CAF - WORKING PAPER #2021/03

First version: April 12, 2021(current)

Bimodal Transport Infrastructure and Regional Development: Evidence from Argentina, 1960 - 1991

José Belmar[1] | **Diego Gentile Passaro**[2]

[1]PhD student, Department of Economics, Brown University.
jose_belmar@brown.edu

[2]PhD student, Department of Economics, Brown University.
diego_gentile_passaro@brown.edu

We estimate the impact of railroad and road infrastructure on local economic development through the study of massive transport infrastructure changes in Argentina. Following a World Bank study, 10,000 kilometers of railroads were closed and 18,000 kilometers of paved roads were built between 1960 and 1991. Our empirical strategy relies on instrumental variables that exploit a discontinuity in how experts chose railroad segments to be studied for closure and hypothetical networks connecting main cities. We show that conventional IV estimates can be misleading when omitting potential substitution of different transport modes. We find that *dismantling* railroads had a negative impact on population and industrial production, and shifted the distribution of labor away from agriculture. On the other hand, we find weak evidence of roads construction having a positive impact on the share of employment in manufacturing and non-tradable industries, but no impacts on total population nor industrial production.

KEYWORDS

Regional Development, Structural Change, Railroads, Highways, Transportation, Economic History, Latin America

CAF - DOCUMENTO DE TRABAJO #2021/03

Primera versión: Abril 12, 2021 (actual)

Infraestructura de transporte bimodal y desarrollo regional: Evidencia de Argentina, 1960 -1991

José Belmar[1] | **Diego Gentile Passaro**[2]

[1]Estudiante de Doctorado, Department of Economics, Brown University.
jose_belmar@brown.edu

[2]Estudiante de Doctorado, Department of Economics, Brown University.
diego_gentile_passaro@brown.edu

Estimamos el impacto diferencial de ferrocarriles y caminos sobre el desarrollo económico local por medio del estudio de cambios de gran escala en la estructura de transporte Argentina. Tras un estudio del Banco Mundial, 10,000 kilometros de ferrocarriles fueron cerrados y 18,000 kilometros de caminos pavimentados fueron construidos entre 1960 y 1991. Nuestra estrategia empírica descansa en variables instrumentales que explotan una discontinuidad en como los expertos eligieron los ramales ferroviarios a estudiar para ser cerrados, y redes hipotéticas que conecten las principales ciudades argentinas. Mostramos que al omitir la posible sustitución de modos de transporte, los estimadores de variables instrumentales pueden estar sesgados. Nuestros resultados principales indican que el *desmantelamiento* de ferrocarriles tuvo un impacto negativo en la poblacińńn y la producción industrial local, y redujo la pariticipación del empleo en el sector agrícola. Por otro lado, encontramos evidencia débil de un impacto positivo de la contrucción de caminos sobre la participación de las manufacturas y las industrias no transables en el empleo, pero ningún impacto en población ni producción industrial.

KEYWORDS

Desarrollo Regional, Cambio Estructural, Ferrocarriles, Carreteras, Transporte, Historia Económica, America Latina

1 | INTRODUCTION

Investing in physical transport infrastructure is regarded and recommended as an important policy to foster economic growth.[1] These investments facilitate migration, enable the supply of goods and services across firms and consumers that are located in different places, they increase the flow of ideas across territories, and they presumably shorten the time of engaging in productive activities. Furthermore, recent empirical work has shown that they have sizable positive effects on trade and economic development across a wide range of historical contexts.[2]

Despite these recent empirical results, our understanding of this relationship remains simple and limited. This is clear from the standpoint of policymakers who need to decide whether to invest in physical transport infrastructure (hereafter PTI), but also in what types and how to distribute them across space in order to enhance the regions and sectors that they consider more advantageous. Is it better to invest in unimodal transport networks or to use multiple technologies? Do some transport modes serve better specific sectors than others? What is the role of national and local comparative advantages in shaping the impact of these investments? What are the consequences of substituting one transportation mode for another? Answering these questions is important because different modes of transport may serve different purposes: (i) they are cost-optimal for different kind of goods, (ii) they affect the movement of people, goods, and technology in different ways, and (iii) they interact such that they can be complements or substitutes depending on the network and the spatial distribution of production and factors. For example, Redding and Turner (2015) state that railroads affect production activities more than population, roads affect more the spatial organization of population than of production, and that roads may have larger effects on the organization of economic activity in rural areas than in cities. However, we are not aware of a paper that studies the mode-specific effects of PTI on the spatial allocation of population and different types of production in a common intranational setting.[3]

Assessing the economic effects of both railroad and road infrastructure is not only important but also challenging. First, it requires a historical context in which both railroad and road infrastructure are changing substantially. Second, it requires plausibly exogenous variation for each mode, as they may have heterogeneous effects on different economic sectors.[4]

In this paper, we aim to study the case of Argentina between 1960 and 1991 when following a World Bank study of the PTI (known as the Larkin Plan, hereafter LP) the total railroad network fell from around 43,500 to around 33,500 kms at the same time that the paved road network grew from around 10,000 to 28,000 kms. These changes in the PTI happened in a context of industrialization in an economy with a substantial agricultural

[1] "Transport is a crucial driver of economic and social development, bringing opportunities for the poor and enabling economies to be more competitive." (World Bank, 2020). The World Bank also views it as an important policy: in 2018, they allocated $3.19 billion to 178 transport infrastructure projects, representing 16% of their lending portfolio.

[2] See for example Michaels (2008); Donaldson (2018); Chandra and Thompson (2000); Fajgelbaum and Redding (2018).

[3] Baum-Snow et al. (2017) estimate the effects of intracity road and railroad infrastructure on the decentralization of different types of economic activity from the city center to the periphery. Banerjee et al. (2012) study the effect of transport infrastructure on local GDP level and growth in China from 1986-2006. However, they are not able to study separately the effects of roads and railroads because they do not have exogenous variation for each transport mode.

[4] For example, it is plausible that railroads have a larger effect on agricultural production than on industrial one, as the latter are less subject to international trade.

sector. This setting is unique because it provides important variation in both railroad and road infrastructure within the same historical context. In 1960 Argentina had important fiscal deficits, and the recently nationalized railroad system was under public scrutiny. Importantly, the LP was aimed to asses the long term viability of the current railroad network and to evaluate and design an expansion of the road network. However, due to budget constraints and the vast existing railroad network, only railroad segments that were below some yearly freight weight cutoff were "studied" in detail[5]. Crucially, the likelihood of the LP recommending to close a segment was much higher for those below the cutoff. Furthermore, despite the LP being abandoned after a few months of announcing its implementation[6] the railroad closures in the 60's and 70's were three times more likely for studied segments probably because no new large scale study was undertaken.

We combine spatially-disaggregated economic, demographic and geographic data to study the effects of these transport infrastructure changes on population and the economic structure of Argentinian *departamentos* (heareafter, districts) using plausibly exogenous variation given by (i) a discontinuity in the LP design, and (ii) hypothetical road networks connecting main Argentinian cities in a cost minimizing way.

First, we show that conventional IV estimates can be misleading when omitting potential substitution of different transport modes. When allowing for this substitution, our instrumental variables approach shows that for every 100 kms of railroads *dismantled* in a district, total population fell in 0.29 log-points, urban population fell in 0.26 log-points, and the population born in the same province they live fell in 0.29 log-points, suggesting outflow migration. Consistently, we find negative impacts on the volume of industrial activity as measured by total value of production, total paid wages and employment. When looking at structural change, we find that every 100 kms of dismantled railroad led to a reduction of 4.1% in the share of labor in the primary sector (agriculture and mining), an increase of 4.1% on the share of labor in manufacturing and an increase of 1.8% on the share of labor in non-tradable industries (construction, electricity, water, and gas). On the other hand, we don't find any impact of roads infrastructure on total population, urban population and total economic activity. Nevertheless, we find that every 100 km of railroads construction lead to an increase of 0.10 log points in population born in in the same province, a 1.2% decrease in the share of labor in other services, and weak evidence of this share being gained by manufacturing and business services.

These results are consistent with the idea that railroads serve better products with low value-to-weight ratio, such as wheat, and more generally agricultural products. In our interpretation of this results, dismantling railroads led to a contraction in agricultural exports, igniting a crisis in the local economy, expressed both in decreasing production and out-flow migration. In comparison to the agricultural sector, industrial activities better endured the crisis due to their relative independence from external demand and protection against external competition induced by railroads dismantling, as reflected in the shift of labor shares from agriculture to industry. On the other hand, our results suggest that the substitution for roads failed to counter the decline of the local economy. Nevertheless there is evidence of roads construction being non-neutral by changing migration patterns (from inter- to intra-province immigration) and weak evidence of structural change. The

[5]By "studied" in the LP, we mean that they checked the segment's explicit revenues and costs and analyzed whether it was convenient to close it or not. The cutoffs were 1 million gross tons per year of cargo passing through a given segment, or 500 tons per kilometer per year of cargo entering the network though a given segment. Only 39.6% of the railroad network was studied.

[6]Following the announcement of the first closures the railroad workers started massive strikes and protests. After several months of clashes and social unrest the government decided to abandon the plan.

latter suggest that future research may find impacts of roads on the long-run and/or other outcomes.

This paper contributes in several ways to the literature on the effects of PTI on economic development. First, we provide the first empirical case that allows to separately estimate reduced form effects of both road and railroad infrastructure on local economic development and specialization. Previous work estimating the effects of roads have been conducted either around the US highway expansion during the 60's (Michaels, 2008; Chandra and Thompson, 2000; Duranton et al., 2014) or in the early 2000's in developing countries (Banerjee et al., 2012; Faber, 2014; Jedwab and Moradi, 2016; Baum-Snow et al., 2018). We provide a novel setting that allows to study the effects of both modes of transportation in a recent historical context. Moreover, we show that it is possible to overestimate the impact of a specific transportation mode when we omit potential substitution. In our view, this results can inform a recent literature on optimal infrastructure investments which -to the best of our knowledge- does not account for multiple transport modes (Fajgelbaum and Schaal, 2020).

Second, while most of the literature focuses on the distribution of people and overall economic activity, this article follows recent papers (Fajgelbaum and Redding (2018) and Forero et al. (2019)) and puts special attention on the impact of PTI on productive structure, understood as the relative importance of different economic sectors.

Third, most of the work on the effects of railroad infrastructure has studied cases during the 19th century and the first globalization, both in developed (Donaldson and Hornbeck, 2016) and developing countries (Donaldson, 2018; Fajgelbaum and Redding, 2018; Jedwab and Moradi, 2016). This is the first attempt to study the impact of *dismantling* railroad infrastructure during the second half of the 20th century in a developing country, and the second attempt overall only preceded by Gibbons et al. (2018) who studied the effects of railroad disinvestment in Great Britain during the 50's, 60's and 70's.

Fourth, it contributes to a long-lasting public debate in Argentina regarding the impact of railroads disinvestment. While some have attributed the declining and disappearance of hinterland towns to railroads dismantling, some have counter-argued that these towns were already declining, and closures just followed them. While further research is required to determine the aggregate welfare implications of this policy, our results suggest that the decline of hinterland towns can be at least partially attributed to railroads dismantling.

2 | HISTORICAL BACKGROUND

In 1857, Argentina started the construction of its first railroad track. Since then, Argentina developed one of the largest railroad networks in the world. This network was developed through national capital, foreign investments (predominantly British and French), and with direct participation of the government by extending the network to rural areas that were not profitable and by managing public sales of land. Cause and consequence of these investments was the boom of agricultural exports experienced during the first globalization. During the second half of the nineteenth century and the first decades of the 20th century, Argentina experienced extraordinary growth of international trade, population and aggregate production. [7]. By the start of the First World War, Argentina was the eighth richest country in the world (Fajgelbaum and Redding, 2018).

Starting the 1930s, exports growth slowed down considerably and profit margins of

[7]From 1869 to 1914, Argentinian real exports grew around 500% while the real imports did so around 200% (Fajgelbaum and Redding, 2018)

railroad companies were greatly reduced. Investments on the railroad infrastructure also stalled and soon the aging network experienced increasing operational costs. In addition, railroads started to face competition from the automotive industry. In 1925, Ford produced the first model T of Latin America in Buenos Aires, a milestone for the incipient industrial sector. In 1932, law 11.658 created a National Trunk Road System and a lump-sum tax to fuels such that all proceedings were allocated to fund the construction and improvement of national roads. By 1935, there were almost 3.000 kms of paved roads.

During the Second World War, Argentina remained neutral and its young industrial sector was able to flourish while the factories of the developed world were retooled to supply for the war. Additionally, the war caused food shortages and triggered international prices of Argentinian agricultural exports. During these years, railroads had a "second life" and helped Argentina to accumulate reserves for $1.6 billion (Whitaker, 2013), after years of positive trade balances. But Argentina also had a vigorous young industrial sector, and by 1945 its contribution to the GDP surpassed that of the agricultural sector for the first time (Potash, 1969).

Supported by the unions, the military, and the new industrial bourgeoisie, General Juan Domingo Perón reached the presidency in 1946. His economic policy was based on three pillars: i) strengthen the domestic market, ii) centralized planning and state ownership of strategic economic means, iii) development of the industrial sector (Whitaker, 2013; Pahowka, 2005). The showcase of his new economic plan was the acquisition and nationalization of the private railroad companies in 1947 for $600 million (Pahowka, 2005). The purchase was hugely debated and several economists of the time (notably Raúl Prebisch) pointed out that the price paid was too large given that the railroad infrastructure was crumbling and outdated (Rock et al., 1985).

Starting the 1950s, when the fortuitous advantages from the war faded away, Argentina's economy slipped into an important depression. In 1952, Argentinian agricultural goods were banned from the European markets participating in the Marshall Plan, due to differences between Perón and the US government. Perón acted quickly to protect the nascent industry and the automotive and petrochemical sectors were armored against international competition and focused on the domestic market. The road network continued to grow at a slow steady pace and by 1955 it had 9,000 kms of paved roads. On the other hand, the railroad network was stable at around 43,500kms, but with a shrinking budget, the network suffered from poor maintenance and ageing equipment.

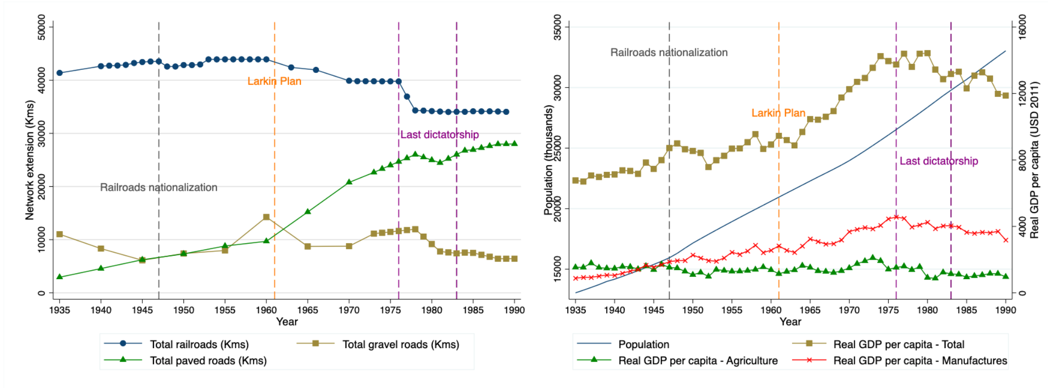

FIGURE 1 PTI, population, and real GDP per capita in Argentina 1935-1991.
Notes: This figure displays the time series of PTI, population, and real GDP per capita in Argentina from 1935 to 1991. The series for railroads comes from López and Waddell (2007). Data from paved and gravel roads comes from Dirección Nacional de Vialidad. Data for population and real GDP per capita comes from Bolt and Van Zanden (2014). Data for industrial and agricultural GDP shares comes from Ferreres (2005). The dashed vertical gray line shows the year of the railroad nationalization (1947). The dashed vertical yellow line shows the year of publication of the LP (1961). The dashed vertical purple lines show the start and end years of the last prevailing dictatorship in Argentina (1976 and 1983).

In 1955 a coup removed Perón from the government, and the country lived a few years of political and economic instability. In 1958, in an election in which Perón was banned, Arturo Frondizi was elected president. His main economic plan was to foster industrialization through foreign direct investment. In 1959, decree 3693 was passed to promote investments in the national automotive industry, and the construction of 23 new factories were approved after one year. In addition, under his government several multinational oil companies landed in the country and the production of oil tripled in only four years, reaching self-sufficiency. In concert with his industrial and oil policies, Frondizi's administration was very aggressive in promoting the construction of roads. In 1958, the lump-sum tax for fuels was replaced by a %35 tax[8], and in 1960 law 15.274 instituted a sales tax on vehicles. In both cases, the funds were permanently allocated to the construction and improvement of roads. As for railroads, initially the president wanted to modernize the crumbling infrastructure. However, the huge investment required to achieve a modern network while keeping its current size was almost prohibitive (López and Waddell, 2007). By 1960, the annual railroad losses were of around $280 million per year and they accounted for almost 80% of the federal government deficit (Keeling, 1993).

In this context, the Argentinian government agreed with the World Bank the arrival of a team of experts to work together with government agencies on evaluating the long term economic and strategic viability of the national PTI. The team was led by General Thomas Larkin, an American expert in military logistics that was condecorated for his efforts supplying combat troops during the Second World War. He had experience as a consultant for the modernization of railroad systems in France, Germany, and Japan. Between 1959 and 1961, the World Bank mission scrutinized the profitability of the current railroad and road network. In a world with cheap and stable oil prices, and given the poor material and financial condition of the railroad system, it is not that surprising that their main recommendations (known as the Larkin Plan) were the closure of around 15,000 kms of railroad tracks (roughly 32% of the existing network), the firing of around 70,000

[8]It was implemented by *Decreto Nacional* 505/1958.

railroad employees (from a total of around 200,000), the disposal of the entire fleet of steam vehicles, the closure of most national railroad workshops and factories, and the aggressive improvement and construction of paved roads, in some cases to substitute railroad segments but also to create new connections in the network (Larkin, 1962).[9] These measures were announced by the president in 1961, but after massive protests and strikes exploded in many cities the LP was aborted. In the short term, only 1,000 kms of tracks were closed but closures continued at a slow and steady pace until 1966 when they stabilized totaling around 4,000 kms. However, in 1976 a violent military coup struck the country, and without union resistance the military government decided the immediate closure of over 6,000 kms of railroad tracks.

A key feature of the LP is that despite being abandoned after a few months of its announcement, it was the only long term comprehensive study of the national PTI for more than 30 years. Therefore, it is likely that the recommendations made by the plan influenced the national transport policy for decades ahead[10]. As observed in Figure 1, the LP marked a trend break in the kilometers of both railroads and paved roads. For railroads, it marked the start of a long-term disinvestment and dismantling policy. For roads, it started an almost two decade long period of aggressive construction and improvement. These large scale changes happened in a context marked by population growth, and increasing GDP per capita. From 1960 to 1991[11], the population of Argentina went from 20 to 33 million, and GDP per capita grew around 35% at an average annual rate of around 0.95%. Already by 1986, the number of kilometers of both railroads and the sum of paved and gravel roads stabilized at around 33,000 and 35,000 kms respectively.

3 | DATA DESCRIPTION

This section describes the data we collect to study the evolution of local economies and PTI from 1960 to 1991. We organize the exposition in four subsections: Larkin Plan and the Railroad Network, Roads Network, Census Data, and Geographic Controls.

3.1 | Larkin Plan and Railroad Network

First, we recover data from the Larkin Plan published by the Public Works Ministry in 1961 as "Argentinian Transportation: Long-run Plan"[12]. This publication has three volumes totalling around 3,000 pages. The first volume is an executive summary describing diagnosis, methodology, results and main recommendations, the latter organized by type of transport and time horizon. The details of the plan are described in the next two volumes of approximately 1,000 pages each. We accessed physical copies of the plan at Mariano Moreno National Library, in Buenos Aires. We scanned and digitized two key objects: a map describing the existing operational railroad network in 1960 -our railroad network baseline- and three tables describing the diagnosis and recommendations made for each railroad track by the group of experts. The team divided the railroad network in segments defined as lines between two non-contiguous stations. The tables we digitized contain, for each railroad segment, its name[13], its relative importance for the network (in three

[9]The LP proposed the construction of around 10,000 kms of paved roads in 10 years.

[10]In section 4 we will discuss this assertion in detail.

[11]We finish our sample period in 1991 because there was a census on that year and because we want to abstract from the potentially different effects of the privatizations of large fractions of the railroad and road networks during the 90s.

[12]*Transportes Argentinos: Plan de Largo Alcance* in Spanish

[13]Each segment is named after the stations in its endpoints

categories, high, middle and low), whether these segments were studied in detail or not, and the experts policy recommendation: to maintain, close or to perform a new study.

To convert these images into useful digital objects, we first georeferenced the map with QGIS. The maps in the plan are meant to provide a referential location. In other words, they are not precise enough for our purposes. Hence, we use a present day shapefile and a high definition map of the network in 1946 to improve their accuracy. Later, we identified each segment of the LP in the resulting shapefile. Similarly, we supported this processes with shapefiles of Argentian railroad stations, settlements, towns, and urban zones. All the auxiliary shapefiles mentioned above are produced and published by the Argentinian National Institute Geography [14].

Having mapped the LP information, we were able to compute for each time-invariant district [15], the number of kilometers of the total railroad network in 1960, the number of kilometers of the network under study, and the number of kilometers that were recommended to keep, close or restudy.

Finally, we use the 1981 Annual National Transport Plan published by the military government (Argentina, 1981). We accessed a physical copy of this document in the Library of Railways Museum Foundation, and scanned and georeferenced a map describing the status of the railroads network in 1979. This map indicates which lines were closed since 1977 and which lines were active. When a segment existed in 1960 but it was not drawn in the 1979 map, we assume it was closed before 1977. Moreover, since we know from administrative records that the first round of closures stopped in 1966, we assume the segment closed between 1960 and 1966. We divided LP segments when necessary to accommodate for partial closures [16]. It is important to note that -as shown in Figure 1- the national network didn't suffer changes in its total extension between 1979 and 1986 either. Hence, when describing the status of the network after the closures, we refer to year 1986 to match the timing of our measures for roads infrastructure.

[14] Shapefiles accesible here.

[15] See Census Data subsection for details on geographical units

[16] Both aggregate series and maps indicate that no new railroads were constructed between 1960 and 1979.

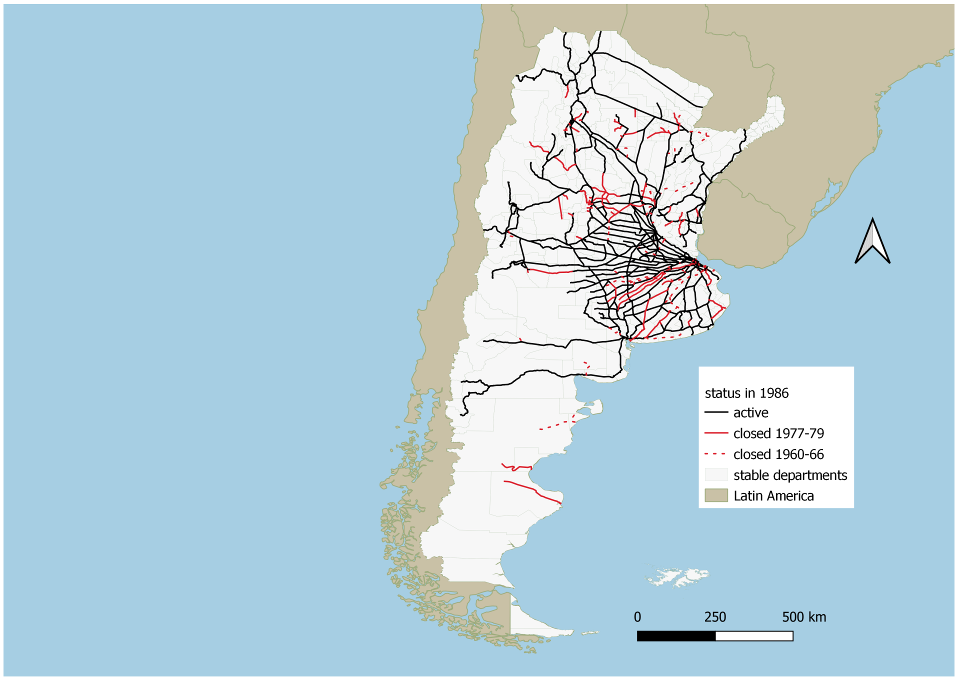

FIGURE 2 Railroads evolution (1960 - 1986).
Notes: This figure displays the status of the 1960 railroad network in 1986. Black solid lines depict segments that were active in 1960 and were still active in 1986. Red lines represent segments that were active in 1960 and not in 1986. Among the red lines, the dashed line depicts segments that closed between 1960 and 1966, and the solid line depicts segments that closed between 1977 and 1979. The data comes from our digitized maps.

The evolution of the railroad network between 1960 and 1986 is illustrated in Figure 2. The figure displays the total railroads network in 1960. Every segment - as defined in the LP - is classified depending on whether it was active in 1986. If not, segments are divided between those closed before the dictatorship (1960-1966) or during the first years of the dictatorship (1976-1979). Panel A of Table 1 displays summary statistics for selected variables measuring railroads infrastructure at the level of stable geographic units as defined by IPUMS (see *Census Data* subsection). These variables are: kms of railroads in baseline (1960), kms of railroads studied in the LP, and kms of railroads active in 1986.

3.2 | Roads Network

To measure changes in road infrastructure we digitized and georeferenced three maps published by the Argentinian Automobile Club[17]. These correspond to years 1954, 1970 and 1986. Although these maps were precise and detailed, we didn't have access to digital objects so we had to georeference and digitize scanned images. Consequently, and just as we did with railroads, we used shapefiles describing the current road network to determine the precise location of historical roads. Importantly, these maps allowed us to classify each route in four decreasing quality categories: paved, gravel, dirt and path. We computed the number of kilometers of roads, in total and per category, for each year. Figure 3 shows the roads network status by 1986 using only paved and gravel roads which are the categories that we

[17] *Automovil Club Argentino* in Spanish, analogous to the AAA in the US.

use for our regression models. Each road is classified based on year of construction[18].

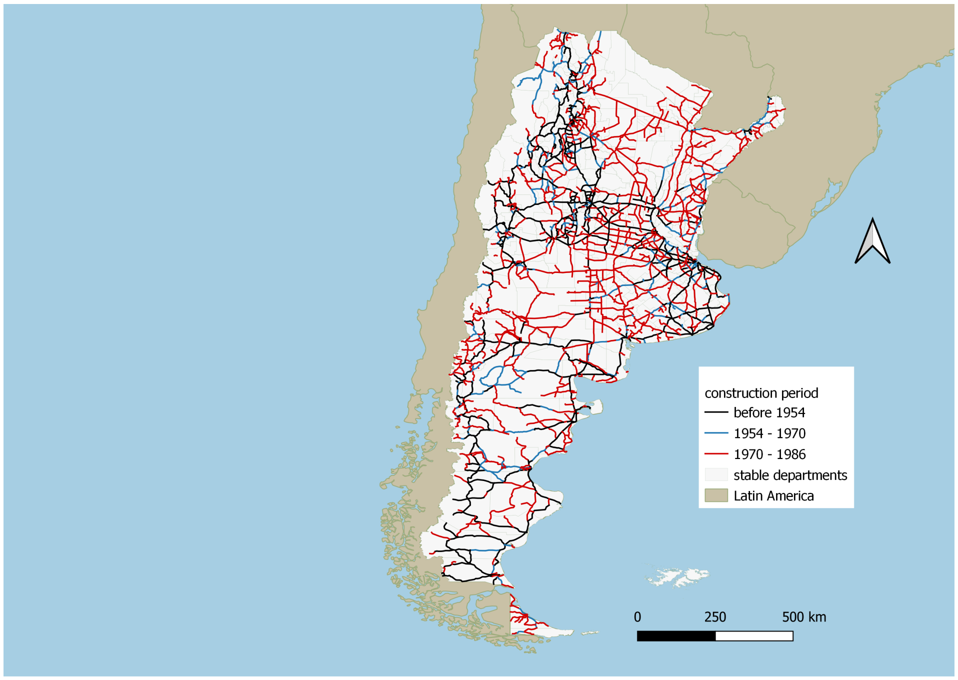

FIGURE 3 Roads in 1986 by year of construction.
Notes: This figure displays the 1986 paved and gravel road network by period of construction. Black lines depict segments that were built before 1954. Blue lines depict segments built between 1954 and 1970. Red lines depict segments built between 1970 and 1986. The data comes from our digitized maps.

Panel A in Table 1 displays summary statistics for selected variables measuring paved and gravel road infrastructure at the level of stable geographic units as defined by IPUMS[19]. These variables are: kms of paved roads at baseline (1954), kms of gravel roads at baseline (1954), kms of paved roads at follow-up (1986), kms of gravel roads at follow-up (1986).

3.3 | Census data

We use demographic and economic variables mostly as outcomes but also to control for pretreatment growth rates. These variables are constructed using sub-national aggregates from different sources. The first corresponds to census data for 1970 and 1991 obtained from Integrated Public Use Microdata Series (IPUMS).

IPUMS homogenizes variables across time to account for definition changes and to harmonize geographical units. Argentina is divided in 23 provinces and the autonomous city of Buenos Aires. Each province, except for Buenos Aires province is divided in *departmentos*, which total 379. Buenos Aires province is divided in 135 *partidos*. The autonomous city of Buenos Aires is divided in 15 *comunas*. Hence, secondary administrative units are called *departamentos*, *partidos* or *comunas*. For simplicity, we refer to them as *districts*. By merging some secondary units, IPUMS provides 314 time-invariant boundary divisions for the period

[18] The map omits a very small amount of roads that appear as paved or gravel in 1954 and 1986 but not in 1970. In principle, it is possible that a few roads where temporarily closed or that they decreased their quality. We abstract from that possibility.
[19] See Census Data subsection for details on geographical units

1970-2010[20]. IPUMS also provides the corresponding shapefile.

The dependent variables constructed from IPUMS data are total population, urban population, number of people living in the same province they were born, and share of workers employed in each economic sector. Unfortunately, IPUMS data for Argentina only starts in 1970, but as the PTI was changing before then, we complement IPUMS with 1960 and 1947 Census data on total and urban population, digitizing tables from the official publications (Instituto Nacional de Estadisticas y Censos, 1960) [21].

Additionally, we include data from industrial and agricultural censuses. Industrial census data correspond to years 1954 and 1985. From this source, we obtain district level data on number of firms, employment, total paid wages and production value. We digitized data from the Agricultural census data in years 1960 and 1988. From this source, we obtain district level data on total area devoted to agriculture and number of farms.

Panels B and C of Table 1 display summary statistics for our main population and labor variables at the level of stable geographic units as defined by IPUMS. In Panel B, we display them for periods before the infrastructure changes, while in Panel C we do so for 1991, our post-treatment period.

3.4 | Geographical controls

We construct geographic controls from different sources. For each stable geographical unit we calculate average potential wheat yield from FAO/IIASA (2011), average caloric suitability measures from Galor and Özak (2016), average elevation from USGC (1996), average ruggedness from Nunn and Puga (2012), and distance to Buenos Aires (using QGIS). Summary statistics for the main variables are displayed in Panel D of Table 1.

4 | EMPIRICAL STRATEGY

Our objective is to identify the different effects of railroad and road infrastructure on population and production. We proceed in the following way. First, for each district we compute the total length of roads and railroads in kilometers around 1960 and around 1986, once the length of both networks has stabilized (see Figure 1). Now define the following:

$$\begin{aligned} \Delta\theta_{dp}^{rail} &= rail_{dp}^{1986} - rail_{dp}^{1960} \\ \Delta\theta_{dp}^{road} &= road_{dp}^{1986} - road_{dp}^{1954} \end{aligned} \tag{1}$$

Where $rail_{dp}^{t}$ is the length of railroads in kilometers in district d in province p at year t, and $road_{dp}^{t}$ is the length of roads in kilometers in district d in province p at year t.

Now, given that we do not have all outcome measures exactly at 1960, we superindex *pre* to be some year before or equal to 1960, and *post* be some year after or equal to 1986. Define $\Delta \ln y_{dp} = \ln y_{dp}^{post} - \ln y_{dp}^{pre}$ for some outcome variable (i.e. population, number of industrial establishments, agricultural production, etc). Then we can estimate the following

[20] In their classification they collapse the 15 *comunas* into one unit that we drop from our analysis.

[21] The census questionnaire included questions on industrial classification of main employment. Nevertheless, the corresponding results were not published at the level of secondary administrative units. As of today, we don't have access to the 1960 and 1947 Census microdata. To the best of our knowledge, they do not exist anymore.

TABLE 1 **Descriptive Statistics (selected variables)**

Panel A: Infrastructure

	Mean	SD	Min	Max
Railroads 1960 kms	137	140	0	780
Railroads 1986 kms	107	118	0	780
Larkin plan kms - studied	67	98	0	767
Roads 1954 kms - paved	34	50	0	302
Roads 1954 kms - gravel	69	159	0	1,241
Roads 1986 kms - paved	167	151	0	1,018
Roads 1986 kms - glavel	91	258	0	2,592

Panel B: Outcomes at baseline

	Mean	SD	Min	Max
Population 1960	43,597	78,346	1,069	663,443
Urban pop 1960	37,961	78,427	0	650,056
Migration 1970 %	93	6	57	100
Primary labor 1970 %	37	22	0	89
Secondary labor 1970 %	22	12	1	56
Tertiary labor 1970 %	41	15	5	89

Panel C: Outcomes at follow-up

	Mean	SD	Min	Max
Population 1991	102,959	162,732	14,848	1,291,921
Urban pop 1991	80,844	149,651	0	1,135,110
Migration 1991 %	82	10	33	97
Primary labor 1991 %	24	18	0	75
Secondary labor 1991 %	19	9	2	51
Tertiary labor 1991 %	57	14	20	97

Panel D: Geographic Controls

	Mean	SD	Min	Max
Elevation mts	420	689	4	4,000
Ruggedness mts	58	93	4	587
Wheat pot. yield - tons/ha	161	324	0	1,915
Area km2	8,166	11,767	27	85,003
Dist. to Buenos Aires kms	670	435	5	2,319

Notes: This table displays summary statistics for the main variables that we use throughout the paper at the level of stable geographic units as defined by IPUMS. Panel A displays our main PTI variables. Panel B displays our main outcomes at baseline. Panel C displays our main outcomes at follow-up. Panel D displays geographic variables that we use as controls.

model:[22]

$$\Delta \ln y_{dp} = \alpha_p + \beta_1 \Delta\theta^{rail}_{dp} + \beta_2 \Delta\theta^{road}_{dp} + \lambda X_{dp} + \epsilon_{dp} \quad (2)$$

Where X_{dp} is a vector of time invariant geographic controls. Our coefficients of interest are β_1 and β_2. To interpret causally those coefficients from a OLS estimation would require the assumptions that conditional on geographic features, X_{dp}, the changes in railroad and road infrastructure at the district level are randomly assigned within the province.

4.1 | Instrumental Variables

It is possible that, conditional on X_{dp}, the placement or dismantling of transport infrastructure is not randomized across space within provinces during the pre and post periods. In fact, given that policymakers likely take economic factors into account when deciding to build roads or to dismantle a railroad track, this assumption seems unlikely to hold. To address these concerns we instrument $\Delta\theta^{rail}_{dp}$ and $\Delta\theta^{road}_{dp}$.

4.1.1 | Instrument Definitions

Our first instrument is motivated by the LP. We think that it is possible to use this plan to study the effects of PTI on economic development because the plan's explicit objective was to reduce public deficit and allow for profitability in the network as a whole. Furthermore, the plan was developed forecasting a world with low oil prices and long-run national industrialization. Both of those predictions were dramatically wrong, suggesting that planners had poor forecasting capacity to target locations depending on future economic development. In figure 4 we display the evolution of the spot nominal price for the crude oil WTI barrel from 1950 to 1991. The huge spike starting in 1973 and lasting well through the 80s was extremely hard to predict. We refer to Bogliaccini (2013) to a discussion of the deindustrialization process that consolidated during the 80s in Latin America and Argentina.

[22]Given that the Argentinian census of 1960 does not have microdata available we had to digitize outcomes at the district level manually from PDFs (see details in section 3). Therefore, for many outcomes we estimated the equation given by:

$$\Delta \ln \tilde{y}_{dp} = \alpha_p + \beta_1 \Delta\tilde{\theta}^{rail}_{dp} + \beta_2 \Delta\tilde{\theta}^{road}_{dp} + \lambda X_{dp} + \epsilon_{dp}$$

where:

$$\Delta \ln \tilde{y}_{dp} = \ln y^{post}_{dp} - \ln y^{1970}_{dp}$$

and where:

$$\Delta\tilde{\theta}^{rail}_{dp} = rail^{1986}_{dp} - rail^{1970}_{dp}$$
$$\Delta\tilde{\theta}^{road}_{dp} = road^{1986}_{dp} - road^{1970}_{dp}$$

Essentially, when we use changes in outcomes from 1970 to some post year close to 1986 we only use changes in road and railroad infrastructure that happened between 1970 and 1986.

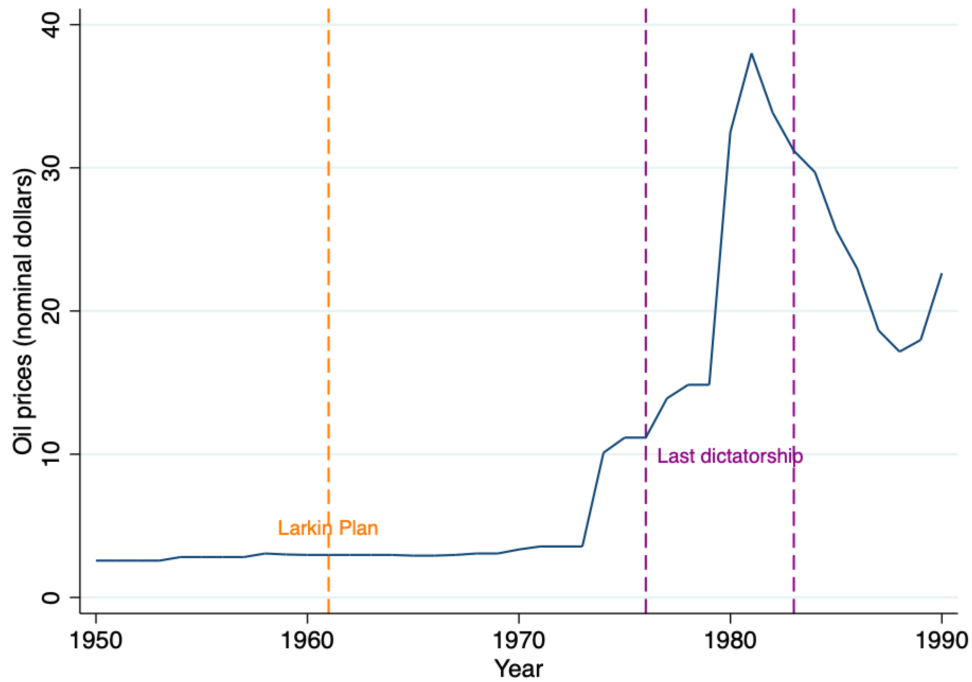

FIGURE 4 Nominal spot price of WTI crude oil barrel from 1950-1991.
Notes: This figure displays the time series of the nominal spot price of a WTI crude oil barrel from 1950 to 1991. Data comes from the Federal Reserve Bank of St.Louis and it is annualized by assigning the price of January to the year. The dashed vertical yellow line is shown in 1961, the year of the publication of the LP. The dashed vertical purple lines are shown in 1976 and 1983, the start and end years of the last prevailing dictatorship in Argentina.

The LP studied the railroad and road infrastructure and proposed a railroad and a road network. In this paper we don't use the LP proposals that were never fully implemented, but instead we exploit the fact that some railroads had a higher likelihood of closing due to some arbitrariness induced by the LP. Due to resource constraints, the group of experts did not study the costs and revenues for the whole universe of railroad segments. Instead, they collected two basic statistics for each of them:

- Traffic density: freight quantity that passes through the segment in a year.
- Originated traffic: freight quantity that enters the railroad network through a segment (normalized by the segment's length) in a year.

Then, each segment's revenue and cost structure were studied in detail if and only if traffic density was less than 1 million gross tons a year or if originated traffic was less than 500 tons per kilometer per year[23]. A study of a railroad segment consisted on comparing costs and revenues of railroad operations based on forecasts of the cost of transporting production and based on how expensive it would be to transport production through roads in case of closure. After studying a segment the possible recommendations were to keep

[23]There were some exceptions to this cutoff rule. Some segments were recommended to close despite not being studied. Those segments correspond to segments that were previously studied during the elaboration of some previous study or those that were too short in length. They constitute a very small fraction of the network.

it, to close it, or to study it further in the future. Importantly for identification, *we don't construct our instrument based on whether the segment was recommended to close, a measure more likely to be targeted, but instead we construct it based on whether segments were studied or not which was determined through ad-hoc cutoffs*.

We define our first instrument as:

$$S_{dp} = \mathtt{studied}_{dp} \tag{3}$$

Where $\mathtt{studied}_{dp}$ is the kilometer length of railroads that were studied by the LP in 1960 in district d in province p. Figure 5 shows, on the left, a table from the LP displaying the fraction of the network that was studied, and on the right, shows that studied segments had a much higher probability of being closed in the two decades after the study.

Ferrocarril	Kilómetros		
	Líneas que requirieron estudio	Líneas que no requirieron estudio	Total
Roca (incl. Patagónicos)	3.104	6.332	9.436
Mitre	2.290	4.165	6.455
San Martín..................	1.889	2.755	4.644
Sarmiento	1.521	2.324	3.845
Total trocha ancha	8.804	15.576	24.380
Urquiza	1.488	2.070	3.558
Belgrano	6.822	8.468	15.290
Total general	17.114	26.114	43.228

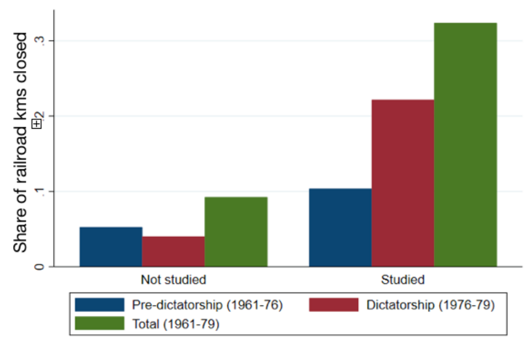

FIGURE 5 Larkin Plan - Studied railroads and share of closures by studied status from 1961 to 1979. *Notes*: On the left we display a table from the LP hardcopy that shows the number of kilometers of railroad that were either studied or not within each railroad line as well as the totals. On the right we display the share of railroad kilometers that were closed from 1961 to 1979 by studied status. The data comes from our digitized maps.

Figure 6 shows the spatial distribution of studied railroad segments. Segments that were selected for a study constituted almost 40% of the network and were dispersed throughout the country. For reference, we also show in Figure 7 the recommendation made by the LP for each segment.

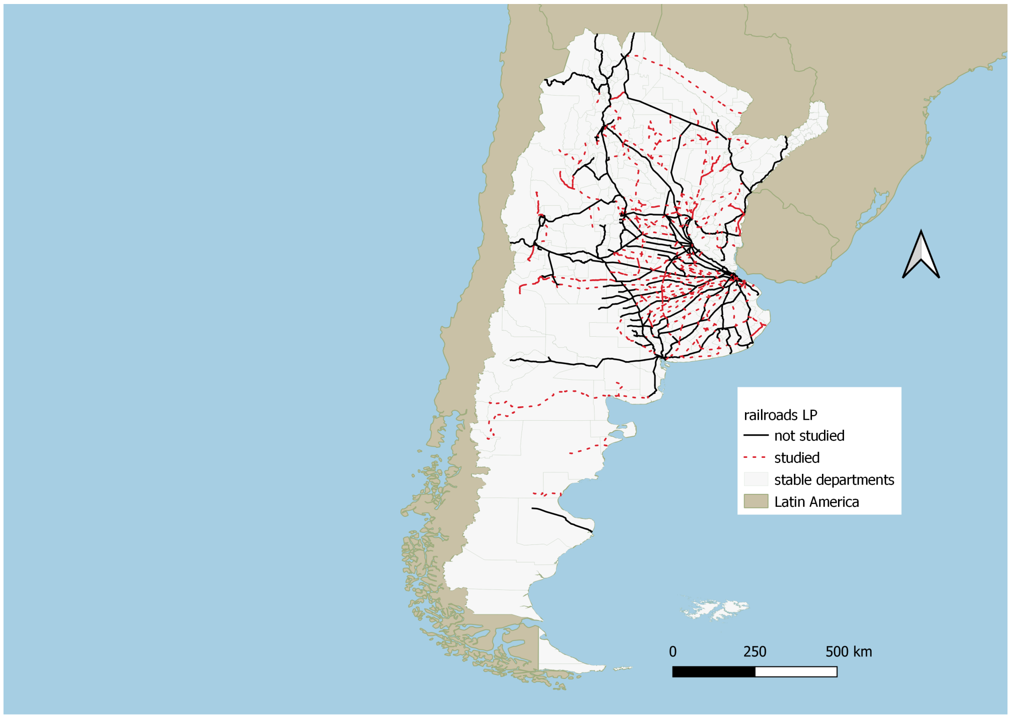

FIGURE 6 Railroad tracks by LP studied status. *Notes*: This figure displays the LP studied status for the 1960 railroad network. Black solid lines depict segments that were not studied, while red dashed lines represent segments that were studied by the LP. The data comes from our digitized maps.

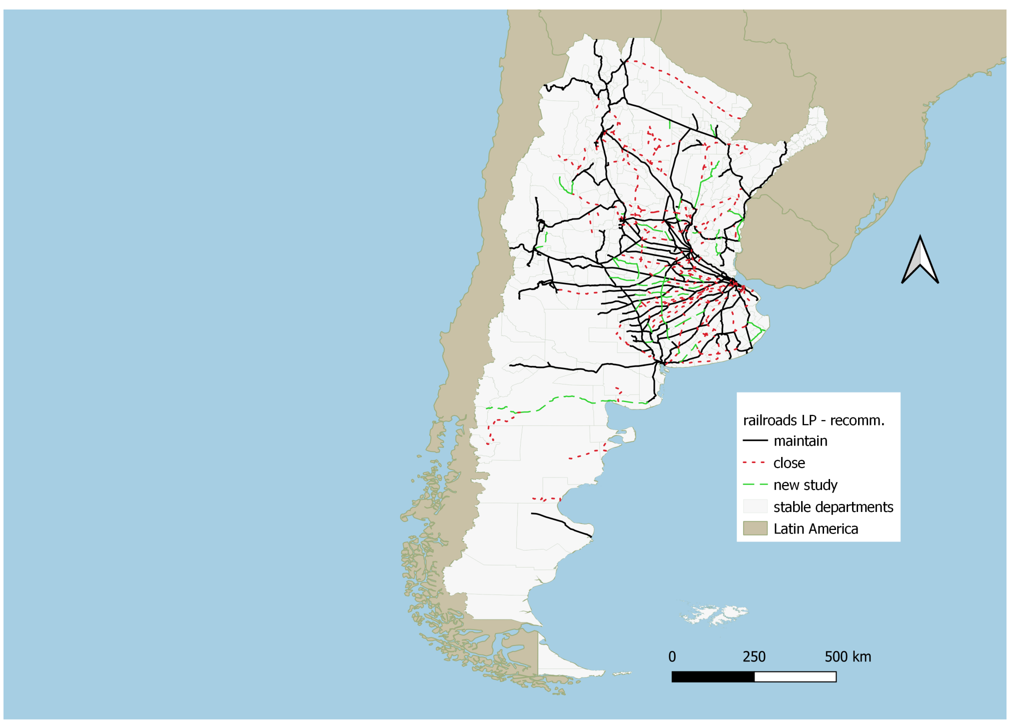

FIGURE 7 Railroad tracks by LP recommendation. *Notes*: This figure displays the LP recommendation status for the 1960 railroad network. Black solid lines depict segments that were recommended to remain active. Red dashed lines represent segments that were recommended to close. Green dashed lines depict segments that were recommended to be studied again in the future.

For our second instrument we build minimum spanning tree networks as in Faber (2014); Perez (2017); Alder (2016). We construct two hypothetical road networks that connect in a single graph all 23 province capitals plus the top 5 regional cities for each region[24] constituting a total of 50 cities, as most are both a capital and a top 5 regional city.

Our first hypothetical network is the result of applying Kruskal's minimum spanning tree algorithm (Kruskal, 1956) to bilateral Euclidean distances between the selected cities. We refer to this network as Euclidean spanning tree network and is depicted in solid orange straight lines in Figure ??. We think of this network as the one that a social planner would have built if the only objective was to connect all capitals minimizing the total euclidean length of the network.

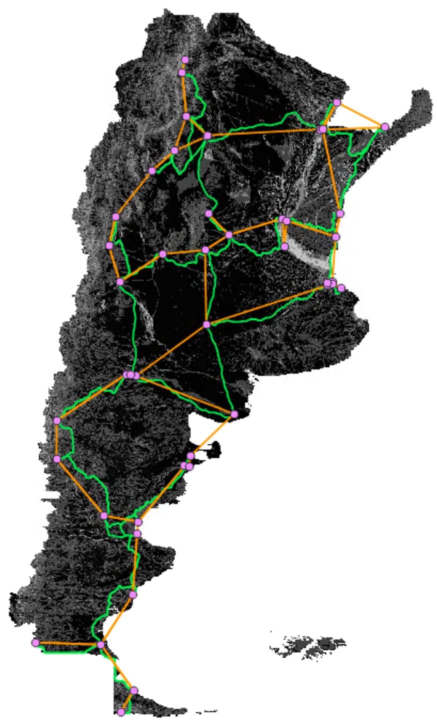

FIGURE 8 Road building costs, selected cities, and hypothetical networks. *Notes*: This figure displays the targeted cities, our two hypothetical networks, and the construction cost assumed for one of them. The purple dots depict the targeted cities, chosen as the all 23 province capitals plus the top 5 regional cities for each Argentinian region as defined in footnote 24. The orange solid line depicts the result of applying Kruskal's minimum spanning tree algorithm (Kruskal, 1956) to bilateral Euclidean distances between the targeted cities. The green solid line depicts the result of applying Dijkstra's algorithm (Dijkstra et al., 1959) to compute all bilateral least cost paths along the construction cost layer given by the equation in footnote 25 between the targeted cities, and then applying to the resulting least cost paths Kruskal's minimum spanning tree algorithm to obtain a continuous graph that minimizes those construction costs. We depict in grey scale the pixel by pixel construction cost implied by the equation in footnote 25, where the lighter coloring implies more expensive construction costs.

[24]We defined 9 regions, each one corresponding to a set of adjacent provinces with similar geography as defined in the publication of the 1960 Population Census. The *Central* region corresponds to the provinces of Cordoba and Santa Fe. The *Chaco* region corresponds to Chaco, Formosa and Santiago del Estero. The *Cuyo* region corresponds to Mendoza, San Juan and San Luis. The *Mesopotamia* region corresponds to Corrientes, Entre Rios and Misiones. The *North West* region corresponds to Catamarca, Jujuy, La Rioja, Salta and Tucuman. The *Pampa* region corresponds to Buenos Aires and La Pampa. Finally, the *Patagonia* region corresponds to Neuquen, Rio Negro, Chubut, Santa Cruz and Tierra del Fuego

We also constructed a version of this hypothetical network that instead of computing bilateral distances computes least cost paths minimizing construction costs taking into account the geography of the terrain[25]. In this case we first use Dijkstra's algorithm (Dijkstra et al., 1959) to compute all bilateral least cost paths along the construction cost layer between the points we want to connect. Then we apply Kruskal's minimum spanning tree algorithm to obtain a continuous graph that minimizes construction costs. The resulting network is described by green solid lines in Figure ??

Intuitively, this hypothetical networks exploit the inconsequential units approach: districts that lie along them are more likely to receive roads just due to their cost minimizing location without being targeted for economic reasons.

Our second instrument based on the either hypothetical network is defined as:

$$H_{dp} = \mathrm{hypothetical}_{dp} \tag{4}$$

Where $\mathrm{hypothetical}_{dp}$ is the length of hypothetical roads in kilometers that belong to district d of province p.

4.1.2 | Controls and Discussion of Identification

Controls Our specification includes province fixed effects. Therefore, identifying variation comes from comparing changes in outcome variables for districts within the same province. To address the concern that larger districts within a province are, all things equal, more likely to have more kilometers of studied railroads and more likely to be along the minimum spanning tree networks, we control for district land area. Because within a province districts closer to Buenos Aires could have different rates of economic growth due to their proximity to the capital and main trade hub, we control directly for distance to Buenos Aires. Because different districts within a province may have different comparative advantages and different geographic characteristics, we control for mean ruggedness, mean elevation, and mean suitability for growing wheat.

Identification assumptions Conditional on the above mentioned controls, our first instrument assumes that, within a province, having more kilometers of studied railroads does not affect economic or sectorial development other than through the probability with which a district gets a railroad dismantled or a road built. As this is a novel instrument, we discuss extensively the plausibility of this assumption below.

Our second instrument, assumes instead that, within a province, lying along the euclidean spanning tree network does not affect economic development other than through the increased probability of getting more kilometers of roads built or railroads dismantled. We refer to Faber (2014) for a deeper discussion of these instruments.

Intuitively, our first instrument should be more important for generating exogenous variation for the kilometers railroads and the second for the kilometers of roads.

Discussion of LP instrument The main threat to our LP instrument is that even though

[25] To define road construction costs we use the same equation as in Faber (2014):

$$c_i = 1 + \mathrm{Slope}_i + 25 * \mathrm{Developed}_i + 25 * \mathrm{Wetland}_i + 25 * \mathrm{Water}_i$$

Where i is a pixel of land, c_i is the construction cost, Slope_i is the terrain slope of pixel i, $\mathrm{Developed}_i$ is an indicator for whether pixel i is built, $\mathrm{Wetland}_i$ is an indicator for whether pixel i is a wetland, and Water_i is an indicator for whether pixel i is a body of water. All GIS layers used to construct the construction cost raster were taken from the National Geography Institute from Argentina and are available here.

the cutoffs to determine eligibility into study may be somehow arbitrary, the variation in the total studied railroad tracks (in kms) could come from segments with traffic statistics way above or below the cutoffs. If this was the case, it could be that the Larkin commission ended up studying railroad segments in already declining districts, and not studying those related to already growing districts. If so, our instrument would be correlated with previous trends in local development, which can generate mean reversion bias. If additionally, previous trends correlate with outcome growth, our instrument would violate the exclusion restriction.

We address this concern in tables 2, 3, 4, 5 and 6. Table 2 splits the districts sample into two groups: those without studied segments and those with at least one studied segment. For each group, it describes previous trends in variables that we consider good proxies of economic development: changes from 1947 to 1960 in the logarithm of population and urban population respectively. In the case of population, the implied changes in logs are the same across the groups up to the second decimal. In the case of urban population, the implied changes in logs are similar but slightly higher for districts without studied segments. For both proxies of economic development, the third and fourth columns show that the difference in previous trends are very small and statistically insignificant.

TABLE 2 Balance in previous trends by treatment status 1960-1947

	No studied segments	At least one studied segment	Diff-in-means	SE Diff-in-means
Change in log pop 1960-1947	-.23715	-.2439866	.0068366	.0807696
Number of districts	80	199	279	.
Change in log urb pop 1960-1947	.4922051	.469289	.0229161	.0766837
Number of districts	62	173	235	.

Notes: This table displays changes in the logarithm of population and urban population between 1947 and 1960 at the constant district level for two subgroups. The first column does it for the group of districts that have no studied segments. The second does it for the group of districts that have at least one studied segment. The third column displays the difference in means between the two groups, and the fourth column the standard errors associated to these differences.

Tables 3 and 4 complement the previous exercise with a set of regressions. In table 3, the dependent variable is the change in log population from 1947 to 1960. The first column uses a dummy that equals one when the district has at least one studied segment. Hence, this specification is analogous to the mean test reported in Table 2. It is possible that dividing the sample in terms of whether they have at least one studied segments is not the relevant comparison to make. Alternatively, column 2 uses a dummy variable that equals one for districts for which the total of studied railroad tracks (in kms) was above the in-sample median. Column 3 uses dummy variables for each quartile of studied railroad tracks (in kms), using the first quartile as base category. Finally, column 4 uses a third-degree polynomial on the number of studied railroad tracks (kms). None of the estimated coefficients is statistically significant. Table 4 shows the same regressions but using the change in log urban population from 1947 to 1960 as dependent variable. Again, none of the estimated coefficients is statistically significant. Altogether, these results suggest that the number of studied kilometers at the district is not correlated with previous trends in economic development.

Finally, we show placebo tests in tables 5 and 6. In these tables we use our IV strategy just as described above but using changes from 1947 to 1960 as dependent variables, for total and urban population respectively. If our instrument violates the exclusion restriction because

TABLE 3 Change in the log population 1960-1947

	(1)	(2)	(3)	(4)
At least one studied segment=1	-0.0274			
	(0.0771)			
Studied kms above median=1		-0.119		
		(0.0742)		
Quintile of studied kms=2			0.0976	
			(0.142)	
Quintile of studied kms=3			-0.0529	
			(0.0940)	
Quintile of studied kms=4			-0.00332	
			(0.0980)	
Quintile of studied kms=5			-0.0968	
			(0.103)	
Studied railroad tracks (kms)				-0.00153
				(0.00209)
Studied railroad tracks (kms) square				0.00000585
				(0.0000172)
Studied railroad tracks (kms) cube				-3.27e-12
				(4.58e-08)
Studied railroad tracks (kms) fourth				-1.05e-11
				(3.57e-11)
R-squared	0.425	0.431	0.430	0.441
Observationss	279	279	279	279

Notes: The dependent variable is the change in the logarithm of population from 1947 to 1960. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. From column to column we change the independent variable. In column (1), it is a dummy for whether the district has at least one segment studied. In column (2), it is a dummy for whether the district is above median in terms of studied railroads in kms. In column (3), they are a set of dummies for the quartiles of studied railroads in kms. In column (4), they are a fourth degree polynomial in the number of studied railroads in kms.

of a correlation with previous development trends, we would expect the IV coefficients associated to infrastructure changes to be statistically significant.

The first column of each table show OLS results. We observe that the OLS coefficients for railroads are positive but statistically insignificant. On the other hand, the OLS coefficient for roads is negative, indicating that roads where improved or constructed also in declining districts. Columns (2) and (3) in tables 5 and 6 show IV estimates. We observe that all the IV coefficients are statistically insignificant, regardless of which dependent variable we use (population or urban population) and which hypothetical network we use to construct the instrument for roads (euclidean or least cost paths). Lack of significance could also be due to weak instruments. We argue first that the Cragg-Donald F statistics are quite high in all

TABLE 4 Change in the log urban population 1960-1947

	(1)	(2)	(3)	(4)
At least one studied segment=1	0.0937			
	(0.0849)			
Studied kms above median=1		-0.0258		
		(0.0799)		
Quintile of studied kms=2			0.0976	
			(0.153)	
Quintile of studied kms=3			0.0840	
			(0.101)	
Quintile of studied kms=4			0.146	
			(0.107)	
Quintile of studied kms=5			0.0304	
			(0.113)	
Studied railroad tracks (kms)				0.00197
				(0.00283)
Studied railroad tracks (kms) square				-0.0000213
				(0.0000296)
Studied railroad tracks (kms) cube				7.02e-08
				(0.000000103)
Studied railroad tracks (kms) fourth				-7.06e-11
				(1.09e-10)
R-squared	0.239	0.235	0.243	0.237
Observationss	235	235	235	235

Notes: The dependent variable is the change in the logarithm of urban population from 1947 to 1960. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. From column to column we change the independent variable. In column (1), it is a dummy for whether the district has at least one segment studied. In column (2), it is a dummy for whether the district is above median in terms of studied railroads in kms. In column (3), they are a set of dummies for the quartiles of studied railroads in kms. In column (4), they are a fourth degree polynomial in the number of studied railroads in kms.

cases, and second that effects do show up in section 5 when using the same instruments but for the periods when the actual changes in infrastructure happened. Altogether, the evidence presented in this subsection suggest that the correlation of the LP instrument with previous trends in local economic development is, if anything, not a first order problem. Additionally, in section 5 we show that our results are robust to controlling directly for previous trends in economic development.

4.1.3 | First Stage

First stage results for our main specification are reported in Table 7. Panel A shows the results for the change in railroads (kms). Columns (1) and (2) define the endogenous

TABLE 5 Placebo: Change in log population 1960-1947

	(1) OLS	(2) IV EUC	(3) IV LCP
Change in kms of railroads 1986-1960	0.000590	0.00163	0.00191
	(0.000601)	(0.00133)	(0.00142)
Change in kms of roads 1986-1954	-0.000660***	0.000251	0.000357
	(0.000148)	(0.000293)	(0.000341)
Cragg-Donald (multivariate) F-stat		29.8847	24.1826
Observations	279	279	279

Notes: The dependent variable is the change in the logarithm of population from 1947 to 1960. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Column (1) corresponds to ordinary least squares (OLS). Column (2) corresponds to our instrumental variables (IV) strategy when the hypothetical network is constructed using euclidean distances. Column (3) corresponds to our instrumental variables (IV) strategy when the hypothetical network is constructed using least cost paths.

TABLE 6 Placebo: Change in log urban population 1960-1947

	(1) OLS	(2) IV EUC	(3) IV LCP
Change in kms of railroads 1986-1960	0.000895	0.000473	0.000753
	(0.000600)	(0.00104)	(0.00112)
Change in kms of roads 1986-1954	-0.000371**	0.000385	0.000597
	(0.000174)	(0.000353)	(0.000431)
Cragg-Donald (multivariate) F-stat		49.6255	31.6105
Observations	235	235	235

Notes: The dependent variable is the change in the logarithm of urban population from 1947 to 1960. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Column (1) corresponds to ordinary least squares (OLS). Column (2) corresponds to our instrumental variables (IV) strategy when the hypothetical network is constructed using euclidean distances. Column (3) corresponds to our instrumental variables (IV) strategy when the hypothetical network is constructed using least cost paths.

variable as changes in railroads from 1960 to 1986. In both cases the instrument for railroads is the number of studied railroad tracks. In column (1) the instrument for roads uses the hypothetical network based on euclidean distances while column (2) uses the hypothetical network based on least cost paths. Columns (3) and (4) replicate (1) and (2) but using changes in railroads from 1970 to 1986 instead. The estimated coefficients associated to the roads instruments are small and statistically insignificant. On the other hand, the instrument for railroads has a sizable negative and statistically significant impact on railroads dismantling. One extra kilometer of studied tracks is associated to 0.34 kilometers of dismantled tracks between 1960 and 1986 (columns 1 and 2). On the other hand, one extra kilometer of studied railroads is associated to 0.22 kilometers of dismantled tracks between 1970 and 1986. We attribute this difference the fact that 4,000 kilometers of tracks were closed from 1960 to 1970 so by 1970 there were less tracks that were candidates to close.

Panel B shows the results for the change in paved and gravel roads (kms). Columns (1) and (2) define the endogenous variable as changes in roads from 1954 to 1986. In both

TABLE 7 **First stage regressions**

Panel A: Change in railroads (kms)

	(1) Change in kms of railroads 1986-1960	(2) Change in kms of railroads 1986-1960	(3) Change in kms of railroads 1986-1970	(4) Change in kms of railroads 1986-1970
Euclidean MST network (kms)	0.00195		0.0398	
	(0.0362)		(0.0297)	
Least-cost MST network (kms)		0.000462		0.0108
		(0.0243)		(0.0200)
Studied railroad tracks (kms)	-0.341***	-0.341***	-0.224***	-0.215***
	(0.0324)	(0.0314)	(0.0266)	(0.0258)
F-stat from testing both instruments = 0	24.51	24.51	14.41	14.12
Observations	311	311	311	311

Panel B: Change in paved and gravel roads (kms)

	(1) Change in kms of roads 1986-1954	(2) Change in kms of roads 1986-1954	(3) Change in kms of roads 1986-1970	(4) Change in kms of roads 1986-1970
Euclidean MST network (kms)	1.440***		0.997***	
	(0.108)		(0.0896)	
Least-cost MST network (kms)		0.857***		0.621***
		(0.0771)		(0.0617)
Studied railroad tracks (kms)	0.372***	0.542***	0.219***	0.326***
	(0.0969)	(0.0997)	(0.0802)	(0.0798)
F-stat from testing both instruments = 0	75.25	61.76	66.77	60.77
Observations	311	311	311	311

Notes: This table shows OLS estimates of the first stages. In Panel A, the dependent variable is the change in kilometers of railroads from 1960 to 1986 (columns 1 and 2) and from 1970 to 1986 (columns 3 and 4). In Panel B, the dependent variable is the change in kilometers of railroads from 1954 to 1986 (columns 1 and 2) and from 1954 to 1986 (columns 3 and 4). The instruments in odd-numbered columns correspond to kilometers of studied tracks and kilometers of hypothetical network when this is constructed using euclidean distances. The instruments in even-numbered columns correspond to kilometers of studied tracks and kilometers of hypothetical network when this is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires.

cases the instrument for railroads is the number of studied railroad tracks. In column (1) the instrument for roads uses the hypothetical network based on euclidean distances while column (2) uses the hypothetical network based on least cost paths. Columns (3) and (4) replicate (1) and (2) but using changes in roads from 1970 to 1986 instead. The estimated coefficients associated to the roads instruments are positive, sizeable, and statistically significant. On the other hand, the instrument for railroads has a sizable positive and statistically significant impact as well.

Interestingly, districts with more studied kilometers of track are also more likely to receive the construction roads after 1954. We interpret that as evidence of substitution from railroads to roads, as the LP recommended. On the other hand, districts that lie along our hypothetical networks are more likely to receive paved and gravel roads after 1954. We observe that in both periods, the point estimate for the least-cost hypothetical network is smaller than the one for the euclidean hypothetical network, which can be attributed to the fact the the number of kilometers is unambiguously higher when using cost-distances. On the other hand, for each instrument the coefficient is smaller for period 1970-1986 than for period 1960-1986. This can be attributed to the fact that mechanically there were more roads built from 1960 to 1991 than from 1970 to 1991.

Our first stage results are reassuring as they have quite large F-statistics[26] and each of our instruments shift differently the endogenous variables. Hence, under the exclusion restrictions described in section 4.1.2, we will be able to recover the causal effects of railroads and roads separately.

5 | RESULTS

This section presents causal estimates of reduced form impacts of railroads and roads on population, urbanization, migration and labor in each major economic sector. First, we show that conventional IV estimates can be misleading when omitting potential substitution of different transport modes. When allowing for this substitution, our instrumental variables approach implies that for every 100 kms of *dismantled* railroads in a district, total population fell in 0.29 log-points, urban population fell in 0.26 log-points, and the population born in the same province they live fell in 0.29 log-points, indicating outflow migration. Consistently, we find negative impacts on the volume of industrial activity as measured by total value of production, total paid wages and employment. When looking at structural change, we find that every 100 kms of dismantled railroads led to a reduction of 4.1% in the share of labor in the primary sector (agriculture and mining), an increase of 4.1% on the share of labor in manufacturing and an increase of 1.8% on the share of labor in non-tradable industries (construction, electricity, water, and gas). On the other hand, we don't find any impact of roads infrastructure on total population, urban population and total economic activity. Nevertheless, we find that every 100 km of railroads construction lead to an increase of 0.10 log points in population born in in the same province, a 1.2% decrease in the share of labor in other services, and weak evidence of this share being gained by manufacturing and business services.

In addition, we test whether the coefficient associated to roads (β_2 in equation 2) is higher or equal than the coefficient associated to railroads (β_1 in equation 2). This test aims to shed light on the consequences of the substitution of railroads by roads proposed by the LP. Assuming the substitution is 1 km of roads for 1 km of railroads, the net average impact of the substitution will be given by $\beta_2 - \beta_1$. In most specifications we reject the null

[26]For reference, in Table 7 we report F statistics that test on each model whether both instruments are jointly equal to zero. As we have two endogenous variables and two instruments, it turns out that the relevant statistics for assessing the strength of the instruments in the equation system is the Cragg-Donald multivariate F statistic. In Section 5 we report those along with our instrumental variable estimates.

hypothesis ($\beta_2 - \beta_1 \geqslant 0$), which overall suggest the substitution proposed by the LP was not conducive to local development. For simplicity, we will refer to this test as *LP test* in what follows.

5.1 | Population, urbanization and migration

First we present estimates for population, urban population and migration. Since LP was published in 1961, the term $\Delta \ln y_{dp}$ ideally should reflect changes between 1960 and 1991. Unfortunately IPUMS data starts on 1970. We address this issue in two ways. On one hand, we digitized data on total population and urban population for 1960[27]. On the other hand, we also exploit differences in infrastructure for the period 1970-1991 [28]

Table 8 shows results for the change in logarithm of population from 1960 to 1991. Correspondingly, changes in the infrastructure network are constructed using period 1960 - 1986. As most tables in this section, Table 8 is organized in 6 columns. All regressions control for province fixed effects and include the following geographic controls: distance to Buenos Aires, ruggedness, elevation, agricultural suitability for wheat, and logarithm of total land area. Columns 1 and 2 show OLS estimates. Columns 3 and 4 show IV estimates when the hypothetical network is constructed using euclidean distances. Columns 5 and 6 show IV estimates when the hypothetical network is constructed using least-cost paths. Columns 1, 3 and 5 include only geographic controls. Columns 2, 4 and 6 include change in log urban population from 1947 to 1960 to control directly for previous trends in local economic development.

TABLE 8 Change in log population 1991-1960

	(1) OLS	(2) OLS	(3) IV EUC	(4) IV EUC	(5) IV LCP	(6) IV LCP
Change in kms of railroads 1986-1960	0.00129***	0.000853*	0.00294***	0.00254***	0.00354***	0.00285***
	(0.000470)	(0.000451)	(0.00102)	(0.000778)	(0.00109)	(0.000836)
Change in kms of roads 1986-1954	-0.000141	-0.000211	-0.000169	-0.000153	0.0000721	0.0000856
	(0.000136)	(0.000155)	(0.000237)	(0.000260)	(0.000271)	(0.000316)
Change in log urban population 1960-1947		0.0730		0.0586		0.0625
		(0.0508)		(0.0528)		(0.0535)
P-value for testing β_2 >= β_1			.0003	.0001	.0001	0
Cragg-Donald (multivariate) F-stat			35.6208	49.6014	29.4324	32.137
Observations	311	235	311	235	311	235

Notes: The dependent variable is the change in the logarithm of population from 1960 to 1991. Columns (1) and (2) report OLS estimates. Columns (3) and (4) report IV estimates when the hypothetical network is constructed using euclidean distances. Columns (5) and (6) report IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Even-numbered columns control for the change in the logarithm of urban population from 1947 to 1960.

Column 1 shows a positive and statistically significant correlation between railroads infrastructure and population, and a statistically non-significant correlation between roads

[27] See details in section 3
[28] See details in footnote 22.

infrastructure and population. When controlling for the previous change in log urban population (column 2), the point estimate for the coefficient associated to railroad infrastructure decreases, while the coefficient for roads remains small and statistically insignificant. Our IV strategies (columns 3 to 6) all show a positive and statistically significant impacts of railroads on population with point estimates ranging from 0.0025 to 0.0035. Additionally, we observe that controlling for previous changes in log urban population decreases point estimates regardless of which hypothetical networks we use to construct the instrument for roads. However, our results are robust to this control. In our preferred specification (column 6), the point estimate indicates that every 100 kms of dismantled roads led to a decrease of 0.29 log-points in population. This is equivalent to one standard deviation in railroads dismantling lowering log-population changes in 0.42 standard deviations.

When comparing IV estimates with OLS estimates for railroads, we observe that the corresponding coefficients are much higher for the former, suggesting negative selection bias. One possible explanation for this would be mean reversion bias. While we we fail to detect a correlation between railroads closures and previous trends in population and urban population growth [29], it is still possible that railroads closure were more prevalent in districts suffering from transitory negative shocks.

On the other hand, the corresponding estimates for roads are small, statistically insignificant and change signs depending on the hypothetical network instrument we use. Not surprisingly, all *LP-tests* fail to reject the null hypothesis suggesting the substitution of railroads for roads lead to a decrease of local population.

TABLE 9 Change in log urban population 1991-1960

	(1) OLS	(2) OLS	(3) IV EUC	(4) IV EUC	(5) IV LCP	(6) IV LCP
Change in kms of railroads 1986-1960	0.000187	-0.0000136	0.00175	0.00185**	0.00271**	0.00264***
	(0.000525)	(0.000509)	(0.00114)	(0.000882)	(0.00123)	(0.000965)
Change in kms of roads 1986-1954	-0.0000821	-0.000192	-0.000119	-0.000248	0.000240	0.000345
	(0.000155)	(0.000175)	(0.000261)	(0.000294)	(0.000305)	(0.000364)
Change in log urban population 1960-1947		0.0717		0.0520		0.0615
		(0.0573)		(0.0597)		(0.0616)
P-value for testing β_2 >= β_1			.0331	.0038	.0103	.0026
Cragg-Donald (multivariate) F-stat			32.8523	49.1762	27.0955	31.9865
Observations	286	234	286	234	286	234

Notes: The dependent variable is the change in the logarithm of urban population from 1960 to 1991. Columns (1) and (2) report OLS estimates. Columns (3) and (4) report IV estimates when the hypothetical network is constructed using euclidean distances. Columns (5) and (6) report IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Even-numbered columns control for the change in the logarithm of urban population from 1947 to 1960.

Table 9 shows the same results as Table 8 but using urban population as dependent variable. In this case, OLS estimates are small and statistically insignificant for both railroads and roads. Again, IV estimates for railroads are always larger than OLS estimates. The IV

[29] See tables 6 and 5. In both cases, the OLS estimates associated to railroads is positive but statistically insignificant.

point estimates for railroads range from 0.00175 to 0.00264, all being statistically significant at conventional levels except for when we don't control for previous development trends and use the euclidean hypothetical network (column 3). In our preferred specification (column 6), the point estimate indicates that every 100 kms of dismantled roads led to a decrease of 0.26 log-points in urban population. This is equivalent to one standard deviation in railroads dismantling lowering log-urban population changes in 0.39 standard deviations. It's worth noting that this magnitude is smaller but statistically indistinguishable from the coefficient for total population. Consequently, we fail to identify an impact on urban population as share of total population.[30] We also note that our results are robust to using changes only in paved roads instead of changes in paved plus gravel (as we do in all the tables presented).[31]

Our estimates for the impact of roads on urban population are always small, statistically insignificant and change sign between specifications. Again, we reject *LP-test* in every IV specification.

To illustrate why it is important to have both modes of infrastructure in one regression, we present Table 10. There we show for our preferred specification, results for models estimating the changes in transport infrastructure in population using changes from one mode at a time, and with an interaction between modes. Column (1) shows the result of estimating only using changes in railroad infrastructure. Column (2) uses only changes in road infrastructure. Column (3) is our baseline specification and coincides with column (6) in Table 8. Column (4) also adds an interaction between changes in trasnport modes and we instrument it with the interaction of our instruments.

We point out that if we were only estimating out of changes in road infrastructure we would conclude that roads have a negative effect on population. However, columns (2) and (3) show that actually roads seem to have a zero effect. We attribute negative effects in (2) to the fact that places where roads were built also were more likely to have trains dismantled. Ignoring that force would lead to biased results.

As for column (4), we point out that we have a problem of weak instruments as the instruments correlate very highly with their interaction (even if we use both the euclidean and least-cost-paths along with the LP instrument). We are currently in the quest for at least another instrument.

For the remaining tables in this subsection and subsection 5.3 we use variation in the dependent variable between 1970 and 1991 because we don't have migration or labor outcomes by sector for 1960. Correspondingly, we exploit changes in infrastructure between 1970 and 1986. In interest of completeness, we replicate Table 8 using variation for 1970-1991 [32]. The results are displayed in Table 11. Overall, we observe that our IV estimates are qualitatively the same as in Table 8. In the case of railroads, the associated coefficients to be slightly larger in magnitude. In our preferred specification (column 6), the point estimate indicates that every 100 kms of dismantled roads led to a decrease of 0.28 log-points in population. This is equivalent to one standard deviation in railroads dismantling lowering log-urban population changes in 0.47 standard deviations. It's worth noting that the both standardized and non-standardized estimates are practically the same as in Table 8. Again, *LP-test* in panel B is always rejected.

Table 12 closes this subsection showing results for migration. The dependent variable is the change in logs of the number of people that were born in the same province that

[30]Results available upon request
[31]Results available upon request.
[32]We cannot do the same for urban population because this variable is not available for 1970

TABLE 10 Attempt at estimating if railroads and roads are complements or subtitutes

	(1) Change in log population 1991-1960	(2) Change in log population 1991-1960	(3) Change in log population 1991-1960	(4) Change in log population 1991-1960
Change in kms of railroads 1986-1960	0.00272***		0.00285***	0.00341*
	(0.000688)		(0.000836)	(0.00192)
Change in kms of roads 1986-1954		-0.000515**	0.0000856	-0.000829
		(0.000254)	(0.000316)	(0.00364)
Change in kms of railroads 1986-1960 × Change in kms of roads 1986-1954				-0.00000386
				(0.0000145)
Change in log urban population 1960-1947	0.0612	0.0724	0.0625	0.0671
	(0.0529)	(0.0516)	(0.0535)	(0.0548)
Cragg-Donald (multivariate) F-stat	76.815	56.291	32.137	.4574
Observations	235	235	235	235

Notes: All regressions use 2SLS with our studied instrument for the changes in railroads and the least-cost-paths for the changes in roads. The district is the unit of observation, and they all include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. In columns (1) and (2), we use changes in gravel and paved roads. In columns (3) and (4), we only use paved roads.

TABLE 11 Change in log population 1991-1970

	(1) OLS	(2) OLS	(3) IV EUC	(4) IV EUC Network	(5) IV LCP Network	(6) IV LCP Network
Change in kms of railroads 1986-1970	0.000924**	0.000711*	0.00260***	0.00221***	0.00312***	0.00276***
	(0.000375)	(0.000382)	(0.000919)	(0.000824)	(0.00101)	(0.000935)
Change in kms of roads 1986-1970	-0.000219**	-0.000224*	-0.000281	-0.000320	-0.0000834	-0.0000400
	(0.000109)	(0.000122)	(0.000197)	(0.000210)	(0.000226)	(0.000261)
Change in log urban population 1960-1947		0.101***		0.0858**		0.0872**
		(0.0337)		(0.0355)		(0.0364)
P-value for testing β_2 >= β_1			.0004	.0006	.0002	.0004
Cragg-Donald (multivariate) F-stat			29.9293	30.5257	23.428	20.4473
Observations	311	235	311	235	311	235

Notes: The dependent variable is the change in the logarithm of population from 1970 to 1991. Columns (1) and (2) report OLS estimates. Columns (3) and (4) report IV estimates when the hypothetical network is constructed using euclidean distances. Columns (5) and (6) report IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Even-numbered columns control for the change in the logarithm of urban population from 1947 to 1960.

they live at the time of the census. Our unit of analysis is the district, and provinces are a administrative units of higher order, analogous to states in the USA [33]. Columns 1 and 2 show that correlations are statistically insignificant for both roads and railroads. Our IV estimates relying on the euclidean hypothetical network lead to positive but statistically insignificant coefficients for both roads and railroads (columns 3 and 4). When using the the least-cost path network, point estimates increase, becoming statistically significant for both types of infrastructure. In our preferred specification (column 6), the point estimate for railroads indicates that every 100 kms of dismantled roads led to a decrease of 0.29 log-points in the number of people living in the same province they were born. This is equivalent to a standardized coefficient of 0.23. On the other hand, the coefficient for roads indicates that every 100 kms of constructed roads lead to an increase of 0.10 log-points in the number of people living in the province they were born, equivalent to a standardized coefficient of 0.36. Again, *LP-test* in panel B is always rejected.

In the case of railroads, the coefficient is consistent with our estimates for population and urban population. It is worth noting that the point estimate for same-province-born population is around half the size of the coefficient for total population during the same period. This is consistent with heterogeneous effects on migration. It is possible that the migration decision is more elastic for people with ties in other regions of the country.

In the case of roads, the positive impact on same-province-born population occurs at the same time we don't observe any impact on total nor urban population. There are two non-exclusive explanations for these impacts to happen at the same time. On one hand, if we we abstract from changes in fertility and mortality rates, the results presented here would imply a shift in migration patterns, substituting inter province migration with intra-province migration. This is consistent with each transport mode having differential impacts on migration. Railroads would facilitate long-distance movements both through direct provision of long-distance transportation and through their impact on local economies. On the other hand, it's possible that roads serve better migration of people in closer distances through the densification of the local transport network.

Alternatively, it could also be the case that roads increased demand for human capital, pushing fertility rates downward and attracting migration. If this is the case, we would expect to find positive impacts of roads in better qualified jobs and schooling. We are currently working in devising a way to distinguish these hypothesis.

5.2 | Production

In this section we explore production quantities. In columns 1 and 2 of Table 13 we estimate the impacts of infrastructure changes on agricultural production as measured by the change of the logarithm of total cultivated area and total number of farms, both variables obtained from agricultural censuses of years 1960 and 1988. In column 3 to 6 we estimate impacts on the change of the logarithm of total value production, paid wages, number of employees and number of firms as measured in the industrial censuses of years 1954 and 1986[34]. The estimates associated to railroads are not statistically significant for agricultural area, number of farms and number of industrial firms. Nevertheless, they are positive and statistically significant for total production value, number of employees and number of firms. Their values indicate that every 100kms of dismantled railroads led to changes of 0.89 log points in production value, 0.72 log points in paid wages and 0.6 log points in number of employees. The corresponding standardized estimates are 0.38, 0.31 and 0.31. On the other hand, the

[33] In our working sample, we have 23 provinces. Each of them has different amount of districts, ranging from 2 to 92, the median value being 7, the average 13.5 and standard deviation of 18.0

[34] In industrial censuses the universe corresponds to all manufacturing firms located in Argentinian territory.

TABLE 12 Change in log number of people that live in the province they were born 1991-1970

	(1) OLS	(2) OLS	(3) IV EUC	(4) IV EUC Network	(5) IV LCP Network	(6) IV LCP Network
Change in kms of railroads 1986-1970	-0.000403	-0.000621	0.00245	0.00167	0.00367**	0.00290*
	(0.000661)	(0.000656)	(0.00161)	(0.00140)	(0.00182)	(0.00163)
Change in kms of roads 1986-1970	-0.0000203	0.000292	0.000266	0.000416	0.000732*	0.00103**
	(0.000193)	(0.000209)	(0.000345)	(0.000356)	(0.000405)	(0.000453)
Change in log urban population 1960-1947		0.314***		0.297***		0.300***
		(0.0578)		(0.0603)		(0.0634)
P-value for testing β_2 >= β_1			.0728	.1697	.0365	.0987
Cragg-Donald (multivariate) F-stat			29.9293	30.5257	23.428	20.4473
Observations	311	235	311	235	311	235

Notes: The dependent variable is the change in the number of people that live in the province they were born from 1970 to 1991. Columns (1) and (2) report OLS estimates. Columns (3) and (4) report IV estimates when the hypothetical network is constructed using euclidean distances. Columns (5) and (6) report IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, and distance to Buenos Aires. Even-numbered columns control for the change in the logarithm of urban population from 1947 to 1960.

estimated impacts for roads are all small and statistically insignificant. Finally, our *LP test* is rejected in all variables associated to the manufacturing sector.

TABLE 13 Changes in agricultural and industrial production

	(1) Change in log agricultural area (ha) 1988-1960	(2) Change in log number of farms 1988-1960	(3) Change in log total value of production 1985-1954	(4) Change in log total paid wages 1985-1954	(5) Change in log number of employees 1985-1954	(6) Change in log number of firms 1985-1954
Change in kms of railroads 1986-1970	0.000249	-0.00130	0.00891**	0.00724*	0.00632*	0.00269
	(0.00236)	(0.00134)	(0.00429)	(0.00413)	(0.00362)	(0.00165)
Change in kms of roads 1986-1970	0.000378	-0.000382	0.000357	0.000620	0.000668	0.000350
	(0.000655)	(0.000372)	(0.00119)	(0.00115)	(0.00101)	(0.000459)
Change in log urban population 1960-1947	-0.165*	0.187***	-0.0827	0.243	0.300**	0.423***
	(0.0936)	(0.0532)	(0.167)	(0.161)	(0.141)	(0.0642)
P-value for testing β_2 >= β_1	.5245	.7786	.0124	.0357	.0397	.0554
Cragg-Donald (multivariate) F-stat	19.6118	19.515	20.4473	20.4473	20.4473	20.4473
Observations	226	225	235	235	235	235

Notes: The dependent variable changes from column to column. From column 1 to 6, they correspond respectively to the change in the logarithm of total cultivated area, the change in the logarithm of total number farms, the change of the logarithm of total value production, the change of the logarithm of paid wages, the change of the logarithm of number of employees and number. All regressions correspond to IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, distance to Buenos Aires and the change in the logarithm of urban population from 1947 to 1960.

The absence of impacts on the outcomes associated to the agricultural sector (columns 1 and 2) may be explained by the rigidity of land as a production factor. It is possible that adjustments in the agricultural sector are made through quantities and prices of flexible factors (e.g. labor), total production, and the production mix. We expect to digitize new data in the near future to assess this possibilities. On the other hand, the negative impacts of railroads dismantling on production value, labor and wages in manufacturing are consistent with the evidence of negative impacts on population, urban population and the number of people born in the same province they live. These can be seen as the economic mirror of population decreases. It's worth noting that the corresponding coefficient have similar magnitudes.

5.3 | Structural change

This subsection explores potentially different impacts of infrastructure changes in the productive structures of local economies, this is, in the distribution of economic activity across industries. We approach the question of structural change by studying the impact of infrastructure changes on the distribution of labor across economic sectors. For this, we rely on IPUMS homogenization which aims to match the main industry groups defined in ISIC revision 4 (UNSD, 2008). IPUMS define fifteen groups which in turn we aggregate to six: agriculture and mining, manufacturing, non-tradable industries (electricity, water, gas and construction), education, health and public administration, business services (transportation, communications, wholesale, retail, hotels, restaurants and real state), and other services. In table 14 we show the distribution of labor across this sectors for years 1970 and 1991. In columns 1 and 2, we show the national aggregates (excluding Buenos Aires city). Columns 3 and 4 result from calculating labor shares at the district level and then taking simple averages. The comparison of columns 1 and 2 show that in the aggregate, the share of workers in agricultural and industrial sectors fell between 1970 and 1991, while the share of workers in services increased. We observe that the biggest changes happened in agriculture and education, health and public administration.

In columns 3 and 4 we observe similar trends. The participation of primary and industrial sectors decreases over time and the participation of services raises. Additionally, we observe that labor shares in agriculture and mining are considerably higher than those reported in the first two columns. The opposite happens with the labor share in manufacturing. These figures are consistent with manufacturing being concentrated in big agglomerations and agricultural activity being highly dispersed over the territory, in low populated areas. This suggests that our main specification may fail to detect structural change if this is heavily concentrated in a small number of big districts. Hence, we proceed in two different ways. First, we use the same specification as for population, urban population and migration. Second, we replicate this exercise weighting each observation with its population in 1960.

The first set of results (unweighted) are reported in Table 15. None of the estimates is statistically significant and we fail to reject the LP-test. The second set of results (weighted by urban population in 1960) are reported in Table 16. For railroads, we observe that dismantling roads shifted the composition of labor from agriculture and mining towards manufacturing and non-tradable industries (electricity, gas, water and construction). More precisely, every 100 kms of dismantled roads caused a fall of 4.1% in the labor share of agriculture and mining, an increase of 4.1% in manufacturing and an increase of 1.8% in non-tradable industry. The corresponding standardized coefficients are 0.21, -0.27 and -0.12 respectively. In the case of roads, the only statistically significant coefficients corresponds to the impact on other services, indicating that every 100 kms of new roads lead to a 1.2%

TABLE 14 Evolution of productive structure

	National		Average district	
	(1)	(2)	(3)	(4)
	1970	1991	1970	1991
Agriculture and mining	.19	.09	.37	.24
Manufacturing	.21	.18	.13	.12
Non-tradable industries	.1	.08	.09	.07
Education health and public	.13	.24	.12	.24
Business services	.25	.28	.2	.21
Other services	.11	.13	.09	.12

Notes: This table shows the estimated share of total labor in each of six industrial aggregates defined by aggregating some IPUMS classifications (IPUMS aggregated categories in parenthesis when corresponds): agriculture and mining, manufacturing, non-tradable industries (electricity, water, gas and construction), education, health and public administration, business services (transportation, communications, wholesale, retail, hotels, restaurants and real state), and other services. Columns (1) and (2) correspond to national aggregates (excluding Buenos Aires city). These were computed by taking population-weighted averages of labor shares at district-level. Columns (3) and (4) describe the average district (excluding Buenos Aires). These were computed by taking unweighted averages of labor shares at district-level.

decrease in the share of labor in other services. The corresponding standardized coefficient is -0.5. The rest of the coefficients are not statistically significant, making it hard to conclude which sectors increased their participation. Nevertheless, the magnitude and sign of the point estimates suggest those sectors might be manufacturing and business services. In fact, the corresponding point estimates are considerably higher in magnitude than those in other sectors and sum up to the point estimate for other services.

Overall, our results indicate that dismantling railroads had a negative impact on population, urban population, migration and industrial activity. Together with this, our weighted regressions suggest that these movements changed the distribution of labor in favor of industrial sectors and at the expense of labor in the primary sector (agriculture and mining). These results are consistent with railroads serving better products of low value-to-weight ratio, such as wheat or other agricultural products. Given that most agricultural production in Argentina goes to cities and/or external markets, the dismantling of railroads may have played the role of an increase in trade barriers. More precisely, dismantling railroads possibly led to a contraction of agricultural exports, reflected in a negative impact on labor share in the primary sector (Table 16). In face of the crisis of the local economies, people migrated more. This may explain our results in population, urban population and migration (tables 8, 9, 11 and 12). Finally, it's important to note that the positive impact of dismantling railroads on the share of industrial labor suggest these sectors were able to better endure the shock as a consequence of being relatively less dependent on external demand and/or benefiting from tacit protection against external competition on tradables.

The LP suggested to replace closed railroads with roads and highways. Hence, one may argue that it is hard to evaluate the plan itself given that many railroads were not fully substituted by roads. Nevertheless, we fail to detect impacts of roads construction in population, urban population and industrial activity. These results, together with the rejection of our *LP-test* in the corresponding cases, suggest that even in places where rails were fully substituted for roads, this substitution was non-neutral in terms of the size of the local economy. If we accept the hypothesis of local crises being originated by a negative shock to agricultural production, our results suggest that road construction failed

TABLE 15 Change in share of labor by industry 1991-1970

	(1) agriculture and mining	(2) manufacturing	(3) non-tradeable industry	(4) education, health, and public services	(5) business services	(6) other services
Change in kms of railroads 1986-1970	0.0000641	-0.000140	-0.000123	0.000175	0.0000677	-0.0000429
	(0.000304)	(0.000223)	(0.000145)	(0.000310)	(0.000227)	(0.000176)
Change in kms of roads 1986-1970	-0.0000242	0.0000405	-0.00000132	0.0000328	-0.0000105	-0.0000373
	(0.0000847)	(0.0000621)	(0.0000404)	(0.0000864)	(0.0000632)	(0.0000490)
Change in log urban population 1960-1947	0.00781	-0.0135	-0.00363	-0.0287**	0.0210**	0.0170**
	(0.0119)	(0.00868)	(0.00565)	(0.0121)	(0.00884)	(0.00685)
P-value for testing β_2 >= β_1	.372	.8193	.8277	.3036	.3491	.5141
Cragg-Donald (multivariate) F-stat	20.4473	20.4473	20.4473	20.4473	20.4473	20.4473
Observations	235	235	235	235	235	235

Notes: The dependent variable in each column corresponds to the change in the share of labor in each of six industrial groups. The industrial groups result from by aggregating some IPUMS classifications (IPUMS aggregated categories in parenthesis when corresponds): agriculture and mining, manufacturing, non-tradable industries (electricity, water, gas and construction), education, health and public administration, business services (transportation, communications, wholesale, retail, hotels, restaurants and real state), and other services. All regressions correspond to IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, distance to Buenos Aires and the change in the logarithm of urban population from 1947 to 1960. Observations are not weighted.

TABLE 16 Change in share of labor by industry 1991-1970 - weighted

	(1) agriculture and mining	(2) manufacturing	(3) non-tradeable industry	(4) education, health, and public services	(5) business services	(6) other services
Change in kms of railroads 1986-1970	0.000414*	-0.000410**	-0.000183*	0.000108	0.000254	-0.000182
	(0.000242)	(0.000197)	(0.000101)	(0.000248)	(0.000185)	(0.000135)
Change in kms of roads 1986-1970	-0.00000217	0.0000553	0.00000653	-0.000000907	0.0000608	-0.000120**
	(0.0000932)	(0.0000757)	(0.0000390)	(0.0000956)	(0.0000712)	(0.0000519)
Change in log urban population 1960-1947	0.00639	-0.00699	-0.0103***	-0.0316***	0.0261***	0.0164***
	(0.00925)	(0.00751)	(0.00387)	(0.00948)	(0.00706)	(0.00515)
P-value for testing β_2 >= β_1	.0339	.9941	.9767	.3206	.1337	.6880
Cragg-Donald (multivariate) F-stat	25.8505	25.8505	25.8505	25.8505	25.8505	25.8505
Observations	235	235	235	235	235	235

Notes: The dependent variable in each column corresponds to the change in the share of labor in each of six industrial groups. The industrial groups result from by aggregating some IPUMS classifications (IPUMS aggregated categories in parenthesis when corresponds): agriculture and mining, manufacturing, non-tradable industries (electricity, water, gas and construction), education, health and public administration, business services (transportation, communications, wholesale, retail, hotels, restaurants and real state), and other services. All regressions correspond to IV estimates when the hypothetical network is constructed using least-cost paths. All regressions have the district as the unit of observation, they include province fixed effects, and control for mean elevation, mean ruggedness, area in squared kilometers, average suitability for wheat, distance to Buenos Aires and the change in the logarithm of urban population from 1947 to 1960. Observations are weighted by urban population in 1960.

to compensate for the negative effects of railroad dismantling even in places were railroads were fully substituted by roads. Nevertheless, we observe effects that suggest that while not changing the size of the local economy, roads construction was non-neutral in other margins. On one hand, we observe an increase in the amount of inhabitants being born in the same province, which together with the absence of impacts on population and urban population, suggest that migration flows shifted away from inter province to intra-province flows. This is consistent with connectivity gains being different across transportation modes: on one hand, railroads tend to connect a few distant points, while roads densify local transport networks[35]

Finally, it is important to note that we find that roads shifted labor shares away from "other services", together with weak evidence of manufacturing and business services gaining importance. This is consistent with roads serving better products with high value-to-weight ratio. While this effect was not sufficient to prevent a decay, it might have long-run effects we have not addressed so far [36].

6 | CONCLUSION

Investing in physical transport infrastructure is regarded and recommended as an important policy to foster economic growth. As recent empirical work has shown, such investments have sizable positive effects on trade and economic development across a wide range of historical contexts.[37] Nevertheless, our understanding of this relationship remains simple and limited, specially in regard to the relative role of different transport modes in fostering economic transformations conducive to long-run development.

In this paper, we study the case of Argentina between 1960 and 1991 when following a World Bank study of the national transport infrastructure, the total railroad network fell from around 43,500 to around 33,500 kms at the same time that the paved road network grew from around 10,000 to 28,000 kms. We combine spatially-disaggregated economic, demographic and geographic data to study the effects of those changes on population, agricultural production, and industrial production in Argentinian districts using plausibly exogenous variation given by (i) a discontinuity in the LP design, (ii) and hypothetical road networks given by a least-cost paths connecting province capitals to Buenos Aires.

First, we show that conventional IV estimates can be misleading when omitting potential substitution of different transport modes. When allowing for this substitution, our instrumental variables approach shows that for every 100 kms of railroads *dismantled* in a district, total population fell in 0.29 log-points, urban population fell in 0.26 log-points, and the population born in the same province they live fell in 0.29 log-points, indicating outflow migration. Consistently, we find negative impacts on the volume of industrial activity as measured by total value of production, total paid wages and employment. When looking at structural change, we find that every 100 kms of dismantled railroad led to a reduction of 4.1% in the share of labor in the primary sector (agriculture and minig), an increase of 4.1% on the share of labor in manufacturing and an increase of 1.8% on the share of labor in

[35] In turn, this conclusion suggest cross-district externalities in the case of roads. While this is not taken into account in this paper, we expect to internalize this externalities in the near future by incorporating measures of market access as firstly introduced by Donaldson and Hornbeck (2016)

[36] For example, by making the primary sector less profitable, the infrastructure-led crises could have forced local economies to engage in activities of higher social value in the long-run. Arguably these activities are more prevalent in the secondary sector, as suggested by Galor and Mountford (2008)

[37] See for example Michaels (2008); Donaldson (2018); Chandra and Thompson (2000); Fajgelbaum and Redding (2018).

non-tradable industries (construction, electricity, water, and gas). On the other hand, we don't find any impact of roads infrastructure on total population, urban population and total economic activity. Nevertheless, we find that every 100 km of railroads construction lead to an increase of 0.10 log points in population born in in the same province, a 1.2% decrease in the share of labor in other services, and weak evidence of this share being gained by manufacturing and business services.

These results are consistent with the idea that railroads serve better products with low value-to-weight ratio, such as wheat, and more generally agricultural products. In our interpretation of this results, dismantling railroads led to a contraction in agricultural exports, igniting a crisis in the local economy, expressed both in decreasing production and out-flow migration. In comparison to the agricultural sector, industrial activities better endured the crisis due to their relative independence from external demand and protection against external competition induced by railroads dismantling, as reflected in the shift of labor shares from agriculture to industry. On the other hand, our results suggest that the substitution for roads failed to counter the decline of the local economy. Nevertheless there is evidence of roads construction being non-neutral by changing migration patterns (from inter- to intra-province immigration) and weak evidence of structural change. The latter suggest that future research may find impacts of roads on the long-run and/or other outcomes.

This paper contributes in several ways to the literature on developmental impacts of public infrastructure. First, we provide the first empirical case that allows to estimate reduced form effects of both road and railroad intranational infrastructure on local economic development and specialization, in a common historical context and space. In this framework we show that conventional reduced form strategies may be misleading when omitting potential substitution across transportation modes. Second, while most of the literature focuses on the distribution of people and overall economic activity, this article puts special attention on structural transformation of local economies. Third, this is the first paper in estimating the impact of dismantling a railroad network in a development country, only preceded by Gibbons et al. (2018), who study similar closures in the UK.

In terms of policy implications, this paper emphasize the importance of considering the relative impacts of different transport modes when deciding transport infrastructure investments. More precisely, it highlights the differential effects of transport modes on structural transformation, migration flows and local development. We believe future research should integrate this insight and our empirical results with recent progress in the theory of optimal transport investment (Fajgelbaum and Schaal, 2020). In our view, research in this direction would not only enrich economic literature but also offer powerful tools for policy design.

REFERENCES

Alder, S. (2016) Chinese roads in india: The effect of transport infrastructure on economic development. *Available at SSRN 2856050*.

Argentina, R. (1981) Política ferroviaria 1979-1981. *Tech. rep.*, Ministerio de Economía, Secretaría de Estado de Transporte y Obras Públicas.

Banerjee, A., Duflo, E. and Qian, N. (2012) On the road: Access to transportation infrastructure and economic growth in china. *Tech. rep.*, National Bureau of Economic Research.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A. and Zhang, Q. (2017) Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, **99**, 435–448.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q. and Brandt, L. (2018) Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 103124.

Bogliaccini, J. A. (2013) Trade liberalization, deindustrialization, and inequality: Evidence from middle-income latin american countries. *Latin American Research Review*, 79–105.

Bolt, J. and Van Zanden, J. L. (2014) The m addison p roject: collaborative research on historical national accounts. *The Economic history review*, **67**, 627–651.

Instituto Nacional de Estadisticas y Censos, A. (1960) Censo nacional de población, vivienda y agropecuario 1960. 1–103.

Chandra, A. and Thompson, E. (2000) Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, **30**, 457–490.

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

Donaldson, D. (2018) Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, **108**, 899–934.

Donaldson, D. and Hornbeck, R. (2016) Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, **131**, 799–858.

Duranton, G., Morrow, P. M. and Turner, M. A. (2014) Roads and trade: Evidence from the us. *Review of Economic Studies*, **81**, 681–724.

Faber, B. (2014) Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, **81**, 1046–1070.

Fajgelbaum, P. and Redding, S. (2018) Trade, structural transformation and development: Evidence from argentina 1869-1914. *revised version of NBER Working Paper*, **20217**.

Fajgelbaum, P. and Schaal, E. (2020) Optimal transport networks in spatial equilibrium. *Econometrica*, **88**, 1411–1452.

FAO/IIASA (2011) Global agro-ecological zones (gaez v3.0).

Ferreres, O. J. (2005) *Dos siglos de economía argentina, 1810-2004: historia argentina en cifras*. Editorial El Ateneo.

Forero, A., González, F., Gallego, F. and Tapia, M. (2019) Railroads, specialization, and population growth in small open economies: evidence from the first globalization. *Working Paper*.

Galor, O. and Mountford, A. (2008) Trading population for productivity: Theory and evidence. *The Review of Economic Studies*, **75**, 1143–1179.

Galor, O. and Özak, Ö. (2016) The agricultural origins of time preference. *American Economic Review*, **106**, 3064–3103.

Gibbons, S., Heblich, S. and Pinchbeck, T. (2018) The spatial impacts of a massive rail disinvestment program: the beeching axe.

Jedwab, R. and Moradi, A. (2016) The permanent effects of transportation revolutions in poor countries: evidence from africa. *Review of economics and statistics*, **98**, 268–284.

Keeling, D. J. (1993) Transport and regional development in argentina: Structural deficiencies and patterns of network evolution. In *Yearbook. Conference of Latin Americanist Geographers*, 25–34. JSTOR.

Kruskal, J. B. (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, **7**, 48–50.

Larkin, T. (1962) Transportes argentinos plan de largo alcance. *Tech. rep.*, Ministerio de Obras y Servicios Públicos, Grupo de planeamiento de los transportes.

López, M. J. and Waddell, J. E. (2007) *Nueva historia del ferrocarril en la Argentina: 150 años de política ferroviaria*. Lumiere.

Michaels, G. (2008) The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, **90**, 683–701.

Nunn, N. and Puga, D. (2012) Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, **94**, 20–36.

Pahowka, G. (2005) A railroad debacle and failed economic policies: Peron's argentina. *The Gettysburg Historical Journal*, **4**, 6.

Perez, S. (2017) Railroads and the rural to urban transition: Evidence from 19th-century argentina. *Job Market Paper*, 1–103.

Potash, R. A. (1969) *The Army & Politics in Argentina: 1945-1962; PerÃ³n to Frondizi*, vol. 2. Stanford University Press.

Redding, S. and Turner, M. A. (2015) Transportation costs and the spatial organization of economic activity. In *Handbook of regional and urban economics*, vol. 5, 1339–1398. Elsevier.

Rock, D. et al. (1985) *Argentina, 1516-1982: from Spanish colonization to the Falklands War*. Univ of California Press.

UNSD (2008) *International Standard Industrial Classification of all Economic Activities (ISIC) (No. 4)*. United Nations Publications.

USGC (1996) Gtopo30. *Mineral Yearbook*.

Whitaker, A. P. (2013) *The United States and Argentina*. Harvard University Press.

World Bank (2020) Transport overview. URL: https://www.worldbank.org/en/topic/transport/overview.