CAF - WORKING PAPER #2022/17

The Long and Winding Roads: Roads, Inequality, and Growth in Colombia

Luis Quintero[1] | **Guillermo Sinisterra**[2]

[1]Assistant Professor, Carey Business School, Johns Hopkins University.
leq@jhu.edu

[2]Assistant Professor, Departamento de Ciencias Contables, Universidad Javeriana .
gsinist@javeriana.edu.co

We measure the impact on economic performance and land ownership inequality in municipalities of a large scale program of road network expansions and improvements that occurred in Colombia between 1993 to 2012. The treatment is measured as improvements in market access to incorporate network effects. We find that roads improve market access, and this increases both municipality GDP and land ownership inequality. We address endogeneity concerns by using two instruments. First, using detailed geographical information we create a least cost path counterfactual for the Colombian road networks linking important cities in 1938. We use this least cost path to calculate a counterfactual market access measure that is determined by exogenous topographic accidents. Next, we build an alternative market access measure which focuses on semi-random market access changes stemming from increased exposure to markets of smaller cities which were not determinant in defining the road network shape.

KEYWORDS
Development, Inequality, Instrumental variables, Least cost Path, Market Access, Regional economics, Roads

Authors thank David Bernal, Laura Carolina Correal and Julio Cesar Vega for excellent research assistance. We also thank Victoria Gomez, Rafael Cubillos for providing useful information. We especially appreciate funding from CAF's research program *Transport Infrastructure for the Development of Latin America*. We thank Matthew Turner, Diego Puga, Elizabeth Villadecans, Yves Zenou, Mark Parkridge, Edward Glaeser, Pablo Sanguinetti, Carlos Juncosa, Lian Allub, Jorge Pérez, Luis Baldomero, and all attendants to the workshop in urban and regional economics at Universidad Javeriana, the Eureka seminar from VU university, the ITEA annual conference, the Cities and Development seminar, and CAF's Infrastructure for Development Workshop for useful comments. We are grateful for CAF funding as part of their Transport Infrastructure for the Development of Latin America grant program.

CAF - DOCUMENTO DE TRABAJO #2022/17

Esta versión: November 2, 2022.

Caminos largos y sinuosos: carreteras, desigualdad y crecimiento en Colombia

Luis Quintero[1] | **Guillermo Sinisterra**[2]

[1]Assistant Professor, Carey Business School, Johns Hopkins University.
leq@jhu.edu

[2]Assistant Professor, Departamento de Ciencias Contables, Universidad Javeriana .
gsinist@javeriana.edu.co

Este artículo mide el impacto en el desempeño económico y desigualdad en la propiedad de la tierra de un programa a gran escala de ampliación y mejora de la red de carreteras que sucedió en Colombia entre 1993 y 2012. Incorporamos efectos de red midiendo el tratamiento como cambios en el acceso de mercado que resultan de las carreteras. Encontramos que las carreteras mejoran el acceso al mercado, y esto aumenta tanto el PIB del municipio como la desigualdad en la propiedad de la tierra. Dos instrumentos abordan los problemas de endogeneidad. El primer instrumento es un acceso de mercado contrafactual determinado por los caminos de menor costo entre las ciudades más importantes en 1938. Se utiliza este camino de menor costo para calcular una medida contrafactual de acceso al mercado que está determinada por accidentes topográficos exógenos. El segundo se centra en cambios semi-aleatorios de acceso al mercado derivados de la mayor exposición a los mercados de las ciudades más pequeñas que no fueron determinantes en la definición de la forma de la red de carreteras.

KEYWORDS

Desarrollo, Desigualdad, Variables instrumentales, Caminos de mínimo costo, Acceso al mercado, Economías regionales, Redes Víales

1 | INTRODUCTION

Transport infrastructure is widely considered as crucial for economic development. Effects of transportation infrastructure on different economic variables have been largely studied, especially for developed countries. In 1993, with a new legal framework that welcomed private investment in the infrastructure sector as builders and operators of roads, Colombia accelerated the growth and modernization of a road network that had largely been built starting in 1950. This resulted in a more than half of the existing network in 1993 either being built or transformed through significant improvements, since 1993. In this work, we estimate the impact of the massive improvement in market access this had for Colombian cities and towns on their economic development as captured by overall GDP and GDP for different sectors, as well as income and wealth inequality measures.

Estimation of the impact of such infrastructure development holds multiple methodological challenges. The role of roads and other transportation infrastructure is relevant insofar as it reduces transportation costs by effectively making it easier to reach other markets. The true effects on economic variables will be the result of multiple interactions with all nodes of the road network. This general equilibrium effects are not well captured by indicator variables of whether the new roads pass through focal locations or by the distance separating them. Instead, we calculate complete road network topologies and use infrastructure changes over time to calculate a panel of travelling times between all cities and towns in Colombia. With that, we capture the general equilibrium effects of the infrastructure by constructing time varying market access measures.

Another important issue is endogeneity. Transportation infrastructure is costly and its construction if often the result of some optimization analysis, where the connectivity of important places is targeted. This is evident in Colombia where, until 1960, roads were built to link demand centers concentrated in the main cities. As a rule, these were department (1st sub-national division) capitals. If the past trends or the expectation of cities strong future behavior in terms of our outcome variables, economic growth or inequality measures, played a role in the planning of the infrastructure, its appearance could be spuriously interpreted as having a causal impact. We implement multiple methods to deal with this endogeneity and compare their performance. First, we directly test that prior economic growth or inequality changes did not determine the probability of road construction or transformation. Then, we focus the analysis on subsamples of incidental cities, those that have been plausibly affected exogenously by the infrastructure due to their location. Following this idea, we also use the quasi exogenous component of the market access, the market access that the infrastructure affects through changing access to incidental cities. This instrument allows us to include all cities in the analysis.

Furthermore, to deal with the possibility that incidental cities were included in the policy maker's objective function when planning the treatments, we further exploit exogenous variation in the construction costs implied by local topography to build a least cost path road network. We define important cities and towns in 1938 as the network nodes and then use detailed engineering cost estimates to determine the path between the nodes. With this counterfactual network, we calculate travel times assuming road quality and improvements happened as and when they did in the actual network. This least cost path market access instrument shifts the location of improvements to areas that minimize construction costs while keeping them joining the same cities or towns.

Through improvements in market access, results show that roads have an important positive causal impact on the total GDP, as well as on its industrial, agricultural and services subcategories. The associated growth of services and industry sectors are stronger than the growth of agricultural GDP, suggesting a redistribution of resources across sectors. The infrastructure also increases the municipal development index, a measure that extends economic activity measures with institutional, quality of life and poverty reduction dimensions. Also, as the economy grows, income inequality is reduced by the increase in market access. However, wealth, as captured by owned land, is further concentrated. The roads increase land property in fewer larger lots and fewer wealthier owners. One possible explanation is the development of larger scale agro-industrial production, which requires more capital investment and is more efficient at larger scales. Our results highlight the importance of infrastructure as drivers of economic growth. At the same times warns about possible effects that may increase wealth concentration.

2 | RELATED LITERATURE

Literature on the impact of transportation infrastructure on development is abundant for developed countries, but much less so for developing countries. Berg et al. (2017) provides a detailed summary of papers that analyze the impact of large scale transportation infrastructure on developing countries economic outcomes. Most studies focus on the effects of transport infrastructure on trade. Duranton (2015) and Duranton and Turner (2012) show that roads increase trade for Colombia and the US respectively. There is an extensive literature showing positive impacts of roads on economic development outputs, including Jedwab and Storeygard (2020) which find the effect of roads on population density in African cities using the market access as independent variable. Donaldson (2018) searches for the effect of railroads on trade cost, prices and income in India and Donaldson and Hornbeck (2016), for the effect of railroads on the agricultural sector in the United States. Banerjee et al. (2012) Faber (2014) and Baum-Snow et al. (2017) look for the effect of roads on GDP and other development variables in China taking into account different time horizons and interventions. Also Duranton (2015) finds that the lack of roads is a major impediment to trade using Colombian data. Within this literature there are some recent developments that allow seeing heterogeneous effects on different regions as in Chandra and Thompson (2000) and Herzog (2020) for roads the effect of roads in the USA.

Our contribution to this branch of literature expands in several ways. First we provide evidence about the effects of roads on land and income inequality which has not been studied for developing countries. Second in our research design we include new instruments that combine historical, statistical and geographical features as source of exogenous variation for the market access. This instruments are novel, we compare results among them to obtain more robust outcomes. On top of that, we compare the results of the full sample of municipalities with the results obtained taking out big cities. Third we measure neighborhood effects following a similar specification that the one used in List et al. (2019) to account for spillover effects of infrastructure on nearby municipalities which complements current literature. There is also a branch of the literature that measures welfare effects that imply general equilibrium settings, includes papers as Allen and Arkolakis (2019) for the total road network in the United States, Tsivanidis (2018) which measures effects for introducing a bus rapid transit system in Bogota, Colombia, and Donaldson (2018) the railroads development in India.

There is a branch of the literature that analyses tertiary roads, lead by a recent paper by Asher and Novosad (2020) that measures the effect of an annual investment of US 40 billion in rural roads in India, they do not find important effects on GDP for small municipalities but do find that people were moving out of the agricultural sector to favor the industrial sector. The small or null growth is explained by the inability of small towns to take full advantage of transport cost reductions given the lack of other inputs such as human capital, technology or agglomeration economies. This analysis however focuses on tertiary roads, which we do not take into account. For these reasons, we believe that effects of tertiary roads on treated municipalities are important but smaller than the ones of main roads.[1] Despite the fact we do not develop a model to measure welfare effects, our calculation of the market access using travel speeds, and the validation of the results with actual GPS data, allows to account for several of the problems to measure the effects of roads. One is congestion, which will be included in the GPS data, the other one is relocation of economic activity which will be evident with the measure of spillover effects and the results on sector GDP. We also expand the current analysis by calculating heterogeneous effects in different regions and places within the country.

Despite the dominance of roads as a transport mode for freight and the big investments done, to our knowledge there are few studies about the effect of roads on economic development for Colombia. One of them is presented by Baldomero-Quintana (2020) who argues that infrastructure shapes comparative advantage trough their close relationship with the domestic trade costs in the Colombian case. Such studies can be used, for example, as a tool to prioritize new roads or the maintenance of the existing ones.

On the theoretical side, Getachew and Turnovsky (2015) highlights that roads devoted to reducing the needs of the poorest agents tend to reduce inequality. However, roads favoring the wealthier owners of capital tend to increase inequality. According to their model, in the end, distributional consequences will depend on the substitutability between private and public capital in production. Our empirical results show the effect of national roads on inequality and can support one of the channels proposed here. Also, Ferreira (1995), introduces credit constraints to show that long-term inequality might remain despite infrastructure investment. In this line of thinking, poor people do not have access to credit, therefore, they cannot have the required start-up capital necessary to invest in new projects. The middle class has partial access to credit and can start a new investment. However, only the richer people can take full advantage of the infrastructure and the resources necessary to develop wealth.

Some of the empirical work in this area try to measure macroeconomic effects of infrastructure on inequality as Barro (1990) and Calderón and Chong (2004) do.[2]

3 | A LARGE SCALE TRANSPORTATION INFRASTRUCTURE PROJECT IN A DEVELOPING COUNTRY

In 1993, new legal frameworks were established in Colombia to allow for public private partnerships where the private sector was allowed to participate in the construction and operation of road infrastructure through road concessions. Three generations of infrastructure contracts were allocated through public bidding processes in 1994, 1997, and 2001 respectively. The first generation of road concession started operations in 1996. Most of the roads intervened and built were designed to connect the Bogotá, the country's capital, to other markets. The second generation mainly improved the access of Cali, one of the largest cities located in the sugar cane cluster in the southwest. The third generation focused on connecting important regional production centers with large internal markets. All had a strong focus on the main cities of the country. Contracts in each new investment generation focused on shifting risk from the government to the private sector. This reduced the public

[1] Other papers focusing on the effects of tertiary roads include: Khandker et al. (2009) and Khandker and Koolwal (2011) for Bangladesh, Dercon (2004) for Ethiopia Gibson and Olivia (2010) for Indonesia and Mu and Van de Walle (2011) for Vietnam

[2] Calderón and Servén (2014) contains a detailed review of this branch of the literature, including both empirical and theoretical developments

costs of investing in these projects, which accelerated significantly the number of miles built and improved. The construction and expansion of roads has been approached as the main way of reducing transportation costs in the country.[3] In addition to these three waves of concessions, we evaluate wave 0, the roads developed by the National Institute of Roads (INVIAS). Figure 1 shows the location of the national road network and the new and improved roads we analyze. Red segments are concession roads; green segments are roads improved by INVIAS. The clear blue lines are the rest of the road network.

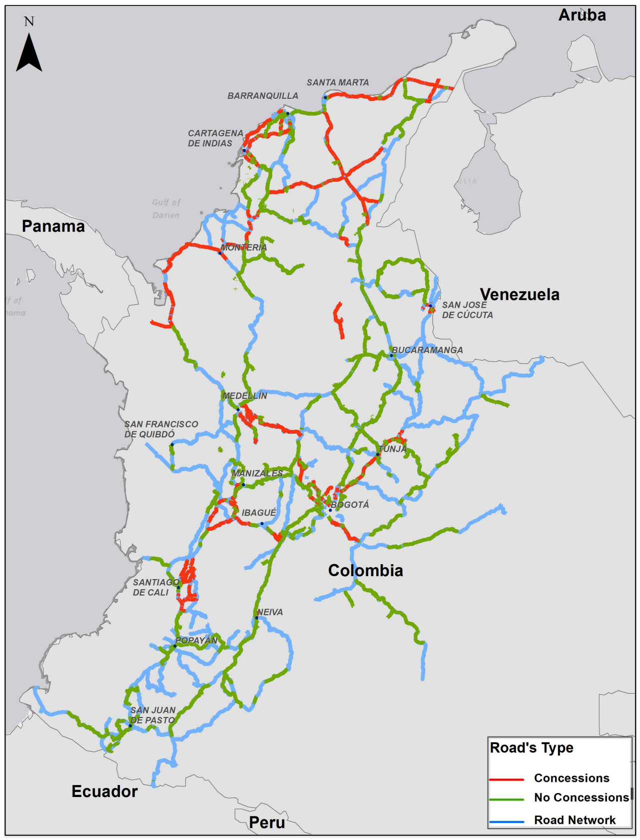


FIGURE 1 National road network vs roads concessions waves. *Source:* Invías and author's calculations

We evaluate the impact of this ambitious plan of road improvements and construction on the universe of municipalities in Colombia between 1993 to 2012 (with different analysis periods available for different outcome variables). A total of 6,963 km of road were built or transformed, which amounts to 36% of the current road network (56% of the network in 1993 when concessions began). Among these, 4,463 km of roads were allocated by concession, and 2500km of roads built by the public sector directly.[4] This represents a significant growth in the transportation infrastructure network compared to recent history. At the beginning of the last century, Colombia had not developed a road system. Even by 1994, its development was lagging. For instance, while the US had 24% of its roads paved, Colombia had only

[3]Although in the 40s', almost 33% of the cargo was moved by rail, a transport mode with economies of scale and cheaper operational costs, this transport mode was abandoned in favor of trucks. There have been several attempts to develop additional transportation modes for cargo like rail or river, but they never get the political capital funded and developed.

[4]A few of the public sector roads finished during our study period began as early as 1986.

0.24% (World Bank, 1994). The main network was built in the first half of the century, in 50 years only 2,500km of paved roads were built, mainly designed to connect department capitals (Pachón and Ramírez, 2006). In comparison to historical road development, the growth since 1994 was much faster and massive (see appendix F).

To have a better understanding of the distribution of interventions across time, Figure 2 shows the number of municipalities affected by a road between 1990 and 2012. Peaks in construction are seen in 1995-98 in 2004 and 2008. Each peak corresponds to a different generation of concessions. Dates are determined in this plot by date of finalization of the road, while construction beginning dates are used for the regression analysis. Finally, in Table 1 we can see the distribution of the road interventions by type. Out of 617 road segment interventions rehabilitation plays a big role, moreover, divided highways are the second most important intervention.

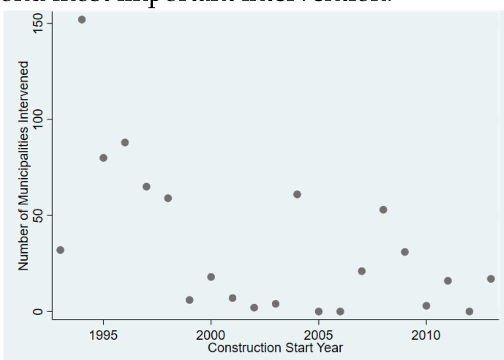


FIGURE 2 National road network vs roads concessions waves. Dates correspond to road finalization times.
Source: Invias and authors' calculations

TABLE 1 Road interventions by type

	Invias	1G	2G	3G
Divided Highway	4%	19%	39%	25%
Rehabilitation	51%	64%	50%	36%
3 Lane Road	3%	4%	-	29%
New Road	33%	7%	-	8%
Bridge/Variant	10%	6%	-	2%
Total	617	96	18	131

Data on interventions allow us to see that about 15.6% of the municipalities were directly in contact with improvements, and 35% were in those neighborhoods. In Table 2 we can see the number of intervened municipalities in each of the different generations.

TABLE 2 Proportion of interventions

	Number of Municipalities Intervened	**(%) of Municipalities intervened**	**Number of close Municipalities affected**
Invías	389	34.67%	2417
G1	72	6.42%	496
G2	17	1.52%	121
G3	98	8.73%	450

Note: We consider close municipalities as municipalities in a radio of 35km

4 | DATA

Colombia has 1103 municipalities which are our unit of observation.[5] We build a database with the universe of the municipalities from 1986 to 2014, using different information sources. Table 3 shows descriptive statistics for all used variables.

[5]Some of the municipalities are not connected to the national road network because they are located in the Amazon region or in islands, therefore, they are excluded from the estimation

4.1 | Outcome variables

The main objective of this work is to quantify the effects of infrastructure on economic development and inequality. Our main output variables will be the GDP with its correspondent sector divisions, the municipal development index, and three different measures inequality measured by GINI coeficients. One of the main sources of information was the municipality panel from the *Centro de Estudios para el Desarrollo Económico (CEDE)*. From this database, we obtained total GDP by year and also GDP by sector (industrial, services, and agricultural) taken from the national bureau of statistics from 2000 to 2009. Since there is not official data for the municipality GDP, the CEDE database uses the data from Torres et al. (2012). They compute it by using the taxes share of each municipality of the department's total tax revenue, $\mathrm{GDP}_i = \frac{tax_i}{\sum_{i \in j} tax_i} * \mathrm{GDP}_j$. i indexes municipalities and j departments that contain municipalities. According to the specific taxes used, this measure also provides a division of the municipal GDP into agricultural, services and industry sectors. Additionally, we use the municipal GDP constructed by the National Planning Department (DNP). This series gives us a longer panel (2000-2016) but associated sectoral GDP divisions are not available. This GDP measure is also built with taxes, although the weights given to different taxes is unknown as it is proprietary. However, when comparing the periods for which the municipal CEDE GDP and DNP longer panel GDP are both available, we find that the variable values are practically identical (within year correlations of the variables go from 0.98 to 0.99).[6] Finally, we measure development more broadly with a municipal development index, which includes social and financial variables of each municipality. This index includes information about population, access to public services, municipal investment per person, municipal tax per person, poverty, illiteracy, and percentage of the population attending school.

Income inequality measured by the income Gini index for 1993 and 2005, calculated using census data. Land and land ownership inequality Gini indices are also constructed using Agustin Codazzi Institute data between 2000 and 2012. The land ownership's Gini considers inequality of land value ownership across different owners. Land Gini indices measure inequalty in land size across distinct properties.

4.2 | Treatment variables

The literature has commonly used discrete variables to determine whether new infrastructure directly serves the focal area, often measuring directly if the infrastructure passes or is within a certain distance of a location. This measure is very precise but overlooks the network general equilibrium effect of most transport infrastructure: the effect on local economies happens through reducing the costs of trade, effectively bringing the city *closer* to other markets. A reduction in travel times between any two locations will affect the rest of nodes connected in the same network. Our main treatment variable is market access which captures access to other markets in the focal economy and will both be affected by nearby improvements in the infrastructure and, less strongly, by more remote changes in the infrastructure connected to the same road network. Gibbons et al. (2019) highlights two advantages of this measure: First, the treatment shows continuous variation in ways not directly proportional to the distance to the new roads or road improvement. Second, it allows to track the intensity of treatment for all municipalities in the network, even those remote to the improvements.[7]

[6]As a robustness test, we calculate our own measure of GDP using equal weights across taxes and use in robustness tests included in the appendix.

[7]In the appendix we discuss the results to distance to the closest infrastructure for the reduced form regressions

4.2.1 | Roads network and travel times

The road network is the official network reported by the Agustin Codazzi Geographical Institute for 2018. Invias provides road categories for each road stretch and associates it with different speeds according to their engineering design. On average, secondary roads had a speed of 50km/h, double lane roads 100km/h, within-cities-roads 40km/h, two way roads 70km/h, and three lane roads 80km/h. The network map was used to identify the location of current infrastructure and their type. Using the speeds for each road from Invias, the traveling times matrix between all municipalities are calcualted. Starting and ending points in the trips are located at the centroid of the main urban center. These traveling times, together with 1985 census population as a proxy for market size, are used to calculate market access (see section 5.1.1. Not changing the population allows to calculate the effects of changes in the traveling speed generated by transport improvements.

To generate traveling time matrices for years previous to 2018, we review all concession contracts and INVIAS archival information about changes to road infrastructure that were not included in concession roads.[8] These documents allowed us to impute back changes to speeds for each road stretch. For instance, if we found that a road was widened in 2009, we construct the counterfactual road for that equivalent stretch in 2008 by removing that road modification and reducing the implied improvement in speed. This new lower speed is used for that road stretch for all previous years unless another modification is found in a previous year, i.e 2003, in the documents. If that happens, the additional speed modification is made for that year backwards, and so on. This means that every historical change to the infrastructure changes the speed for the stretch in which the modification happened and affects the overall traveling times matrix for all municipalities in the network. When the modification is a new road, the whole network topology is recalculated to re-draw paths between municipalities so they follow the available roads in each point in time.

The engineering estimated speeds are homogeneous across roads with identical construction and dimensions. We further allow for speed changes due to geographical features, especifically due to terrain ruggedness, as in Nunn and Puga (2012). We further penalize speeds for high slopes. This means travel speeds in our matrix are not symmetrical. In a place with such a difficult geography (see section B), this additional nuance improves our market access calculation as a proxy for transportation costs.

We test our estimated travel times for the current network contrasting them with actual GPS measured traveling times using data from Garmin GPS provided by HERE technologies. We find our estimates are satisfactory, with differences between the estimated and observed times of less than 20% in the large majority of the trips. Differences in relative times across trips are even smaller across methods. The Figure A1 shows the distribution of errors.[9]

Interventions considered as affecting speeds are pavement, construction of second or third lane, road expansion to become a dual highway, construction of a new bridge, new ring road, and new roads. Therefore, it will be possible to calculate the effects of road

[8] Both sources of information were rich in details about road characteristics. Unfortunately, they were only available in hard copies, and digitization of this information was time consuming. We see making this information available to researchers with this work as a partial contribution as well.

[9] HERE technologies is the world leading location platform company. They track over 160 million vehicles all over the world, using Garmin devices. Data is mostly obtained in Colombia from GPS installed in freight trucks. Many trucks will not follow the shortest track between two points because of efficiency gained by servicing different points in the network in a single trip. As a consequence, some of the GPS traveling times we use to test will be longer than our estimated speeds from road characteristics obtained from contracts and road specifications, as the latter assume shortest trips between all points when defining the trip distances matrix. Indeed, we identify that some of the large differences between the GPS and the estimated matrix times were given by truck routes meant to cover several regions instead of going straight away from point single origin to single destination.

improvements in general but also to differentiate for the heterogeneous effects of different interventions, which is rare in the literature.

4.3 | Control Variables

We incorporate controls that might determine changes in outcome variables, additional to the effect of changes in market access. We include have the average score in high school standardized test *Saber 11* and average years of education to proxy for human capital quality. We use homicides per capita as a proxy of institutional strength. We include altitude to control for natural characteristics correlated with economic development. We include the literacy rate in 1951 to control for historical human capital. We also control for the rurality index that measures the percentage of the population that lives in rural areas in each municipality.[10]

Urban population density is calculated for the urban center of the municipality to control for unobserved agglomeration economies that may affect local production and development. The urban population density was calculated by using adjusted projections from the census 2005 census from DANE and the area of each municipality calculated using night lights. This allows us to account for growth and focus on actual urban footprints.[11]

5 | METHODOLOGY

This section outlines the methods we use in our empirical analysis as well as in some variable calculations.

5.1 | Research Design

We will estimate different versions of the following equation:

$$Y_{it} = \alpha * MA_{it} + \beta * X_{it} + a_i + c_t + \varepsilon_{it} \tag{1}$$

Y_{it} denotes the output variables in our case will take the logarithms of total GDP, agricultural GDP, industry GDP, service GDP, municipal development index, land GINI, owners GINI, and the income GINI. The vector X_{it} contains the control variables including geographic, historical, socio-economic, and institutional variables that might determine GDP (see section 4.3).

a_i, c_t are the municipality and time fixed effects respectively, so we can account for unobserved location-specific characteristics and time trends. Treatment$_{it}$ represents the treatment variable, the logarithm of market access.[12] In this case, α is the coefficient of interest and will capture the effect of the treatment on the output variable, which can be interpreted as an elasticity or semi-elasticity depending on the outcome variable.

In order to deal with endogeneity issues we use different strategies: First, we use the instruments for market access described in section 5.2. Also, since roads were designed predominantly to join the most important cities, those that were expected to have the strongest

[10] We also include additional variables as robustness tests: math and language results in standarized tests, temperature, annual rainfall, historical violence indicators like terrorist attacks. All results are robust to inclusion of these additional controls.

[11] Data on night lights are available in https://ngdc.noaa.gov/eog/dmsp/download_radcal.html.

[12] In the appendix we also calculate the same set of regressions using as explanatory variable the distance to the infrastructure which is often used as a treatment in this literature

TABLE 3 Variables Sources and Descriptive statistics

Source	Description	N	mean	sd	min	max	Years
Output variables							
DNP	GDP Long panel (m)	12,318.00	194,063.32	614,308.93	2.01	11535289.64	2000-2016
CEDE	GDP (m)	9,427.00	197.56	567.37	1.70	10,459.65	2000-2009
	Agricultural GDP(m)	9,427.00	23.80	29.06	0.01	353.89	2000-2009
	Industrial GDP (m)	9,427.00	65.52	224.35	0.12	4,035.58	2000-2009
	Services GDP (m)	9,427.00	92.35	305.29	0.09	6,458.35	2000-2009
	GINI Index	899.00	0.45	0.03	0.39	0.57	1993 and 2005
	Land GINI Index	12,318.00	0.69	0.11	0.02	0.96	2000-2012
	Owners GINI Index	12,318.00	0.71	0.10	0.02	0.97	2000-2012
	Municipality development Index	10,386.00	39.71	19.25	0.05	92.29	2000-2010
Control Variables							
CEDE	Altitude in m	12,318.00	1,156.80	1,204.33	2.00	25,221.00	All
	Total quantity of literates in 1951	12,318.00	4,277.82	21,500.28	0.00	431,654.00	All
	Standarized tests(normalized)	12,242.00	47.66	3.13	33.98	76.87	2000-2014
Population(DANE); Homicides(CEDE)	Homicides per capita	9,375.00	0.00	0.00	0.00	0.01	2003-2014
Authors' calculations; population(DANE); nightlights urban footprint(NOAA)	Population density (inhabitants per km^2)	12,318.00	113.97	410.64	0.20	8,781.83	1985-2017
Treatment Variables							
Authors' calculations (see section 5.1.1)	Market Access without municipalities in a radio of 35km	11,523.00	4,519,185.85	922,507.45	0.00	6,577,668.00	1993-2012
Instrumental variables							
Authors' calculations (see section B)	Market Access with least cost path	11,870.00	3,988,555.41	941,490.31	1,348.42	6,537,936.00	1993-2012
Authors' calculations	Market Access with semi-random component	11,575.00	2,430,063.81	580,516.02	0.00	4,397,569.00	1993-2012

development, the estimated effect of the improvements in market access in equation 1 will suffer from the most bias precisely for these main cities. Thus, besides full sample regressions, we also run regressions where the sample excludes the most important cities in the country and cities within 35km of them.

5.1.1 | Market Access Calculation

We use a continuous measure of treatment using a market access gravity equation as in Henderson and Wang (2007):

$$MA_{it} \equiv \sum_{k \in j | j \neq i}^{N_j - 1} \frac{n_{k85}}{t_{ik}^{\alpha}}. \tag{2}$$

We use census population in 1985, n_{k85}, as our proxy variable for market size in equation 2. By keeping population unchanged, we focus on the effect of changes in travel times due to roads improvements. The denominator uses the travel time between all municipalities that are connected in the network i and j, as described in section 4.2.1.

To calculate the decay parameter α in the denominator, which is the elasticity of trade to distance, we use the 2013 Commodity Flow Survey (CFS) for Colombia, from the Colombian Ministry of Transport, which measures bilateral trade flows among Colombian municipalities, following Duranton (2015) for Colombia and Duranton et al. (2014) for the US. This is a survey of trucks on major Colombian roads. Trucks are pulled aside on the road and information is collected about weight, origin, destination, and nature of the cargo, including products and their value. The estimated elasticity was -0.34 (see Table A.2), which means that distance between municipalities is an impediment to trade. Contrast, for instance, with elasticities for the US estimated with the same methodology, of between 1.64 and 1.91 (Duranton and Turner, 2012), or between 1.17 and 1.41 (Herzog, 2020).

Duranton (2015) estimates this elasticity to be -0.60 for Colombia when using the 2011 version of the survey. The 2013 version we use includes about 90.000 additional observations. Results are robust to changing the value of the elasticity to -0.60.

Table 3 shows some descriptive statistics for the market access. Figure 3 shows the percentage change in market access from 1993 to 2016. This gives an idea of the main winners with infrastructure during the period of study. Although the center of the country has important changes, the Caribbean coast in the north and some places in the east also present big changes in their market access. There are areas of the country that are unconnected to the national road and stayed that way. Figure 4c show the Market access for 2016, as you can see in the distribution for the three measures we use, the market access is considerably larger in the center of the country.

5.2 | Endogeneity concerns

According to Allen and Arkolakis (2019) there are three concerns to measure welfare effects of infrastructure: The routing problem, which shows that routes chosen are not exogenous and improvements in one road segment change trading cost for the entire network. The economic problem because road improvements by changing trade costs can also change the location of economic activity, generating winners and losers through general equilibrium relationships. And congestion, which also changes endogenously with road improvement and trade costs. Fortunately, we deal with some of these problems in our empirical strategy.

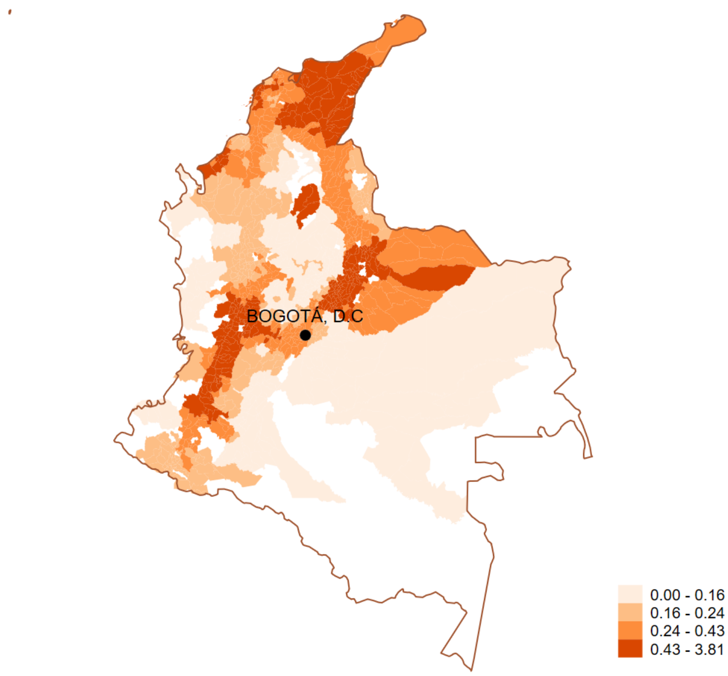
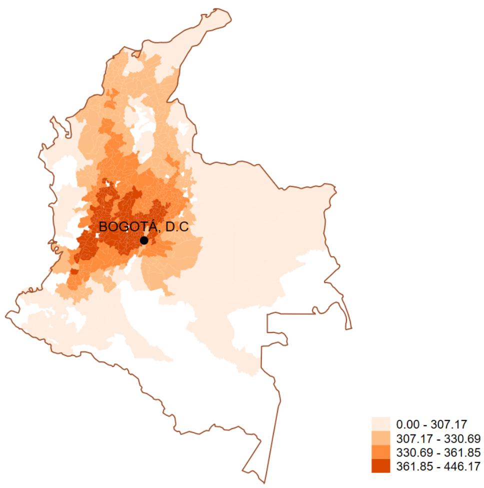
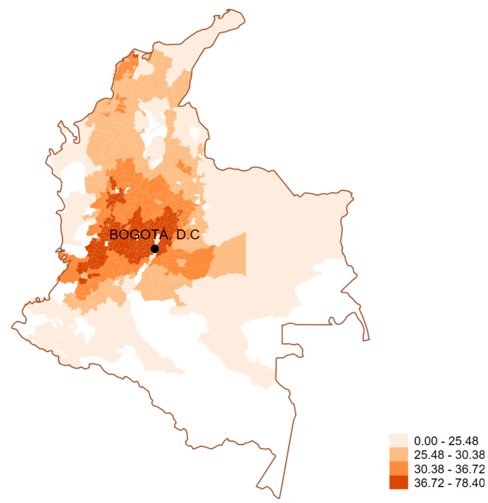


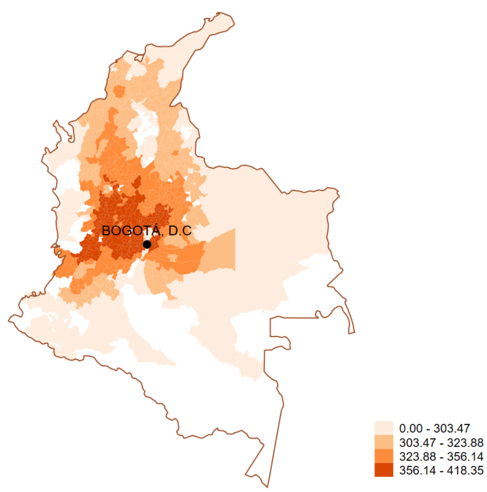
FIGURE 3 Percentage Change in Market access from 2000 to 2016 by municipalities. For this calculation we use 1985 population as constant



(a) MA^{α_1}



(b) $MA^{\alpha_1}_{SRC}$



(c) $MA^{\alpha_1}_{SRC,35km}$

FIGURE 4 Market Access in 2016

Note: To compute this market access we use an α of -0.34 and divide values by 10,000

Source: Authors' calculations

TABLE 4 Summary of Endogeneity concerns and estimation strategies

Endogeneity Concern	**Empirical Strategies**
1. Road network and improvements located to benefit main cities	A. Selection tests Chandra and Thompson (2000) B. Estimation of impact using only incidental cities C. IV: Quasi-random component of market access as an instrument (eq 3).
2. Road network path is chosen endogenously	A. IV: market access using least-cost-path counterfactual roads based on differential construction costs across space.Market access measures built with pre-colonial roads are tested as in Duranton (2015). Results upon request.
3.Spillovers between local municipality market effects and market access effects	A. Calculate market access only with municipalities that are farther than 50 km from the focal location B. Fix population of market access to initial period. C. Include market access of neighboring populations as controls.

The routing problem, is partially solved by using the market access, which takes into account how road improvements reduce travel times for the entire network. We could convert these reductions in travel costs. Market access also takes into account the variation of infrastructure over time, allowing to understand how the improvements might affect economic activity, which is one of the main output variables of this work. Finally, the fact that we verified travel times with the ones got from HERE technologies means that road congestion would be taken into account in our calculations. Also to take into account the economic problem we calculate spill over effects on nearby municipalities, which can gives partial view of winners or losers with the intervention.

Another common endogeneity concern is that new roads are not exogenous. In general, they are built to link important places, but they affect unimportant places. The fact that important places are taken into account to determine which roads are built will generate a simultaneous determination of the treatment and the output. However unimportant places receive road in a rather exogenous way. This will be one of the key problems facing the result, and we deal with it in several ways. First, in our must trusted specification we excluded from the estimation the capital and the second biggest cities in all *departments* by 1951. These cities are the ones that intuitively might be endogenous with roads construction. So in our reduced sample we only account for small and intermediate cities, sometimes called incidental cities.

A similar concern is that only faster-growing cities will be assigned roads, i.e. growth was generating the roads and not the other way around. To tackle this concern, we used a test similar to the one proposed by Chandra and Thompson (2000). Table 5 and 6 show the results of these tests for lags of our outcome variables for first differences, and also in levels. These tables report the coefficient of the lags in a logit model with random effects on the probability of the municipality getting a new road or road improvement inside their administrative borders, specifically estimating ψ and μ coefficients in $\mathrm{Pr}(\mathrm{Road}_{it}) \equiv \Lambda(\sum_{t=1}^{3} \psi_t \Delta Yt - (t-1))$ and $\mathrm{Pr}(\mathrm{Road}_{it}) \equiv \Lambda(\sum_{t=1}^{3} \mu_t Yt - (t-1))$. Overall, all results show either not statistically significant or with negligible economic magnitudes. These results suggest the endogeneity stemming from allocation of roads to locations with large levels or growth rates of our outcome variables is weak or non existent. These grant some validity to the OLS results. However, some coefficients show statistical significance despite small coefficients. Also, this test does not guarantee that selection is not done conditional on unobservable expectations of future behavior of the outcome variable. We develop instrumental variables to assuage such endogeneity concerns.

To deal with the simultaneous determination of the market access and the output variables, we also developed a market access instrument that subtracts from the regular market access a market access calculated for the big cities only. The new market access

TABLE 5 Selection tests using lagged first differences of outcome variables as determinants of the probability of receiving a road or road improvement

	TotalGDP		AgriculturalGDP		ServicesGDP		IndustrialGDP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First lag	-5.71e-08	-0.00000547	0.000000548*	0.0000562*	4.54e-08	0.00000755	-5.03e-08	-0.00000574
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Second lag	8.97e-08***	0.00000992***	0.000000276	0.0000264	0.000000137	0.0000148	0.000000114***	0.0000146***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	7939	7943	7939	7943	7939	7943	7939	7943
Random effects	No	Yes	No	Yes	No	Yes	No	Yes

	GDPLongPanel		Land Gini		Owners GINI		Dev Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First lag	0.00000271	0.000606	0.0333	5.209*	-0.00256	-6.998*	-0.0000541	-0.00346
	(0.000)	(0.001)	(0.019)	(2.125)	(0.069)	(3.309)	(0.000)	(0.007)
Second lag	0.00000662*	0.00195*	0.0101	0.921	0.0245	1.755	0.000226**	0.0217*
	(0.000)	(0.001)	(0.021)	(4.138)	(0.017)	(3.299)	(0.000)	(0.010)
N	14764	14770	9633	9638	9320	9325	8873	8877
Random effects	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: We do not include the GINI since we have information for only two years and lags cannot be implemented.

TABLE 6 Selection tests using lagged levels of outcome variables as determinants of the probability of receiving a road or road improvement

	TotalGDP		AgriculturalGDP		ServicesGDP		IndustrialGDP	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First lag	2.10e-08	0.00000309	5.55e-08	0.00000476	2.77e-08	0.00000367	7.57e-08	0.00000966*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Second lag	-1.83e-08	-0.00000257	1.25e-08	0.00000280	-1.92e-08	-0.00000227	-7.48e-08	-0.00000915*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	8873	8877	8873	8877	8873	8877	8873	8877
Random effects	No	Yes	No	Yes	No	Yes	No	Yes

	GDPLongPanel		Land Gini		Owners GINI		Dev Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First lag	0.00000377	0.000938	0.0212	2.845	0.0567**	7.234	0.000126	0.0112
	(0.000)	(0.001)	(0.021)	(4.540)	(0.021)	(3.790)	(0.000)	(0.007)
Second lag	-0.00000147	-0.000109	-0.00758	-0.447	-0.0203	-1.856	0.000279***	0.0419***
	(0.000)	(0.001)	(0.021)	(4.453)	(0.019)	(3.599)	(0.000)	(0.010)
N	14764	14770	10555	10560	10135	10140	9804	9809
Random effects	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: We do not include the GINI since we have information for only two years and lags cannot be implemented.

measure will contain only the variation generated by incidental cities. As said before, incidental cities are supposed to be exogenous, therefore only the semi-random component of the market access remains.

These methodologies are supported by current literature; Redding and Turner (2015) show some ways to address endogeneity concerns using instrumental variables. These can be summarized in planned route IV, as was proposed by Baum-Snow (2007) for the US and Baum-Snow et al. (2017) for China, and also used in Herzog (2020) and in a similar setting Hsu and Zhang (2014) for Japan. They also include the historical route IV, which is used in Duranton (2015) which proposed indigenous roads and the road system in 1938 for Colombia, or Duranton and Turner (2012) which uses both 1947 planned roads and 1898 map of railroads. Finally, they mention the inconsequential places approach based on the selection of sub-samples that are exempt from endogeneity, given that the presence of new infrastructure in these locations can be considered random. Faber (2014) uses this kind of variation to explain the effect of new highways on the production of Chinese cities. It is a similar approach to Banerjee et al. (2012), which uses inconsequential places to measure the effect of new roads in economic development in China. A more traditional contribution by Chandra and Thompson (2000) also uses this approach to account for the effects of the highway system in the United States on firms profits. Jedwab and Moradi (2016) and Jedwab and Storeygard (2020) use an instrumental variables approach to calculate effects of transportation infrastructure on Ghana and other 39 African countries from 1950 to 2015. The instruments developed in this work are new to the literature, being a clear contribution to how to think about sources of exogenous variation. We show that the statistical behavior of these instruments is better than those usually used.

A different strand of the literature uses structural models to estimate welfare effects, such as Allen and Arkolakis (2019) calculating the welfare effects of road interventions for the US road network. Additionally, Donaldson and Hornbeck (2016)calculate welfare effects of railroads on agricultural productivity for the US, updating well-known results by Fogel (1964). Our work uses elements of several of these methodologies to allow for a more reliable identification of the effects of road interventions on output and inequality. We used indigenous routes suggested by Duranton (2015), and we built a hypothetical transport infrastructure to link main Colombian cities in 1938 (Donaldson and Hornbeck, 2016), by finding the least cost path using Colombian geography, and used it as an instrument for actual infrastructure. We also take the main cities out of the estimation.

5.2.1 | Semi-random component of market access

As said before, infrastructure is built to connect big agglomeration centers, therefore any estimation of the effect of the infrastructure on output variables including those centers might be biased and inconsistent. In order to regain the ability to say something about the impact of roads on the main cities or nodes, we look at the change in market access experienced by the main cities that correspond to the unplanned access to incidental cities gained or increased from new roads or upgrades. We calculate it as a difference between MAs:

$$z_{it} = MA_i - MA_{\text{Main},\, i}{}^{13} \tag{3}$$

This measure will take out from the market access the endogenous component, which is given by the effect of big cities, allowing us to use it as an instrument for market access.

[13] $MA_{\text{Main}i}$ is the market access calculated without incidental cities. This strategy is a novel contribution that builds on a strategy used by Herzog (2020).

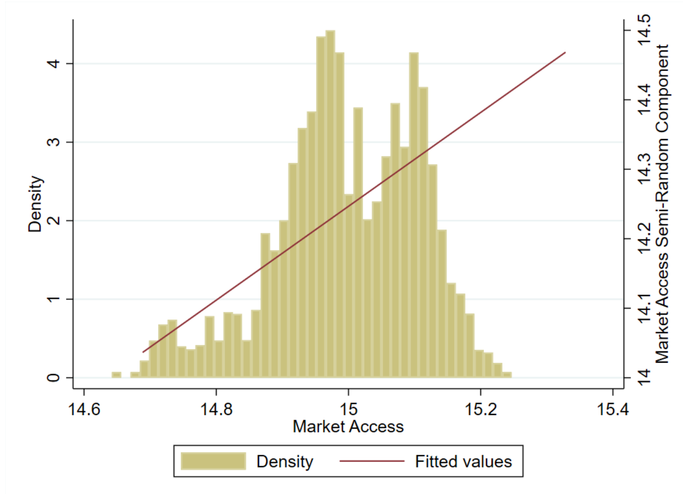


FIGURE 5 Market Access and Market Access Semi-Random Component

Note: Both Market access variables are in logarithms, in yellow you can see the distribution of the Market access and in red the OLS regression line of the Market Access vs the Market Access Semi-random Component.

5.2.2 | Market access least cost path

We used the calculated least cost path as was done by Bogart et al. (2020), used as a source of exogenous variation for distance to train stations to calculate the effect of railroads on population location and occupations between 1851 and 1891 in England. In this analysis they also exclude node cities from the main estimating sample as we did. Also Berger (2019) for Sweden uses the LCP calculated using geographical characteristics. However, as discussed before, distance measures do not account for network effects. Furthermore, we used the least cost path to create a new market access following the procedure discussed above and replacing the least cost path segments for the actual roads to calculate distances. For the rest of the network, not connected to the least cost path, mostly small cities, we used the same information calculated before. So this measure of market access is a source of exogenous variation for all roads connecting main cities. A detailed explanation of the steps followed to calculate the least cost path instrument is found in section B. Figure 6 shows this instrument presents significant variation.

TABLE 7 Regressions first stage with different instruments

	(1) Market Access	(2) Market Access	(3) Market Access
Market Access LCP	0.38***		-0.01***
	(0.005)		(0.000)
Market Access SRC		0.79***	0.80***
		(0.002)	(0.002)
N	10155	10223	10080

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities.

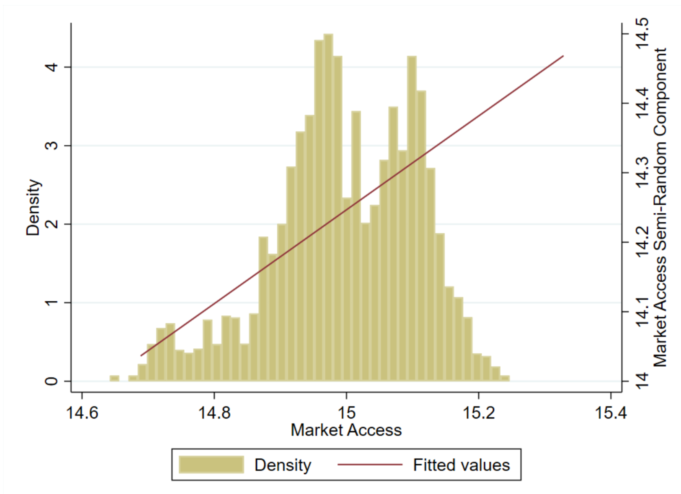


FIGURE 6 Market Access and Market Access least cost path

Note: Both Market access variables are in logarithms, in yellow you can see the distribution of the Market access and in red the OLS regression line of the Market Access vs the Market Access least cost path.

6 | RESULTS

6.1 | Instrument Performance

First stage result reported in table 7 show that the instruments proposed in this paper are strong predictors of market access, the treatment.Significant coefficients in the first stage and large F statistics suggest our instrument is relevant. Figures 5 and 6 illustrate the first stage results through the fitted line that shows a positive relationship between the instrument and the treatment conditional on controls, i.e. a relevant first stage. Table 7 shows the details of the first stage. The calculated instrument are a good predictor of contemporaneous municipality market access.

Next, we explore the validity of our instruments by inspecting how well they satisfy the assumptions of conditional independence, monotonicity, and the exclusion restriction.

The conditional independence assumption requires that the instrument provides a variation exogenous to the characteristics of the units receiving treatment. To assess this

TABLE 8 1 stage regressions for subsamples divided by quantiles

	(1) Quantile 1	(2) Quantile 2	(3) Quantile 3	(4) Quantile 4
IV: MA-SemiRandomComponent				
Market Access SRC	0.825***	0.792***	0.727***	0.724***
	(0.006)	(0.007)	(0.007)	(0.005)
N	1206	1206	1231	1383
IV: MA-Leastcostpath				
Market Access LCP	0.470***	0.304***	0.545***	0.647***
	(0.018)	(0.017)	(0.018)	(0.012)
N	1187	1196	1224	1378

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities.

assumption, we regress the instrument on tenants' observable characteristics following Bhuller et al. (2020), a procedure comparable to a test of random assignment of treatment in a randomized controlled trial. A few characteristics remain significant in explaining the instrument, although the economic size is negligible in most. Between the two, the least cost path seems to perform better in this test.[14]

Conditional random assignment of municipalities to locations with different instrument values (either predicted market access by least cost path networks, or increases access to incidental cities, -the semi random component of the market access-) validates the causal interpretation of the effect of the instrument on our outcomes of interest. However, our question is not about the effect of these instruments but about the effect of market access. To interpret the IV estimates as causal effects of the market access, we need to make sure the the exclusion restriction condition is satisfied, namely, that the market access implied by the least cost path network, should affect a municipality's economic outcomes only through changing the probability of occupying a rent-stabilized unit. The same argument must be made for the SRC instrument. We include a series of control variables, including municipality fixed effects, and get robust 2SLS coefficients on the treatment.

If the effect of market access improvements from road expansion is constant across municipalities, conditional independence and the exclusion restriction guarantee our estimate gives a causal effect. Otherwise, we can still perform estimation, but monotonicity must be assumed. Although different values of the instrument does not need to affect all potential municipalities, it should affect all those it does affect in the same direction (Cunningham, 2021), i.e. being located in a place that lowers the cost of building a road should not reduce your probability of getting the road, conditional on observables. To test this monotonicity assumption, we show first-stage estimates are nonnegative for multiple subsamples following Bhuller et al. (2020). Table 8 shows the first stage results when subsamples are divided by quartiles from the market access distribution. Coefficients are positive and of a similar magnitude. Table 14 performs a similar exercise dividing the data into regional subsamples, while table 10 divides observations into rural and urban municipalities. As before, results are nonnegative and stable.

[14] This is analog to the analysis of Table 5 and 6, following Chandra and Thompson (2000), but for the instruments instead of the treatment variables. Detailed results are available upon request

TABLE 9 1 stage regressions for subsamples divided by regions

	(1) Andes	(2) Caribbean	(3) Pacific	(4) Orinoquia
IV: MA-SemiRandomComponent				
Market Access SRC	0.716***	0.236***	0.935***	0.661***
	(0.005)	(0.012)	(0.001)	(0.015)
N	3597	1018	772	342
IV: MA-Leastcostpath				
Market Access LCP	0.023***	0.069***	0.953***	0.010
	(0.001)	(0.012)	(0.006)	(0.016)
N	3543	1036	742	342

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities.

TABLE 10 1 stage regressions for subsamples divided by urban and rural municipalities

	(1) Rural	(2) Urban
IV: MA-SemiRandomComponent		
Market Access SRC	0.789***	0.779***
	(0.003)	(0.008)
N	3818	2040
IV: MA-Leastcostpath		
Market Access LCP	0.211***	0.770***
	(0.007)	(0.006)
N	3812	1998

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities.

6.2 | Impact of roads on economic performance

Table 11 shows OLS and IV regressions using the entire sample of municipalities. All specifications include fixed effects of time and municipality and controls by human capital, geography, demographics, institutions and history.[15] The last specification using both instruments also clusterizes standard errors at the departments level. That is our preferred specification. Columns show results for different outcome variables. IV results shows that market access growth coming from the road expansions increases total and sectoral GDP, with a higher increase shown by the industrial sector GDP. This sector, together with agriculture, has the most tradeable type of activity and is expected to reap the most benefits from transportation costs reduction. Finally, municipal development increases with market access. Higher market access improves the potential of the municipalities in different areas.

Income inequality can be affected as a consequence of the growth in economic activity. Factor productivity can increase with the growth in the scale of local sectors, increasing wages and return to entrepreneurial capital. Equilibrium prices should also increase if other markets have higher demand for the locally produced goods, or if the losses from transportation are lower. The effect on within city income inequality depends on whether lower income people benefit disproportionately from this increase in economic market access. Our results find a small point estimate with no statistical significance. This suggests the growth in economic activity is not changing the distribution of these gains across workers of different income levels.

As described in section 4, land ownership's Gini considers inequality of land value ownership across owners. Land Gini indices measure inequality in land size across distinct properties. Our main results show that land gini and owners gini are increased by market access, suggesting an increase in rural land concentration in two directions: bigger parcels and fewer owners. Faguet et al. (2020) explore the effects of land reforms on development in Colombia. One of the interesting conclusions of this work is that cities in which land owning elites were already established saw much lower levels of ownership changes in attempted reform initiatives. In particular, plot sizes were smaller, land distribution was more disperse, and government investment was lower. While this work does not consider the effects of roads, it seems that roads might have deepened into this result helping to increase land concentration.

Table 12 shows the results for a reduced sample of cities All the effects on GDP are slightly higher when main cities are excluded from the sample, showing that the marginal growth effects in small municipalities are higher than the ones in big municipalities. Higher agglomeration economies and bigger production size in big cities imply smaller coefficients. This can be seen in Table 12 where the inequality coefficients, also hold.

6.2.1 | Spatial fadeout

We aim to establish if the effects of roads are diluted as distance to infrastructure increases. Data show that they are and dilution rate is fast. For that we analyze the effect of market access on municipalities that are in a buffer of certain number of km from any intervention (20km, 25km 30km, 35km and 40km). Table 13 shows OLS and instrumental variable results of market access on GDP using those subsamples. The preferred specification shows effects of infrastructure through market access on total GDP that decrease with distance to the infrastructure, as expected. It is interesting that effects of market access up to 20km of the infrastructure are 2.6 times as big as the average effect and three times the effect of those

[15] Variables include standarized test scores *Saber 11*, altitude above sea level, homicides per capita, population density, rurality index and number of literate people in 1951.

TABLE 11 Regressions Summary- full sample

	(1) GDPLongPanel	(2) TotalGDP	(3) AgriculturalGDP	(4) ServicesGDP	(5) IndustrialGDP	(6) GINI	(7) OwnersGINI	(8) LandGINI	(9) DevIndex
OLS: Market Access									
Market Access	-0.052	0.203***	0.253***	0.364***	0.075*	-0.069*	0.057***	0.129***	1.776***
	(0.030)	(0.023)	(0.028)	(0.026)	(0.034)	(0.034)	(0.012)	(0.011)	(0.375)
N	10881	6909	6909	6909	6909	106	9217	8639	7898
IV: MA-Leastcostpath									
Market Access	0.262***	0.551***	0.305***	0.638***	0.718***	-0.054	0.133**	0.199***	2.516***
	(0.052)	(0.041)	(0.048)	(0.042)	(0.062)	(0.091)	(0.050)	(0.014)	(0.636)
N	10672	6776	6776	6776	6776	98	9052	8499	7746
F 1st stage	5021	2831	2831	2831	2831	9	644	73833	3374
IV: MA-Semi-Random Component									
Market Access	1.421***	2.238***	2.366***	2.602***	2.385***	-0.078*	0.080***	0.132***	27.008***
	(0.148)	(0.105)	(0.128)	(0.120)	(0.158)	(0.036)	(0.013)	(0.012)	(1.762)
N	10762	6834	6834	6834	6834	106	9117	8539	7812
F 1st stage	136791	83345	83345	83345	83345	717	131653	204321	96764
IV: MA-Semi-Random Component and Leastcostpath									
Market Access	1.445**	1.851***	1.769**	1.703*	2.705***	-0.079	0.109	0.167***	21.784***
	(0.453)	(0.476)	(0.504)	(0.633)	(0.730)	(0.040)	(0.079)	(0.039)	(3.506)
N	10586	6722	6722	6722	6722	98	8982	8429	7684
Sargan test ^{a}	.2211	.2166	.7402	.2078	.2054	.3233	.6672	.021	.4503

Standard errors in parentheses. ^{a} Sargan Test p-value. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities.

TABLE 12 Regressions Summary- Reduced sample

	(1) GDPLongPanel	(2) TotalGDP	(3) AgriculturalGDP	(4) ServicesGDP	(5) IndustrialGDP	(6) GINI	(7) OwnersGINI	(8) LandGINI	(9) DevIndex
OLS: Market Access									
Market Access	-0.026	0.224***	0.276***	0.388***	0.089**	-0.060	0.073***	0.132***	1.753***
	(0.030)	(0.022)	(0.027)	(0.026)	(0.034)	(0.036)	(0.013)	(0.014)	(0.373)
N	10331	6559	6559	6559	6559	101	8760	8202	7499
IV: MA-Leastcostpath									
Market Access	0.306***	0.588***	0.341***	0.679***	0.743***	-0.048	0.140**	0.186***	2.491***
	(0.053)	(0.040)	(0.047)	(0.041)	(0.063)	(0.097)	(0.052)	(0.015)	(0.631)
N	10155	6447	6447	6447	6447	95	8625	8092	7371
F 1st stage	4745	2674	2674	2674	2674	9	597	70542	3188
IV: MA-Semi-Random Component									
Market Access	1.501***	2.284***	2.508***	2.709***	2.426***	-0.069	0.096***	0.140***	27.324***
	(0.152)	(0.105)	(0.128)	(0.121)	(0.166)	(0.037)	(0.013)	(0.014)	(1.800)
N	10223	6491	6491	6491	6491	101	8670	8112	7421
F 1st stage	125553	76443	76443	76443	76443	645	120280	128618	88781
IV: MA-Semi-Random Component and Leastcostpath									
Market Access	1.631**	1.947***	1.955**	1.844*	2.777**	-0.070	0.124	0.157**	22.090***
	(0.473)	(0.522)	(0.561)	(0.672)	(0.786)	(0.040)	(0.079)	(0.046)	(3.708)
N	10080	6400	6400	6400	6400	95	8565	8032	7317
Sargan test ^{a}	.2395	.2397	.7555	.2364	.2127	.4085	.8186	.0417	.4726

Standard errors in parentheses. ^{a} Sargan Test p-value. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression excludes main cities.

TABLE 13 Fade-out Regressions

	(1) 20km	(2) 25km	(3) 30km	(4) 35km	(5) 40km
OLS: Market Access					
Market Access	4.301***	3.445***	3.242***	2.026***	1.824***
	(0.110)	(0.113)	(0.106)	(0.106)	(0.109)
N	2676	3240	3825	4397	4762
IV: MA-Leastcostpath					
Market Access	5.373***	5.810***	5.033***	4.294***	4.726***
	(0.233)	(0.348)	(0.303)	(0.388)	(0.413)
N	2618	3168	3746	4318	4669
F 1st stage	598	362	499	365	388
IV: MA-Semi-Random Component					
Market Access	4.297***	3.575***	3.224***	1.933***	1.914***
	(0.112)	(0.115)	(0.110)	(0.110)	(0.114)
N	2648	3198	3783	4355	4720
F 1st stage	107315	82046	44301	45749	48114
IV: MA-Semi-Random Component and Leastcostpath					
Market Access	5.040***	3.837***	3.330***	1.826*	1.765*
	(0.561)	(0.940)	(0.881)	(0.753)	(0.714)
N	2611	3147	3725	4297	4648
Sargan test p-value	.6957	.2418	.2518	.2507	.2544

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in table 3, they also include municipality and year fixed effects, for the two instruments regression we include clustered errors at department level. This regression includes incidental cities. The dependant variable is the GDP

that are only 20km further away.

6.2.2 | Regional Heterogeneity

We also test for subsamples using a conventional division of the country into *natural regions*, set by using geographic features of territories. According to these criteria, the country is divided into 5 regions: *Andean*, which gathers all municipalities on top of the Andes; *Caribbean*, composed by all municipalities next to the Caribbean ocean; *Pacific*, composed by the municipalities next to the pacific ocean in the west part of the country; the *Orinoquia*, where all the plain field in the east of the country is located; and finally the *Amazon*, which is basically the amazon jungle. The last two regions have a small number of municipalities, which affects the precision of the results.

Results in Table 14 show important heterogeneity among regions. As expected, in the *Andean* region, where most of the country's economic activity is located, market access has a positive and significant effect on total GDP. The size of the coefficient is very similar to the one obtained for the reduced sample before. It is, however, surprising that in the Caribbean region the relationship between market access and GDP turns negative and significant for some specifications; a similar pattern arises in the *Orinoquia* region. In the pacific region market access has a big positive impact on GDP that doubles the one found in the *Andean* region.

TABLE 14 Regional regressions

	(1) Andes	(2) Caribbean	(3) Pacific	(4) Orinoquia
OLS: Market Access				
Market Access	1.560***	-0.009	3.635***	-1.856*
	(0.152)	(0.018)	(0.220)	(0.723)
N	4042	1151	823	370
IV: MA-Leastcostpath				
Market Access	-80.173	0.026	0.453	-0.036
	(241.658)	(0.024)	(0.578)	(0.977)
N	3979	1130	795	370
F 1st stage	0	9163	183	448
IV: MA-Semi-Random Component				
Market Access	1.683***	-2.774*	3.883***	-2.861***
	(0.185)	(1.193)	(0.225)	(0.736)
N	4042	1095	823	370
F 1st stage	8557	431	20309	10839
IV: MA-Semi-Random Component and Leastcostpath				
Market Access	1.715*	-2.970	3.884**	-2.504
	(0.592)	(5.463)	(0.540)	(1.728)
N	3979	1095	795	370

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for all regression we use the GDP as dependent variable.

7 | CONCLUSION AND POLICY DISCUSSION

Transport infrastructure affects GDP and inequality. Results show that total GDP, services GDP, industrial GDP, and Agricultural GDP increase with road improvements that increase market access. This result is robust to all our specifications. This result has been well established in the literature for other regions. In this study we improve on the measurement of this effect by implementing novel and refined instrumental variables. A clear policy lesson here are the large gains in economic activity from transportation network improvements that increase market access. This impact affects different economic sectors, so there should not be opposing interests across sectors.

We find that market access has slightly stronger effects on GDP in small and intermediate cities (reduced sample), while land inequality had a higher increase when big cities were included. Also effects are heterogeneous in different parts of the country. Only the Andean and the Pacific regions show positive significant effects of market access on GDP. The effect of the pacific region is almost twice as big as the one in the Andean region. The Pacific region has a much more serious issue of infrastructure deficit. Larger gains in the Pacific region suggest there may be decreasing marginal returns of infrastructure. This suggest that infrastructure improvements that aim to maximize impact should focus on areas that currently have low levels of infrastructure and that are found in locations where improvements can bring the access to larger markets.

Another important finding of this research is that the type of improvement or infrastructure matters. New roads, third lanes, double lanes, bridges, and variants have bigger and stronger impacts on output variables. Spillover effects of GDP on neighboring municipalities show that municipalities in the control group would have higher GDPs (industrial and services) if they are benefited with a road improvement, this effect is bigger the closer they are to the road. Also this benefit would imply better income distribution but generates a higher concentration of land (See A.4 in the Appendix.

At the same time, while income inequality tends to decline with new road expansion, it does not show a strong statistical relationship with market access. However, land inequality increases significantly with market access, showing that a process of concentration of rural land is accelerated by the growth in market acess. One possible explanation is that exploitation of land has increasing returns to scale with the growth of market access. As a consequence, improvements in market access make it efficient for landowners that already own large areas of land to bid more to augment their holdings.[16] In terms of policy, this suggests that despite positive effects on economic activity, investment in transportation infrastructure has distribution effects that may concentrate wealth in fewer hands. Potential redistribution mechanisms could be contemplated. If land transfers respond to efficiency gains, this could be beneficial both for the landowners that see their holdings increase, as for those who sell. In places with poor land property rights, however, care should be taken so that these process do not result in displacement of owners that are not compensated fairly.

[16] Despite instruments passing all validity tests, one could still worry that there was some anticipation effects that led to land concentration. Faguet et al. (2020) finds that Colombian elites were better informed and had influence in different decisions of local governments when road improvements were decided. As a consequence, they could have had speculative motives to perform land banking.

REFERENCES

Allen, T. and Arkolakis, C. (2019) The welfare effects of transportation infrastructure improvements. *Tech. rep.*, National Bureau of Economic Research.

Asher, S. and Novosad, P. (2020) Rural roads and local economic development. *American economic review*, **110**, 797–823.

Baldomero-Quintana, L. (2020) How infrastructure shapes comparative advantage.

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.

Barro, R. J. (1990) Government spending in a simple model of endogeneous growth. *Journal of political economy*, **98**, S103–S125.

Baum-Snow, N. (2007) Did highways cause suburbanization? *The quarterly journal of economics*, **122**, 775–805.

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. (2016) *Highways, market access and urban growth in China*. SERC, Spatial Economics Research Centre.

Berg, C. N., Deichmann, U., Liu, Y. and Selod, H. (2017) Transport policies and development. *The Journal of Development Studies*, **53**, 465–480.

Berger, T. (2019) Railroads and rural industrialization: Evidence from a historical policy experiment. *Explorations in Economic History*, **74**, 101277.

Bhuller, M., Dahl, G. B., Løken, K. V. and Mogstad, M. (2020) Incarceration, recidivism, and employment. *Journal of Political Economy*, **128**, 1269–1324.

Bogart, D., You, X., Alvarez, E., Satchell, M. and Shaw-Taylor, L. (2020) Railways, population divergence, and structural change in 19th century england and wales1.

Calderón, C. and Chong, A. (2004) Volume and quality of infrastructure and the distribution of income: an empirical investigation. *Review of Income and Wealth*, **50**, 87–106.

Calderón, C. and Servén, L. (2014) *Infrastructure, growth, and inequality: an overview*. The World Bank.

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.

Cunningham, S. (2021) *Causal Inference*. Yale University Press.

Dercon, S. (2004) Growth and shocks: evidence from rural ethiopia. *Journal of Development economics*, **74**, 309–329.

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. (2015) Roads and trade in colombia. *Economics of Transportation*, **4**, 16–36.

— (2016) Agglomeration effects in Colombia. *Journal of Regional Science*, **56**, 210–238. URL: `http://onlinelibrary.wiley.com/doi/10.1111/jors.12239/abstract`.

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

Duranton, G. and Turner, M. A. (2012) Urban growth and transportation. *Review of Economic Studies*, **79**, 1407–1440.

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

Faguet, J.-P., Sánchez, F. and Villaveces, M.-J. (2020) The perversion of public land distribution by landed elites: power, inequality and development in colombia. *World Development*, **136**, 105036.

Ferreira, F. H. (1995) Roads to equality: wealth distribution dynamics with public-private capital complementarity. *LSE STICERD Research Paper No. TE286*.

Fogel, R. W. (1964) *Railroads and American economic growth*. Johns Hopkins Press Baltimore.

Getachew, Y. Y. and Turnovsky, S. J. (2015) Productive government spending and its consequences for the growth–inequality tradeoff. *Research in Economics*, **69**, 621–640.

Gibbons, S., Lyytikäinen, T., Overman, H. G. and Sanchis-Guarner, R. (2019) New road infrastructure: the effects on firms. *Journal of Urban Economics*, **110**, 35–50.

Gibson, J. and Olivia, S. (2010) The effect of infrastructure access and quality on non-farm enterprises in rural indonesia. *World Development*, **38**, 717–726.

Hallmark, S. L. and Isebrands, H. (2005) Evaluating speed differences between passenger vehicles and heavy trucks for transportation-related emissions modeling. *Journal of the Air & Waste Management Association*, **55**, 1441–1450.

Henderson, J. V. and Wang, H. G. (2007) Urbanization and city growth: The role of institutions. *Regional Science and Urban Economics*, **37**, 283–313. URL: `https://ideas.repec.org/a/eee/regeco/v37y2007i3p283-313.html`.

Herzog, I. (2020) National transportation networks, market access, and regional economic growth. *Tech. rep.*

Hsu, W.-T. and Zhang, H. (2014) The fundamental law of highway congestion revisited: Evidence from national expressways in japan. *Journal of Urban Economics*, **81**, 65–76.

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.

Jedwab, R. and Storeygard, A. (2020) The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960-2010. *Tech. rep.*, National Bureau of Economic Research.

Khandker, S. R., Bakht, Z. and Koolwal, G. B. (2009) The poverty impact of rural roads: Evidence from bangladesh. *Economic development and cultural change*, **57**, 685–722.

Khandker, S. R. and Koolwal, G. B. (2011) Estimating the long-term impacts of rural roads: a dynamic panel approach. *World Bank Policy Research Working Paper*.

List, J. A., Momeni, F. and Zenou, Y. (2019) Are estimates of early education programs too pessimistic? evidence from a large-scale field experiment that causally measures neighbor effects.

Mu, R. and Van de Walle, D. (2011) Rural roads and local market development in vietnam. *The Journal of Development Studies*, **47**, 709–734.

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

Pachón, and Ramírez, M. T. (2006) *La infraestructura del transporte en Colombia durante el siglo XX*. Fondo de Cultura Económica.

Redding, S. J. 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.

Safford, F. and Palacios, M. (1998) Colombia: Fragmented land, divided society. edition.

Safford, F., Palacios, M. et al. (2002) *Colombia: fragmented land, divided society*. Oxford University Press, USA.

Torres, F. S., España, I. et al. (2012) Urbanización, desarrollo económico y pobreza en el sistema de ciudades colombianas 1951-2005. *Tech. rep.*, Universidad de los Andes-CEDE.

Tsivanidis, N. (2018) The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio. *Unpublished manuscript*.

World Bank (1994) *World Development Report 1994 : Infrastructure for Development*. Oxford University Press.

A | TRAVEL TIMES

We use gps data from Here.com to validate our estimated travel times. We find that 80% of the data lies with an absolute error lower than 25% and that the error is centered around zero. Larger differences in some routes could come from gps times that come mainly from truck and not car routes. First, truck routes follow routing considerations other than minimization of travel time, especially when they have to service multiple dropoff locations. Second, they are slower on average (Hallmark and Isebrands, 2005). In contrast, night routes, also more often taken by trucks than by cars, can present lower gps times than the times predicted for cars. The error is defined as $\varepsilon = \frac{\frac{Realtime}{Max(Realtime)} - \frac{Estimatedtime}{Max(Estimatedtime)}}{\frac{Realtime}{Max(Realtime)}}$

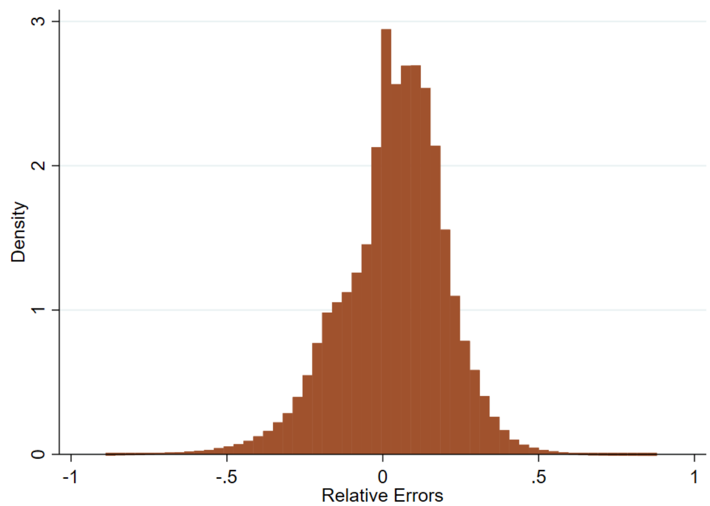


FIGURE A1 Relative errors in travel times

Note: This figure shows the density of the relatives errors, we compare the travel times computed with our network with the travel times obtained from Here.com.

Source: Authors' calculations

B | DETAILS ON THE CALCULATION OF THE LEAST COST PATH INSTRUMENT

As said, one of the instruments for distance to infrastructure was the least-cost path which was developed by calculating the least construction cost to link main cities in the country in 1938. We took consecutive pairs of cities from south to north of the country, and used detailed information on Colombian geography such as body waters (lagoons, swamps, rivers, etc.), terrain roughness, protected areas and dense jungle among other geographic characteristics. Engineering information provided by INVIAS would rank the relative difficulty, and therefore cost increase, of each of these features for construction.

Colombia has a uniquely broken geography (Safford et al., 2002), which makes road transportation inefficient and costly. For example, while driving by car from Bogota to Cali in Colombia entails about the same distance as driving from Miami to Daytona Beach in Florida, US, at 65km/h, the first journey takes 9 hours and the second takes about 3 hours 45 minutes. This broken geography gives strong variation to constructions costs over space.

Colombian geography is unique due to the fact that the Andes is divided into three different mountain chains within it, from its border with Ecuador that goes in the north-northeast direction. This generates natural geographic divisions among regions that have consequences on economic and social outputs. The historical dispersion of most of its population in isolated mountain pockets delayed transport infrastructure and an integrated national market. Roads were created to connect these important but isolated hubs. In the first stages, they did not connect areas with a smaller population and less access to markets. For example, about 56% of the country's territory represented mostly by the plain fields on the eastern region, and the Amazon in the southeast do not have national-level roads to connect them to the rest of the country because the construction costs there are high and the population is still low. This region accounts for less than 19% of the total population, and also, the region has some of the lower per-capita income in the country. Since Colombia is in the tropic, it does not have seasons. However, there is a noticeable variation in weather from perpetual snow mountains to the Amazon jungle to temperate weather regions where most of the country's GDP is generated. For example, only 23 cities of the departments (out of 1103) produce 70% of the national GDP. From the historical point of view, this result is that tempered weather regions had better agricultural productivity, less incidence of sickness, and are located in the high-but-plain parts of the mountain chains making construction and development easier than in extreme weather regions. All these geographical features have made it more costly to integrate the country using transport infrastructure.

Colombia is also an interesting study case, given its unique geography and its institutional arrangement. The country is crossed by the Andes which is divided into 3 separate branches. This feature creates natural heterogeneity in small areas, from snow-capped mountains to extensive plain fields and rain forests. This diversity directly impacts agricultural productivity; the highest agricultural productivity was obtained in elevated plateaus with temperate climate. Spanish colonial development was concentrated in these areas. Besides high agricultural productivity, this climate implied low incidence of tropical diseases, while the altitude provided natural protection against indigenous tribes (Safford et al., 2002). This concentrated has fragmented development, with highly heterogenous regions. Accordingly, we expect heterogeneity in the impact of roads in different regions. We test for this heterogeneity in our analysis.

To develop our counterfactual least-cost path road network, we use granular geographical information available and the information on the main cities in each department to the 1938 census. Back in 1938, the country was divided into administrative units ordered by their importance and were *departments*, *intendencias*, and *comisarías*. All these units contained smaller units called municipalities. Departments contained around 97% of the country's population, and their city capitals were the main economic agglomerations in that area. There were 15 departments: Antioquia, Atlántico, Bolivar, Boyacá, Caldas, Cauca, Cundinamarca, Choco, Huila, Magdalena, Nariño, Norte de Santander, Santander, Tolima y Valle del Cauca. To build this counterfactual, we took the capital and second main city according to census information. After locating these cities in the map, we start calculating the least cost path between them using Dijkstra's algorithms and linking pairs of cities from south to north. In order to use this function it is necessary to include the different weights that each geographical characteristic has. For this calculation, we considered: terrain slopes which make harder to built roads. Water bodies such as rivers, lakes, lagoons, swamps, reservoirs, and wetlands. In general, the presence of these areas implies a higher construction cost for roads or the impossibility of construction. We also take into account protected areas because of environmental reasons.Specific weights to these characteristics were assigned according to the criteria given by INVIAS' engineers that were consulted about the difficulty that each of these features implied for construction. We used their experience about Colombian terrain, roads construction and supervision in order to define the specific weights. We

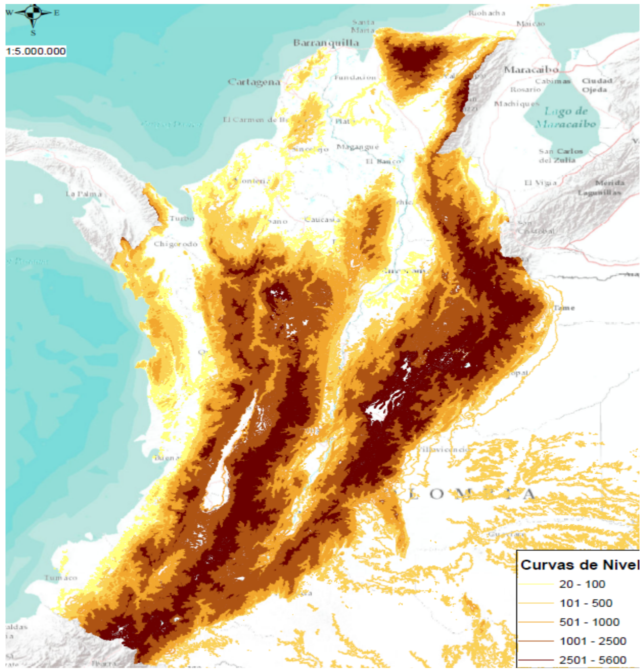


FIGURE A2 Uneven Colombian Geography

consult three different engineers and answers were alike. Table A.1 shows a summary of the different weights applied to Colombian geography.

The least cost path found does not have any restrictions in their calculation such as limiting the number of km. Actually, in some places of the country actual roads are pretty far away from the least cost path.

This instrument has two arguments for its validity. The first one is historical. The current road system started construction in 1951. Thus,selecting population in 1938 to determine the main cities could diminish endogeneity concerns. The second argument is geographic information and technology to calculate the least-cost path, which is more exogenous than the actual roads but is related to it. The least-cost path is a beneficial instrument if geography is not related to our outcome variables, unfortunately in our case, it is related to main cities. That is why we take them out of the econometric analysis. However, small towns that are crossed by the least cost path are exogenous to the infrastructure assignment because they do not influence building roads decisions.

The Figure A4 shows the comparison of the least-cost path in green versus actual roads in red. In this closeup, we can see that the least-cost path in many regions of the country is fairly exogenous, even though the first stage statistic confirms that it is a good predictor for actual roads. The least-cost path tool in ArcGis determines the least cost way to get from a destination point to a source, taking into account different rasters containing relevant information to improve the accuracy of the calculation, and each raster has to be weighted according to their impact on the cost. This allows the researcher to include various features to make the calculation either more realistic or adjusted to the theoretical model.
Technically, the least-cost path analysis in GIS means that the eight neighbors of a raster cell are evaluated. The generated path moves to the cells with the smallest accumulated or cost value. This process is repeated multiple times until the source and destination are connected. The completed path is the smallest sum of raster cell values between the two points, and it has the lowest cost. The Figure A3 shows an example of the mechanism of the least-cost path linking two places.

TABLE A.1 Least cost path determinants with weights given by engineering information

Type	Weight
Slope 0-3%	1
Slope 3-7%	2
Slope 7-12%	3
Slope 12-25%	4
Slope 25-50%	5
Slope 50-75%	6
Slope $> 75\%$	7
Bog	6
River	9
Slope 6	9
Protected Area	∞
Reservoir	∞
Wetland	∞
Swap	∞
Lagoon	∞

In practical terms, protected areas, reservoirs, wetlands, lagoons, and swamps will make the road deviate from surrounding these areas, increasing the construction cost. At the same time, rivers and moderately and strongly steep slopes are still an option but at a higher cost. The table shows the geographic features taken into account and their respective weights. We also use this least-cost path to build counterfactual market access for each municipality used as the source of exogenous variation for the market access.

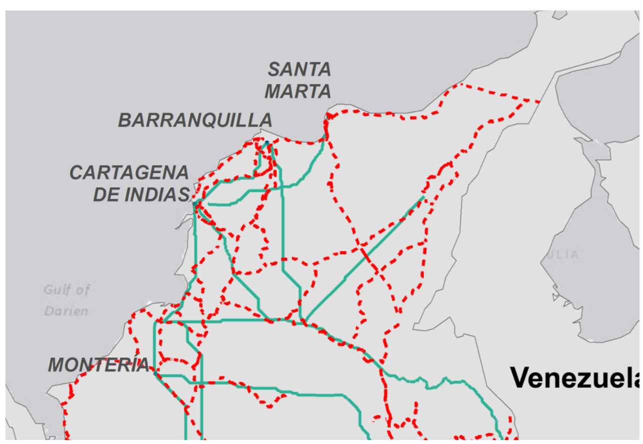
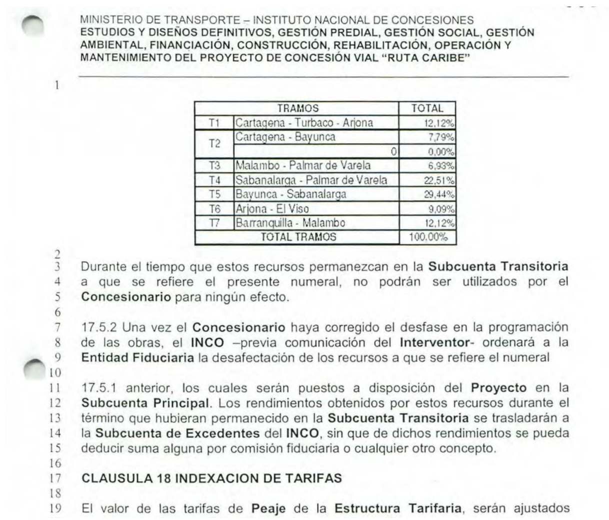


FIGURE A4 Zoom of the comparison between the least cost path (green) vs actual roads(doted red)

Source: Authors' calculations and INVIAS

FIGURE A3 Calculated least cost path connecting main cities
Source: Authors' calculations


MINISTERIO DE TRANSPORTE – INSTITUTO NACIONAL DE CONCESIONES
ESTUDIOS Y DISEÑOS DEFINITIVOS, GESTIÓN PREDIAL, GESTIÓN SOCIAL, GESTIÓN AMBIENTAL, FINANCIACIÓN, CONSTRUCCIÓN, REHABILITACIÓN, OPERACIÓN Y MANTENIMIENTO DEL PROYECTO DE CONCESIÓN VIAL "RUTA CARIBE"

	TRAMOS	TOTAL
T1	Cartagena - Turbaco - Arjona	12,12%
T2	Cartagena - Bayunca	7,79%
	0	0,00%
T3	Malambo - Palmar de Varela	6,93%
T4	Sabanalarga - Palmar de Varela	22,51%
T5	Bayunca - Sabanalarga	29,44%
T6	Arjona - El Viso	9,09%
T7	Barranquilla - Malambo	12,12%
	TOTAL TRAMOS	100,00%

Durante el tiempo que estos recursos permanezcan en la **Subcuenta Transitoria** a que se refiere el presente numeral, no podrán ser utilizados por el **Concesionario** para ningún efecto.

17.5.2 Una vez el **Concesionario** haya corregido el desfase en la programación de las obras, el **INCO** –previa comunicación del **Interventor**- ordenará a la **Entidad Fiduciaria** la desafectación de los recursos a que se refiere el numeral 17.5.1 anterior, los cuales serán puestos a disposición del **Proyecto** en la **Subcuenta Principal**. Los rendimientos obtenidos por estos recursos durante el término que hubieran permanecido en la **Subcuenta Transitoria** se trasladarán a la **Subcuenta de Excedentes** del **INCO**, sin que de dichos rendimientos se pueda deducir suma alguna por comisión fiduciaria o cualquier otro concepto.

CLAUSULA 18 INDEXACION DE TARIFAS

El valor de las tarifas de **Peaje** de la **Estructura Tarifaria**, serán ajustados

FIGURE A5 Sample of Colombian road contract

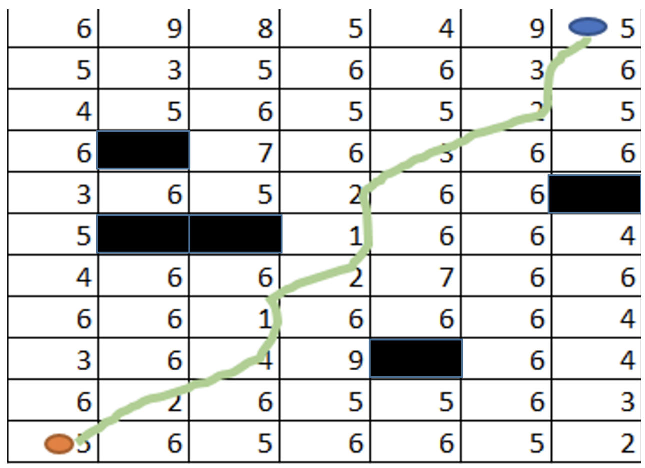


FIGURE A6 Least cost path mechanism
Source: Authors' calculations

C | INSTRUMENT BASED ON INDIGENOUS ROADS

This approach was first used by Duranton (2015) as an instrument for Colombian roads. Its validity relies on the fact that native indigenous were circulating across the country long before the Spanish colonization. It is expected not to have any influence on current GDP or inequality. However, given Colombian geography, these corridors for natives are a good predictor for actual roads. The Figure A7 shows the indigenous roads in the country. These roads belong to different indigenous tribes that were separated because of geography. These roads also include rivers and were not necessarily used by the Spanish after the conquer, giving additional exogeneity to the variable. We use the distance to indigenous roads as an instrument for current roads.

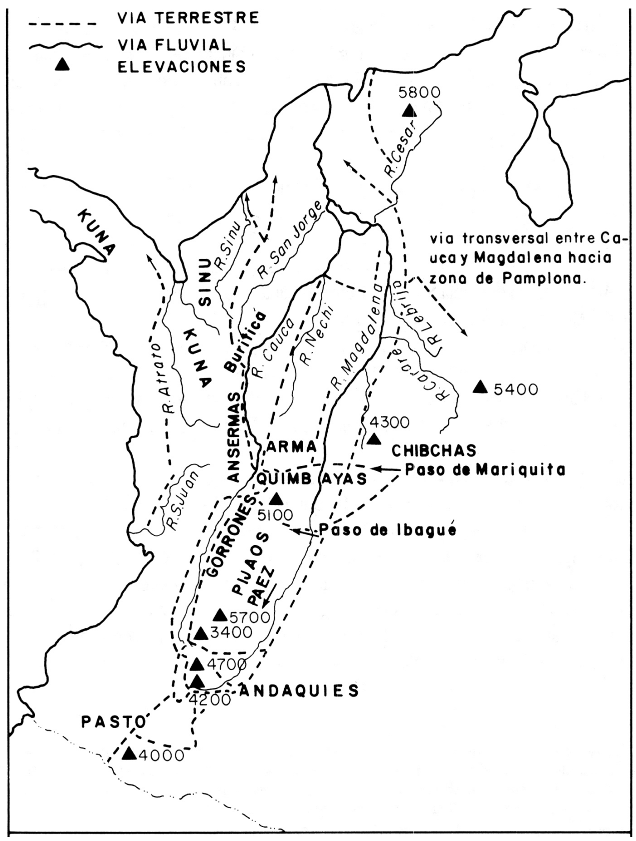


FIGURE A7 Pre-colonial (indigenous) roads in the country including river transportation
Source: Safford and Palacios (1998)

TABLE A.2 Elasticity of trade wrt distance, 2013

	Ln Trade		Ln Trade (no zeros)	
Ln Travel Time	-.34661	-	-.2926	-
SE	(.01413)	-	(.0024)	-
Observations	17121	-	987094	-
Ln Distance	-	-.33779	-	-.2737
SE	-	(.01472)	-	(.0024)
Observations	-	17121	-	987094
Origin & destination FE	Yes	Yes	Yes	Yes

Notes: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for all regression we use the GDP as dependent variable.

D | CALCULATION OF MARKET ACCESS DECAY PARAMETER

To calculate the decay parameter α in the denominator, which is the elasticity of trade to distance, we use the 2013 Commodity Flow Survey (CFS) for Colombia, from the Colombian Ministry of Transport, which measures bilateral trade flows among Colombian municipalities, following Duranton (2015) for Colombia and Duranton et al. (2014) for the US. Table A.2) shows the estimated elasticity.

E | INSTRUMENTS FOR DISTANCE TO THE INFRASTRUCTURE

Distance to the infrastructure is a common treatment used to measure the effects of infrastructure. However, this measure does not consider the effects that road improvements or new roads might have on places that are not close to the improvement because of network effects. This restriction makes the interpretation of distance results limited, but we calculate it to contrast our more believable estimate using market access. Given that distance to the infrastructure is endogenous, we use two instruments as a source of exogenous variation: The distance to the calculated least cost path network and the distance to the pre-colonial (indigenous) roads.

TABLE A.3 Distance descriptives

Source	Description	N	mean	sd	min	max	Years
Treatment Variables							
Invias and authors' calculations	Distance to the nearest intervention in km	12,318.00	32.51	49.94	0.00	804.91	1986-2019
Instrumental Variables							
Safford and Palacios (1998) and authors' calculations	Distance to indigenous roads in m	12,318.00	72,551.85	86,035.11	8.22	1,058,399.79	All
Authors' calculations, see Section B	Euclidean Distance to the least cost path in m	12,318.00	33,257.87	65,366.33	0.21	952,126.62	1984-2019

E.1 | Results on distance

In table 12 we show the regressions of the effects of distance to the infrastructure on GDP and inequality. In this case, the variables distance to the infrastructure included refers to bridges, ring roads, double lane, third lane, and new roads, which are considered main interventions[17]. It is clear that once we take municipal level fixed effects, total GDP increases with distance to transport infrastructure, giving a counterintuitive result. This, however, is consistent in all specifications even when we do not take into account big cities. The total GDP's elasticities to distance are 1.397 using the reduced sample. These estimates include only indigenous roads as an instrument, the size of the coefficient is slightly reduced if we use both instruments instead. Two main conclusions arise. First, the effects are important on total GDP and agricultural, service, and industrial GDPs. Second bigger cities can take more advantage of the benefits of the improved infrastructure. However, these results are not taking into account any network effects that might affect many cities in a more complex way. To measure network effects, there is our measure of market access.

Some works such as Baum-Snow et al. (2016) find similar counterintuitive results in China for some of the prefectures of study. In this study most important cities in each region seem to get all the benefits of infrastructure, while hinterland prefectures get little of negative results. Furthermore, Duranton (2016) found a negative relationship between wages and market access for Colombia, which he justifies as a consequence of a natural diversification and competition on the goods market that comes with higher market access that leads to lower prices and therefore lower wages.

E.2 | Calculation of spillover effects

Whith this specification we would like to explore the effect of different intensities of a treatment on total GDP. We separate the effect of municipalities are located on the road and all the close neighbors. To estimate spillover effects, we will estimate the following difference-in-difference type panel equation, following List et al. (2019).

$$Y_{it} = \beta * \text{treated}_{it} + \zeta * \mathtt{Neighbor35}_{it} + d_i + f_t + \varepsilon_{it} \tag{4}$$

Where $\mathtt{treated}$ are municipalities with direct contact with the new infrastructure and neighbors35 represent the rest of neighboring municipalities within a range of 35km from the infrastructure. Variables d_i and f_t represent location-specific and time-specific fixed effects respectively. Given that we keep the whole sample in this regressions, β will be the average effect of taking a municipality in the control group and putting it on the road.

E.2.1 | Results on Spillover effects

Table A.4 show that there are important spillover effects in most of our outcome variables. Regarding total GDP, services GDP and Industrial GDP to impact a control municipality with a road would increase its production substantially. The effect is not so clear for agricultural GDP which shows a negative effect but not significant. Also a road intervention would reduce income inequality, but would increase the concentration of property. Also in terms of the municipal development index it shows an important positive effect. It also shows an important effect of interventions on neighbors up to 35km from the intervention, in this case coefficients for industrial and services GDP are smaller in size but signs remain. Agricultural GDP, however shows a positive and significant effect of bringing a road to a control municipality compared with these group. Main conclusions of this section are that municipalities in the control group would have higher GDPs (industrial and services) if they are benefited with a road improvement, this effect is bigger the closer they are to the road. Also this benefit would imply better income distribution but generates a

[17] This specific measure does not include repavement which is a minor intervention. Results are robust when repaving is taken into account

higher concentration of land. Also municipalities the effect of roads on the municipal development index which takes into account poverty access to public services, education among other variables, is positive and significant in all specifications. As a robustness check of these effects we run a specifications in which we control by the average market access municipalities at 35km from each municipality, trying to understand geographic correlation of the results above. Surprisingly, coefficients remain the very similar in size and sign, but significance increased in all specifications reinforcing the importance of the role of roads on our output variables.

F | INSTITUTIONAL FRAMEWORK: ROAD NETWORK EXPANSION IN COLOMBIA

The country started an institutional change in 1990 to bring private investment to the public roads sector, following other countries such as Mexico and Chile and making the country more competitive for exports. The government of *Cesar Gaviria Trujillo (1990-1994)* established the legal and institutional bases for private participation in infrastructure using concession contracts that were first released in 1993. Due to the lack of experience of the public sector, the first generation of contracts suffered from a lack of planning, insufficient information, and had many re-negotiations. In order to bring private players to the infrastructure market, the government assumed almost all the risks for those contracts which were developed until 1998.

The next *Ernesto Samper (1994-1998)* continued with the structuring of the second generation of roads concessions released in 1996, including a better risk-sharing with the private sector and better planning in the environmental and technical areas[18]. These two generations focused around the capital city and the sugar cluster of the country, respectively. In 2004 the third generation started, which covered a wider area and offered better contractual terms of risk-sharing for the government.

Finally, in 2014 the fourth generation of roads concession started. For that generation, the government created independent agencies to develop the projects' technical, financial, and environmental dimensions.[19] Effects of this last wave of concessions will not be taken into account for this research given that some of those projects are still unfinished.

Between 1950 and 1959 there was a large development push with 4,600 km of roads improved (60%) or built (40%). From 1960 to 1974, the emphasis was put on tertiary and state roads, which increased 7% annually. Finally, between 1975 and 1994, more than 5,600km were paved while the tertiary and state roads were increased or maintained in an amount of 55,000km. *Instituto Nacional de Vias* (INVIAS) currently handles all national and state roads that are not under concession. However, by 1994 Colombia was still one of the less developed countries in terms of road infrastructure, when compared with similar countries and several studies point that as a barrier to economic growth. While the US had 24% of its roads paved, Colombia had only 0.24% (World Bank, 1994).

[18] The national department of planning of Colombia coordinates a national council for economic and social policy. The documents that are produced from this council are called *CONPES* which contain policy briefs for nationwide topics, including the presentation of the detected problem, the policy suggested and the guideline for the implementation and evaluation of the programs, this council includes members of different areas of the government. The first generation of concessions is ruled by the *CONPES* 2597 of 1992 starts *CONPES* 2775 of 1995 that clarifies the rules for private participation in roads concession and the third generation was established in the *CONPES* 3045 of 1998. Each document builds on learning the immediately previous experience, making the risk-sharing fairer for both parts.

[19] The Agencia Nacional de Infraestructura (ANI) and the Agencia Nacional de Licencias Ambientales (ANLA) were created in 2012 to expedite road building by providing investors with up-to-date studies in these areas. The government also was committed to prove bank warranties for the project to make the funding easier.

Figure 1 shows the location of the national road network and the different interventions done by INDIAS and during the concession waves. Red segments are concession roads; green segments are roads improved by INVIAS. The clear blue lines are the rest of the road network.

Instituto Nacional de Vias (INVIAS) currently handles all national and state roads that are not under concession. However, by 1994 Colombia was still one of the less developed countries in terms of road infrastructure, when compared with similar countries and several studies point that as a barrier to economic growth. While the US had 24% of its roads paved, Colombia had only 0.24% (World Bank, 1994).

G | ADDITIONAL RESULTS

TABLE A.4 Spillover effects of roads on nearby municipalities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	GDPLongPanel	TOTALGDP	AgriculturalGDP	ServicesGDP	IndustrialGDP	GINI	OwnersGINI	LandGINI	DevIndex
Control variables									
Treated	0.559*	0.804*	-0.026	1.021*	1.297*	-0.015*	0.019*	0.006	6.810*
	(0.042)	(0.032)	(0.034)	(0.038)	(0.042)	(0.004)	(0.004)	(0.003)	(0.453)
35km or	0.043	0.300*	0.282*	0.569*	0.194*	-0.028*	0.022*	0.006	4.699*
less	(0.037)	(0.028)	(0.029)	(0.033)	(0.036)	(0.003)	(0.003)	(0.003)	(0.389)
N	19074	10970	10967	10970	10970	2086	12938	12332	12066

Standard errors in parentheses
$p < 0.05$, $p < 0.01$, ** $p < 0.001$

Note: These regressions include municipality and year fixed effects.

TABLE A.5 Distance Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	GDPLongPanel	TotalGDP	AgriculturalGDP	ServicesGDP	IndustrialGDP	GINI	OwnersGINI	LandGINI	DevIndex
OLS: Distance									
Distance to	-0.041***	-0.076***	-0.037***	-0.099***	-0.190***	0.001	-0.007***	-0.008***	-0.029
intervention	(0.010)	(0.008)	(0.010)	(0.009)	(0.011)	(0.003)	(0.001)	(0.001)	(0.130)
N	11281	7147	7147	7147	7147	119	9467	8881	8174
IV: Distance to Indigenous roads									
Distance to	1.563***	1.397***	1.728***	1.613***	1.071***	-0.041	-0.051**	0.091*	8.555***
intervention	(0.297)	(0.164)	(0.197)	(0.191)	(0.157)	(0.042)	(0.016)	(0.043)	(1.473)
N	11281	7147	7147	7147	7147	119	9467	8881	8174
F 1st stage	40	97	97	97	97	1	37	9	91

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: These distance regressions are for the reduced sample, which means we exclude incidental cities. we include year and municipality fixed effects

TABLE A.6 Fade-out Inequality

	(1)	(2)	(3)	(4)	(5)
	20km	25km	30km	35km	40km
IV: GINI vs MA					
GINI	-0.109**	-0.113**	-0.111**	-0.112**	-0.110**
	(0.035)	(0.031)	(0.030)	(0.031)	(0.030)
IV: Land GINI vs MA					
Land GINI	0.065	0.041	0.044	0.021	0.022
	(0.053)	(0.047)	(0.050)	(0.056)	(0.060)
IV: Owners GINI vs MA					
Owners GINI	0.184**	0.171**	0.166*	0.153*	0.154*
	(0.062)	(0.057)	(0.063)	(0.066)	(0.069)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All the regressions include the controls we present in Table 3, they also include municipality and year fixed effects, for all regression we use the Market access and the Market access least cost path as instruments.

TABLE A.7 Taxes regressions

	(1)	(2)
	Taxes Income	Industrial taxes Income
OLS: Market Access		
Market Access	0.170***	0.234***
	(0.033)	(0.057)
N	10267	10177
IV: MA-Leastcostpath		
Market Access	0.670***	0.460***
	(0.059)	(0.099)
N	10091	10004
F 1st stage	4448	4401
IV: MA-Semi-Random Component		
Market Access	1.990***	3.519***
	(0.161)	(0.275)
N	10160	10073
F 1st stage	124809	122550
IV: MA-Semi-Random Component and Leastcostpath		
Market Access	2.337***	3.659**
	(0.592)	(1.144)
N	10017	9930
Sargan test p-value	.1734	.4449

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: These regressions are for the taxes income, we use the reduce sample.