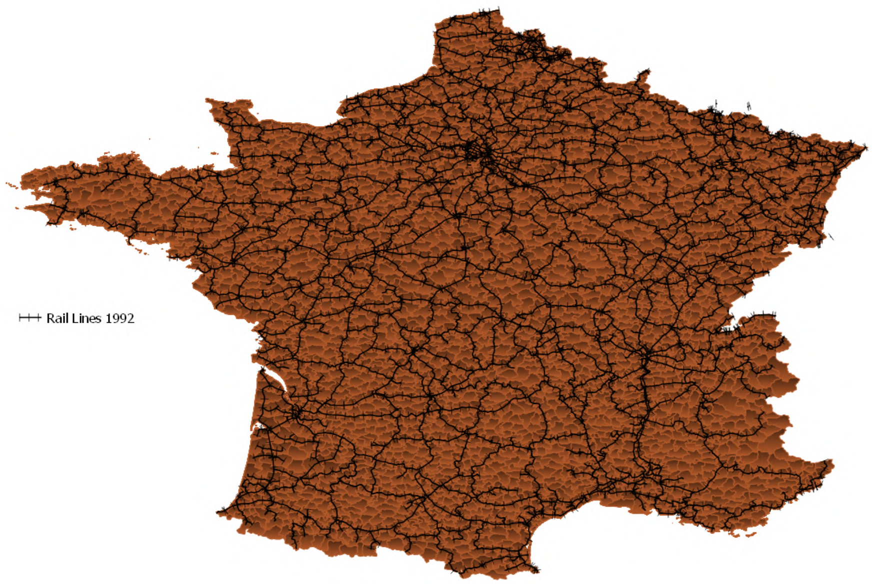
(**d**) 1880



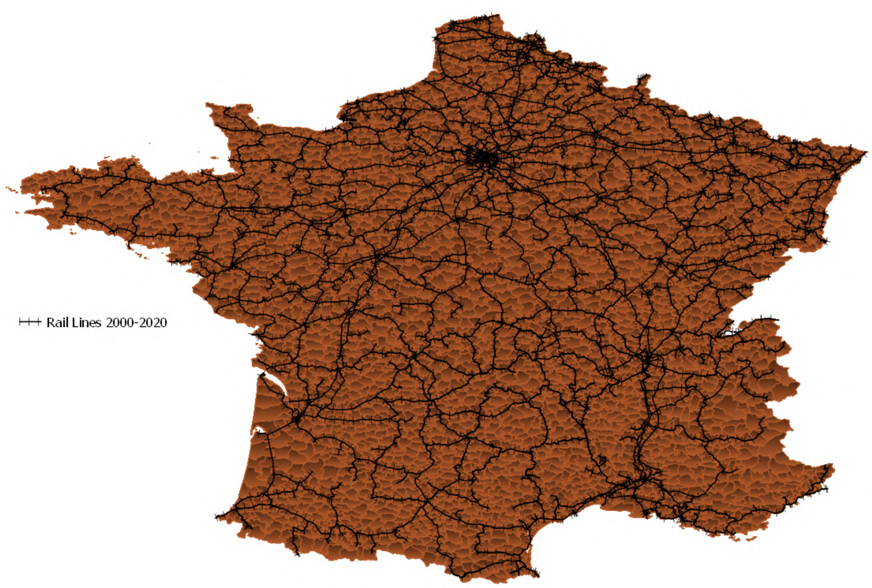
(**e**) 1890

Notes: This figure shows the expansion of the communes with rail stations. Source: Mimeur et al. (2018).

Figure C3: Modern rairoad network



(**a**) rail lines until 1992



(**b**) rail lines 2000-2020

Notes: Modern railroad network. Source: DIVA-GIS (2020) and Jeansoulin (2019).

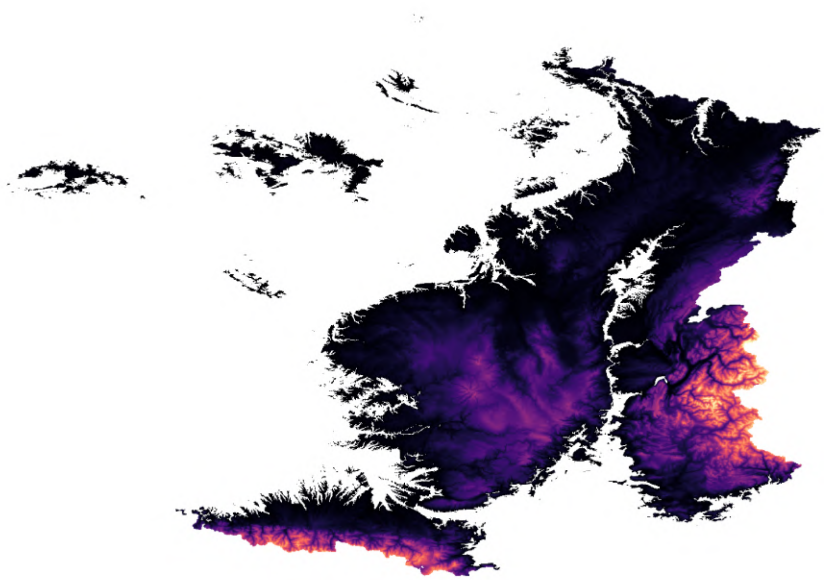
Figure C4: canal network and ports

(**a**) canal network in 1878

(**b**) Ports in 1877

Notes: The first graph presents the canal network of France in 1878. The second includes the ports in 1877. Source: University of Chicago (2020) and David Rumsey (2020).

Figure C5: areas with high elevation



Notes: Cantons of France with high elevation. Source: DIVA-GIS (2020).