World Development 127 (2020) 104739

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

The Brasília experiment: The heterogeneous impact of road access on spatial development in Brazil

Julia Bird [a], Stéphane Straub [b,*]

[a] Vivid Economics, United Kingdom
[b] Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

ARTICLE INFO

Article history:
Accepted 23 October 2019
Available online 14 November 2019

JEL classification:
O18
N76
N96
R40
R11
R12
F15

Keywords:
Infrastructure
Roads
Agglomeration externalities
Brazil

ABSTRACT

This paper studies the impact of the rapid expansion of the Brazilian road network, which occurred from the 1960s to the 2000s, on the growth and spatial allocation of population and economic activity across the country's municipalities. It addresses the problem of endogeneity in infrastructure location by using an original empirical strategy, based on the historical natural experiment constituted by the creation of the new federal capital city Brasília in 1960. It highlights long term center-periphery agglomeration effects and shows heterogeneous effects of roads depending on the characteristics of metropoles they lead to and on the location of the municipalities themselves, in line with predictions in terms of agglomeration economies.

1. Introduction

A number of recent studies using spatially disaggregated data within countries have shown that transport improvements are a major driver of changes in local development outcomes. Other papers have analyzed the spatial redistribution of activities across places, such as large metropolitan centers vs. small peripheral regions, and city centers vs. suburban areas (see Berg, Deichmann, Liu, & Selod, 2017; Redding & Turner, 2015, for recent reviews of these literatures).

The existing empirical knowledge is however still limited on how this spatial redistribution effects depend on key characteristics of the locations being connected. This paper addresses this issue, providing evidence of how the effect of new transport corridors on local outcomes varies according to the characteristics of the cities being connected and of the main end points on these corridors. It uses as a source of heterogeneity across places some of the main parameters highlighted by the urban and economic geography literature, namely market size, human capital endowment, quality of amenities, and industry specialization.

It does so using an "historical natural experiment" to study the impact of the rapid expansion of the Brazilian road network on the growth and spatial allocation of population and economic activity across the country's municipalities between 1970 and 2000. Our specific contribution is to use the radial nature of the new road system to shed light on the heterogeneity of the impact of a long-term reduction in transport cost, depending on the characteristics of urban centers roads are connecting to. The estimation of the causal impact of a change in road access, combined with the heterogeneity in the characteristics of the State capitals that each road connects to, provides us with a set of quasi-experimental mini-laboratories to test the effect of improved infrastructure on changes in population and production.

Brasília, Brazil's current federal capital city, was built from scratch between 1956 and 1960 in a previously unpopulated area. The location was selected because of its geographic centrality, at the initiative of then President Juscelino Kubitschek, who wanted to shift the country's center of gravity away from the Southern coastal region. The following decades were characterized by one of the largest post war infrastructure development

* Corresponding author.
E-mail address: stephane.straub@tse-fr.eu (S. Straub).

https://doi.org/10.1016/j.worlddev.2019.104739
0305-750X/

programs worldwide, during which Brazil paved over 150,000 km of roads.[1]

As this national road construction program connected the new capital to other main population and economic centers, the resulting radial highway system also incidentally reached other inland municipalities along the way. Proximity to the roads built after the creation of Brasília was a key factor in explaining the subsequent changes in local access to major economic centers. However, whether intermediate municipalities were close to or far from the new corridors was mostly due to luck rather than to their specific economic or geographic characteristics.

Our empirical strategy is based on superimposing onto a map of Brazil eight straight lines, coinciding with the subsequent shape of the new radial highway system which connects the country's new capital to State capitals and ports chosen according to their population size and economic importance in 1956, the year of the decision to build Brasília. We then create a municipality-level distance index capturing proximity to the lines and, excluding lines' end points, use it to instrument the subsequent municipality-level improvement in road access over time, assessing the impact of these improvements on local-level changes in population and GDP, as well as GDP per capita. As such, we rely on a combination of the inconsequential unit approach pioneered by Chandra and Thompson (2000) and the use of planned routes proxied by straight lines based on historical preconditions as instruments, as first introduced by Banerjee, Duflo, and Qian (2012).

This allows us to solve the main difficulty inherent to eliciting the impact of roads, namely their potential non random placement. Roads are likely to be allocated to specific locations according to observed or unobserved characteristics that are not orthogonal to their development potential. For example, they may be prioritized in fast growing municipalities or in those with suitable geographic characteristics, in which case their estimated impact would be upwardly biased. Alternatively, policymakers may want to cater to the needs of lagging regions, with opposite effects. Finally, examples of infrastructure works allocated for political reasons rather than economic rationales abound,[2] potentially biasing estimates towards zero.

Our empirical analysis exploits successive census data between 1970 and 2000, aggregated at the municipality level, together with a composite measure of the cost of access from each individual location to its State capital in each decade from the late 1960s to the 1990s.

We first show the absence of pre-1960 differences in levels and trends of the densities of population and GDP, and GDP per capita between municipalities that were subsequently close or far from the radial lines resulting from the construction of Brasília, which vindicates the first part of our identification strategy. On the other hand, the reduced form results show that proximity to these lines strongly predicts subsequent increases in population, GDP, and GDP per capita between 1970 and 2000.

Using this distance to the lines to instrument the change in access to State capitals in a 1970–2000 long difference framework shows that road access improvements translated to significant population and GDP increases. Consistently with the fact that the creation of Brasília led to new roads running through previously underdeveloped areas, the 2SLS estimates are larger than the OLS. Moreover, using a semi-parametric specification, we show that average effects are stronger around the main urban centers within 200 km of the State capitals.

Finally, we use several empirical strategies to document how these effects vary according to the characteristics of the urban centers to which the radial roads connect. As expected, effects appear to hold for very large urban centers, especially São Paulo, and are conditional on end points having good amenities, as captured by high levels of piped water access, high levels of human capital, and higher shares of non-agricultural activities. Similarly, effects are stronger for roads leading to cities with a high manufacturing to service ratio.

This paper pertains to the recent strand of literature mentioned above that tackles the issue of transportation infrastructure impacts using spatially disaggregated data. It has for example documented specific positive impacts of better infrastructure access on trade (Donaldson, 2018; Michaels, 2008), firms' location and efficiency (Datta, 2012; Ghani, Goswami, & Kerr, 2016; Rothenberg, 2013), income levels (Banerjee et al., 2012; Storeygard, 2016), urban growth (Duranton & Turner, 2012), population (Atack, Bateman, Haines, & Margo, 2010), structural transformation (Aggarwal, 2018; Asher & Novosad, Forthcoming; Qin & Zhang, 2016), land values (Donaldson & Hornbeck, 2016) and ecological impact (Damania, Russ, Wheeler, & Barra, 2018), among others. Most papers rely on the inconsequential unit approach pioneered by Chandra and Thompson (2000) or on the use of planned routes as instruments. Specifically, Banerjee et al. (2012) was the first paper to use straight lines based on historical preconditions to provide an exogenous measure of access to modern transportation corridors.

Our work also relates to a growing body of applied work that analyzes the impact of transportation investment on the changes in location patterns of agents and economic activity by integrating insights from economic geography models (Baum-Snow, 2007; Baum-Snow, Brandt, Henderson, Turner, & Zhang, 2017; Faber, 2014; Lall, Shalizi, & Deichmann, 2004; Roberts, Deichmann, Fingleton, & Shi, 2012).

Our specific contribution is to document the within-country heterogeneity of the effects of roads investments. The nature of the experiment we analyze provides us with two main levels of evidence. First, we find that average effects in terms of the transport cost elasticity of population and GDP are stronger near urban centers, and decline for locations that are located farther away. We conclude that this captures the fact that the road construction program has led large cities to 'soak' the hinterland through strong agglomeration effects.

The second main and novel dimension of heterogeneity relates to the characteristics of the main road corridors end points. We conclude that the spatial patterns that are generated by large scale transport infrastructure investments are strongly dependent on the type of cities that the roads connect to. We document specific characteristics of the end points, based on urban pull factors stressed in the literature, that shape such benefits. These include market access, industrial specialization, knowledge spillovers, and amenities.[3]

More broadly, our paper also relates to the literature that uses Brazil as a testing ground for the link between improvements in different types of infrastructure and economic outcomes, including Lipscomb, Mobarak, and Barham (2013) on electricity, Chein and Assuncão (2016) on roads, migration and labor markets, Da Mata, Deichmann, Henderson, Lall, and Wang (2007) on city growth, and Reis (2014), who describes the evolution of spatial inequality in Brazil over the long run (1872–2000). Closely related to ours is the paper by Morten and Oliveira (2018), which uses a version of Faber (2014) Euclidean Minimum Spanning Tree (EMST), which overlaps with our radial lines for the parts connecting Brasília to State capitals, to instrument the travel time from different

1 Mitchell (1995) and World Bank (2008). This figure excludes urban roads.

2 See for example Burgess, Jedwab, Miguel, Morjaria, and Padrí i Miquel (2014) and Cadot, Roller, and Stephan (2006).

3 The working paper version (Bird & Straub, 2014) also extensively documented the South-North heterogeneity in the results.

locations and analyze the impact of roads improvements on trade and migration.

The paper is structured as follows. Section 2 discusses the main insights from the urban and economic geography literatures to guide our empirical exercise. Section 3 details the state of Brazilian infrastructure since the 1960s and the relevant institutional facts. Section 4 presents the different sources of data used in the paper. Section 5 introduces the empirical strategy, showing pre-treatment tests and reduced form outcomes. Section 6 presents the main results and develops the heterogeneity tests, Section 7 presents additional robustness tests, and Section 8 concludes. Additional material, results, and robustness checks are provided in the Appendix.

2. The sources of agglomeration economies

We hypothesize that the reduction in transport costs due to the improvement of road infrastructure will spur the growth of cities for which there are strong agglomeration economies, that is cities which are large enough, have a high stock of human capital, a high share of non-agricultural activities, a high industry-to-service ratio, and those with good amenities.

These variables can be traced back to the urban and the economic geography literatures, surveyed among others in Duranton and Puga (2013), Redding and Turner (2015), Rosenthal and Strange (2004), which provide ample evidence on the channels through which agglomeration economies determine the strength of urban areas' pull factors.

The main sources of these agglomeration economies were introduced by Marshall (1920), who is often thought as favoring a localization approach to agglomeration, i.e., the idea that it is the concentration of specific industries and the induced specialization process that give rise to this phenomenon. In addition, the idea that urbanization, and its corollary of diversity, was key has also been considered early on. Empirically, localization was proxied by some measure of industry concentration, whereas urbanization was rather related to the sheer size of cities. In terms of microfoundations, Marshall referred to input sharing, labor market effects, and knowledge spillovers as driving increasing returns in urban areas. To these, the subsequent literature has added consumption, for example of local amenities such as infrastructure, and home market effects, in close relation to the urbanization concept mentioned above. In the last few decades, these channels have been both conceptualized theoretically, and documented empirically. Let us briefly discuss them and present the proxies that we are using in this work.

The Marshallian input sharing channel relies on the existence of scale economies in input production due to the concentration of specific activities, and has been documented for example in Holmes (1999). As for labor markets, contributions have considered that an increase in the probability of good job matches could arise both through the urbanization channel, with city size making such matches more likely, and localization, in which case the main factor is local specialization. In addition, this argument extends naturally to risk sharing (Krugman, 1991), which can again be related to either urbanization if it provides protection against specific industry shocks that are diversified away by the fact that a large array of activities coexist, or specialization, if it makes the probability of an alternative match in the same activity more likely. Given the data at hand, we focus on the share of non-agricultural activities and on the industry-to-service ratio as two alternative measures of industrial concentration and therefore of specialization.

The second channel is market access, going back to Harris (1954) market potential idea and Krugman (1980) home market effect, which relates to the concentration of industries in already large markets. Based on these seminal economic geography contributions, many subsequent empirical papers have analyzed empirically the extent to which large local markets attract production, firms, and population (See Redding & Turner, 2015, for a review). Our proxies of choice for the strength of agglomeration forces along the urbanization and market potential channel are the size of the city at the end point of our radial lines, measured either with population or GDP.

Third, knowledge spillovers are another key aspect of agglomeration forces, and the literature has mostly focused on the transmission of knowledge through workers. As surveyed in Moretti (2004), proxies used in the literature have revolved around some measure of human capital, for example the average education level of the workforce, or the share of college graduates. Here, we rely on a simple average schooling measure of the population at end points.

The final channel considered in the literature goes through the consumption and amenity benefits from living in large cities. These include the existence and quality of public goods, the exclusive availability of specific goods and services, and social interactions. Consistent with previous research showing that some Brazilian localities may have voluntarily withheld water provision to poor neighborhoods as a way to deter in-migration (Feler & Henderson, 2011), we use average water coverage as a measure of a key amenity potentially favoring agglomeration.

To summarize, in our empirical exercise, we use the variables defined above -city size as measured by GDP or population, average schooling level, average water connection levels, and share of non-agricultural GDP and industry-to-service ratio- to test the strength of agglomeration effects across main urban centers in the Brazilian case.[4]

3. Brazilian infrastructure

Brazil is South America's first, and the world's fifth largest country, both by geographical area (over 8.5 million km^2) and by population (close to 200 million). As of 2008, it had just over 1.7 million kilometers of roads, around 10 km per thousand inhabitants, of which only 12% were paved and close to one third concentrated in the Southeast Region. The road sector, especially the highway system, has historically been the primary internal mode of transport for both freight and passengers in Brazil. According to computations by Castro (2004), as of 1999, truck transport by road represented 82.1% of domestic freight output, and 93.6% of related expenses. Over 60% of cargo was transported by road in 2011.[5]

Between 1952, which corresponds to the earliest available aggregate paved road data, and 2000, there was a 471% increase in total road length. In the same period, GDP grew by 883% and GDP per capita by 233%.[6] While in the 1950s, most new connections were between State capitals along the Atlantic coast, from the 1960s new penetration corridors started linking the hinterland main urban centers, e.g., connecting Brasília to São Paulo, Belo Horizonte or Belém.[7]

[4] An equivalent way of thinking about agglomeration in urban centers is to consider the wage-rental price ratio. A higher ratio in the center would correspond to a relative opportunity cost of factors characterized by a higher productivity of labor. The determinants would be similar to those discussed above (See Bird & Straub, 2014, for a formalization along this line).

[5] Seehttp://www.brasil.gov.br/sobre/tourism/infrastructure/roads, Revista CNT No.206 novembro 2012.

[6] This development of the road network was accompanied by a surge in the number of vehicles available, which went from around 6 vehicles per thousand inhabitants in 1945, to 37 in 1970, then more than doubled to 84 in the 1970–1980 decade, reaching 135 in 2000 and 219 in 2011 (Mitchell, 1995), and Ipea data (Ipea's database on Brazilian macroeconomic, regional and social statistics. Brasília).

[7] Castro (2004) and World Bank (2008).

Concomitantly, there was a rapid expansion of the agricultural frontier towards the center-west part of the country, and an increase in the output share of the three less developed macroregions (North, Northeast and Center-west), which went from 17.3% in 1975 to 24% in 1996.

While Brazil remains a quite centralized country,[8] the country's extension and geographical dispersion implies that for municipalities in regions distant several thousands of kilometers from the country's economic core (the States of Minas Gerais, São Paulo and Rio de Janeiro), access to the local State capital is the main issue. We therefore use changes in the cost of access to the local State capitals as our main explanatory variable.

4. Data

4.1. Census data

Brazil is divided into 5 regions, containing 26 states as well as the federal district of Brasília, which in turn contain (in 2010) 5,564 municipalities. Our analysis focuses on the impacts of road access at the municipality level, the smallest level of government and administration within Brazil. There was a large increase in the number of municipalities over the last 50 years: between 1960 and 2010 their number has increased from 2,767 to today's 5,564. To ensure that the geographical focus of our data is consistent over time, we therefore use Minimal Comparable Areas (MCAs), a geographical division of Brazil created by the Institute of Applied Economic Research (IPEA), which boundaries do not change over time.[9]

From the decennial national censuses between 1970 and 2000 (from IBGE, the Brazilian Institute of Geography and Statistics), we extracted economic and social data at the MCA, state, and regional level. This data includes in particular local GDP measures, aggregated and by main sectors,[10] and population figures.

In addition, we use geographical data from IBGE's 1998 Brazilian CIM map (International map of the world at the millionth scale) which was digitized in 2003. This map provides detailed geological and geographical coverage of Brazil, as well the locations of cities and smaller population centers, road infrastructure and ports. From this we were able to locate the major economic centers of 1956, and construct lines from them leading to Brasília. By imposing the geographical boundaries of our MCAs we could then construct an index to measure how close each MCA is to these lines. More detail is given in Section 4.3. In addition, we constructed various indicators such as distance from the coastline, area of MCA, direct distance to the state capital, and percentage of land suitable for development (i.e., not subject to severe flooding, covered by the Amazon, etc.), and included data on municipalities' elevation and slope.[11]

4.2. Road data

The cost of access measure is provided by IPEA. This measure was computed by Newton de Castro (2002) for every municipality in Brazil, and summarize the cost of travel, in terms of quality adjusted kilometers to travel, to the State capital respectively for 1968, 1980, and 1995.

Castro (2002) first identified main traffic nodes across Brazilian municipalities, and for each of these nodes and each of the three dates, the shortest route to the State Capital and São Paulo, with the connecting roads, and their quality. The distances between each node were then calculated, with unpaved roads weighted at 1.5 times that of paved roads due to the increased time of travel, and waterways weighted at 10 times the cost of paved roads.[12] A similar measure of cost of access to São Paulo also exists. See Section 7 for robustness checks using this variable.

This measures provides us with a detailed mapping of the costs of access to State capitals, and how they change over time. These costs are kilometer equivalent, and therefore give us a clear spatial understanding of what they mean in terms of actual distances.[13]

4.3. Distance to the lines

Brasília is located in the Central-West region of Brazil, on the Planalto Central plateau. Following the inauguration of the city, it became necessary to connect it by road to other major cities of the day. The radial highway system, composed of federal highways BR-010 to 080, was built after 1960, either through entire new routes or the radical improvement of existing ones (see Fig. 1).

Note that the radial highways network was of course not immediately fully completed. Due to higher level of development in the Southern part of the country before 1960, radial highways connecting cities in this part of the country were build earlier and were of higher quality. However, connections to cities in the North-West and the North, including Porto Velho, Belém, Salvador, and Fortaleza quickly caught up in the 1980s and 1990s, with only some small interruptions, as can be seen in Figs. 2 and 3.[14] Finally, in the main test we do not include Recife and Porto Alegre, two of the largest cities at the time, as primary end point, as their location implies that any connection to them would go through Salvador and São Paulo respectively.[15] However, in Section 7 below, we perform tests with an alternative axial network, which includes these cities.

In linking Brasília to these cities, the radial network established corridors, which incidentally connected other urban centers along the way. For example, the BR-010, Belém-Brasília Highway, partially built between 1958 and 1960, was the first to connect the Federal District and the State of Goiás, in the center of the country, to the State of Pará in the middle north region. In doing so, it also crossed the States of Tocantins and of Maranhão, connecting local urban centers along the way, while other municipalities were located farther away from the road corridor. However, these differences in distance from the

[8] The Southeast region still represents around 60% of overall GDP, and as of the early 2000s the port of Santos, in the State of São Paulo, accounted for 38% of all import and export activity going through Brazilian ports, serving 13 States almost exclusively and part of the commerce of all 27 States, and moving close to 6.5% of the country's GDP (World Bank, 2008).

[9] IPEA is a government-led research organization linked to the ministry of the economy of the Republic of the Brazil. MCAs aggregate municipalities into the smallest possible boundary-invariant groupings. In this paper, we use MCA 70-00, which covers 3,660 areas, allowing us to compare data at any point between 1970 and 2000. In what follows, we use the terms municipalities and MCAs indistinctly to mean MCA 70-00, unless specified otherwise. We also use an alternative grouping, MCA 40-00, to conduct pre-treatment tests. See Reis, Pimentel, Alvarenga, and Santos (2011) for details on MCAs construction.

[10] A detailed description of how the municipality-level production data was constructed can be found in the Appendix.

[11] We used the openware software Quantum GIS to analyze our spatial data. Following our regressions, data could be re-inputted into QGIS to spatially represent our results. Quantum GIS is an official project of the Open Source Geospatial Foundation (OSGeo) and is licensed under the GNU General Public License.

[12] If multiple routes lay within one municipality, Castro took the average of the travel costs from these nodes as the cost of access measure. If the municipality contained no nodes, he took the travel cost from the node of a neighboring municipality, adding the expected distance from this node weighted by 2 to represent the likely poor quality of any connection.

[13] This measure stops short of being a complete market access measures, as the equivalent weighted distance is not available for all pairs of municipalities.

[14] The maps are based on the DeLorme geometry with imputed quality information coming from Quatro Rodas digitized maps. We thank Matheus Bueno for help with these maps.

[15] This is also the outcome of Morten and Oliveira's (2018) Euclidean Minimum Spanning Tree.

Fig. 1. Radial Roads connecting Brasilia. Source: Authors' elaboration based on map from Ministério dos Transportes, Brazil.

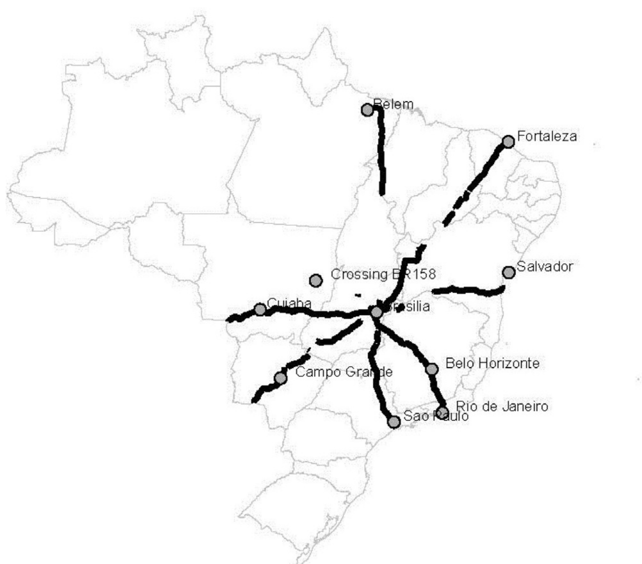


Fig. 2. Radial Roads in 1988. Source: Authors' elaboration based on DeLorme geometry and 1988 Quatro Rodas map.

roads were unrelated to their other economic or geographic characteristics, as placebo tests below confirm.

We capture these differences in proximity to the corridors, by computing for each MCA a distance index to the closest hypothetical lines linking Brasília to a set of 8 major Brazilian cities. To choose the cities, we list the State capitals of Brazil and major towns in order of population in 1956, and select the major centers of the day.[16]

We create successive buffer zones at 10 km intervals around the lines (see Fig. 4), and measure the percentage of each MCA's area within each zone. From this, we compute the weighted sum of the shares of an MCA's area lying in each successive range (see

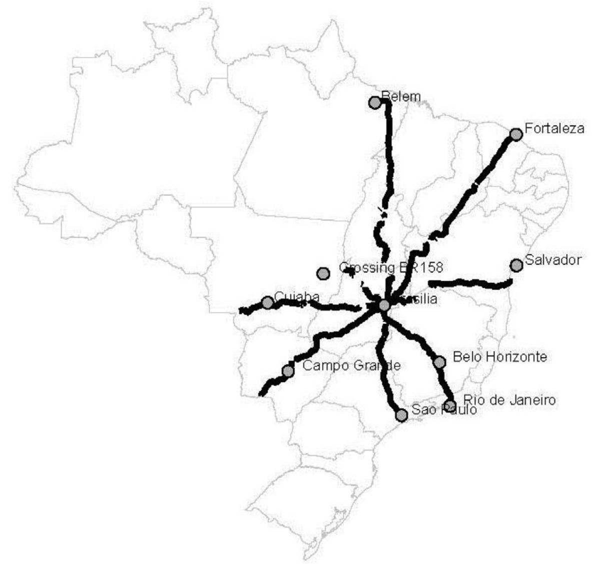


Fig. 3. Radial Roads in 2001. Source: Authors' elaboration based on DeLorme geometry and 2001 Quatro Rodas map.

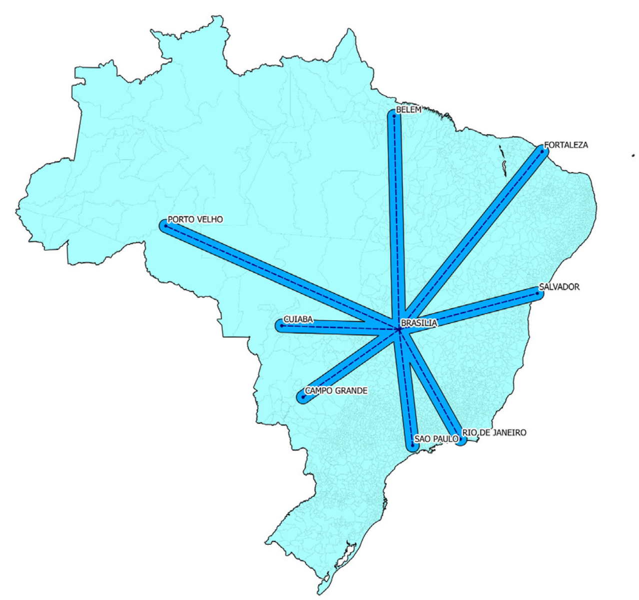


Fig. 4. Buffer Zones. Source: Authors' elaboration.

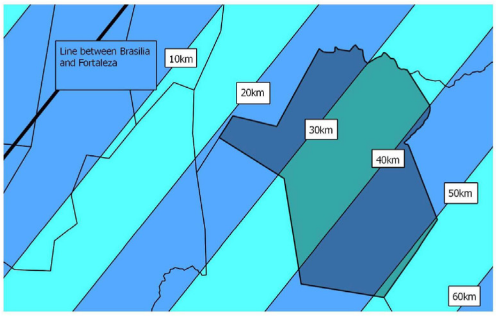
Fig. 5), and take the log.[17] Formally, our distance measure is given by:

$$D_{is} = \sum_j A_{is,j}.d_j, \tag{1}$$

where $A_{is,j}$ is the area of municipality i in State s that falls in buffer j, with $j = \{0-10, 10-20, \ldots\}$ and d_i is the average distance of buffer j from the line ($j = \{5, 10, 15, \ldots\}$).

[16] As explained above, we removed Recife and Porto Alegre from the list.

[17] More specifically, if 20% of an MCA is within 10 km of a line, 40% between 10 and 20 km and 40% between 20 and 30 km, we would calculate $0.2 \times 10 + 0.4. \times 20 + 0.4 \times 30 = 22$ and then take the log. We compute this measure taking into account the distance from all lines, and separately, the distance from the nearest line by constructing the index for all lines independently and taking the smallest value. The latter has the advantage of enabling us to differentiate between lines, and hence connections, by using lines-specific dummies or interactions in our estimations.The two are highly correlated at 0.97. Note also that our distance measure has a 0.99 correlation with the one based on distance to municipalities' centroïds.



MCA: 22 AMC7097 037, with bands around straight lines displayed, allowing the calculation of the area of the MCA within each band.

	Area km^2	Percentage %
0-10km	0	0
10-20km	0	0
20-30km	118	20.8
30-40km	246.1	43.4
40-50km	192.9	34.0
50-60km	0.6	1.8
Total Area	567.6	

MCA: 22 AMC7097 037

$$\text{Index} = (5 \text{ x } 0) + (15 \text{ x } 0) + (25 \text{ x } .208) + (35 \text{ x } .434) + (45 \text{ x } .340) + (55 \text{ x } .018) = 36.68$$

Fig. 5. Construction of distance from the lines.

Table 1
Summary Statistics.

1970	Mean	Std. Dev.	Min.	Max.	No. Obs.
GDP	54273	266546	216	7024682	3644
Population	21762	50052	830	1235030	3644
Cost of access to São Paulo	2521.59	1763.82	25	15343.58	3644
Cost of access to State Capital	650.93	480.87	0	5949.00	3644
Distance from Brasilia	1020.98	424.86	49.538	2843.916	3644
Distance from State Capital	241.10	157.66	0	1365.74	3644
Distance to São Paulo	1090.09	766.91	6.40	3563.88	3644
Area	2091.10	12599.54	8	367284	3644
GDP/cap	1.59	2.065	0.046	51.756	3644
2000	Mean	Std. Dev.	Min.	Max.	No. Obs.
GDP	232331	963631	2464	16510641	3644
Population	38939	107753	795	2238526	3644
Cost of access to São Paulo	1469.86	1106.39	10	10511.91	3644
Cost of access to State Capital	417.53	359.50	0	5949.00	3644
Distance from Brasilia	1020.98	424.86	49.53	2843.91	3644
Distance from State Capital	241.10	157.66	0	1365.74	3644
Distance to São Paulo	1090.09	766.91	6.40	3563.88	3644
Area	2105.082	12685.80	3.6	359750.906	3644
GDP/cap	4.066	4.854	0.641	123.701	3644

Descriptive statistics for the main variables used in the empirical analysis. All variables are observed at the municipality (MCA) and year level. We use census data from 1970 and 2000. GDP and GDP per capita are in constant 2000 Real. Cost of Access is in effective, quality-adjusted kilometers. Distances are in km. Area is measured in squared km.

Once we take the log of this index, we find that the distance from the lines has a bell shaped distribution, slightly skewed towards those more distant from the lines.

Table 1 outlines the main variables used in the analysis.

5. Empirical strategy

5.1. Pre-treatment tests

First, we present placebo estimations of pre-treatment differences in levels and trends between municipalities at different distance from the hypothetical lines that correspond to the radial highway network to be built over the future periods. The geographical unit used in this section is IPEA's 1940 Minimal Comparable Areas (MCA 40-00), a grouping of municipalities into 1,275 areas which are comparable at any point between 1940 and 2000. We estimate:

$$\log(Y_{is}) = \alpha_0 + \alpha_1 \log(D_{is}) + X_{is}'\alpha_2 + \theta_s + \varepsilon_{is}, \tag{2}$$

where D_{is} is the distance of municipality i in State s to the subsequent lines, θ_s is a State fixed effect, and X_{is} includes a set of controls for MCA's initial conditions and fixed characteristics: the 1970 initial values of the dependent variables (GDP, per capita GDP, population), the crow flies distances and squared distances to Brasília and the State capital, and geographical controls, including dummies for whether part of the MCA is in the amazon, whether it is within 50 km of the coast, is coastal, its average slope, and its elevation.

Specifically, we present two sets of results. In Table 2, panel A, we first show estimations in which Y_{is} are the 1950 levels of population, GDP, and GDP per capita. Panel B then uses the changes in these variables between 1950 and 1960.

To account for the fact that MCAs vary by size in a systematic way, being smaller in the South where the lines are also more dense, we normalize population and GDP using area.[18] In addition, we also control for area to account for the possibility that smaller/larger places may have other systematic differences such as quality of governance.

The results in Table 2 show that levels and trends of population and GDP, and GDP per capita are not correlated to municipalities'

[18] When looking at the change, the area normalization is just a fixed effect that is differenced out.

Table 2
Pre-treatment Tests.

	(1) Log Pop 1950	(2) Log GDP 1950	(3) Log GDP pc 1950
Log Distance Line	−0.0478	−0.0787*	−0.0213
	(0.031)	(0.046)	(0.022)
Observations	1249	1249	1249
R^2	0.644	0.640	0.620
	(1) ΔLog Pop 1950–60	(2) ΔLog GDP 1950–60	(3) ΔLog GDP pc 1950–60
Log Distance Line	−0.00417	−0.0225	−0.0183
	(0.008)	(0.020)	(0.017)
Observations	1249	1249	1249
R^2	0.114	0.145	0.125

Pre-treatment tests on population, GDP, and GDP per capita. Panel 1 reports regressions on levels, normalized by area for population and GDP, to account for the fact that 1940 MCAs are much smaller in the South where lines are more dense. Panel 2 report regressions on 1950–1960 changes. Controls include state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, its average slope, and elevation. The geographical unit used is IPEA's 1940 Minimal Comparable Areas (MCA 40-00). Robust standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 3
Reduced form estimations – 1970–2000 changes.

	(1) ΔLog Pop	(2) ΔLog GDP	(3) ΔLog GDP pc
Log Distance from Lines	−0.0725***	−0.0624***	−0.0481***
	(0.012)	(0.017)	(0.011)
	[0.015]	[0.643]	[0.480]
Log Population 1970	0.0761***		
	(0.009)		
	[0.009]		
Log GDP 1970		−0.0912***	
		(0.010)	
		[0.011]	
Log GDP per capita 1970			−0.608***
			(0.020)
			[0.021]
Observations	3644	3644	3644
R^2	0.332	0.195	0.507

Reduced form estimations, with differences in log Population, log GDP and log GDP per capita between 1970 and 2000 regressed directly on the distance from the straight lines. Controls include initial (1970) value of the dependent variable, state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, in the amazon, its average slope, and elevation. The geographical unit used is IPEA's 1970 Minimal Comparable Areas (MCA 70–00). Robust standard errors are in parentheses. Robust 'Conley' standard errors, allowing for spatial dependence up to twice the median distance with neighboring municipalities, in brackets. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

position relative to the subsequent placement of the lines corresponding to the future radial highway network.[19] This strengthens confidence in the validity of the inconsequential unit approach part of identification strategy. We discuss below the second part, which relies on distance from the lines as an instrument for the cost of access.

The results in this section are consistent with the view that until Kubitschek's presidency the idea of moving the capital away from Rio de Janeiro was never given serious consideration by Brazilian politicians (see Smith, 2002). This project can be traced back to José Bonifacio, advisor to Emperor Pedro I, who suggested in 1827 that moving the capital away from the Southeast Region to a more central location and coined the name Brasília. It was then formally written in the 1891 Constitution of the Brazilian Republic. However, it was only in 1955 that the Commission for the New Federal Capital chose the definitive location for Brasília, and it was Kubitschek's urge to see the city built, which led to its completion in three and a half years.

5.2. Reduced form

Our objective is to estimate the long-term effect of improvements in road access on our outcome variables at the local (MCA) level. Consider the following reduced form long difference model:

$$\Delta \log(Y_{is}) = \alpha_0 + \alpha_1 \log(D_{is}) + X'_{is}\alpha_2 + \theta_s + \varepsilon_{is}, \tag{3}$$

where $\Delta \log(Y_{is})$ is the change in the log of the outcome of interest in MCA i and State s between 1970 and 2000, i.e., its growth rate, again estimated as a function of log distance to the lines $\log(D_{is})$ and a set of controls.

Results for this specification are in Table 3. Over the period 1970–2000, municipalities closer to the lines experienced increases in population, GDP and GDP per capita relative to their more distant counterparts (column 1). The respective elasticities are 0.073, 0.062, and 0.048 and are statistically significant at the 1% level.

Their magnitude can be benchmarked (for per capita GDP only) to those in Banerjee et al. (2012), who use a similar specification for China over the period 1986–2003. In our data, comparing the 25th- to the 75th-percentile MCA in terms of distance shows that the latter is 4.2 times further away from a line. The corresponding gaps in population, GDP and GDP per capita are 23.2%, 20.0%, and 11.3% respectively.[20] By comparison, between 1970 and 2000, the total increase in these variables was 64%, 287%, and 136%.[21] The corresponding ratio for per capita GDP in China is 7.9% (Banerjee et al., 2012), hence slightly smaller but in the same order of magnitude.

For population, the differences stemming from the distance to the lines represent over one third of the change over the 1970–2000 period, while for GDP and per capita GDP, the same ratio is 7% and 11.3%. These preliminary results therefore indicate that population movements were a major change attributable to the construction of the radial highway system in Brazil. They also show that the distance to the lines mattered for subsequent outcomes.

5.3. Instrumental strategy: long differences

Eq. (3) is the reduced form of a two-stage strategy using distance to the lines as an instrumental variable to address the potential correlation between the independent variable of interest R_{is}, the (change in) the cost of access to the State capital of MCA i in State s, and the error term related to the non-random placement of roads ($Cov(R, \varepsilon) \neq 0$). Note that to ensure the basic validity of our instrumental strategy, we systematically exclude end points from the sample, as these are obviously not randomly assigned. The results discussed in the rest of the paper are also robust to excluding larger areas of 50 and 100 km around these end points.

Our main second stage equation is given by:

$$\Delta \log(Y_{is}) = \beta_0 + \beta_1 \Delta \log(R_{is}) + X'_{is}\beta_3 + \theta_s + \varepsilon_{is} \tag{4}$$

[19] We also display standard errors adjusted for potential spatial dependence, following Conley (1999).

[20] 0.0725*3.2 = 23.2, 0.0624*3.2 = 20.0, and 0.048*3.2 = 11.3.

[21] The annual growth rates were 2%, 5.1% and 3% for population, GDP and GDP per capita respectively. These rates differ slightly from official rates, as our sample excludes the line end points and a few remote MCAs.

In addition, we want to flexibly account for the potential non-linearities in terms of distance to the main urban centers that are typically expected in economic geography models.[22]

We do this in two ways. First, we interact our cost of access variable $\Delta\log(R_{is})$ with a vector B of dummy variables (b_1 to b_4) taking value 1 if the effective distance of the municipality (i.e., its cost of access) is in the range 0–100, 100–200, 200–500, and more than 500 km respectively.[23] The corresponding equation is:

$$\Delta\log(Y_{is}) = \beta_0 + (\Delta\log(R_{is}) * B)'\beta_1 + X'_{is}\beta_3 + \theta_s + \varepsilon_{is}. \quad (5)$$

Second, in the results section below, we display graphical results from a semiparametric specification that allows for non-linear effects depending on the initial cost of access.

We start by showing first stage regressions for our parametric specification and discuss identification in more details.

5.4. First stage

The corresponding first stage equations are given by:

$$\Delta\log(R_{is}) = \beta_4 + \beta_5\log(D_{is}) + X'_{is}\beta_6 + \theta_s + \varepsilon_{is}, \quad (6)$$

and, for all $j = 1, \ldots, 4$:

$$\Delta\log(R_{is}) * d_j = \beta_4 + (\log(D_{is}) * B)'\beta_5 + X'_{is}\beta_6 + \theta_s + \varepsilon_{is}. \quad (7)$$

The first stage regressions in Table 4 show that the instrument is a strong predictor of the long-term changes in the MCA-level cost of access to the State capital. In columns 1–3, we show simple estimates of the change in access as a function of distance to the line, corresponding respectively to the specifications for population, GDP, and GDP per capita. The more negative the left-hand variable, the more the cost of access decreases, i.e., the municipality gets better connected. The point estimates are positive and strongly significant, lying between 0.021 and 0.024, meaning indeed that places closer to the lines have experienced greater improvements in access to State capitals: a 10% decrease in distance to the line translates into up to a 0.24% improvement in access. Given the 320% gap between the 25th and the 75th distance percentile, the former therefore has a 7–8% greater improvement in access, around one fourth of the country-level average improvement (33%) over the entire period.

In Appendix Table A1, we show the first stage estimations for the MCA-level cost access to the State capital interacted with the four intervals b1–b4.[24] Again, the results are as expected and strongly significant throughout. In each column, the distance interaction corresponding to the dependent variable (i.e., distance*b1 for MCA-level cost of access*b1, etc.) is positive, indicating that within each interval, places closer to the lines experience bigger access improvements. The F-tests are satisfactory and support the validity of the instruments. Note that the negative signs of the other interactions that need to be included, for example in column 2 distance*b2 to distance*b4 are an artifact of the dummy interactions meaning they are equal to zero when the dependent variable is not, and conversely. Interestingly, the main effect gets weaker at larger distances from the lines, from 0.11 up to 200 km, to 0.07 between 200 and 500 km, and 0.02 beyond that.

[22] See for example Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2003) and Combes, Mayer, and Thisse (2008) for textbook treatment.

[23] In terms of distance to their respective State capitals (which are not all endpoints of our lines of choice), these bins correspond to average distances of 40, 70, 155, and 325 km respectively.

[24] These are the first-stage regressions for the population second stage. Those for GDP and GDP per capita are very similar, as the only difference are the controls for the 1970 GDP or GDP per capita values replacing the population one, and are not shown here to save space.

Table 4
First Stage.

	(1) ΔAccess 1970–2000	(2) ΔAccess 1970–2000	(3) ΔAccess 1970–2000
Log Distance from Lines	0.0240***	0.0210***	0.0212***
	(0.006)	(0.006)	(0.006)
Log Population 1970	-0.0382***		
	(0.004)		
Log GDP 1970		-0.0270***	
		(0.003)	
Log GDP per capita 1970			-0.0200***
			(0.007)
Observations	3644	3644	3644
R^2	0.387	0.385	0.373
FS F-stat excluded instruments (p-value)	17.88 (0)	13.49 (0)	13.13 (0)

First stage regressions, with changes in log cost of access to the State capital between 1970 and 2000 regressed on the log distance from the straight lines, for the second stage estimation of population (column 1), GDP (column 2) and GDP per capita (column 3). We use the hyperbolic inverse sine transformation instead of the log for the cost of access transformation, to account for a few null values. Controls include initial (1970) value of the dependent variable, state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, in the amazon, its average slope, and elevation. The geographical unit used is IPEA's 1970 Minimal Comparable Areas (MCA 70-00). Robust standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Identification relies on the fact that municipalities' improvements in their road access to major economic centers over the period of interest depend on their distance to the constructed corridors. Moreover, the excludability condition also requires that the distance to the lines affects the outcomes $\Delta\log(Y_{is})$ only through its impact on the change in the cost of access (i.e., only through road access), conditional on the controls, which include State fixed effects, and MCA-level time invariant aspects X_{is}, including controls for 1970 initial values, and a rich set of geographical controls including area, whether the municipality is coastal, whether it is in the amazon region, its average slope, and its elevation, and a set of distance measures (to Brasília and the State capital) as well as their square, already included in the reduced form. These distance controls are important as they may be correlated with distance to the lines.[25]

As discussed also in Morten and Oliveira (2018), the historical evidence indicates that the location of roads connecting Brasilia was not chosen for economic reasons or were meant to upgrade previously existing corridors that were used for internal trade for example. One specific concern is that these roads could have been built to facilitate trade in agricultural products from the central region. However, we show in Section 7 below that proximity to these roads had no impact on future agricultural GDP growth. We also note, as in Faber (2014), that our estimates are barely affected by the inclusion of initial controls for population or agricultural GDP. Finally, we take a cautious approach and show that all our results hold qualitatively when performing reduced form estimation using only the inconsequential unit approach.

6. Results

6.1. Average effects

We start by looking at the average effect for all Brazil. Table 5, panel A, shows the results from estimating (5) on the whole sample of Brazilian municipalities, where the left-hand side variable is

[25] Certain controls cannot be used simultaneously for or multicollinearity problems. Distances to the federal capital and São Paulo have a correlation of 0.71. Initial levels of infrastructure access are very strongly correlated between them and with initial GDP and GDP per capita.

Table 5
Second stage – 1970–2000 changes.

	(1) Δ Log Pop OLS	(2) Δ Log Pop 2SLS	(3) Δ Log GDP OLS	(4) Δ Log GDP 2SLS	(5) Δ Log GDP pc OLS	(6) Δ Log GDP pc 2SLS
Δ Log Cost Access	−0.0446	−2.850***	−0.111**	−2.775***	−0.0326	−2.197***
	(0.040)	(0.841)	(0.054)	(1.070)	(0.033)	(0.753)
Log Population 1970	0.0577***	−0.0488				
	(0.009)	(0.035)				
Log GDP 1970			−0.116***	−0.190***		
			(0.011)	(0.033)		
Log GDP per capita 1970					−0.626***	−0.677***
					(0.020)	(0.031)
R^2	0.339		0.213		0.517	
Δ Log Cost Access*0–100 km	−0.329**	−3.082***	−0.485**	−2.817***	−0.0382	−1.185**
	(0.153)	(0.873)	(0.202)	(0.890)	(0.133)	(0.496)
Δ Log Cost Access*100–200 km	−0.0928	−2.816***	−0.0509	−2.382***	0.0613	−0.956**
	(0.095)	(0.596)	(0.134)	(0.660)	(0.088)	(0.404)
Δ Log Cost Access*200–500 km	−0.131***	−1.942***	−0.132*	−1.053*	−0.0267	−1.024***
	(0.049)	(0.450)	(0.068)	(0.546)	(0.040)	(0.376)
Δ Log Cost Access*>500 km	0.0240	−2.508***	0.00196	−1.756***	−0.00549	−1.081**
	(0.051)	(0.471)	(0.073)	(0.601)	(0.046)	(0.429)
Log Population 1970	0.0545***	−0.0263				
	(0.009)	(0.020)				
Log GDP 1970			−0.115***	−0.156***		
			(0.011)	(0.019)		
Log GDP per capita 1970					−0.630***	−0.657***
					(0.020)	(0.024)
R^2	0.346		0.214		0.512	
Observations	3644	3644	3644	3644	3644	3644
Test of equality of interactive coeffs		0.16		0.01		0.96

Second stage estimations, with differences in log Population, log GDP and log GDP per capita between 1970 and 2000 regressed on the change in log cost of access to the state capital. We use the hyperbolic inverse sine transformation instead of the log for the cost of access transformation, to account for a few null values. Columns 1, 3 an 5 report the OLS estimations, and columns 2, 4 and 6 report the 2SLS estimations, using the first stage in Tables 4 and A1. Panel 1 provides the results for all municipalities, and Panel 2 shows the regressions using distance bins, with the cost of access interacted with four intervals of distance from the state capital. Controls include initial (1970) value of the dependent variable, state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, in the amazon, its average slope, and elevation. The geographical unit used is IPEA's 1970 Minimal Comparable Areas (MCA 70-00). Robust standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

the change in log of MCA i total population/ GDP/ GDP per capita between 1970 and 2000.

For population (column 1) and GDP per capita (column 5), the OLS outcome shows a small negative and insignificant effect, while for GDP (column 3) it is a small negative and significant at the 5% level. When instrumenting the cost of access with distance to the line, the effects become larger, all negative and statistically significant. Places experiencing larger reductions (a larger negative value of the explanatory variable) have larger population, GDP, and GDP per capita increases. The elasticities are important: 2.9 for population, 2.8 for GDP, and 2.2 for GDP per capita.[26]

Note that the difference between the OLS and the 2SLS estimates is what we expect if roads were assigned to run through previously underdeveloped areas, in this case as a result of the political process which led to the creation of Brasília. OLS results are indeed biased towards zero.

How non-linear are these effects? Panel B presents the same estimations where the cost of access is interacted with the dummy variables b_1 to b_4. Interestingly, in the OLS specification we now get the stronger negative and significant effects on population and GDP for locations at less than 100 km, meaning that within this band there are larger increases in population and GDP for places that experience a larger access improvement. The effects then decrease in size as distance to the capital grows. In Table A3 in the Appendix, we show that these interactive effects are qualitatively similar in the reduced form specifications.

As before, when moving to the 2SLS, coefficients are an order of magnitude larger and systematically negative and significant. For population, the effects are larger for locations at less than 100 km initially. There, a 1% gain in access cost translates to roughly a 3.1% population gain. This decreases as initial distance to State capitals gets larger. Beyond 200 km, the elasticity is only 1.9. For GDP, there is a similar pattern, with an elasticity of 2.8 at less than 100 km, 2.4 between 100 and 200 km, 1.1 between 200 and 500 km, and 1.8 beyond 500 km.

Tests of the equality of these interactions show that we can reject that the interactions are equal for GDP, at the 1 percent level, but not for population and GDP per capita. There thus seems to be a pattern of stronger effects on GDP near urban centers, decreasing beyond that.[27]

Figs. A1 and A2 in the Appendix show semi-parametric estimates for population and GDP changes, based on least square specifications in which initial 1970 cost of access to the State capital is allowed to enter non-parametrically, following Robinson (1988). They also display more negative slopes for places initially closer to the State capitals.

Finally, it is worth mentioning that the coefficients for the 1970 values of population, GDP, and GDP per capita are all negative and significant, meaning that larger and richer municipalities tend to

[26] In Appendix Table A2, we display standard errors adjusted for spatial dependence following Conley (1999). As spatial GMM estimations with State-level variables do not converge, we compare 2SLS and spatial GMM in a specification where State fixed-effects are differentiated out first.

[27] GDP displays some of U-shaped effect, and a test of the equality of coefficients shows that the 500 km coefficient is indeed statistically different from the 200–500 km one.

Table 6
End points interactions – 1970–2000 changes.

	(1) ΔLog Pop	(2) ΔLog GDP	(3) ΔLog Pop	(4) ΔLog GDP	(5) ΔLog Pop	(6) ΔLog GDP	(7) ΔLog Pop	(8) ΔLog GDP
Δ Access to State Capital	2.458***	2.771***	−0.750***	−0.546	0.896**	0.991**	−1.447***	−1.520***
	(0.368)	(0.546)	(0.232)	(0.337)	(0.358)	(0.492)	(0.252)	(0.368)
Δ Access*GDP	−0.163***	−0.171***						
	(0.015)	(0.023)						
Δ Access*Water Access			−0.0640	−0.117				
			(0.084)	(0.120)				
Δ Access*Schooling					−1.433***	−1.353***		
					(0.230)	(0.318)		
Δ Access*Industry/Services							−0.638***	−0.749***
							(0.068)	(0.097)
Observations	3644	3644	3644	3644	3644	3644	3644	3644

2SLS regressions with differences in log Population, log GDP and log GDP per capita between 1970 and 2000 regressed on the change in cost of access to the state capital and the change in cost of access interacted with end point characteristics, instrumented by the distance to the nearest line. Endpoint characteristics are grouped into above and below median levels. Characteristics included are the endpoint GDP, the average rate of water access, the average years of schooling and the manufacturing to services ratio. Controls include initial (1970) value of the dependent variable, state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, in the amazon, its average slope, and elevation. The geographical unit used is IPEA's 1970 Minimal Comparable Areas (MCA 70-00). Robust standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

display smaller gains in population, GDP, and GDP per capita respectively.

What exactly explains this pattern is the next question. It seems that the stronger road-induced agglomeration close to the main metropolitan centers corresponds to the improvement in access to large urban centers themselves. We test this hypothesis further in the next section on heterogeneity below, where we analyze how these average effects vary depending the characteristics of the urban centers the roads are connecting to.

6.2. Heterogeneity

6.2.1. Average line effects

Our discussion in Section 2 relates local the growth pattern to the strength of agglomeration economies in the main connected urban areas. We now propose two different tests of how our results vary across urban centers and their influence areas.

Our first test is based on a simple specification in which the second stage takes the form:

$$\Delta\log(Y_{is}) = \beta_0 + \beta_1\Delta\log(R_{is}) + \beta_2\Delta\log(R_{is}) * W_j + X'_{is}\beta_3 + \theta_s + \varepsilon_{is}, \tag{8}$$

where W_j is one of the proxies for agglomeration externalities discussed in Section 2 above. Specifically, we use the initial characteristics of the endpoint city of the nearest line to each municipality; i.e., alternatively the endpoint GDP (as a proxy for size),[28] the average rate of water access (as a proxy for amenities), average years of schooling of the endpoint population (as a proxy for human capital), and the manufacturing-services ratio or the share of non agricultural activities in GDP (as a proxy for specialization).

Since the interaction introduces an additional endogenous variable, we estimate each first stage equation by including a set of eight instruments generated by interacting distance to the line with the eight line dummies. Table 6 presents the results. For each of the eight lines and corresponding end points of the radial highway system, as characterized by a given value of W_j, the net effect computed as $\hat{\beta}_1 + \hat{\beta}_2 W_j$ value should be interpreted as the average treatment effect for the group of municipalities for which this specific line is the closest one. Hence, municipalities at all distances are pooled together. These results are of course only partially capturing the overall strength of pull factors on different cities, but a full specification including several interactions simultaneously is not possible because of the strong correlations between the different proxies included here.

In columns 1 and 2, the GDP interaction is strongly significant, confirming that bigger cities appear to elicit stronger agglomeration effects on average. Given the value of W_j, this average effect is actually only positive for the biggest of the cities, São Paulo.

In columns 5 and 6, conditioning on schooling shows that the average agglomeration effects on population and GDP are stronger the higher the average level of schooling of end point cities. All eight lines actually display net positive agglomeration effects along this dimension.

Finally, in columns 7 and 8, a strong relationship appears to hold between manufacturing to service ratios and average agglomeration effects. Again, the effect is already positive at baseline and moving from the line with the lowest to that with the highest ratio implies a 40% increase in the overall strength of agglomeration.

These estimates provide the average treatment effect over all the municipalities in a line area of influence, pooling together places where agglomeration is positive and others where it is negative. Next, we intend to differentiate across urban areas while still capturing the spatial decomposition of effects included in specification (5). To that end, we group end points into above/ below median groups according to our agglomeration proxies and estimate our main two-stage specification using a dummy for these groups interacted with $R_{ist} * B$, the cost of access variable in different distance bins.

Given the 1970 characteristics of our end points, we split cities in two groups. The 'low' (respectively 'high') group includes cities with below (respectively above) median water access and schooling, which also corresponds to cities with an above (respectively below) median share of agriculture in GDP. The resulting 'low' group includes Belém, Campo Grande, Ciuaba, and Porto Velho, while the other one includes São Paulo, Rio, Salvador, and Campo Grande.[29]

The resulting coefficients are displayed in Fig. 6, and in Table A4 in the Appendix. We find strong agglomeration effects of roads for places connected to end points with high water access/ high schooling, and a low agriculture share, and mostly null results for other connections. The qualitative findings for this group of end points is very similar so the general ones from Table 5. Effects

[28] Estimations using population as a proxy for size, not included here, yield very similar results.

[29] Alternative criteria, such as GDP, population, and population density, or median manufacturing to services ratio, provide groups that largely overlap. For example using GDP and population leads to substitute Belém for Fortaleza. he effects found using that alternative grouping are very similar.

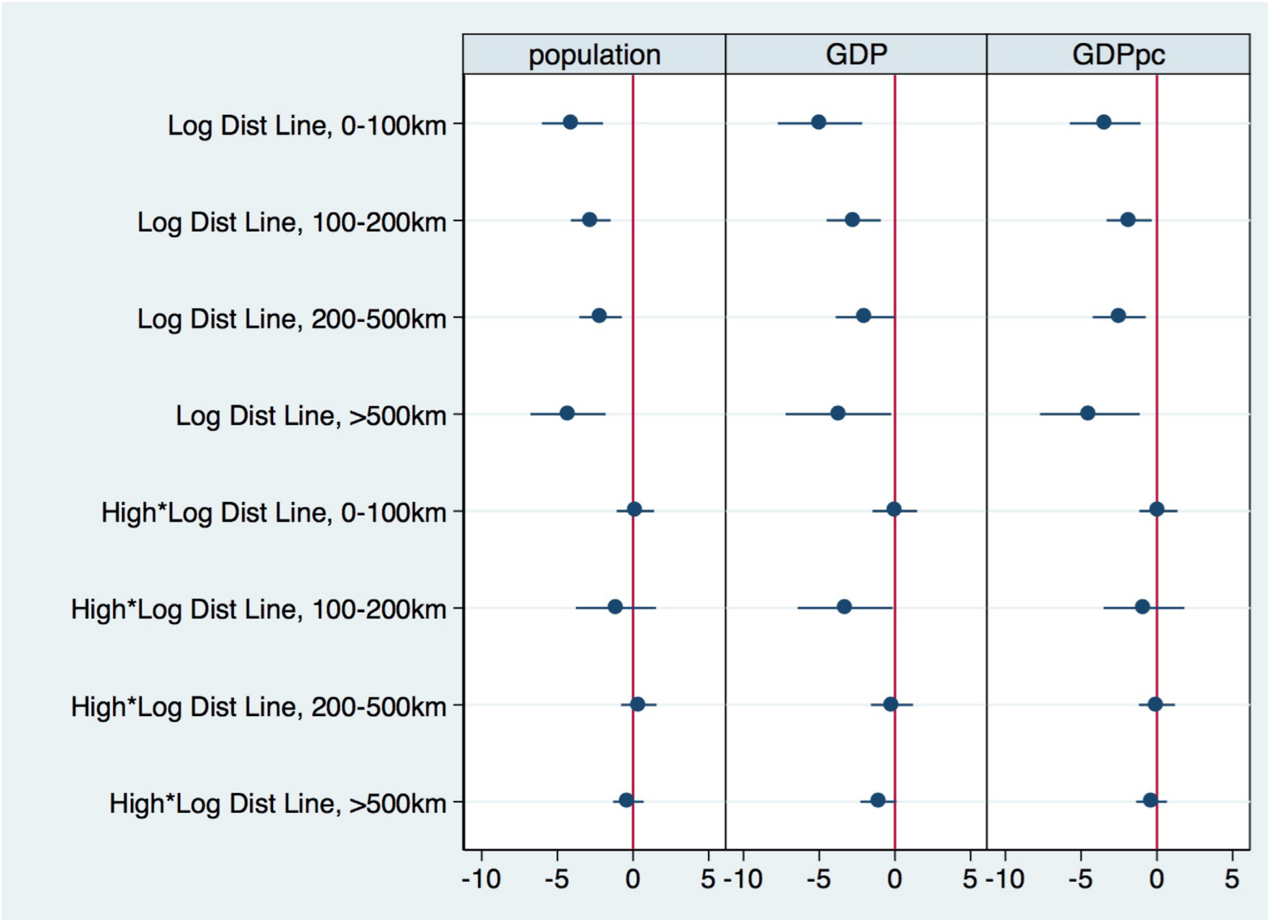


Fig. 6. End point heterogeneity: Low vs. high Agriculture/GDP share.

are positive throughout the whole spatial range and again stronger within 100 km of cities, a bit weaker from 100 to 500 km, and growing again beyond 500 km.

7. Robustness tests

In this Section, we present a number of robustness checks to our main results above.[30] First, given the economic weight of the Southern part of the country described in Section 3, we address the fact that the relevant cost of access measure might be the one linking municipalities to São Paulo rather than the State capitals.

Table 7, columns 1–3, reproduces the main 2SLS results from Table 5 for the sake of comparability. In columns 4–6, we then show that our main results are robust to controlling flexibly for distance to São Paulo.[31] Coefficients are slightly reduced but qualitatively unchanged.

Second, we consider the possibility that the assumed radial highway network used for identification may be spuriously overemphasizing corridors in the Northern part of the country and the relevance of access to the State capitals in that part of the country. We therefore construct an alternative network of axial roads that now excludes the rays to Salvador and Fortaleza, where road connections were likely of lower quality up to the end of our period of interest, and includes additional rays to Porto Alegre in the South, and Recife in the North.[32] The new network places higher weight on the connections in the South, and defines an alternative distance index based on the set of distances from municipalities to these new lines. The overall correlation between the two indexes is 0.41, but breaking it down by regions shows that it is 0.71 for Southern municipalities, and −0.20 for Northern ones.

Using this additional index as a second instrument, we are now able to run a 'horse race' between improvements in the cost of access to the State capitals and to São Paulo as determinants of changes in population and GDP over the 1970–2000 period. Results in Table 7, columns 7 to 9, show that the decrease in the cost access to State capitals remain the main determinants to improvements in population, GDP, and GDP per capita, while the change in the cost access to São Paulo is of the opposite sign and mostly insignificant.

Finally, we perform separate estimations for changes in agricultural, manufacturing, and service GDP, to inquire whether the spatial pattern we uncover maybe due to the differential effect of changes in the cost of access on these sectors. In columns 1–3 of Appendix Table A5, improved access appear to generate increases in manufacturing and service GDP but has no effect on agriculture. Interactions with distance bands, in columns 4 to 6, shows that the manufacturing and service effects concentrate in the first two intervals, i.e., between 0 and 200 km, while the impact on agriculture is not significant throughout.

[30] Note also that Bird and Straub (2014) show that the main results hold when using instead the MCA 40-00 geographical units.

[31] In our main specification we do not systematically include this distance as control, as it is highly correlated with distance to the Federal capital (0.71). As such, multicollinearity is likely to be an issue in these specifications.

[32] We thank an anonymous referee for suggesting this test.

Table 7
Robustness checks – Distance to Sao Paulo~ – 1970–2000 changes.

	(1) ΔLog Pop	(2) ΔLog GDP	(3) ΔLog GDPpc	(4) ΔLog Pop	(5) ΔLog GDP	(6) ΔLog GDPpc	(7) ΔLog Pop	(8) ΔLog GDP	(9) ΔLog GDPpc
Δ Access to State Cap.	−2.850***	−2.775***	−2.197***	−2.209***	−1.643	−1.025*	−1.402*	−3.447*	−1.172*
	(0.841)	(1.070)	(0.753)	(0.855)	(1.009)	(0.605)	(0.761)	(1.773)	(0.708)
Δ Access to São Paulo							1.397	12.45*	−0.626
							(2.846)	(7.014)	(3.115)
Log Population 1970	−0.0488			−0.0212			0.0191		
	(0.035)			(0.036)			(0.030)		
Log GDP 1970		−0.190***			−0.160***			−0.105	
		(0.033)			(0.031)			(0.065)	
Log GDP per cap. 1970			−0.677***			−0.673***			−0.667***
			(0.031)			(0.025)			(0.030)
Observations	3644	3644	3644	3644	3644	3644	3644	3644	3644
Distance to São Paulo				Yes	Yes	Yes			
Alternative Network IV							Yes	Yes	Yes

2SLS regressions with differences in log Population, log GDP and log GDP per capita between 1970 and 2000 regressed on the change in cost of access to the state capital (columns 1 to 3) and to São Paulo (columns 4 to 6). Controls include initial (1970) value of the dependent variable, state dummies, distances to Brasilia and the State capital as well as their square, and a set of geographical controls including area, whether the municipality is coastal, in the amazon, its average slope, and elevation. In columns 1 to 3, distance to São Paulo and its square is an additional control The geographical unit used is IPEA's 1970 Minimal Comparable Areas (MCA 70-00). Robust standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

8. Conclusion

Using a unique historical natural experiment, the construction of Brasília, we have been able to exploit an exogenous impulse in constructing a new radial highway network within Brazil to identify the causal impact of improvements in road access on population and economic activity over three decades.

Our analysis reveal that better access to large urban centers as a result of exogenous proximity to radial roads generated important agglomeration effects in terms of population and GDP growth. These effects were stronger in the immediate proximity of urban centers up to 200 km.

This overall pattern, however, masks striking differences across Brazil. Using several empirical strategies to look at the heterogeneity of these results, we find that these are driven to a large extent by the differences between endpoint characteristics in terms of agglomeration economies related to size, human capital, industrialization, and amenities.

In particular, overall agglomeration effects appear to be stronger around urban areas with above median amenities, as captured by access to water connections, human capital, as measured by the average schooling of the population, and a high share of non-agricultural activities in GDP. They also appear to be stronger for places connected to urban areas with a high manufacturing to services ratio, especially in the 0 to 200 km range. We discuss in detail the conditions for our instrumental strategy to be valid, and we also show that all our results hold in the reduced form based simply on the inconsequential unit approach.

Note that the natural experiment we analyze involves the creation of a new capital and a shift in the spatial structure of the country. One important issue is that our results may capture the joint treatment effect of access to the new highways built after the creation of Brasília and of access to this new capital city itself. While all estimations already control flexibly for distance to Brasília, which is also correlated with distance to the lines, proximity to the new capital may have a non linear impact on local development. Robustness checks, excluding all localities within 200 km of Brasília, show, however, that our main effects are unchanged quantitatively, lending credit to the idea that what we are documenting is indeed the impact of roads.

If anything, in the Brazilian context, amenities and knowledge spillovers, together with specialization, seem to matter more than size alone. In addition, the share of non-agricultural activities seems to be a stronger determinant than the manufacturing-to-services ratio, so the pull of labor to urban areas does not seem to be only manufacture driven but also to correspond to a large share of (likely informal) services.

In addition, these results largely overlap with the heterogeneity between the country's richer and denser South, and its poorer Northern part. In the South, municipalities experiencing a stronger marginal effect of improved road access appear to be concentrated around the main metropolitan areas, while in the North the reverse pattern holds.

The results help to explain how the shape of a highway network impacts economic development. The effects of a highway on local GDP and population depend not only on having improved transport access, but also on where this improved access leads to, and the economic strength of a municipality relative to the surrounding area. Connecting hinterland regions could lead to an increase or decrease in population and GDP in these areas, and these changes can in part be explained by the initial economic characteristics of the end-points and the municipality itself. In follow-up work, we intend to compare these results to those for other developing and emerging countries, to try to understand more systematically the extent to which they vary according to key features of the economic and urban landscape.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper supersedes an earlier version called *The Brasília Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil*. We thank Nicolas Ahmed-Michaux-Bellaire for excellent research assistance, the editor Arun Agrawal, two anonymous referees, and numerous participants in seminars in Amsterdam, Berkeley, EUDN Berlin, Padua, Passau, Stanford, Toulouse, Universidad de Chile and the World Bank for helpful discussion. Support from the World Bank Research Support Budget is gratefully acknowledged. Stephane Straub acknowledges funding from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir Program).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.worlddev.2019.104739.

References

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics, 133*, 375–395.

Asher, S. & Novosad, P. (Forthcoming). Rural roads and local economic development. American Economic Review.

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850–60. *Social Science History, 34*(2), 171–197.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., & Robert-Nicoud, F. (2003). *Economic geography and public policy*. Princeton University Press.

Banerjee, A., Duflo, E., & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China. *NBER Working Paper 17897*.

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics, 122*(2), 775–805.

Baum-Snow, N., Brandt, L., Henderson, V., Turner, M., & Zhang, Q. (2017). Roads, railroads and decentralization of chinese cities. *Review of Economics and Statistics, 99*(3), 435–448.

Berg, C., Deichmann, U., Liu, Y., & Selod, H. (2017). Transport policies and development. *Journal of Development Studies, 53*, 4.

Bird, J., & Straub, S. (2014). The Brasilia experiment: Road access and the spatial pattern of long-term local development in Brazil. *World Bank Policy Research Working Paper WPS6964*.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & Padrí i Miquel, G. (2014). The value of democracy: Evidence from road building in Kenya. *American Economic Review, 105*(6), 1817–1851.

Cadot, O., Roller, L.-H., & Stephan, A. (2006). Contribution to productivity or pork barrel? The two faces of infrastructure investment. *Journal of Public Economics, 90*, 1133–1153.

Castro, N. (2002). Transportation costs and brazilian agricultural production: 1970–1996. Unpublished Manuscript.

Castro, N. (2004). Logistic costs and Brazilian regional development. Unpublished Manuscript.

Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics, 30*(4), 457–490.

Chein, F., & Assuncão, J. (2016). How does out-migration affect labor markets? Evidence from road construction in Brazil. *Brazilian Review of Econometrics, 36*(2), 157–184.

Combes, P.-P., Mayer, T., & Thisse, J.-J. (2008). *Economic geography. The integration of regions and nations*. Princeton University Press.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics, 92*(1), 1–45.

Da Mata, D., Deichmann, U., Henderson, J. V., Lall, S., & Wang, H. (2007). Determinants of city growth in Brazil. *Journal of Urban Economics, 62*, 252–272.

Damania, R., Russ, J., Wheeler, D., & Barra, A. F. (2018). The road to growth: Measuring the tradeoffs between economic growth and ecological destruction. *World Development, 101*, 351–376.

Datta, S. (2012). The impact of improved highways on Indian firms. *Journal of Development Economics, 99*(1), 46–57.

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

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A market access approach. *Quarterly Journal of Economics, 131*(2), 799–858.

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

Duranton, G. & Puga, D. (2013). The growth of cities. CEPR discussion paper 9590.

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

Feler, L., & Henderson, J. V. (2011). Exclusionary policies in urban development: Under-servicing migrant households in Brazilian cities. *Journal of Urban Economics, 69*(3), 253–272.

Ghani, E., Goswami, A., & Kerr, W. (2016). Highway to success in India: The impact of the golden quadrilateral project for the location and performance of manufacturing. *Economic Journal, 126, 591*, 317–357.

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers, 44*, 315–348.

Holmes, T. J. (1999). Localization of industry and vertical disintegration. *Review of Economics and Statistics, 81*(2), 314–325.

Krugman, P. R. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review, 70*, 950–959.

Krugman, P. R. (1991). *Geography and trade*. Boston: MIT Press.

Lall, S. V., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. *Journal of Development Economics, 73*(2), 643–673.

Lipscomb, M., Mobarak, A. M., & Barham, T. (2013). Development effects of electrification: Evidence from the geologic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics, 5*(2), 200–231.

Marshall, A. (1920). *Principles of economics*. London: MacMillan.

Michaels, G. (2008). The effect of trade on the demand for skill – Evidence from the interstate highway system. *Review of Economics and Statistics, 90*(4), 683–701.

Mitchell, B. R. (1995). *International historical statistics: The Americas 1750–1988*. Stockton Press. second revision.

Moretti, E. (2004). Human capital externalities in cities. In J. V. Henderson & J.-F. Thisse (Eds.). *Handbook of regional and urban economics* (vol. 4, pp. 2243–2291). Amsterdam: Elsevier.

Morten, M., & Oliveira, J. (2018). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper 22158*.

Qin, Y., & Zhang, X. (2016). The road to specialization in agricultural production: Evidence from rural China. *World Development, 77*, 1–16.

Redding, S., & Turner, M. (2015). Transportation costs and the spatial organization of economic activity. In G. Duranton, J. Vernon Henderson, & W. Strange (Eds.). *Handbook of urban and regional economics* (pp. 1339–1398). Chapter 20.

Reis, E., Pimentel, M., Alvarenga, A. I. M. & Santos, M. d. C. H. (2011). Í reas mínimas comparíveis para os períodos intercensitírios de 1872 a 2000. I Simposio Brasileiro de Cartografia Histí rica., Parati, RJ, UFMG – CRCH. http://www.ipeadata.gov.br/doc/AMC-1872-2000.doc.

Reis, E. (2014). Spatial income inequality in Brazil, 1872–2000. *EconomiA, 15*(2), 119–140.

Roberts, M., Deichmann, U., Fingleton, B., & Shi, T. (2012). Evaluating China's road to prosperity: A new economic geography approach. *Regional Science and Urban Economics, 42*(4), 580–594.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica, 56*, 931–954.

Rosenthal, S. S., & Strange, W. (2004). Evidence on the nature and sources of agglomeration economies. In V. Henderson & J.-F. Thisse (Eds.). *Handbook of regional and urban economics* (vol. 4, pp. 2119–2171). Amsterdam: North-Holland.

Rothenberg, A. (2013). Transport infrastructure and firm location choice in equilibrium: Evidence from Indonesia's highways. Unpublished Manuscript.

Smith, J. (2002). *A history of Brazil*. Longman.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa. *Review of Economic Studies, 83*(3), 1263–1295.

World Bank. (2008). Brazil, "Evaluating the Macroeconomic and Distributional Impacts of Lowering Transportation Costs". Brazil Country Management Unit, PREM Sector Management Unit. Latin America and the Caribbean Region.