Made
centro de
pesquisa em
macroeconomia
das desigualdades

Working Paper 01.07.2024 nº 023

Highway Infrastructure and Local Outcomes: measuring causal impacts of infrastructure investments using a three-step instrumental variable identification strategy

Victor Medeiros,
Rafael S. M. Ribeiro,
Pedro V. M. Amaral &
Alexandre Q. Stein

This paper provides an original third-step identification strategy using instrumental variables to evaluate the causal impact of highway investments on the local economy. First, we construct a novel national highway dataset at the municipal level in Brazil using the Growth Acceleration Program (PAC) (2007-2018) as a case study. Second, we rely on some of the main infrastructure project costs to propose several costrelated instruments to correct for measurement errors in the road variables. Third, we circumvent the omitted variable bias from the non-random placement of roads by building instruments based on global cost minimization methods, historical plans, and the propensity of a municipality to receive highway interventions. Our identification strategy allows us to identify relevant biases coming from both measurement error and omitted variables. Our preferred estimates point out a reliable road elasticity in the range of 0.011 to 0.017. From this, we calculate a nonbiased return rate to highway infrastructure of 21.3% in Brazil, proving the high rentability of those investments in the developing world context.

Keywords: highway infrastructure; local outcomes; instrumental variables

JEL: O18; R42; R58; C26

MEDEIROS, V; RIBEIRO, R. S. M.; AMARAL, P. V. M.; STEIN, A. Q. Highway Infrastructure and Local Outcomes: measuring causal impacts of infrastructure investments using a three-step instrumental variable identification strategy. São Paulo: Centro de Pesquisa em Macroeconomia das Desigualdades (Made/USP), 2024. (Working Paper, n.23).

Victor Medeiros is a PhD Candidate in Economics at the Federal University of Minas Gerais (UFMG) and Researcher at the Research Center on Macroeconomics of Inequalities - Made/FEA-USP.

Rafael S. M. Ribeiro holds a PhD in Economics, University of Cambrige. Assistant Professor at the Federal University of Minas Gerais (UFMG) and Researcher at the Research Center on Macroeconomics of Inequalities - Made/FEA-USP.

Pedro V. M. Amaral holds a PhD in Economics, University of Cambrige. Assistant Professor at the Federal University of Minas Gerais (UFMG) and Visiting Professor at the University of Chicago.

Alexandre Q. Stein is a PhD Candidate in Economics at the Federal University of Minas Gerais (UFMG).

The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their support. Any errors and omissions present in the paper are the sole responsibility of the authors.

made.feausp@gmail.com
victor-medeiros@cedeplar.ufmg.br
rsmribeiro@cedeplar.ufmg.br
pedroamaral@cedeplar.ufmg.br
queiroz.stein@gmail.com

1. Introduction

The economic effects of transportation infrastructure have been explored by various studies (Aschauer, 1989; Baum-Snow *et al.*, 2017; Baum-Snow *et al.*, 2020; Chandra and Thompson, 2000; Donaldson and Hornbeck, 2016; Duranton and Turner, 2012, 2014; Faber, 2014; Foster *et al.*, 2023a, 2023b; Michael, 2008; Straub, 2011). More interestingly, a substantial part of this literature has provided relevant findings on highway investments' role in local economic activity. Generally, the findings indicate a positive and direct association between road investments and various outcomes, such as economic growth, productivity, employment, poverty alleviation, among other outcomes (Allen and Arkolakis, 2014; Bird and Straub, 2020; Ghani, Goswami, and Kerr, 2014; Holl, 2016; Huang and Xiong, 2018; Lí*et al.*, 2017; Percoco, 2015; Wang, Wu, and Feng, 2020).

Despite the long-standing development of this literature, endogeneity bias remains a persistent empirical challenge (Caldero n and Serve n, 2014; Redding and Turner, 2015; Roberts *et al.*, 2019; Straub, 2011). First, the non-random allocation nature of road policies makes it hard to identify causal impacts. Planners may allocate more resources to regions where infrastructure yields higher returns to foster national economic growth. On the other hand, policymakers may choose to target less developed and more remote regions to promote more equitable economic development across regions. If it occurs, conventional Ordinary Least Squares (OLS) estimates are likely upward biased in the former context, while downward biased in the latter. In addition, highway investment measures often embed measurement errors due to inefficiencies and corruption in the infrastructure project design phase ,as well as in its building and operation stages (Kenny, 2009). Then, road variables are likely to be inflated in the sense that, for many places, we usually observe an overpriced amount of investment *per* kilometer of road. Consequently, the measurement error may obscure the actual impact of roads on economic growth and development, resulting in an underestimation of their profitability.

To solve the endogeneity issue, most empirical studies have used reduced-form estimations under several different instrumental variables (IV) as sources of quasi-random variation in the observed infrastructure (Foster *et al.*, 2023a, 2023b; Redding and Turner, 2015; Roberts *et al.*, 2019). Investigations have proposed instruments based on either planned routes (Baum-Snow, 2007; Bird and Straub, 2020; Duranton and Turner, 2012; Duranton *et al.*, 2014; Frye, 2016; Hsu and Zhang, 2014; Michaels, 2008; Sheard, 2014; Herzog, 2021; Rokickia and Stępniak, 2018), or historical routes (Adler *et al.*, 2020; Baum-Snow *et al.*, 2017; Baum-Snow *et al.*, 2020; Duranton and Turner, 2012; Duranton *et al.*, 2014; Garcia-Lo pez *et al.*, 2015; Holl, 2012; Holl, 2016; Hsu and Zhang, 2012; Lee, 2021; Martın-Barroso, Nunez-Serrano and Velazquez, 2015; Martincus *et al.*, 2017; Percoco, 2015; Rokickia and Stępniak, 2018; Zhang, Hu and Lin, 2020), or infrastructure project costs (Holl, 2012; MartinBarroso, Nunez-Serrano, and Velazquez, 2015; Lu *et al.*, 2022; Medeiros *et al.*, 2021a; Medeiros *et al.*, 2021b; Zhang, Hu, and Lin, 2020) or hypothetical road networks built on a global minimization path intended to connect important localities (hubs) (Faber, 2014; Frye, 2016; Ghani, Goswami, and Kerr, 2014; Huang and Xiong, 2018; Yang, 2018; Xu and Feng, 2022). A relevant part of those studies has combined some kind of IV with the inconsequential unit approach pioneered by Chandra and Thompson (2000).

While this strand of literature has developed interesting and robust identification strategies for measuring the causal highway impact on several outcomes, some open points remain. First, instruments are hard to find in practice, and a highly replicable IV approach is needed. Second, monetary road variables have been avoided by researchers due to measurement error (Caldero n and Serve n, 2014; Straub, 2011). With the increase in public administration transparency worldwide in the last decades, investment and expense flow data might be an important source of information on infrastructure policies' efficacy, efficiency, and effectiveness. In addition, highway investment variables constitute a direct way of measuring infrastructure profitability and providing an easy-to-understand indicator for planners and the society

(Fernald, 1999; Lí*et al.*, 2017; Wang, Wu, and Feng, 2020). Therefore, a reliable identification strategy dealing with measurement errors in highway measures is critical. Third, monetary variables have the advantage of capturing both road provision and quality. Most empirical studies have used measures of physical provision, and empirical approaches to adapt a multitype road intervention setting are still lacking. Fourth, a more careful look into the study design is demanding. For instance, recent investigations have evaluated the impact of national road programs on local outcomes, an approach that considerably reduces endogeneity concerns and has provided reliable results (Bird and Straub, 2020; Faber, 2014; Ghani, Goswami, and Kerr, 2014; Herzog, 2021).

To overcome the mentioned issues and estimate the causal impact of highway investments on local outcomes, we propose an original three-step identification strategy using IVs. The first *ex-ante* step consists of designing a study that minimizes endogeneity concerns. For this, we use data from a national road policy, the Growth Acceleration Program (PAC)[1], launched by the Brazilian Federal Government in 2007. We georeferenced the PAC data to construct a novel granular national highway investment data at the municipal level.

The PAC has some characteristics that make it an interesting case study. First, highway investments were duplicated during the PAC (2007-2018) compared to the previous ten years (Medeiros *et al.*, 2021a). Second, the PAC coincided (up to 2015) with a period of relatively high economic growth in Brazil (Nassif *et al.*, 2015). Third, the program faced severe criticism for its inefficiency and poor budget management, as it exhibited multiple construction delays and required much larger investments than the original estimates (Amann *et al.*, 2016; Raiser *et al.*, 2017). In this sense, this developing economy case study seems appropriate for evaluating the infrastructure-development nexus wherein measurement error is highly expected.

Then, we proceed to our second and third steps to evaluate the causal impact of national highway investments on municipal GDP *per capita* growth in Brazil between 2007 and 2018. In the second step, we instrumentalize our road variable using some of the main geographical, environmental, and human physical costs of infrastructure projects to correct measurement error bias. We rely on the empirical literature and real Brazilian infrastructure projects to develop replicable and reliable cost-related IVs. Conditional on controls, the second-step econometric estimations allow us to identify suitable (strong and exogenous) cost-related IVs.

In the third step, we combine our second step cost-related IVs with several road allocation IVs to correct for omitted variable bias coming from the non-random placement nature of highway policies . We draw upon the specialized literature and adapt it to the Brazilian case to generate IVs treating investment in both new and existing roads. The first set of instruments is intended to treat economically and politically biased road policies for newly connected municipalities. In this case, we use the Least Cost Path - Minimum Spanning Tree (LCP-MST) method (Faber, 2014) combined with cost-related measures to create hypothetical minimized global cost network instruments. In addition, we utilize the Brasília Plan (Bird and Straub, 2020) to generate additional and more specific instruments to treat political bias in road placement. For already connected municipalities in the starting period, we build up a unique instrument using road traffic data to avoid endogeneity coming from municipalities highly prone to receiving road interventions.

Our main results show that our empirical strategy is suitable at identifying road causal impacts on the local economy. Firstly, second step estimates allow us to correct (or mitigate) measurement error and to identify expected downward biased OLS elasticities. Secondly, third step regressions are critical to fix the non-random placement of road bias even after fixing measurement error. In this case, our findings suggest that the PAC prioritizes economic development over regional balance, as its actions favor more prosperous areas , thus upward biasing "free from measurement error" second step estimates. We find a solid third step road elasticity in a range from 0.011 and 0.017, implying a return rate to highway investment of around 21.3% in Brazil. From that, we can infer that Brazil would need to invest 2.5 more in road infrastructure to achieve a suitable road stock of 16% of the national GDP, proving the high rentability of highway investments in the developing economies scenario. Results remain unchanged under several robustness checks.

Our contributions are manifold. First, we develop a novel three-step IV identification strategy that allows us to correct both measurement error and non-random road allocation bias. We believe that our steps and instruments have a high degree of replicability. Thus, we contribute to an extensive strand of the literature using reduced-form equations to estimate the causal impacts of highway investments on the economy (Bird and Straub, 2020; Chandra and Thompson, 2000; Faber, 2014; Herzog, 2021; Michaels, 2008), as well as to the overall empirical literature on infrastructure.

Second, we construct an unpublished granular highway investment data at the municipal level. This data is particularly relevant as geographically detailed data on infrastructure investment is quite scarce (Brooks and Liscow, 2019). In this sense, we contribute to the broad infrastructure-development literature interested in measuring highway investments and their impact on the economy at the local level.

Third, we propose several original and replicable cost-related instruments in our second stage. We rely on real infrastructure projects to propose geographical, environmental, and human physical measures to represent road costs. To the best of our knowledge, our study is the first to jointly use those cost types as IVs in the measurement error bias context. In addition, some of our proposed measures are tested for the first time by our study. We complement a range of works using cost-related IVs for transport measures (Holl, 2012; Martin-Barroso, Nunez-Serrano, and Velazquez, 2015; Lu *et al.*, 2022; Medeiros *et al.*, 2021a, 2021b; Zhang, Hu, and Lin, 2020).

Fourth, we contribute to the empirical literature using LCP-MST instruments for road variables in two ways (Faber, 2014; Frye, 2016; Ghani, Goswami, and Kerr, 2014; Huang and Xiong, 2018; Yang, 2018; Xu and Feng, 2022). First, we construct a cost index based on our infrastructure project cost variables and include it in the minimization process in the LCP-MST method. By doing so, we improve the instruments based on the LCP-MST by including different kinds of road costs. Second, we propose a more replicable way of stablishing the hubs the LCP-MST procedure is connecting. We do this by identifying starting and ending points of roads constructed or improved by the PAC, which we believe is a reasonable approach when a clear policy identification of targeted cities is absent. Similarly, we complement the study by Bird and Straub (2020) by proposing an extension of their Brasília Plan instrument.

Fifth, we attempt to empirically deal with the multitype road intervention setting (building, paving, enhancements, and duplications) of our data. In our econometric specification, we differentiate municipalities already connected by federal roads in 2006 from those not linked. Then, we use LCP-MST and Brasília Plan instruments for non-connected municipalities, while we create a novel instrument for already connected ones. This original instrument, called "*potential road intervention areas*" IV, is calculated by identifying critical points using road traffic data, a measure we believe has not been tested in past studies.

The paper is structured as follows. Section 2 discusses the related empirical literature evaluating the infrastructure-development nexus, focusing on IV identification strategies. Section 3 presents the rationality behind our three-step IV identification approach. Section 4 details the different sources of data used in the paper. Section 5 introduces the empirical model. Section 6 presents the main results, falsification tests, and robustness checks. Section 7 discusses our results considering the return rate to highway investments. Section 8 concludes. Additional material and robustness checks can be found in the Appendix.

2. Related Literature

2.1. Road infrastructure and regional or local economic development: empirical related literature

An extensive strand of literature has examined the relationship between road infrastructure and a number of economic factors, such as output growth (Barzin et al, 2018; Baum-Snow *et al.*, 2017; Baum-Snow *et al.*, 2020; Bird and Straub, 2020; Ke and Yan, 2021; Rokickia and Stępniak, 2018; Zhang, Hu and Lin, 2020), productivity (Ghani, Goswami and Kerr, 2014; Fahardi, 2015; Holl, 2016; Huang and Xiong, 2018; Lí*et al.*, 2017; Martın-Barroso, Nunez-Serrano and Velazquez, 2015; Xu and Feng, 2022; Yang, 2018; Zhang e Ji, 2019), trade (Coşar and Demir, 2022; Duranton *et al.*, 2014; Martincus *et al.*, 2017), population (Adler *et al.*, 2020; Baum-Snow, 2007; Baum-Snow *et al.*, 2017; Baum-Snow *et al.*, 2020; Bird and Straub, 2020; Duranton and Turner, 2012; Faber, 2014; Garcia-Lo pez *et al.*, 2015; Gertler *et al.*, 2019, 2022; Jaworskiy and Kitchensz, 2019; Meijers *et al.*, 2012; Percoco, 2015) and structural transformation (Albalate and Fageda, 2016; Asher and Novosad, 2020; Yang, 2018). Since Aschauer (1989), most studies have shown a significant relationship between infrastructure and economic outcomes. However, results vary critically according to the investigation context and the identification strategy used – associated with endogeneity issues and with whether the study uses aggregate data at country, regional, or local level.

In this paper, we focus on regional and local level studies. Infrastructure is spatial by nature (Ottaviano, 2008; Straub, 2008, 2011), and a more geographically disaggregated view of the theme can clarify some transmission channels. As transport infrastructure buildings serve a limited geographic zone, the results from those policies might be expanding economic growth in some regions and sectors at the expense of others[2]. This is likely the reason why works using regional and local level data provide more heterogeneous results on the role of road investment on economic activity (Foster *et al.*, 2023a, 2023b; Redding and Turner, 2015; Roberts *et al.*, 2019).

Then, we turn to the main empirical issue in the infrastructure-economic development literature: endogeneity bias. First, measurement error bias often occurs, especially when using monetary infrastructure variables. Highway investments take time to mature and suffer from inefficiencies and delays, particularly in the developing world context (Caldero n and Serve n, 2014; Kenny, 2009; Straub, 2011). Second, the placement of roads is not random, as policymakers may target places expected to growth more or promote economic prosperity in underdeveloped regions (Redding and Rossi-Hansberg, 2017; Redding and Turner, 2015). In this sense, we can also expect endogeneity issues coming from omitted variables and reverse causality.

Roberts *et al.* (2019) study a wide range of papers estimating the effects of transportation infrastructure on several outcome variables. Around 63.5% of them used identification strategies to address endogeneity concerns, of which 78% used reduced form estimations, and 52.6% utilized IVs. Therefore, most of the papers rely on IVs, and studying this identification approach in detail is critical.

In the context of IV studies, the main bottleneck is the identification of suitable (strong and exogenous) instruments for highway measurement. In successful applications, this identification strategy has been quite reliable in identifying the causal impacts of road investments on local outcomes (Baum-Snow, 2007; Duranton and Turner, 2012; Faber, 2014; Holl, 2012). Nonetheless, important efforts are still needed, especially in the developing economy context where data is rough and measurement errors are expected to exist in road variables. This issue is accentuated by using monetary variables as infrastructure investments or expense flows if those variables might embed inefficiencies and mask the true relation between transport infrastructure and economic activity. In the following section, we cover the main IV approaches used by the specialized literature.

2.2. Main IV approaches in studies on the impacts of road infrastructure on regional or local development

Most part of the empirical literature on infrastructure and regional or local development has used IVs to identify causal road impacts. These studies can be classified into a few main approaches based on the IV rationale and its measurement form. We draw upon the seminal paper by Redding and Turner (2015), then we update this literature review and propose some advances. To this end, we also rely on Foster *et al.* (2023a, 2023b), which reviewed a massive number of empirical studies on infrastructure and development, including the road sector.

Redding and Turner (2015) identified three main IV strategies. The first one, the planned route approach, is an IV strategy relying on planning maps and old documents as a source of quasi-random variation for the observed infrastructure (Baum-Snow, 2007; Bird and Straub, 2020; Duranton and Turner, 2012; Duranton *et al.*, 2014; Frye, 2016; Hsu and Zhang, 2014; Michaels, 2008; Sheard, 2014; Herzog, 2021; Rokickia and Stępniak, 2018). The rationale behind these IVs is that the planned routes were created to serve purposes quite different from those coming from modern infrastructure development we are trying to test (for example, impacts on modern GDP, population, or employment growth). Planned routes used in this approach are, for instance, the 1947 USA national interstate highway plan and the "Pershing plan" for US data, the Brasilia Road Plan for Brazilian data, Japan's 1987 National Expressway Network Plan applied to Japan, and the project for a motorway network by E. Buszma dating from 1945 and the Resolution of the Automotive Council for the Council of Ministers plan from 1963 for the Polish case. A similar strategy was used by Sheard (2014) applied to the airport sector, using the USA 1944 National Airport Plan as IV.

The second approach, the historical route instrumental variable approach, relies on very old transportation routes as a source of quasi-random variation for observed infrastructure (Adler *et al.*, 2020; Baum-Snow *et al.*, 2017; Baum-Snow *et al.*, 2020; Duranton and Turner, 2012; Duranton *et al.*, 2014; Garcia-Lo pez *et al.*, 2015; Holl, 2012; Holl, 2016; Hsu and Zhang, 2012; Lee, 2021; Martın-Barroso, Nunez-Serrano and Velazquez, 2015; Martincus *et al.*, 2017; Percoco, 2015; Rokickia and Stępniak, 2018; Zhang, Hu and Lin, 2020). The validity of this instrument requires that, conditional on controls, factors that do not directly affect economic activity in the localities of interest at the period of study (mostly at the end of the twentieth century) determine the configuration of these historical networks. In other words, the validity of this identification strategy depends critically on the fact that the historical routes served different aims (for instance, moving agricultural goods to local markets, achieving military, administrative, and commercial goals and so forth) than they have today (for example, to foster economic growth and employment). Several distinct IVs were proposed in this setting, including the 1898 railroad routes and the routes of major expeditions of exploration between 1528–1850 for US data, a market access measure based on railroad network in 1870 for European regions data, the network of Roman roads for Italy data, the old Roman roads, the 1760 Bourbon roads, the 1760 Spanish postal route network and an accessibility to final markets in 1857 in the Spanish case, railway connections in 1952 for Polish data, locations of rail and tram stations in 1897, 1930, and 1960 for South Korea data, road and railroad networks in 1962 and postal routes in 1936 for China, and the PreColumbian Inca road network for the Peruvian case.

The third one, the inconsequential unit approach, relies on choosing a sample that is inconsequential to road allocation, in the sense that unobservable attributes do not affect infrastructure placement (Chandra and Thompson, 2000). Sometimes, this strategy is combined with some planned, historical, or cost-related IV (Banerjee *et al.*, 2012; Bird and Straub, 2020; Faber, 2015; Herzog, 2021; Percoco, 2015). In this approach, routes that sought to connect large centers cross (inconsequentially) small units (as a small city or municipality) across the pathway. Then, we expect that unobserved characteristics of these small units, which are just between targeted hubs, are independent of political and economic reasons.

Another strand of literature has utilized a Least Cost Path - Minimum Spanning Tree (LCP-MST) IV (Faber, 2015; Frye, 2016; Ghani, Goswami, and Kerr, 2014; Huang and Xiong, 2018; Yang, 2018; Xu and Feng, 2022). This identification strategy tries to answer the question of which routes planners would have been likely to build if the sole policy objective had been to connect all targeted city nodes on a single continuous network subject to global construction cost minimization. To this end, an LCP-MST network connecting large cites (hubs) is generated based on a global cost minimization process. The identifying assumption is that this hypothetical highway network should affect city outcomes and the spatial allocation of industries only through the actual highway network, conditional on controls. The LCP-MST approach has been used specially for Chinese data, with some applications for the USA and India.

A fifth approach uses cost-related environmental and geographical IVs as a source of quasi-random variation for highway investments (Holl, 2012; Martin-Barroso, Nunez-Serrano, and Velazquez, 2015; Lu *et al.*, 2022; Medeiros *et al.*, 2021a; Medeiros *et al.*, 2021b; Zhang, Hu, and Lin, 2020). The rationale behind this identification strategy is that geographical costs directly affect road construction and maintenance, although they are exogenous factors in the way they have not direct relation with modern GDP, employment, wages, or productivity growth. These IVs are mainly measured as slope, terrain ruggedness, altitude, or elevation, and they have been used for Chinese, Spanish, and Brazilian data. The literature has also proposed cost-related IVs based on legally protected areas as a *proxy* for environmental costs (Medeiros *et al.*, 2021b), and demographic variables – for instance, populational density – to represent physical human costs as expropriation and interferences (Zhang e Ji, 2019). We are not arguing that those kinds of costs are equal, but that they have similar attributes in how they act as some of the main highway infrastructure costs.

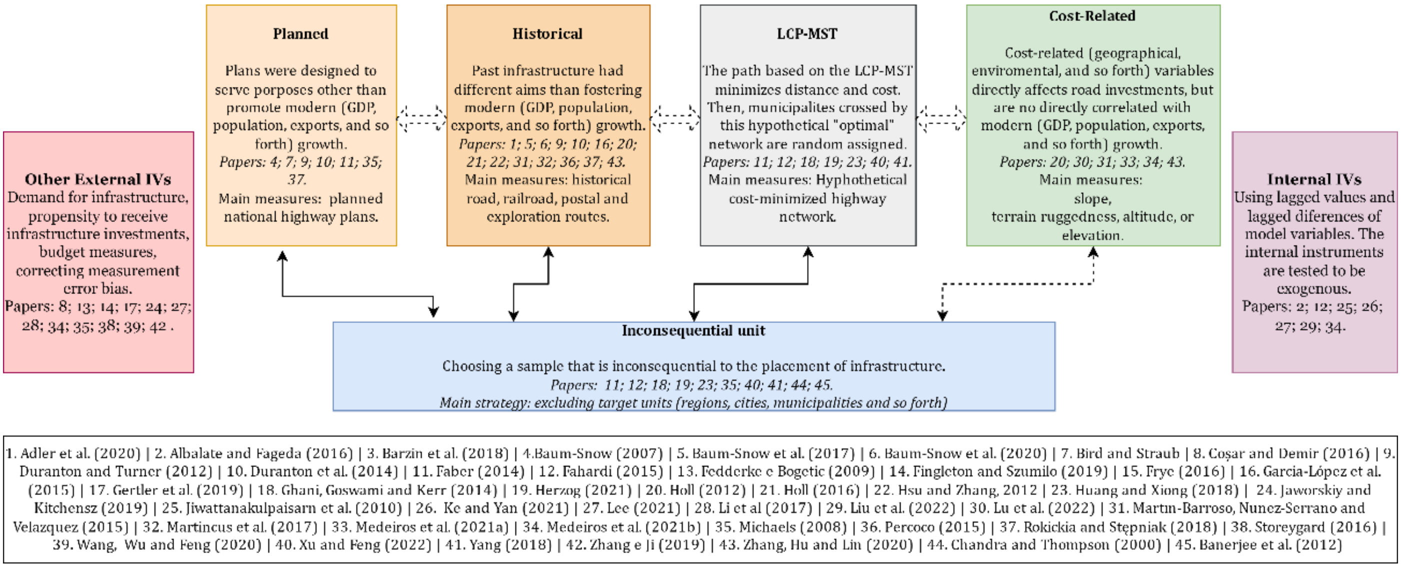
We can also identify another disconnected group of studies using different external IVs with distinct rationales. For instance, Coşar and Demir (2016) used the initial share of expressways as an instrument for provincial access to gateways. Fedderke and Bogetic (2009) constructed a demand for infrastructure IV based on the predicted demand for infrastructure stocks using time series econometric models. Baum-Snow (2007) combined the planned route IV with national construction rates over time as an IV. Storeygard (2016) used oil prices to instrument a road cost variable. Gertler *et al.* (2019) relied on national and provincial road budgets IVs. As we can see, the IVs used in those studies are quite specific, unlike the other 5 approaches described before.

Using panel data models, some studies applied internal instruments (lagged values and lagged differenced values of the model variables) in a GMM estimation setting (Albalate and Fageda, 2016; Bird and Straub, 2020; Fahardi, 2015; Fingleton and Szumilo, 2019; Jiwattanakulpaisarn *et al.*, 2010; Ke and Yan, 2021; Liu *et al.*, 2022; Barzin *et al.*, 2018). Some of those studies combine internal with external IVs (Gertler *et al.*, 2019; Holl, 2012; Medeiros *et al.*, 2021a).

Finally, another strand of literature has used an infrastructure reliance measure as an identification strategy (Li et al, 2017; Medeiros *et al.*, 2021b; Percoco, 2015; Wang, Wu, and Feng, 2020). The rationale behind this strategy is that industries that, for technological reasons, tend to rely less on transportation services (e.g., because they move lighter goods that tend to be less "road intensive" (Duranton *et al.*, 2014), or because their employees do not need to travel long distances for business trips) act as a sort of control group for the "treated" ones, i.e., those that rely more on transportation services. This approach is particularly suitable when using sector or firm level data. Some investigations have used IVs to correct for potential measurement errors in the infrastructure reliance measures (Li *et al.*, 2017). Medeiros *et al.* (2021a) combine the infrastructure reliance approach with internal and external IVs to estimate the road impact on productivity in Brazilian economic sectors.

Figure 1 summarizes the main identification approaches – focusing on those using IVs – used by the infrastructure-development literature. The arrows indicate that there exists some convergence between some of the identification approaches – for instance, some studies used both historical and cost-related IVs, or planned and historical IVs, or some combination of the inconsequential unit approach with some external IV. On the other hand, in most cases, the studies classified as "other external IVs" propose specific IVs and are not much related to different frameworks. Similarly, the internal IVs, while combined with external IVs in a few cases, are mostly unrelated to the historical, planned, LCP-MST, or cost-related approaches.

Figure 1. Related literature: main IV approaches used in infrastructure-regional economic development studies



Source: authors' elaboration

2.3. Open points and potential progress

The strand of literature using IV identification strategies has provided important advancements in identifying the causal impacts of highway investments on local economic activity. However, we can raise some gaps in the literature that need careful attention.

First, strong instruments are hard to find in practice. This issue comes from the lack of data in many cases, especially in developing countries, wherein more disaggregated data on highways is poor. In addition, whilst planned routes, historical networks, and LCP-MST IVs have been proved to be strong predictors of road measures in many applications, their replicability might be limited. In most part of countries around the world, there is no information on historical routes or past road plans we could rely on to develop a potential exogenous instrument. Regarding LCP-MST IV, it might be quite tricky to identify cities (hubs) targeted by planners. In developing countries, road projects often target areas with high transport demand, but the absence of strategic government planning hinders the identification of potential hubs and the creation of artificial networks applying the LCP-MST method.

Second, measurement error bias from highway measures is often overlooked. This issue is critical when using monetary measures as investment or expense flows, as long as they are likely embedding infrastructure projects inefficiencies, corruption, and its long time to mature nature. Whilst several studies have argued against the use of monetary variables (Caldero n and Serve n, 2014; Straub, 2011), highway investment flows constitute a direct way to measure infrastructure profitability and to provide an easy-to-understand indicator for planners and the society. Also, monetary variables capture quality related investments, which is absent in several applications using physical measures such as road length or access. Finally, the number of granular datasets and information about infrastructure investment allocation has grown worldwide in the last decades due to the increasing demand for transparency in public administration. Using this data is an obvious way to evaluate the efficiency, effectiveness, and cost-benefit of public spending (Brooks and Liscow, 2019; Kornfeld and Fraumeni, 2022). Then, providing a robust identification strategy dealing with measurement errors in monetary road variables is an important contribution to the empirical literature.

Third, empirical strategies that deal with different road building types are scarce. Governments might invest in constructing new roads in poorly connected or isolated regions or improving existing roads in localities with a higher expected return to transportation infrastructure. Those different kinds of interventions have particular costs and will likely impact economic activity heterogeneously. This implies that the economic return on road investment is diverse across space. Therefore, if data is available to conduct this type of analysis, identification strategies handling road intervention heterogeneity is critical and an important tool to provide novel and reliable results on causal road impacts on the local economy.

Finally, an important and not widely used *ex-ante* step in the identification strategy is to rely on a study design that alleviates endogeneity concerns. Some interesting cases can be found in the literature evaluating the causal impact of national highway policies on local outcomes (Baum-Snow *et al.*, 2017; Bird and Straub, 2020; Faber, 2014; Herzog, 2021). Those studies evaluate the impact of road investments made by high level governments on economic activity in smaller units such as districts, cities, and municipalities. This design reduces endogeneity issues as we do not expect small localities to directly influence national government decisions. Then, the policy exogeneity is rather more plausible compared with studies evaluating road policies' impacts on economic activity at the same administrative level. In addition, it allows us to work with a substantially larger number of observations, even when excluding central cities, by using the inconsequential unit approach.

In the next section, we describe in detail our novel identification strategy trying to solve the mentioned issues. In a study design evaluating the impact of a national road program on local economic activity, we combine cost-related instruments fixing measurement error bias in highway investment variables with LCP-MST, political and propensity to receive road intervention instruments correcting the nonrandom placement of road bias. Whenever it is feasible, we construct IVs that we believe are quite replicable around the world. By doing so, we provide a reliable empirical approach to identifying road impacts on local outcomes. In addition, our strategy allows studies to properly work with monetary highway variables, an issue that has been largely neglected so far.

3. Measuring causal impacts of road investments on local outcomes: a new three-step proposal combining cost-related and non-random placement IVs

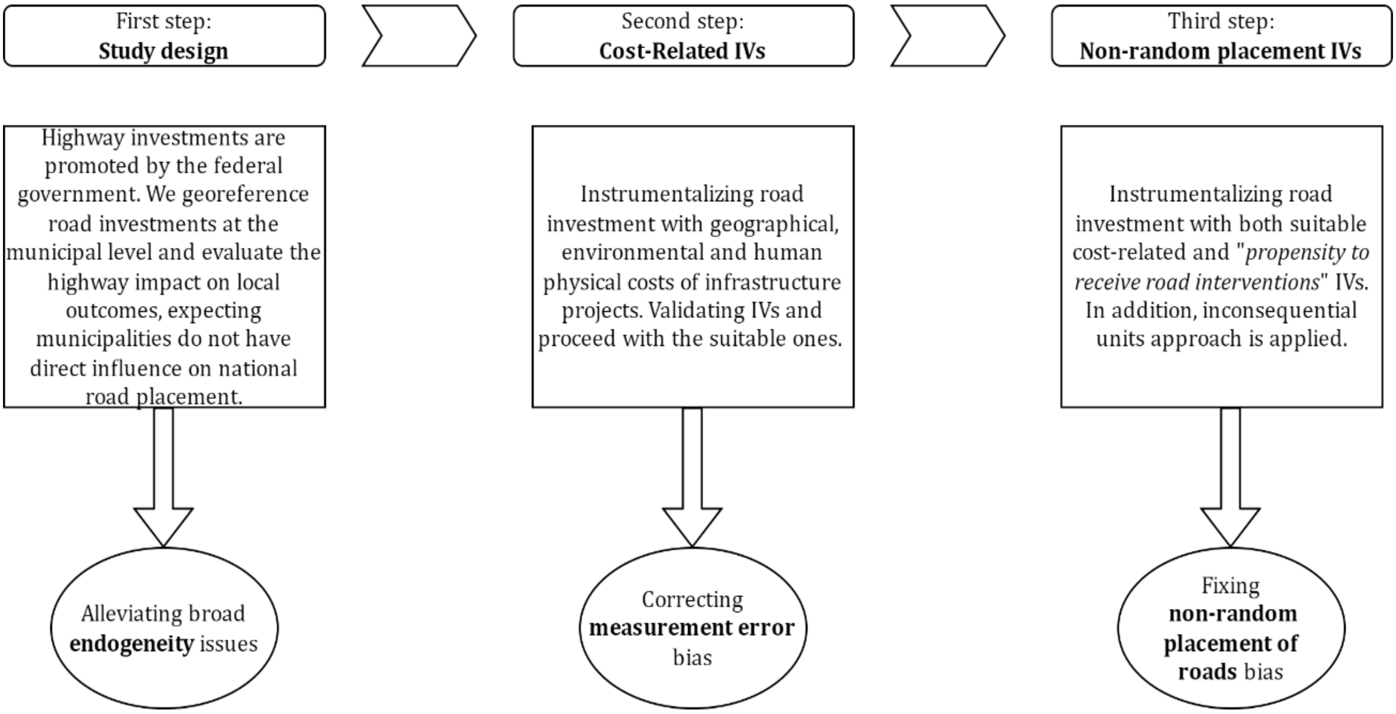
To deal with the critical issue of measuring the causal impacts of road investment on economic activity, we develop a novel empirical strategy in a three-step setting. Our main goal is to provide a reliable identification strategy robust to both measurement error and omitted variable bias.

First, we describe our study design evaluating the impacts of a national level highway policy on local outcomes. This first-step alleviates some concerns about endogeneity between road investments and economic development, as municipalities are expected to have no direct influence on federal road allocation (Herzog, 2021).

Second, we correct measurement errors from inefficiencies in road projects (Caldero n and Serve n, 2014; Straub, 2011) using the main related geographical, environmental, and human physical costs as IVs for highway investments. In this second-step, we statistically validate several replicable cost-related IVs and use them in the succeeding step.

Third, we fix omitted variable issues from non-random placement of highways by combining our previous two steps with IVs related to what we call "*the local propensity to receive road interventions*". We utilize three main IV types in our third step. First, we use the LCP-MST global cost minimization process (Faber, 2014). To calculate our hypothetical road network, we identify potential hubs (targeted municipalities) based on the start and end points of the highways receiving PAC interventions, which we believe can be easily replicated for several applications. Then, we generate a hypothetical road network minimizing not only the Euclidean distance between hubs, but also geographical, environmental, and human physical costs. To the best of our knowledge, no study has applied the LCP-MST minimizing environmental and human physical costs. Second, we also try IVs based on Brazilian historical plans (Bird and Straub, 2020), which alleviate concerns about politically biased road allocation. Third, we use proximity, which we name "*potential road intervention areas*", to instrumentalize highway investments in municipalities already connected by federal roads in the start period[3]. To this end, we use traffic intensity data to establish cities that are highly likely to receive road interventions, being an original and highly replicable instrument. Complementary to the proposed IVs, we also use the inconsequential unit approach pioneered by Chandra and Thompson (2000) to exclude likely targeted cities. Figure 2 sums up our empirical approach.

Figure 2. Three-step empirical approach: identifying causal impacts of highway investment on local outcomes



Source: authors' elaboration.

3.1. Study design

Our main goal is to estimate the causal impacts of highway investments on local outcomes. To this end, we rely on a national road program (a program within the PAC) implemented by the Brazilian Federal Government between 2007 and 2018.

By desegregating national investments at the municipal level, a procedure we describe in detail in the data section, we can evaluate the local impacts of an extensive and aggregated road policy. This study design is important to alleviate endogeneity concerns between infrastructure investments and economic activity (Faber, 2014; Herzog, 2021), as we do not expect municipalities to directly influence the Brazilian federal government's decisions to place highways across the country.

Currently, Brazil has 5,570 municipalities in 27 states. While we might expect some important and direct political influence of the state level governments on the federal government investment priorities potentially biasing econometric estimates, this issue is quite reduced at using municipal level data.

One evident way that municipalities might influence federal government decisions is by their population and economic activity. Central cities may receive priority treatment from federal infrastructure policies to attend massive populations and foster national economic growth. However, most part of the Brazilian municipalities are medium or small sized. In 2006, around 4,986 (89,5%) out of 5,570 municipalities had less than 50,000 inhabitants. Only 130 (2,3%) municipalities had a population above 200,000. On average, a municipality had 0,018% of the national population, indicating that the municipal population is unlikely to affect federal public policy decisions directly in the vast majority of cases.

Second, elected federal deputies may request policies more directly from the federal government. In this regard, the representatives may act in favor of the municipalities that gave them the most votes. Each Brazilian State elects a certain number of federal deputies by nominal votes. The deputies may demand financial resources for projects in specific regions and cities, as well as influence legislations and norms that will impact infrastructure policies. In this sense, if we observe a high share of votes for a federal deputy coming from a specific municipality, we can suppose that the deputy is devoting greater efforts to serving his electorate. Nonetheless, this argument seems to hold for a few large municipalities, especially the 27 state capitals. For instance, in the 2006 election in São Paulo – the most populated Brazilian state –an elected federal deputy received an average of 162,827 nominal votes. Only 17,7% of the municipalities of São Paulo had an electorate greater than that in 2006, indicating that, in most cases, a reasonable number of municipalities is needed to elect a candidate. In addition, 434 federal deputies were nationally elected, a considerable number that raises doubts about the political strength of a single deputy to directly influence federal decisions in favor of a specific city.

In this sense, evaluating the impact of federal road policy on local outcomes strongly alleviates broad endogeneity issues. First, it is unlikely that small and medium sized municipalities have any direct political power to influence federal policies in their favor. Second, even if we believe a few large and potentially targeted cities influencing directly federal decisions, we can use the inconsequential unit approach to exclude them, which allows us to keep a large number of "non-targeted" observations.

3.2. Cost-related IV rationality: avoiding (or alleviating) measurement bias of monetary highway investments variables

A way of dealing with measurement error (and maybe omitted variable) bias of road measures is choosing reliable cost-related IVs. Conditional on controls, cost-related IVs – for instance, geographical, environmental, and physical human costs as expropriation and interferences – might affect the outcome variable only through the highway variable (Holl, 2012; Lu *et al.*, 2022; Martín-Barroso, Nunez-Serrano, and Velazquez, 2015; Zhang, Hu, and Lin, 2020). We may expect local terrain ruggedness to make a road project unfeasible, which is expected to affect a region or city's economic development. Nonetheless, this observable characteristic will unlikely directly impact GDP, population, or another outcome variable growth. Similarly, environmental costs – as the extent or existence of legally protected areas – and physical human costs – as expropriation and interferences due to highly urban density – directly affect the feasibility and success of an infrastructure project but are unexpected to directly affect outcome variables.

This kind of IV may avoid (or alleviate) endogeneity bias in two ways. First, they may solve endogeneity issues related to omitted variables bias commonly found in infrastructure-economic development studies, as proposed by past studies (Holl, 2012; Lu *et al.*, 2022; Martín-Barroso, Nunez-Serrano, and Velazquez, 2015; Zhang, Hu, and Lin, 2020). Second, which we consider more reasonable, they might act as a corrective instrument for measurement error bias of highway investment variables. This is particularly relevant when using monetary highway variables as investment flows in developing economies wherein a high inefficiency in allocating infrastructure investments is expected (Caldero n and Serve n, 2014; Straub, 2011).

Then, instrumentalizing the road variable by the main infrastructure costs may reduce both inefficiency bias - it is expected that more costly locations tend to have greater delays in buildings, making infrastructure investment impacts on outcome variables unclear - and road variable (specially the monetary ones) measurement error bias - for instance, more geographical, environmental and physical human costly locations may obviously demand a higher level of investment *per* length of the road, requiring the construction of tunnels, bridges, expropriation and compensation payments, which might be reflected into unclear economic returns.

To better elucidate how cost-related IVs can affect road investments, we use some emblematic Brazilian cases. First, we briefly describe the iconic example of the BR381/262 highway segment crossing the states of Minas Gerais (MG) and Espírito Santo (ES) to contextualize how cost-related inefficiencies occur in practice. The highway is noted for its poor quality and high traffic accident incidence and death rates, and buildings in many road segments have been demanded for decades. According to the CNT 2022 Highway Survey, one of the most critical segments of the highway, between the municipalities of João Monlevade and Martins Soares in the state of Minas Gerais, is in the 345º position in terms of road quality, out of 510 segments evaluated. According to data from the Federal Highway Police (PRF), the BR-381 is the fourth highway in number of accidents and the fifth in accidents with deaths. Due to its critical safety and quality condition, the highway segment is known as the "Highway of Death".

The BR-381/262 (MG/ES) segment is also emblematic because of its several risks and project complexity. The road crosses areas with high population and urban infrastructure density, demanding huge additional monetary resources and time efforts to solve expropriation and interference conflicts. Several delays occurred in the building schedule due to expropriation and evacuation disputes, as many households did not accept the expropriation terms offered by the Brazilian government. Those issues indirectly affected the population's welfare through their direct impact on the efficacy and efficiency of the investments in the highway. The conflictive status of the road required a long time to be solved, and supplemental money was needed to attend the impacted population and maintain unfinished road segments that were not being effectively used by the society.

In 2022, the Brazilian Federal Government tried to grant the BR-381/262 (MG/ES) segment. The result was the auction cancelation due to the lack of private partner interest. It was the fourth time that the “Highway of Death” auction was canceled. The road sector private agents argued that the project was too risky – as it required massive investments and could suffer from several external interferences –, and it would require higher financial compensation to be taken.

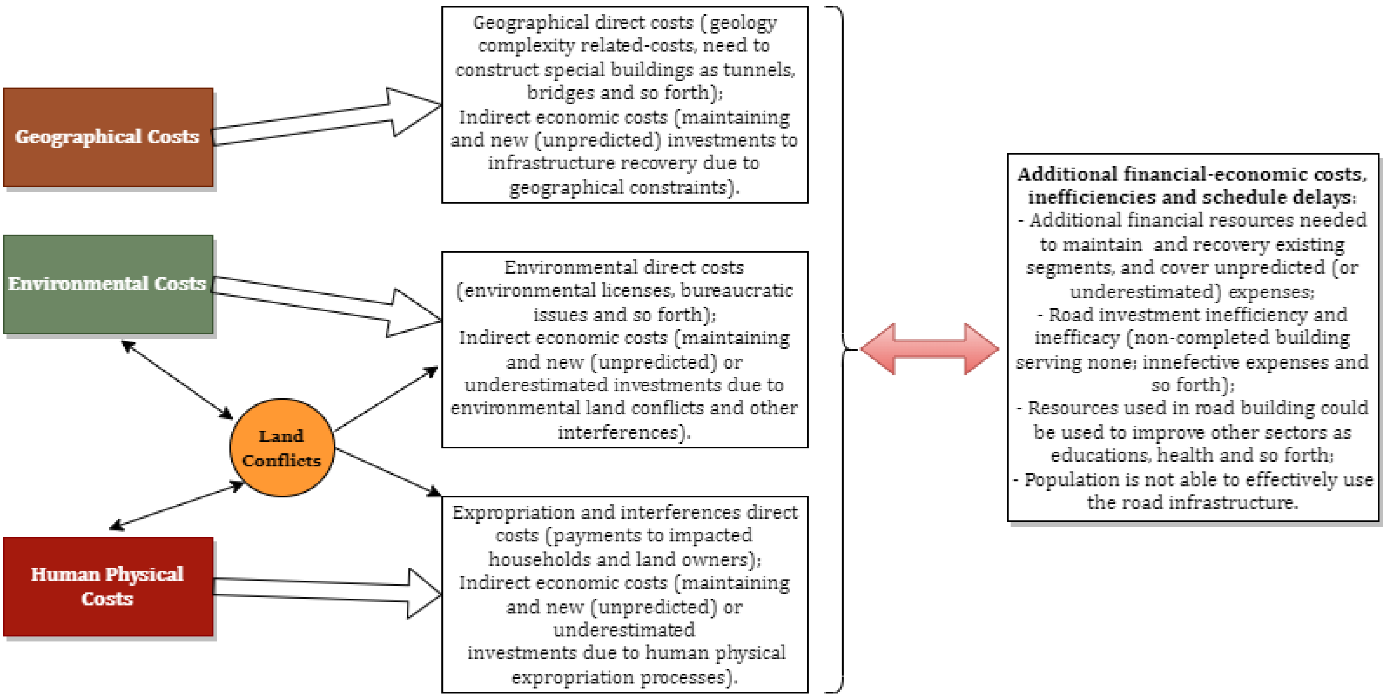
A second interesting case is the BR-163 north segment between the states of Mato Grosso (MT) and Para (PA). This segment was constructed in the 1970s, but it has never been completely paved until recently. The typical road segment picture was kilometers of stuck trucks for weeks during raining periods. As those trucks mostly carried perishable beans to the northern Brazilian ports, they suffered several economic losses strictly associated with poor road conditions.

The BR-163 north segment is also symbolic because it crosses a large legally protected indigenous area. To mitigate potential negative environmental and social impacts of road paving – as deforestation, illegal extractivism and so forth –, the Brazilian government released the “BR-163 Sustainable Plan” (*Plano BR-163 Sustentável*) in 2006. The plan was designed to improve the institutional quality of civil society organizations in the region in terms of its monitoring, evaluation, and information systems, as well as expanding the mechanisms of social participation and control.

Due to its critical environmental sensibility, the BR-163 project received massive demands from environmental and indigenous organizations. To cover some of those requirements, the Environment and Renewable Natural Resources (IBAMA) listed 13 conditions to be met by the National Highway Infrastructure Department (DNIT) in paving the road. Among them, it included a study on the need to build passages for animals, a Monitoring Program for Water Courses, an implementation schedule for the Fire Fighting Program, and that the DNIT should inform IBAMA and the National Institute for Colonization and Agrarian Reform (INCRA) on the occurrence of *quilombola* communities found in the influence of the highway segment. In practice, those needs implied additional economic and bureaucratic costs. In addition, the buildings in the BR-163 north segment were paralyzed several times due to native people's interventions claiming environmental protection and conservation. It incurred in several schedule delays and conflicts over land rights. Even in 2020, some BR-163 north segments were not fully paved.

Another segment of the BR-163 in the Serra da Caixa Furada in the state of Mato Grosso (MT) presented huge geological issues in a duplication building. In 2014, it occurred a landslide on the road, which implied 5 years of complete stoppage in the building. Due to its geology complexity, the building resumption and the reconstruction of the destructed highway demanded several additional financial efforts. The investments applied in the initial stage were rather inefficient as the population could not use the incomplete road duplication. In addition, more money was required to recover the collapsed highway segment and to maintain the already constructed (but not used) road segments. Figure 3 summarizes how geographical, environmental, and physical human costs may impact road infrastructure projects.

Figure 3. Geographical, environmental, and human physical costs in infrastructure projects



Source: authors' elaboration.

The high infrastructure project risk heterogeneity in Brazil, which resulted in empty auctions as in the "Highway of Death" case, forced the Brazilian road sector authorities to rethink the assumptions guiding the rate of return methodology used to establish tariffs and economic-financial balances in the sector. Recent resolutions (6.002/2022, 6.003/2022 and 6.004/2022) on the calculation methodology of the Weighted Average Cost of Capital (WACC) were approved in December 2022 by the National Land Transportation Agency (ANTT).

The pivotal new resolution aimed to include the main risk components in road infrastructure projects to calculate the federal highway sector WACC. To do so, the ANTT established several measurable indicators related to traffic demand, geographical, environmental, and physical human costs of road projects. For instance, the responsible actors must report the road length crossing legally protected areas, urban areas, and hilly areas. Also, the ANTT 6.002/2022 resolution included measures related to road project expropriation and interferences costs, illegal urban occupation, and traffic demand risk.

Whilst the new resolutions focus on new road concession projects – including both brownfield and greenfield projects –, they provide us some reliable (and measurable) indicators that can be adapted to predict the feasibility and correct measurement errors of different types of road investments, as building, paving, duplication and enhancements. As the new ANTT methodology relies on real infrastructure projects, they are based on computable and highly replicable measures. Relying on the infrastructure-development literature (Holl, 2012; Lu *et al.*, 2022; Martín-Barroso, Nunez-Serrano, and Velazquez, 2015; Medeiros *et al.*, 2021b; Zhang, Hu, and Lin, 2020) and measures based on real Brazilian infrastructure project costs, in the next sections we propose some cost-related IVs to overcome endogeneity issues in road investment variables.

In short, we propose several geographical, environmental, and human physical infrastructure project costs to correct measurement errors in highway investment variables. As several inefficiencies are likely to occur in the developing country context, we might expect the highway measure to be inflated, implying a higher value *per* kilometer of road. It puzzles the relationship between road infrastructure and economic activity as a larger number of roads (in monetary terms) is needed to generate an additional unit of output, tending to downward biased OLS road elasticity estimates. By instrumentalizing our highway variable by strong and exogenous cost-related instruments, we intend to fix (or alleviate) measurement error bias.

3.3. Non-random allocation IVs rationality: avoiding (or alleviating) omitted variable bias of highway investment

Even when using cost-related IVs to instrumentalize highway investments and potentially correct measurement errors, omitted variable bias from non-random road placement may persist. Correcting for measurement errors from infrastructure project inefficiencies is just a (important) part of the problem, but it is likely not enough to eliminate endogeneity bias related to the propensity of certain localities to receive federal highway interventions.

First, governments might allocate roads to underdeveloped regions to promote regionally balanced economic growth or directing highway investments to more developed localities – wherein the expected return to road investment is higher – to foster national economic growth. If it occurs, naıv e OLS regressions would be underestimated in the former context, while overestimated in the latter.

To correct those issues, several studies have relied on LCP-MST instruments to generate exogenous sources of variation for highway measures based on a global minimization cost network. We can easily apply the LCP-MST method to Brazilian data as well as to several other economies. In a more general and replicable way, we can stablish targeted points based on road network data. The starting and ending points of a road receiving federal investments are clearly potential hub candidates. From these core cities, we could calculate LCP-MST hypothetical networks and generate instruments for highway measures. For instance, the capital Brasília is the starting point for all radial roads, whilst for longitudinal, transversal, or diagonal roads, starting and ending points represent economic, touristic, or even political hubs. In this way, this approach might provide a comprehensive and highly replicable way of stablishing hubs in studies where targeted cities are not so clear. Another way to identify potential hubs is through historical Census data – for instance, highly populated cities hundreds of years ago – or historical plans. Then, the same LCP-

MST procedure might be applied to generate IVs to correct the non-random placement of road investments.

In addition, the LCP-MST allows us to include the main infrastructure costs in the minimization cost path. In other words, we might be interested in minimizing the global network cost not only based on Euclidean distances, but also on the main geographical, environmental, and human physical costs. As we argued in the previous section, Brazil has several infrastructure costs and risk constraints, and including those characteristics in the minimization LCP-MST procedure can enhance our constructed instrument and empirical strategy. Some studies have used raster data to include geographical costs in the minimization path (Faber, 2014), an approach we improve by adding environmental and human physical indicators.

Second, governments might be guided by political reasons, such as to equalize the territory by expanding the road network. If political allocation bias occurs, we could expect OLS estimates to be biased towards zero. In Brazil, we have the Brasília Plan case proposed by Bird and Straub (2020) to instrument local road improvements in a politically biased road policy context. We also have the Brasília JK Road Cruise, an extension of the Brasília Plan and includes some additional targeted cities. In both cases, the rationale behind the instruments is that the national Brazilian government in the 1950s and 1960s aimed to connect the whole country, having the new capital Brasılia as the central point of the network, and municipalities in the way among Brasília and the endpoints were incidentally connected. In this sense, we can use those historical plans to generate exogenous instruments – based on the distance from the hypothetical lines connecting cities targeted by the plans to Brasília – for highway investments.

Third, by using highway investment flows disaggregated by intervention types – as building, paving, duplications, enhancements and so forth – as we will propose in this study and describe in further detail in the next section, we also need to correct endogeneity bias from the propensity to a locality already connected by a highway in the start period to receive a road intervention. In the Brazilian case, we might anticipate the federal government prioritizing road segments with critical traffic intensity or traffic accidents. In addition, in the context of underdeveloped economies like Brazil, poor institutions, scarce economic resources, and low planning capacity tend to narrow extensive planned road networks around the country, confining infrastructure interventions just to attend infrastructure demanding localities. If it occurs, municipalities crossed by those critical segments are an obviously intended group, and an upward bias in OLS estimates is expected. We can generate "*potential road intervention areas*" instruments based on traffic safety or traffic intensity data, which we believe are available for several countries. The rationality behind this instrument is that, conditional on controls, municipalities already connected by roads in the start period and farther to "*potential road intervention areas*" are more likely to (inconsequentially) receive highway investments to reduce traffic levels and accidents in the critical areas and its surroundings. However, this "luck" at receiving a federal road intervention would be unrelated to economic or political reasons, providing us with a potentially suitable instrument.

In short, we propose a range of LCP-MST, historical and "*potential road intervention areas*" instruments to correct for the non-random placement of roads. As governments might be guided by political and economic reasons, we could observe downward or upward bias in OLS regressions. By using a "free from measurement error" road variable, our proposed non-random allocation instruments have the rough aim of correcting the remaining omitted variable biases.

4. Data

4.1. National highway investments

To measure the impact of national highway investments on local (municipal) economic activity, we constructed a new dataset of national investment flows at the municipal level from 2007 to 2018. To this end, we use two main publicly available datasets.

The first one concerns data on investment flows in the highway sector of the Federal Government's Growth Acceleration Program (PAC). This dataset includes annual information on road investment flows for each of the 27 Brazilian states, including a brief description of each intervention. As an illustrative example, it is reported whether there was a building intervention on Highway 1 between Local X (it can be the name of a municipality, the number of km of a road or other local as a port, a connection with another road infrastructure as state or municipal and so forth) and Local Y. The data allows us to differentiate interventions in road construction, paving, duplication, enhancements, or maintenance.

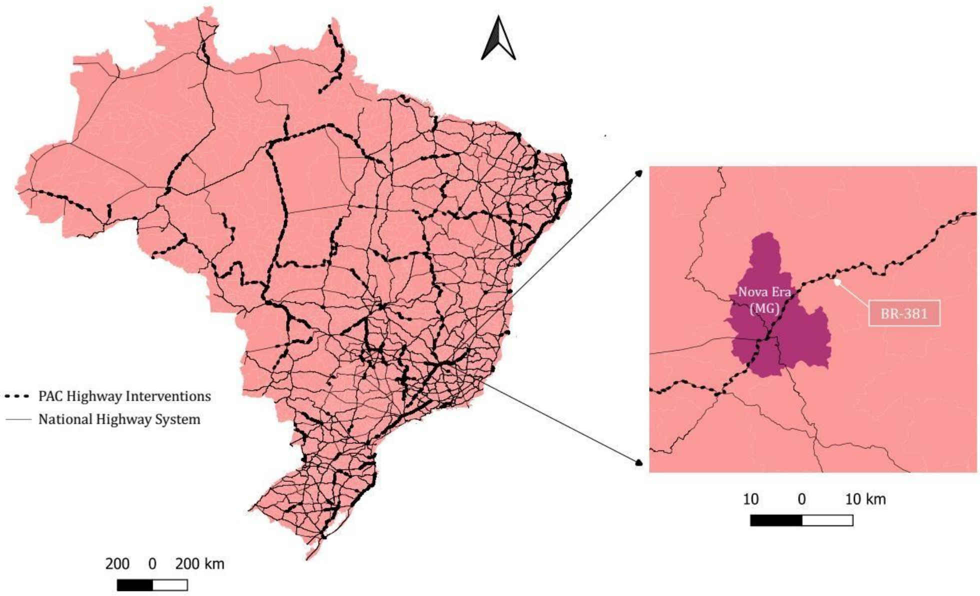
The second one refers to the National Highway System (SNV) georeferenced road data made available by the National Highway Infrastructure Department (DNIT). From this data, it is possible to identify the length of each road segment, its condition (paved, duplicated and so forth), and the places (municipalities) crossed by each of the intervention road segments.

The PAC dataset is not georeferenced, which implies we have restricted information about whether and to which extent a municipality is crossed by an intervention. To create a national highway investment dataset at the municipal level, we combine the PAC's data description of each intervention and the georeferenced SNV data.

The first step was to identify the treated highway codes and their starting and ending points from the intervention description of the PAC data. Next, the PAC treated highways were linked to the SNV geolocalized data using the highway code and their starting and ending points. It should be noted that the starting and ending points of the two datasets are not fully compatible, which made it necessary to match them one by one manually.

Second, we calculated the total PAC intervention road length by municipality and used it to measure the share of the road length in the municipality in relation to the total intervention road length. As we have investment data only by intervention, we use the measured share to compute the highway investment by municipality. It should be noted that the maintenance intervention descriptions barely describe the state and the highway code. In this sense, it was not possible to geolocate this type of investment at the municipal level, and they were excluded from further analysis. Therefore, the investments refer to building and paving, as well as duplications and enhancements. Figure 4 exhibits the PAC highway interventions.

Figure 4. Federal Highway Investments: Municipal Data Construction

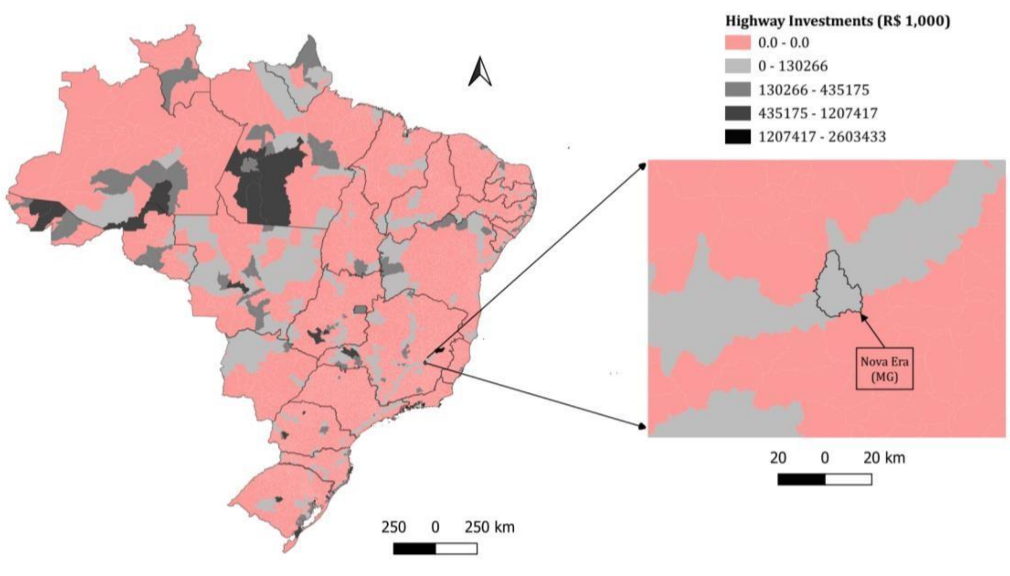


Source: authors' elaboration.

To illustrate how the proportion was measured, we show in Figure 4 the example of the municipality of Nova Era, which was crossed by a PAC intervention in the highway BR-381. The total road length covered by one of the buildings in the BR-381 was 163 km, while the total investment in it in 2009 was R$ 10.8 million. Around 25 km (15%) of the intervention road segment crosses Nova Era. Multiplying the road intervention proportion of 15% for the Nova Era by the total intervention highway investment, the municipality directly received around R$ 1.6 million in investments in 2009. Then, the same procedure was performed annually for all interventions and treated municipalities.

Figure 5 shows the sum of PAC highway investments distributed by municipality between 2007 and 2018. The program directly reached out to 703 municipalities, totaling R$ 77.3 billion, excluding investments in road maintenance. This data is particularly relevant as granular data on infrastructure investment is quite limited (Brooks and Liscow, 2019). In most cases worldwide, data is aggregated at the national level, and it is difficult to standardize.

Figure 5. Federal Highway Investments by Municipality

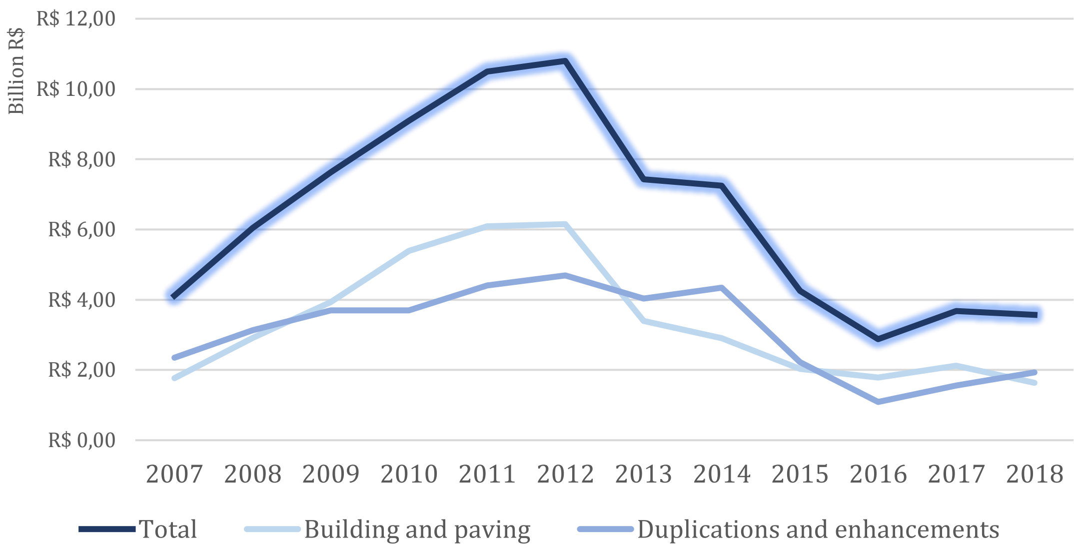


Source: authors' elaboration. Black lines are state boundaries.

Figure 6 shows the highway investment path during the PAC using the constructed data. The federal highway investments more than duplicated in the PAC's period compared to the previous decade (Medeiros *et al.*, 2021b). Our data similarly follows the PAC investments growth pattern[4], presenting a constant rise during the first years. From 2015, the highway PAC investment abruptly dropped following several political and economic Brazilian crises.

Whilst our municipal data seems to be rather representative and provides a novel and important source to measure local impacts of national road investments, it has some practical issues. First, as our investment measure is based on road lengths crossing municipal boundaries, large area municipalities might overestimate highway investments. For instance, a municipality can receive a wide road segment, but its economic center is far from the nearest road. Whether it occurs, we could observe and highway investment overestimation. This problem is critical in the North and Mid-West regions. Second, as exposed by several studies (Caldero n and Serve n, 2014; Straub, 2011), monetary measures such as ours are more likely to be constrained by institutional issues such as inefficiency, corruption, and harmful bureaucracy. In this sense, a suitable identification strategy is needed to measure the causal highway investment impact on local outcomes using this data.

Figure 6. PAC Highway Investments (2022 R$ billion): Georeferenced Municipal Data



Source: authors' elaboration. Notes: the values do not account for investments in maintenance.

As robustness checks, we will also try two additional road variables. The first is a dummy variable assuming value one if the municipality received a road investment during the PAC period, and zero otherwise. The second one is the road length growth rate between 2006 and 2018. In this case, we use 2006 data from the 2007 National Transport Logistics Plan (PNLT) and 2018 data from DNIT[5].

4.2. Infrastructure project cost-related IVs

To construct the cost-related IVS we use a set of variables representing environmental, geographical and expropriation costs at the municipal level. The next subsections describe the proposed variables in detail.

4.2.1. Environmental costs

At the environmental scope, we use georeferenced data of legally protected areas[6] available in the National Registry of Conservation Units (CNUC), maintained by the Ministry of the Environment (MMA). Then, we merge this data with the municipality boundaries shapefile to identify whether a legally protected area intersects a municipality. Our variable is a dummy which assumes value 1 if the municipality is intersected by a legally protected area, and 0 otherwise.

Second, we utilize the environmental embargo terms data of the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) inspection system. Environmental embargoes represent penalties applied to prevent an exploratory activity from continuing. The embargos also serve to prevent ongoing damage and promote environmental recovery. For instance, penalties are applied to prevent activities with the potential to damage the environment, such as deforestation, pollution, and hunting. The application of environmental embargoes by IBAMA occurs mainly in cases where the degradation or damaging activity involves permanent legally protected areas. To generate our variable, we first aggregate the number of embargoes in the five previous years (2002-2006)[7] from the PAC by municipality. Hence, we create a dummy variable assuming a value 1 if there was an environmental embargo in the municipality during this period, and 0 otherwise.

We also use the share of forest area to the total municipality area in 2006 as a third option. This variable came from MAPBIOMAS (Souza *et al.*, 2020) land cover and land use data, which can be obtained at the municipal level. While this variable can capture an important cost of infrastructure building, it may lack variation as Brazil has an enormous forest area. In this sense, weak instrument issues might be expected.

It is important to point out that the institutions accountable for the data are utterly independent from municipalities. The legally protected areas data are administered by the Brazilian (federal) environmental protection system and are controlled by the Chico Mendes Institute for Biodiversity Conservation (ICMBio), as part of the

National System of Nature Conservation Units (SNUC). Similarly, IBAMA is a federal agency related to the MMA. In this sense, concerns about the environmental IVs' exogeneity condition are quite reduced.

The rationale behind the environmental IVs is that they impact the feasibility and the efficiency of highway investment. Conditional on controls, we can expect environmental issues causing delays and rising infrastructure project costs, which may (indirectly) affect outcome variables.

4.2.2. Geographical costs

To construct geographic-related IVs we rely on a few studies measuring infrastructure effects on local outcomes using this kind of identification approach (Holl, 2012; Lu *et al.*, 2022; Martin-Barroso, Nunez-Serrano, and Velazquez, 2015; Medeiros *et al.*, 2022; Zhang, Hu, and Lin, 2020). All those works utilize some measure based on slope, elevation, ruggedness, or altitude.

Our preferred measure is the share of the municipality area with a slope above 20%, which corresponds to hilly areas. This variable is highly related to road construction in the world and in Brazil, as DNIT defines maximum values for slope to be applied to the construction of highways and local roads. The slope is characterized by the relation between a gradient and a corresponding distance on a very small scale, which is unlikely to affect any development outcome directly at the aggregated municipal level. To calculate this variable, we use slope raster data from the National Institute for Space Research (INPE), which allows us to count the number of slope pixels above the 20% cutoff. Then, we generate the share of hilly pixels in relation to the total pixels as our main geographic IV.

We could try several other geographical variables as additional IVs. The first candidate is the municipal average altitude. The second is an elevation-based measure, calculated as the percentage of non-plain areas in the total municipality area (Lu *et al.*, 2022). Both variables are made available by INPE. Other variables could be average rain, temperature, the proportion of mass water in the total municipal area, longitude, and latitude. All those variables are available in a public dataset from the IBGE and the National Information System on Water Resources (SNIRH) of the National Water Agency (ANA).

However, it is important to note that all those potential IVs are much more likely to violate the exclusion restriction than our preferred slope measure. Geographical variables measured as average altitude, elevation, rain, or even temperature have been used to instrumentalize several other independent variables. The most common case is institutions, in which we can cite the study of Iasco-Pereira, Romero and Medeiros (2021) using latitude, longitude, temperature, rainfall and altitude as IVs to municipal institutional quality in a growth equation like ours. If geographical IVs affect economic growth through pathways other than highway building, exclusion restriction is likely to be violated (Felton and Stewart, 2022). In addition, whilst the slope variable has a strict relationship with highway building norms in Brazil, some of the other variables may suffer from an excess of generality. For instance, even in a high (low) altitude or elevation municipality, there exist unlimited (limited) areas with adequate slopes to construct highways.

Like the environmental cost IVs, we expect an indirect impact of geographic costs on outcome variables through the infrastructure measure. Geology complexity tends to raise the level of road investments as well as its efficiency. On the other hand, geographical cost is more easily predicted in the study design phase of the infrastructure project, which might imply that hilly regions should be avoided by planners following engineering guidelines. Then, the relationship between geographical instruments will depend on the opposite forces of the higher costs in geographically complex areas and the ability of planners to avoid those areas.

4.2.3. Expropriation and interferences (human physical) costs

To quantify expropriation (human) costs we use urban infrastructure building, demographic, and land conflict variables. In the urban context, our preferred variable is the share of urban infrastructure[8] building in relation to the total municipality area. For this, we use land use and land cover data made available by MAPBIOMAS (Souza *et al.*, 2020), extracted at the municipal level. Our second measure is populational density, measured as the number of inhabitants divided by the area (km²) of the municipality in 2000. The source is the 2000 Demographic Census of the Brazilian Institute of Geography and Statistics (IBGE).

Those urban related demographic variables represent only a part of the problem. Policymakers will likely avoid highly dense areas precisely to avoid the economic costs of expropriating households and properties. Most of the federal highway extensions run through rural areas. However, these areas can also present different cost levels related to human expropriation and interference issues. As seen in Section 3, there are localities where the population has invaded highway segments claiming their land rights. In this sense, there are also additional costs arising from expropriation and interference in the context of rural and isolated regions.

To represent expropriation and interference costs in the rural context, we use data from the CEDOC Dom Toma s Balduíno of the Pastoral Land Commission (CPT) on land conflicts. More specifically, we get data on households involved in land conflicts in 2006. From that, we aggregate the number of households involved in land conflicts and create a dummy variable that assumes value 1 if the municipality presents one or more land conflicts, and 0 otherwise. This variable is a proxy for expropriation and interference costs in the rural context, as it captures the propensity of a municipality to be impacted by some dispute in rural areas. We argue that the greater the number of households involved in conflicts in the rural context, the greater tends to be the expropriation costs and the larger tends to be the propensity of the population to interfere in potential road buildings crossing the municipality.

All expropriation and interference costs are likely to violate the exclusion restriction for some reason. First, urban agglomeration –sometimes measured as population density, which in turn is highly correlated with urban infrastructure building– can promote economic growth. Second, rural land conflicts might come from disputes on land rights but also involve economic interests such as extractivism and deforestation. If it occurs, we might observe a direct human expropriation and interference effect on economic outcomes. It increases our concern about exclusion restriction and makes critically necessary a careful model specification including several control variables to block potential instrument outcome confounding.

The rationality of these IVs is that physical human costs – as expropriation and interferences – will demand more resources and time to build road projects, conditional on controls. In this way, they also affect outcome variables trough infrastructure. In this case, however, we need to be further careful in including control variables as physical human costs are expected to be highly correlated with demographic variables. To the best of our knowledge, urban infrastructure has not been tested as an urban cost-related instrument before as well as our land conflict variable is a novel measure of expropriations and interferences in the rural context.

4.3. Non-random road allocation IVs

We use three sets of instruments to represent the "*local propensity to receive road interventions*". The first ones are constructed using the LCP-MST method. The second ones are based on the Brasília Plan. The third one is based on traffic intensity data.

4.3.1. LCP-MST IVs

To construct our LCP-MST hypothetical road system, we first need to point out the hubs that the network is supposed to connect through a minimization cost process (Dijkstra, 1959; Kruskal, 1956)[9]. Our preferred strategy considers the starting and ending points of highways receiving PAC investments as hubs. To identify those central cities, we use SNV georeferenced data and generate a dummy variable assuming value 1 if a municipality is a starting or an ending point of a PAC intervention road, and zero otherwise. For instance, the BR-262 presents the municipalities of Cariacica (ES) and Corumba (MS) as starting and ending points. Cariacica is part of the Metropolitan Region of Vitoria (ES), the capital of the state of Espírito Santo (ES), a coastal area containing one of the most important ports of the country (the Vito ria Port). On the other hand, Corumba is a hinterland municipality in the Mid-West region bordering Bolıv ia. In addition, Corumba is one of the most important and richer cities in the Mato Grosso do Sul (MS) state. This example elucidates how the starting and ending points approach might raise hubs of different kinds. Corumba seems to represent a mixture of a political hub – due to its proximity to an international border – and an economic hub – due to its economic importance to the regional economy. In a similar rationality, we can identify several other (74 ending or starting points) hubs related to economic, political, touristic or a combination of those factors and others. This approach gains relevance specially in country contexts wherein road policies have no clear direction, making it hard to predict which places governments aimed to connect.

Second, we also try to stablish hubs based on the centrality of cities. For this end, we use the Regions of Influence of Cities (REGIC/IBGE) survey of 2007. The REGIC classifies municipalities in order of their influence on other cities in terms of goods and services provided – for instance, the existence of universities, airports, health facilities and so forth. Then, municipalities are classified into five classes: metropolis; regional capitals; sub-regional centers; zone centers; and local centers. Local centers are municipalities with less influence on others, while metropolis are the most central Brazilian cities. In 2007, the two up classes were represented by 75 (1,4%) municipalities out of 5,280 municipalities evaluated, a similar number compared to our ending and starting point hubs. To better illustrate the survey hierarchy, Vito ria (ES) is classified as the regional capital (second class) and Corumba (MS) as the zone center (fourth class). Based on the REGIC, we create a hub dummy variable assuming value 1 if the municipality is a metropolis or a regional capital, and zero otherwise. It is important to note that this approach considerably modifies the considered hubs in comparison with using the starting and ending points of intervention roads, suggesting that this survey captures more local-related influence that might be far from the federal government's aims. In addition, this approach is more likely to exclude potential political hubs from being hubs if they are more regionally isolated, as the survey is based on urban centrality and the linkages among cities.

Third, we use population historical data from IBGE to stablish what we call "historical" hubs. We create a dummy variable assuming value 1 if a municipality had a population above 50,000 in 1920, and zero otherwise. This condition is observed for 98 municipalities, which are then considered hubs. If current infrastructure policies are just following a path dependence dynamic - represented by historical variables as the population in 1920- aiming to serve historically important cities with restrained demand for transportation infrastructure, the considered highly populated historical cities are obvious hub candidates. In this case, we obviously observe a concentration of hubs in coastal areas, as those were first populated due to their access facilities and proximity to ports.

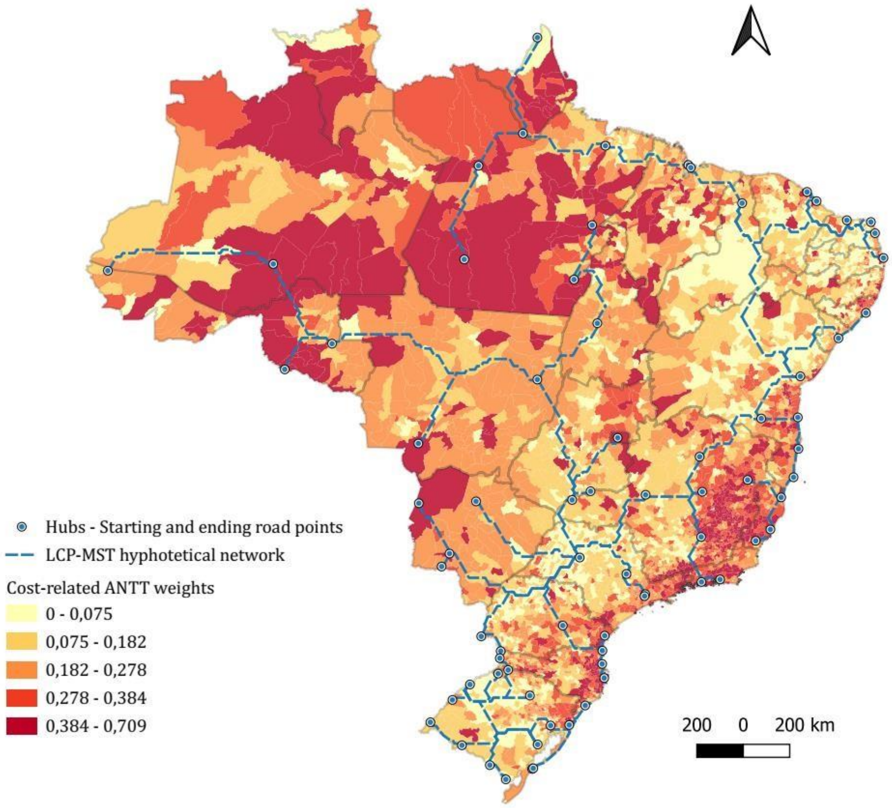
Next, we use the LCP-MST approach to generate hypothetical road networks minimizing the costs of connecting all the stablished hubs for each of our three proposed groups of hubs. First, we generate simpler hypothetical networks based on Euclidean distance (see Figure A1 in Appendix A). Then, we improve the Euclidean distance-based hypothetical networks by including our preferred geographical, environmental, and human physical costs in the LCP-MST minimization cost path.

To create an infrastructure project cost weight, we rely on the ANTT resolutions discussed in earlier sections. ANTT stablished several geographical, environmental, and human physical indicators, among others. For each one of those cost-related measures, the ANTT methodology provided weights capturing their importance on the feasibility of transport projects. We filter the indicators related to geographical, environmental, and human physical costs, and recalculate the importance (weight) of each cost on project feasibility. Environmental costs received a weight of 23.43%, geographical costs 42.68%, and human physical costs 33.89%. Then, we multiply our preferred cost-related measures – sloped area for geographical costs, a weighted average of legally protected areas and embargos for environmental costs, and a weighted average of urban infrastructure and land conflicts for human physical costs – of each cost type by their respective ANTT weights. Our cost index ranges from 0 to 1, and the higher its value the more it costs to build roads. Finally, the cost-related weights interact with the Euclidean distance and this interaction is used to calculate the LCP-MST network for each one of the three potential hub approaches.

It is important to note that those novel infrastructure cost-related hypothetical networks are potential instruments for correcting (or alleviating) both measurement error and omitted variable bias from non-random allocation of road investments, constituting an important contribution of this work. We also try additional cost measures as robustness checks.

From the hypothetical networks, we create our non-random allocation LCP-MST instruments. The instruments are measured as the log form of the distance from the municipality center to the nearest hypothetical road segment. Figure 7 shows the cost-related ANTT LCP-MST network using starting and ending road points as hubs. REGIC and Historical networks can be seen in Figures A2 and A3 in Appendix A.

Figure 7. LCP-MST hypothetical road networks: starting and ending road points using cost-related weights in the cost minimization path



Source: authors' elaboration.

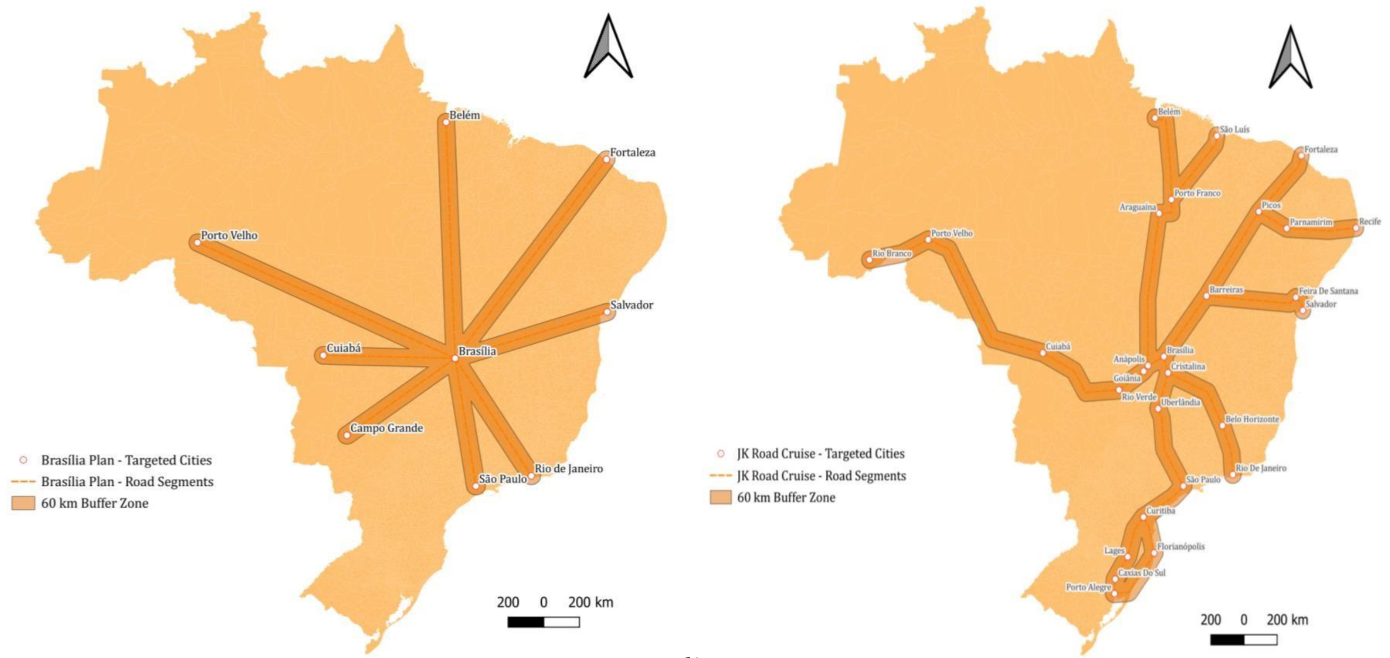
4.3.2. Political IVs

We rely on Bird and Straub (2020) to construct political IVs based on historical plans. Bird and Straub constructed a hypothetical radial (straight line) network connecting the capital Brasília to eight important cities around the country. By linking the capital to those cities, the radial network established corridors, which incidentally connected other localities along the way. As described in section 2.3, the rationality behind this instrument is that, conditional on controls, the Brasília Plan was designed to attend quite different (political) purposes than modern productivity growth.

Our first political IV is the distance from a municipality center to the nearest Brasília Plan segment. We replicate the same index proposed by Bird and Straub (2020) based on buffer zones around the straight lines and the shares of each municipalities' area within each zone.

Our second political IV is based on the Juscelino Kubitschek (JK) Road Cruise, an extension of the Brasília Plan. We digitalized old maps to construct our second historical instrument. The JK Road Cruise's main goals were the same as described by Bird and Straub (2020) for the Brasília Plan, i.e., to serve the future Brazilian capital and to connect the whole country to spread the governmental power on the territory as well as to expand the consumption of domestically produced goods and services (Brasil, 1956a, 1956b, 1958). The JK Road Cruise included more hubs than the Brasília Plan, likely indicating the political bias at developing some specific regional areas not directly connected to Brasília. We follow the procedure in Bird and Straub (2020) and apply it to the JK Cruise network, generating 60 km buffer zones around the lines. Then, our two political IVs are distance-based indices from the municipalities' centers to the lines weighted by the municipal area shares in the buffer zones. Figure 8 exhibits our two historical networks based on the Brasília Plan.

Figure 8. Historical IVs: the Brasília Plan



(b)

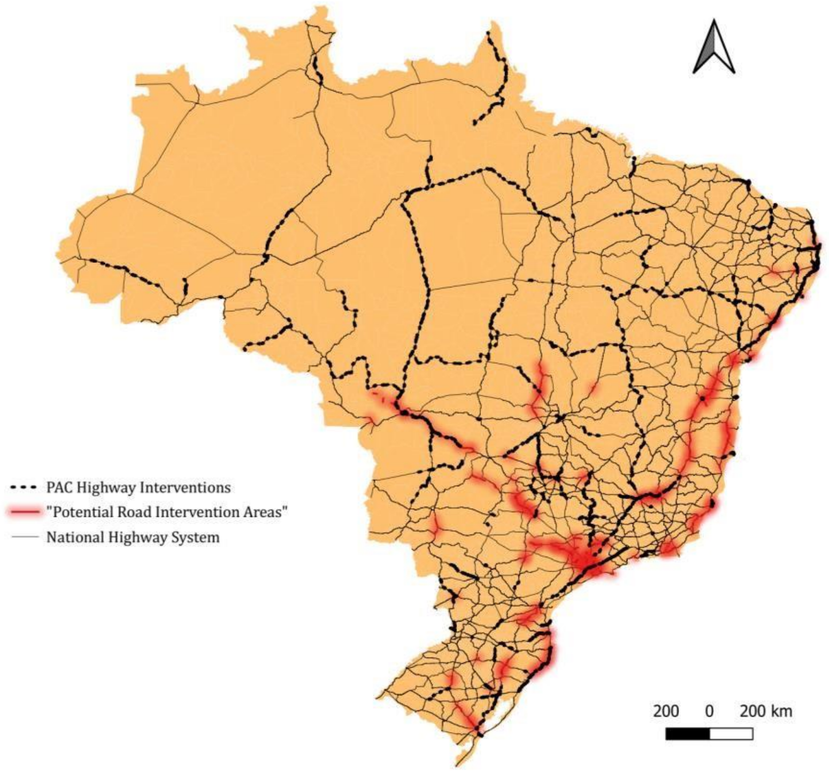
Source: (a) Bird and Straub (2020); (b) authors' elaboration.

4.3.3. "Potential Road Intervention Areas" IV

An additional econometric issue with using highway investment variables is related to intervention heterogeneity. The vast empirical literature on infrastructure, growth and productivity has mainly examined the impact of constructing new roads (Foster *et al.*, 2023a, 2023b). However, in our data, around half of the total federal highway investment was devoted to duplications and enhancements in localities already connected by the national road network in the PAC starting year. To overcome this issue, we calculate an original instrument based on traffic intensity data on federal roads.

First, we use 2007 PNLT data on traffic intensity to identify "*potential road intervention areas*". In this data, road segments are based on the classification A-F of traffic intensity commonly used in transportation engineering studies in Brazil. Based on the vehicle flows on federal roads, highway segments are classified from A (light traffic) to F (heavy traffic, including huge traffic jams). Then, we consider heavy traffic roads those segments classified as D, E and F, being of potential federal government focus. Figure 9 shows the "*potential road intervention areas*".

Figure 9. Potential Road Intervention Areas



Source: authors' elaboration.

Next, our instrument is calculated as the distance from the municipality center to the nearest heavy traffic road segment. The rationale behind this IV is that, conditional on controls, municipalities already connected by federal roads in 2006 and located close, but not so close to "*potential road intervention areas*" are more likely to (inconsequentially) receive highway investments to reduce traffic levels in the critical areas. In this case, including demographic controls is a critical condition to ensure IV validity, as we can expect heavy traffic roads crossing urban agglomerations to be directly impacted by events involving people, such as accidents paralyzing roads and the need for traffic signals reducing car travel speed.

In Figure 9 we can visualize two patterns relating PAC road interventions and the "*potential road intervention areas*". First and foremost, some heavy traffic areas were targeted by the PAC. In this case, we can rely on the inconsequential unit approach and exclude aimed cities. Second, there are interventions close, but not so close to the "*potential road intervention areas*", likely indicating that highway investments were in some part (incidentally) directed to surrounding cities to reduce traffic congestion in the critical points.

4.4. Dependent and control variables

Our main dependent variable is the Gross Domestic Product (GDP) *per capita*. We made this choice based on the related literature in which we observed a massive use of GDP as one of the interest variables[10]. In addition, GDP *per capita* can be used as a *proxy* for productivity, which allows us to calculate the return rate on highway investments and provide a more interpretable result for policy purposes. As robustness checks, we also use employment, firms, and wages as labor market measures.

To avoid omitted variable bias, we include an extensive set of controls (see Table A1 in Appendix A for more details). Our control variables are based on the same group of studies described previously in Figure 1. We also include some controls to capture specificities related to the Brazilian context.

First, we include the initial level of the dependent variable as a control for the municipality development level. Municipalities with different levels of development can present different returns to infrastructure investments. In addition, this variable captures convergence effects (Coscíand Mirra, 2017). It is an important control as policymakers may act to promote balanced economic growth or to foster national growth by targeting developed regions. We also include the share of poor people to control for policies oriented to poverty alleviation, which is a characteristic of most of the PAC period.

Second, as road infrastructure is served by the federal and state level governments, we include state fixed effects to control for regional infrastructure policies. This variable also controls for other types of state fixed effects - as institutional, geographical, environmental, and political. This control is also relevant because the PAC realized several investments in road maintenance. However, this road investment data is only available at the state level and cannot be disaggregated by municipality. In this sense, this fixed effect controls for omitted variable bias of other types of PAC road investments.

Third, the municipality area is included to control for territorial size. This control is critical as our main infrastructure variable is based on the PAC road length crossing a municipality area. Then, including this variable is important to not confuse the true highway investment impact on outcomes with the highway impact in a municipality that has just a large area, then receiving more investments because it has a longer highway extension. In addition, this variable also accounts for the possibility that smaller/larger places may have other systematic differences, such as institutional quality (Bird and Straub, 2021). As an additional municipal size variable, we include the formal workforce measured as the log of the number of formal workers in 2007.

Fourth, we include one important control related to the agricultural sector. The PAC period coincided with a period of basic product price appreciation. Suppose governments aimed at meeting the existing demands of their growing economic sectors. In that case, investments in infrastructure may have been partly due to the performance of the agriculture sector in the period (Medeiros *et al.*, 2021b). Then, municipalities specialized in agriculture may have been more likely to receive federal road investments. In this sense, we used the agricultural share in GDP in 2007 as a control.

Fifth, infrastructure investments are seen as an important determinant of export performance (Coşar and Demir, 2016; Duranton *et al.*, 2014). Exporting municipalities may have received a higher priority in the allocation of national infrastructure investments in the period of good performance in the international market. Then, we include the share of exports of each municipality in the national exports in 2007 as a control.

Sixth, we include a set of controls related to complementary infrastructures. Municipalities well connected with federal and state road networks, as well as port and rail infrastructures, might be able to benefit more from national road investments. For example, municipalities served with high quality state road networks can observe a greater impact of federal highway investments, as these roads may complement other well-established infrastructures. Then, we include the distance (km) to the nearest port, railroad, and state road in 2006 as controls.

Seventh, we control for the historical propensity of a municipality to receive federal road investments. Our variable is the number of railway stations in 1920, the main transportation modality in that period. This variable controls for the municipalities' propensity to receive highway investments because they are historically well located in the country's transportation network. This control is particularly important to the validity of our physical human (interferences and expropriation) IVs as well as our LCP-MST IV based on historical hubs, as they are correlated with demographic and historical variables. In other words, controlling for this variable also alleviates the problem of non-random allocation of federal investments based on past infrastructure. We also control for the distance to the capital Brasília to overcome political related policy aims.

Finally, we include some controls related to the municipal social and institutional background. Those controls are relevant, specially to avoid validity issues on our geographical and expropriation IVs. Geography is the key determinant of climate and natural resource endowments, and it can also play a fundamental role in the disease burden, transport costs, and the extent of diffusion of technology from more advanced areas that societies experience. It, therefore, exerts a strong influence on agricultural productivity and the quality of human resources (Rodrik, Subramanian and Trebbi, 2004). In addition, if expropriation and interferences are related to land conflicts influenced by economic activity as extractivism, a direct effect on economic growth may exist. In this sense, we first include the Index of Municipal Institutional Quality (IQIM)[11] to control for municipal institutions. Second, we include the population share with master's or doctoral degrees as a proxy for local social development. Table A1 in Appendix A summarizes our variables and their respective data sources. Table A2 shows descriptive statistics.

5. Econometric specification

In this section, we present our empirical approach. The *ex-ante* first-step consists of our study design based on disaggregated highway investment data, which evaluates the impact of a national road program on local (municipal) outcomes and alleviates broad endogeneity issues between road investment and economic activity. From that, we apply the next two steps described in detail below.

5.1.1. Correcting measurement error bias (the second step): testing costrelated IVs

Our second step uses the main geographical, environmental, and human physical costs of infrastructure projects as instruments for our highway investment measure.

As we described before, the rationality of this step is to correct for measurement error bias from several inefficiencies in road investment, scuh as poor engineering, and socio-environmental project design. Those inefficiencies lead to measurement errors, as highway investment flows do not fully represent the available road infrastructure for use by the population.

To overcome this issue, we test for suitable cost-related IVs for highway investments. Our main objective is to estimate the long-term effect of improvements in highway networks on our outcome variables at the municipal level. Our main second stage equation is given by:

$$\Delta Y_{is} = \beta_0 + \beta_1 \Delta HighwayInvestments_{is} + \beta_2 X_{is} + \theta_s + \varepsilon_{is}, \quad (1)$$

where $HighwayInvestmentsis$ is the log form of the federal highway investments, $\beta 0$ is a constant term, $\beta 1$ is the elasticity of highway investment to our dependent variable, Yis is our dependent variable for municipality i in state s, Xis is a vector of control variables, θs is a state fixed effect, $\beta 2$ is a vector of parameters related to the control variables and εis is the idiosyncratic error term.

The term Δ represents the change in our dependent variable between 2007 and 2018, i.e., ΔYis is the GDP *per capita* growth rate. As our highway variable is an investment flow measure, the highway change is the sum of the investment flows between 2007 and 2018.

In our second step, the corresponding first stage equation is given by:

$$\Delta HighwayInvestments_{is} = \beta_3 + \beta_4 CostRelatedIV_{is} + \beta_5 X_{is} + \theta_s + \epsilon_{is}, \quad (2)$$

where $CostRelatedIV_{is}$ is a vector containing our cost-related IVs, and β_3, β_4 and β_5 are parameters to be estimated. We run equations 1 and 2 using two-stage least squares (2SLS) and test for IV strength using robust first-stage F-statistics (Imbens, 2014). This approach allows us to identify suitable cost-related IVs for highway investments and to hopefully correct for measurement error bias coming from infrastructure project inefficiencies.

5.1.2. Fixing omitted variable bias (the third step): including non-random allocation IVs

In our third step, we include a set of "*local propensity to receive road interventions*" (non-random road placement) IVs. Then, we rewrite the first-stage equation as follows:

$$\Delta HighwayInvestments_{is} = \beta_6 + \beta_7 CostRelatedIV_{is}^{suitable} + \beta_8 NonRandomAllocationIV_{is} + \beta_9 X_{is} + \theta_s + \epsilon_{is}, \quad (3)$$

where $CostRelatedIV_{is}^{suitable}$ is a vector of suitable cost-related IVs tested in the second step, $NonRandomAllocationIV_{is}$ is a vector of non-random allocation IVs, and β_7 and β_8 are they respective parameter vectors.

As we described earlier, the vector of non-random allocation IVs includes several LCP-MST, political and "*potential road intervention areas*" measures. The first two kinds are more related to the construction of new roads, as they predict a hypothetical road network based on global cost minimization or political reasons. On the other hand, our "*potential road intervention areas*" IV works for those municipalities that were crossed by a federal road in the start period, i. e., they are strictly related to duplications or enhancements of existing roads. To capture those heterogeneities in road investment, we expand our third step first-stage equation as follows:

$$\begin{aligned}\Delta HighwayInvestments_{is} = \beta_{10} + \beta_{11} CostRelatedIV_{is}^{suitable} &+ \\ \beta_{12} D_{is}^{New} \times NonRandomAllocationIV_{is}^{LCP-MST} &+ \\ \beta_{13} D_{is}^{New} \times NonRandomAllocationIV_{is}^{Political} &+ \\ \beta_{14} D_{is}^{Existing} \times NonRandomAllocationIV_{is}^{InterventionAreas} &+ \\ \beta_{15} X_{is} + \theta_s + \epsilon_{is}\end{aligned} \quad (4)$$

where $DisNew$ is a dummy variable assuming value 1 whether a municipality was not connected by a federal road in 2006, and zero otherwise, $DisExisting$ is a dummy variable assuming value 1 whether a municipality was crossed by a federal road in

2006, and zero otherwise, $NonRandomAllocationIVisLCP\text{–}MST$ is a vector of LCP-MST IVs, $NonRandomAllocationIVisPolitical$ is a vector of politically based IVs, $NonRandomAllocationIVisInterventionAreas$ is our "*potential road intervention areas*" IVs, and $\beta 11$ to $\beta 14$ are their respective parameters to be estimated.

By running equations 4 and 1, we expect to correct both measurement error and omitted variables bias of road investments. In addition, by differentiating for new and existing roads in the start period, we provide a novel strategy to estimate the causal impact of highway investments in a multitype road intervention setting.

6. Empirical results

6.1. Correcting measurement error bias and selecting suitable cost-related IVs: the second step

Before we start the econometric analysis, we present some correlations and preliminary tests on our instruments. Figure B1 in Appendix B shows the correlation matrix including the highway investment variables and the proposed instruments. Highway variables are positively associated with environmental and expropriation IVs and negatively with geographical IVs. Obviously, in most cases, IVs correlate more strongly with other IVs of the same type. For instance, urban infrastructure and population density are highly correlated. The same occurs (in lower magnitude) between legally protected areas and environmental embargos or sloped areas and altitude. The lower correlation among cost-related IVs of different types suggests that those variables capture different spectrums of infrastructure project costs. It is important to note that the forest area variable presented a very poor correlation with the highway variables, which places some caution on the following analyses.

Table B1 in Appendix B shows naıv e OLS estimations on the relationship between highway investments and cost-related IVs. We include the full set of controls[12]. Results corroborate the relationships observed in the correlation matrix. As we can see, the estimated coefficients for forest area and populational density are not significant, likely indicating a weak IV issue. Column 10 includes all IVs together, and results remain unchanged. Columns 11-15 includes combinations of our preferred instruments for environmental, expropriation and geographical costs. Results suggest that all our preferred cost-related IVs are significantly correlated with the highway investment variables.

In Table 1, we propose a set of specifications based on our preferred cost-related IVs. As we believe different kinds of costs affect infrastructure investment in different forms and magnitudes, we start by including at least one measure of each cost type by specification. We have not included forest area and population density in Table 1 regressions as they did not influence highway investments. To check instrument strength, we report KP Wald F and *effective* F (Olea and Pflueger, 2013) statistics. Both statistics are robust to heteroskedasticity and weak instruments, and the *effective* F statistic works suitably even in multiple instrument settings (Andrews, Stock and Sun, 2019), as proposed by our identification strategy.

The first stage regressions show that our cost-related IVs are strong predictors of the long-term changes in national highway investments at the municipal level. Nonetheless, some IVs seem to violate exclusion restrictions or unconfoundedness.

Table 1. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: Second Step 2SLS IV Regressions

	1	2	3	4	5	6	7	8	9	10	11
Second stage											
Log Highways Investments	0.0287***	0.0243**	0.0074	0.0024	0.0251***	0.0336***	0.0415***	0.0347**	0.0259***	0.0339***	0.0240***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
First stage											
Legal Protected Areas	0.4580*** (0.10)		0.2859*** (0.10)	0.3914*** (0.10)	0.4478*** (0.10)	0.4997*** (0.10)	0.5096*** (0.10)		0.4255*** (0.10)	0.4719*** (0.10)	
Environmental Embargos		0.3876*** (0.11)						0.4292*** (0.11)	0.3390*** (0.11)	0.3721*** (0.11)	
Sloped Area	-1.4738*** (0.33)	-1.2936*** (0.33)			-1.4695*** (0.33)	-1.7850*** (0.33)	-1.7880*** (0.33)	-1.5968*** (0.33)	-1.5083*** (0.33)	-1.8155*** (0.33)	
Altitude			-0.4280*** (0.06)								
Non-plain Areas				-28.3482*** (9.77)							
Urban Infrastructure	5.0701*** (1.22)	5.2007*** (1.22)	5.0793*** (1.19)	5.0443*** (1.23)	5.0867*** (1.22)				4.9258*** (1.21)		
Land Conflict					0.5528** (0.25)	0.5469** (0.25)					
Cost Index 1											0.0160*** (0.00)
Cost Index 2											0.0329*** (0.01)
Observations	5466	5466	5331	5466	5466	5478	5478	5478	5466	5478	5466
KP Wald F Statistic	21.128	17.582	28.017	15.729	17.215	18.282	24.594	18.772	18.012	19.699	27.275
Effective F Statistic	20.857	18.677	29.628	14.354	16.429	15.538	24.599	18.671	18.795	20.054	33.915
2SLS critical value for tau=5%	21.250	21.256	22.085	25.517	23.039	17.848	3.238	4.208	22.554	14.091	14.888
2SLS critical value for tau=10%	13.299	13.292	13.639	15.597	13.984	11.258	3.118	3.636	13.806	9.009	9.888
R^2	0.17	0.19	0.24	0.23	0.19	0.15	0.10	0.14	0.19	0.15	0.19

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

In columns 1 and 2 we include our preferred measures for each infrastructure cost type. In both cases, *effective* F statistics are not large enough compared to the 2SLS critical value for τ equal to 5%, but they are using τ equal to 10%. Next, we try several other IV combinations in columns 3-10.

In columns 3 and 4 we replace sloped areas for altitude and non-plain areas, respectively. In column 5 we include the land conflict variable together with urban infrastructure, whilst in column 6 we exclude urban infrastructure and maintain land conflict. The low *effective* F statistic in columns 4, 5 and 6 points to weak instruments issues and likely estimation bias. We then investigate the reasons for those issues. In addition, the non-significant and quite smaller highway parameters in columns 3 and 4 raise concerns in using altitude and non-plain areas as IVs for highway investments.

Regarding geographical IVs, our main suggestion is the inadequacy of the altitude – at least as measured, being the municipal average altitude – and the non-plain area variables as IVs. Geography variables, such as average altitude, are used as controls in income equations like ours to capture the direct geography impact (through agricultural productivity, for instance) and the indirect impact (through the institutional quality channel) on economic growth. In our estimations, the direct impact seems to be non-negligible – and, whether it occurs, geography affects both income and highway investments, and estimation bias is expected. This finding cautions against using the average altitude measure as IV in studies about infrastructure and development in underdeveloped countries such as Brazil. On the other hand, our preferred measure (sloped area) is the proportion of hilly areas to the total municipal area, which means that even municipalities located in high (low) average altitude areas might have a large (small) proportion of their area geographically suitable for road investments at lower costs. The differences between these two geographical variables can be seen in the far from high correlation (0.18) between them (Figure A1), suggesting that they predict geographical issues quite distinctly. In addition, the non-plain area presents quite a small variation (mean around 0.99 and standard deviation around 0.01), as Brazil is characterized by small parcels of non-plain areas, and this lack of variation might be affecting econometric results.

On the expropriation and interferences variables, results put caution about using this kind of infrastructure project cost as IV. Urban infrastructure seems to be a suitable but slightly weak IV for highway investments, as its inclusion in the model lowers *effective* F statistics below the critical values in some specifications (it can be seen comparing columns 10 and 9, for instance). Even conditional on several controls, urban infrastructure is likely to impact economic growth directly. Similarly, land conflicts seem to have a non-negligible direct effect on growth, which is likely related to economic related land disputes such as extractivism.

Next, in columns 7 and 8 we try combinations of our preferred environmental and geographical cost IVs, as they seem to be more plausibly exogenous in comparison to the human physical costs. Both specifications are satisfactory, presenting *effective* F statistics suitable at the 5% level. In column 9 we include urban infrastructure together with our suitable IVs based on columns 7 and 8. We can note that, when including the urban infrastructure measure, *effective* F statistic decreases to a level smaller than the 2SLS critical values when choosing τ equal to 5%. This finding corroborates the results obtained in the previous columns and raises some caution in using expropriation and interference measures as IVs to road investments.

Finally, we also try some composite cost indexes (Cost Index 1 and Cost Index 2) based on dimensionality reduction methods[13], as cost types may have some complementary characteristics. In addition, working with a smaller number of instruments might alleviate weak IV and overidentification concerns. We generate composite indexes using our preferred measures – legally protected areas, environmental embargos, sloped areas, and urban infrastructure. Cost Index 1 can be interpreted as an environmental cost index, whilst Cost Index 2 can be seen as a geographical-expropriation index. Results in column 11 show a significant effect of both cost indexes on highways and *effective* F statistics considerably increase.

Then, we turn our analysis to the IVs parameters signals and magnitude. Environmental IVs positively affect federal highway investments. This result was expected as more environmentally costly places might demand a higher level of investments *per* kilometer of road. Similarly, the higher the expropriation and interference costs, the higher the value of investments in federal highways tends to be demanded. On the other hand, geographical costs negatively affect highway investments. One possible explanation is that geographic costs can be more easily predicted from engineering studies in the design phase of the infrastructure project. In this sense, its negative sign may represent the fact that planners are avoiding these regions. On the other hand, expropriation costs have a highly unpredictable aspect as they are impacted by legal disputes involving people. In the same way, environmental costs might raise additional bureaucracy due to environmental licenses not initially planned in the infrastructure project design, as well as lead to new legal disputes involving native populations claiming their rights in legally protected areas.

Regarding the signal and magnitude of the highway investment parameter in the second stage, it corroborates an extensive number of previous studies that found a positive highway investment impact on local outcomes. Based on our specifications with suitable *effective*-F statistics, we find a second step highway investment elasticity between 0.02 and 0.03, which is in the range between 0 and 0.06 found by Foster *et al.* (2023b) based on a huge number of infrastructure econometrics studies. In Foster *et al.* paper, the average elasticity for the transportation sector is 0.03, which brings our estimates even closer to the basis.

Finally, we can discuss the direction of the measurement error bias by comparing our estimations in Table 1 with OLS regressions (see Table B5 in Appendix B). As we expected, the OLS elasticities (around 0.004) are downward biased. Whereas inefficiencies and delays due to geographical, environmental, and human physical issues occur, measurement error takes place by inflating the highway investment variable. In other words, more economic resources are needed to construct or improve a kilometer of road that will be finally used by the population. This measurement error leads to an underestimation of the strength of the (positive) association between transportation infrastructure and economic activity.

By applying our second step identification strategy, we fixed (or alleviated) the expected measurement error underestimation bias. In this sense, our second step parameters may be understood as a kind of "*free from measurement error*" elasticity. Nonetheless, two obvious empirical issues remain. First, if our instruments are not fully or truly capturing the main infrastructure costs leading to inefficiencies, some measurement error bias might remain. This issue seems more troublesome for human physical IVs. Second, non-random allocation bias might exist even after correcting for measurement error in highway variables. The next section presents several tests on the latter bottleneck.

6.2. Solving non-random allocation bias: the third step

Now, we evaluate if the highway investment impacts found so far remain unchanged when correcting for both measurement error and non-random road allocation bias using our proposed LCP-MST, historical and intervention areas IVs. We use our preferred second step (and hopefully "*free from measurement error*") specifications in columns 10 and 11 of Table 1 as our starting point for the third step regressions. Then, we test several specifications using our non-random allocation IVs.

Table 2 summarizes the third step econometric results. We start by inserting our preferred[14] LCP-MST IV using starting and ending road points as hubs (columns 1 and 2). Next, we try our "*potential road intervention areas*" IV (columns 3 and 4) and Brasília Plan (columns 5 and 6) separately. In columns 7 and 8 we include our full set of non-random allocations and suitable cost-related IVs as proposed in equation 4. Finally, in columns 9 and 10 we test a Non-random Allocation Index composed by our preferred LCP-MST, historical and intervention areas IVs[15]. We use the inconsequential unit approach in all specifications, excluding hubs, historical cities and/or critical traffic points.

Table 2. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: Third Step 2SLS IV Regressions

	1	2	3	4	5	6	7	8	9	10
Second stage										
Log Highways Investments	0.0165***	0.0162***	0.0126**	0.0122***	0.0110**	0.0110**	0.0124**	0.0124***	0.0129***	0.0127***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
First stage										
Legal Protected Areas	0.4597***		0.3555***		0.4232***		0.3389***		0.3423***	
	(0.10)		(0.10)		(0.10)		(0.10)		(0.10)	
Environmental Embargos	0.3312***		0.3478***		0.3547***		0.3383***		0.3384***	
	(0.11)		(0.11)		(0.11)		(0.11)		(0.11)	
Sloped Area	-1.3606***		-1.6060***		-1.4378***		-1.4601***		-1.4676***	
	(0.31)		(0.31)		(0.32)		(0.30)		(0.29)	
Cost Index 1		0.2586***		0.2274***		0.2389***		0.2177***		0.2190***
		(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
Cost Index 2		0.3571***		0.4427***		0.3756***		0.3882***		0.3896***
		(0.07)		(0.07)		(0.07)		(0.07)		(0.07)
LCP-MST Starting and Ending Road Points	-0.3875***	-0.3918***					-0.1193***	-0.1204***		
	(0.02)	(0.02)					(0.04)	(0.04)		
Potential Road Intervention Area			0.3252***	0.3294***			0.0893*	0.1025**		
			(0.02)	(0.02)			(0.05)	(0.05)		
Brasília Plan					-0.3170***	-0.3176***	-0.1421***	-0.1314***		
					(0.01)	(0.01)	(0.03)	(0.03)		
Non-random Allocation Index									-0.4994***	-0.5035***
									(0.02)	(0.02)
Observations	5402	5391	5190	5178	5469	5457	5126	5115	5126	5115
KP Wald F Statistic	113.896	156.650	97.221	137.336	123.195	169.934	72.567	89.717	106.925	147.383
Effective F Statistic	104.053	117.056	96.535	108.326	108.318	122.313	73.324	78.795	104.841	112.493
2SLS critical value for tau=5%	20.184	21.202	19.398	20.783	20.688	21.922	24.187	26.112	20.345	21.008
R^2	0.22	0.22	0.24	0.24	0.23	0.23	0.23	0.23	0.23	0.23

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01

All our proposed non-random road allocation IVs are strongly correlated with the highway investment variable. First-stage results show a consistent correlation between instruments and highway investments and suggest that our identification strategy seems solid at identifying the causal impacts of road investments on local economic activity. *Effective*-F statistics severely increase by including non-random allocation IVs, even after solving (or alleviating) measurement error bias. In this sense, our findings restate the critical role in fixing endogeneity bias coming from the non-random placement of highway investments, specially using monetary variables. In addition, results indicate that our full econometric strategy following equations 1 and 4 is suitable for measuring causal road impacts on local productivity in a context of road intervention heterogeneity. In contrast, findings cautioned using onlycost-related IVs in contexts of non-random road policy allocation and seem to indicate cost-related IVs solving (or alleviating) just a (relevant) part of the problem.

Now, we turn our analysis to the IV parameters' direction and magnitude. Starting with our LCP-MST IV, we find an expected negative relationship with highway investment, indicating that municipalities far from the hypothetical LCP-MST network received a smaller amount of road investments. The same holds for our Brasília Plan IV, and the interpretation is quite similar, suggesting that localities far from the plan lines received less investments. On the "*potential intervention road areas*" IV, the positive parameters suggest that, for already highway connected municipalities in 2006, the greater the distance from a critical traffic area, the greater the investment received. This result points out that planners seem to avoid critical road segments by directing intervention to road segments not so close to critical areas. It might be related to the potential correlation between population density and other human physical costs and critical traffic points, which is likely impacted by higher economic costs and inefficiency in those segments. Then, critical road segments appear to be a strategic (and likely endogenous) feature in road policies and using the inconsequential unit approach is overarching in this way.

Regarding second stage results, elasticities are stable in a range from 0.011 to 0.017. In other words, a one percent increase in federal highway investment boosts GDP *per capita* by 0.011-0.017 percent. It is important to note that the third step elasticities are not negligibly lower than those in the second step (between 0.02 and 0.03). This result suggests that, even correcting for measurement error bias, a nonrandom allocation overestimation bias might remain, and our identification strategy is important to rule those issues out.

On the direction of the bias, an analysis through the PAC is quite difficult. The program had several and different goals, including connecting underdeveloped regions to promote regionally balanced growth as well as attending developed regions with restrained demand for transport investments to foster national economic growth. Complementary, those motivations might be mixed up with political, predetermined historical infrastructure and/or touristic reasons, which is rather hard to distinguish.

Nonetheless, we can provide some elucidation starting from our second step estimates (Table 1). Suppose that our preferred second step regressions fully correct for measurement error bias, but do not so for non-random allocation bias. Then, we have a (potentially biased) start point elasticity around 0.02 and 0.03, which we compare with our smaller third step elasticities (0.011-0.017). The apparently overestimated second step parameters suggest that, "*free from measurement error bias*", PAC planners targeted comparatively more places with higher expected returns to infrastructure investments and/or higher expected traffic demand (Baum-Snow *et al.*, 2017; Faber 2014). It seems to be true as almost half of the total PAC highway investments were allocated to road enhancements and duplications in already connected municipalities in 2006. Regarding the other half, we expect a mix of underdeveloped oriented policy – specially in the poorly connected North region – and higher expected returns to infrastructure policy – mainly in export and agriculture specialized regions. Medeiros *et al.* (2021) found that highway investments have higher returns for low-technology-intensive sectors such as agriculture and mining in Brazil. In this sense, the economically more developed oriented nature of the PAC seems to be stronger than its regionally balanced growth aims, upward biasing estimates "*free from measurement error bias*" but not solving non-random allocation issues. Indeed, this interpretation is more suggestive than a statement, as we do not completely know whether cost-related IVs are solving some share of the non-random allocation bias or if they are entirely fixing measurement error bias. We continue this discussion in section 7.

6.3. Robustness Checks

In this section, we present several robustness checks to our main results shown in the previous two topics. First, we propose a falsification test based on a novel planned road sample exercise. Second, we try other commonly used dependent variables. Third, we test two different measures of road infrastructure. Fourth, we run additional robustness checks using the limited information maximum likelihood (LIML) estimator, bootstrap standard errors, and excluding potential outliers.

6.3.1. Falsification test

If our instruments are valid, they should affect the outcome only through the highway investment variable. Therefore, cost-related and non-random road allocation IVs should have no effect on local economic outcomes that are not in the highway investment pathway. A way to test the exclusion restriction is by using falsification tests based on alternative study samples (Felton and Stewart, 2022; Pizer, 2016). A useful falsification sample would not be exposed to the treatment (highway investment), but it would be subjected to all the potential confounders that might be correlated with the instrument and the outcome, like demographic, economic, complementary infrastructure, and institutional features.

To test the exclusion restriction, we specify a sample closely related to our study population that is not receiving highway investments. First, we selected all municipalities that had federally planned roads in 2006. We extracted this data from the same 2007 PNLT shapefile. Second, we use the 2018 SNV shapefile to compare[16] whether those planned roads in 2006 were still classified as planned in 2018. Next, we excluded those municipalities that presented planned roads for both years and received PAC investments, as it could capture an inefficiency or a long-time to build infrastructure issue.

Following the DNIT road terminology, planned highways are those roads that do not physically exist yet. Practically, planned roads represent a hypothetical highway acting as a guideline intended to meet potential traffic demand. In this sense, our rationality behind this falsification test is that places (hypothetically) crossed by planned roads were similarly demanding road infrastructure interventions, likely exposed to the same potential confounders as the full sample. As none of those municipalities received federal highway interventions, the falsification test is performed by including the IVs in an alternative specification of the outcome equation. Table C2 in Appendix C summarizes the results. We run the same specifications as in Table 2 using the planned roads sample and include our preferred cost-related IVs separately as well (columns 1 and 2). The results show that GDP *per capita* growth is not correlated with our proposed cost-related IVs in the falsification sample between 2007 and 2018[17]. Conditional on controls, results suggest that our suitable cost-related and non-random allocation IVs do not violate exclusion restrictions. These results increase confidence in our identification strategy.

6.3.2. Outcomes

In this section, we estimate the impact of federal highway investments on employment, firms, and wages. This test is important to validate our empirical strategy and to confirm the influence of highway investment on local outcomes. We use the econometric specification in column 10 of Table 2 and apply it to each one of the dependent variables using formal labor market data from RAIS/MTE.

Table C3 in Appendix C shows the results. For all tested dependent variables, the cost-related and non-random allocation IVs are strong predictors of national highway investments. In addition, the signal and significance of the first stage coefficients remained quite similar, as well as remained positive the road impact on outcomes. Our results suggest that our identification strategy is suitable not only for the GDP *per capita* variable, but it might also be suitable for predicting the highway investment impacts on several other labor market local outcomes.

6.3.3. Highway investment measure

Several studies correctly argue that monetary variables - such as the investment flows used in our study - may contain several measurement errors (Caldero n and Serve n, 2014; Kenny, 2009; Straub, 2011). To increase the reliability of our empirical strategy, we also tried a dummy (intervention) measure of road infrastructure. Related studies have used similar measures to capture access to highways (Asher and Novosad, 2017; Frye, 2016; Michaels, 2008; Percoco, 2015; Zhang, Hu and Lin, 2020). Our dummy variable assumes value 1 if the municipality was crossed by a PAC highway intervention, and zero otherwise. If there are too high measurement errors in our preferred investment flow variable, the intervention dummy variable might alleviate the problem as it no longer contains monetary values. On the other hand, the intervention variable gives the same weight to all municipalities receiving highway investments, which is a not negligible empirical issue. Second, we try a road length variable following a vast strand of literature (Baum-Snow *et al.*, 2020; Duranton *et al.*, 2014; Foster *et al.*, 2023a, 2023b; Straub, 2011). Tables C4 and C5 in Appendix C summarize the results using the same specifications following Table 2. As expected, the highway intervention parameters are positive and significant in all specifications, and the same holds for road length. The cost-related and non-random allocation IVs work in the same way as using continuous highway investment flows, corroborating previous estimates.

6.3.4. Additional robustness checks

To raise confidence in our main estimates, we run the same specifications of Table 2 using the LIML estimator (Anderson and Rubin, 1949). In the overidentification scenario, proposed by our identification strategy, 2SLS estimates might be biased towards OLS as the bias is proportional to the degree of overidentification (Angrist and Krueger, 2001). In overidentified models, LIML is approximately unbiased because the median of its sampling distribution is generally close to the estimated population parameter. In addition, we also provide bootstrap confidence intervals. Young (2022) finds that bootstrapped confidence intervals perform better in real-world settings as heteroscedasticity and weak IV assumptions are likely violated. Tables D1 and D2 in Appendix D exhibit the results. In both cases, the findings remain unchanged, indicating that our main estimations are reliable.

Next, we try several specifications excluding potential outliers. In this set of estimations, we use column 10 of Table 2 as our benchmark specification. Results are described in Table D3 in Appendix D. In column 1, we drop all municipalities of the state of Sa o Paulo. Sa o Paulo is the richest Brazilian state, representing more than 30% of the national GDP. In addition, Sa o Paulo presents the best road infrastructure in the country, which is in substantial part privately managed. The major share of those high-quality roads is the responsibility of the Sa o Paulo state government, with federal roads a small fraction of the total. The result might be a small fraction of the PAC highway investment directed to a high road demanding state, and an underestimation bias could be expected. In column 2, we exclude all municipalities in the states of the North region. Those municipalities are characterized by large territorial areas, which might bias our highway measure variable as it depends on the road length crossing the municipalities. In column 3, we exclude both Sa o Paulo and Northern municipalities. In column 4, we consider highway investment values smaller than R$ 50 million to be zero. This test is important as we relied on road length crossing municipal areas to construct our highway flow measure, and short road segments (and consequentially small investment values) might be poorly capturing a highway intervention. Finally, in column 5 we exclude municipalities in the top 1 and bottom 1 percentiles of GDP *per capita* growth. In general, findings remain almost unchanged. The most noticeable variation comes from the exclusion of Sa o Paulo municipalities. In columns 1 and 3, the elasticity is around 0.017, whilst our benchmark estimate is 0.012. This result suggests that municipalities of Sa o Paulo might be slightly downward biasing our estimates.

7. Assessing the return rate to highway investments: how large can the bias be?

Now, we return to our main issue: endogeneity bias. We can ask some questions. First, is there any remaining bias in our second step elasticity? If there is any, then what is the magnitude of such a bias? To answer these questions, we compare our preferred third step estimates with second step and OLS naıv e estimates (Table 1 and Table B5 in Appendix B, respectively).

The OLS estimates returned a highway investment elasticity of around 0.004, while the second step elasticity is close to 0.025. On the other hand, the highway investment elasticity estimated using our full third step IV identification strategy is in the range between 0.011 and 0.017. Our results suggest a meaningful bias in OLS regressions and point to a remaining non-random allocation bias in our second step estimates.

To better illustrate the bias, we calculate the return rate to highway investment in Brazil comparing the OLS and our preferred 2SLS-IV estimates. To do so, we follow the equation below (Fernald, 1999; Wang, Wu, and Feng, 2020):

$$RR = \beta_{highway} * \frac{GDP}{Highway\ Stock} \tag{5}$$

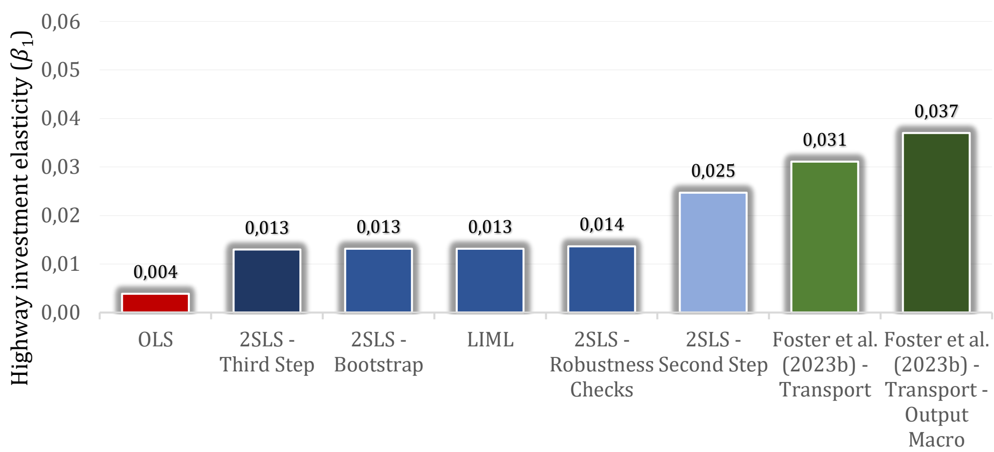
Where RR is the return rate to highway investments, $\beta highway$ is the predicted elasticity following equations 1 to 4 in the econometric specification section, GDP is the national GDP and Highway Stock measures, in monetary terms, how much the national highway infrastructure stock is worth. As a proxy for the highway stock, we follow Medeiros *et al.* (2021) and use the sectoral infrastructure stock estimated by Frischtak and Moura o (2017). The study provides a quite stable road stock of 6.0% of the GDP. The Brazilian GDP was around R$ 9.9 trillion in 2022. Considering the road stock share in GDP constant over time, the actual Brazilian highway stock is supposed to be approximately R$ 594 billion, implying a GDP/Highway Stock ratio of approximately 16.7. This ratio is very close to the Chinese one of 19.6 used by Wang, Wu, and Feng (2020).

In Figure 10, we summarize our results regarding elasticities and return rates. Then, Figure 10 shows us the elasticity and return rate by estimation group, which is measured by the average considering all their respective estimates. In addition, we include Foster *et al.* (2023b) transport sector elasticities as benchmarks for comparison as well as the return rates found by Lí*et al.* (2017) and Wang, Wu, and Feng (2020) for China, and Medeiros *et al.* (2021) for Brazil.

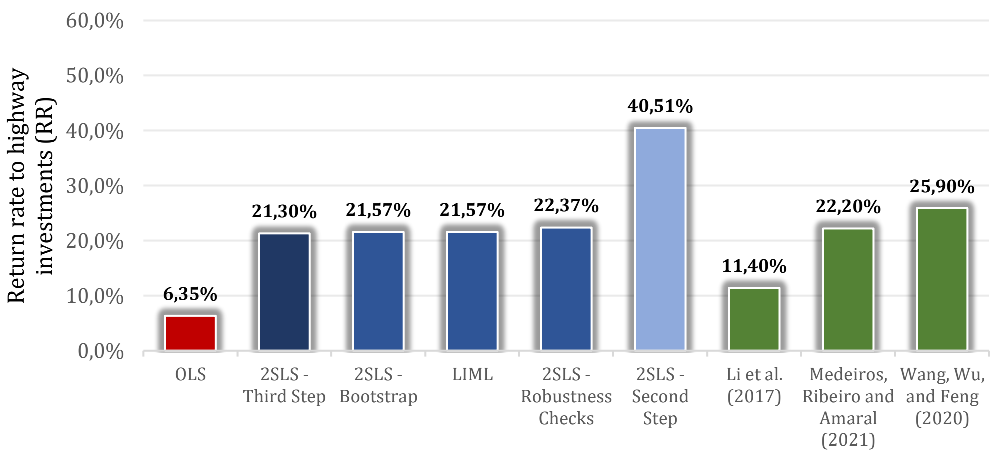
Using the naıv e OLS highway elasticity, we generate an underestimated RR of 6.35%. In contrast, using the 2SLS estimates based on the second step cost-related IVs identification strategy, we calculate an RR of around 40.51%, which we consider rather high even for a developing and lacking highway infrastructure economy such as Brazil. Nonetheless, when we turn to our third step estimations – including all robustness checks we have run –, we find a stable average elasticity around 0.013, implying a return rate to highway investment of approximately 21.3%. This result suggests that correcting both measurement error and non-random allocation bias is critical to identifying robust infrastructure investment elasticities and return rates. In addition, findings put much caution on instrumentalizing highway investment variables only by cost-related measures in cases wherein road policies are likely guided by economic, political or need for transportation reasons as seems to have occurred in Brazil during the PAC.

Whilst the methodology and data are different and making a full comparison among studies is impossible, our calculated RR is in line with those ones calculated by Líet *al.* (2017) and Wang, Wu, and Feng (2020) for the Chinese case, and specially with Medeiros, Ribeiro and Amaral (2021) for the Brazilian case. To evaluate the rentability of highway investments in Brazil, we can take the Social Discount Rate (TSD) of 8.5% calculated by the Ministry of Economy (2021) and broadly used to evaluate infrastructure projects in the country. To reduce our third step rate of return of 21.3% to the threshold of 8.5%, Brazil would need 2.51 times more highways, which implies a road stock of 15.1% of national GDP. This finding corroborates the estimates by Frischtak and Moura o (2017) and Medeiros, Ribeiro and Amaral (2021) predicting an ideal road stock of 13.5% and 16% of GDP, respectively. Considering even smaller long-run real rates of return from 4% to 5% worldwide, Brazil would need to invest 4.26 times more in highway infrastructure, strengthening the high rentability of transport investments in the country during the PAC period.

Figure 10. Elasticities (a) and return rates to highway investments (b): Brazil, 2007-2018



(a)



(b)

Source: authors' elaboration.

8. Concluding remarks

Using a novel three-step IV identification strategy, we evaluated the causal impact of national highway investments on local economic activity in Brazil between 2007 and 2018. We relied on several original infrastructure project costs and non-random road placement instruments as sources of quasi-random variation of observed highway infrastructure to correct for both measurement error and omitted variable biases. We proved that our preferred IVs are conditionally exogenous and strong predictors of road investments. In addition, we argue that a substantial part of our proposed instruments might be replicated in several applications worldwide.

Our main results showed that our empirical strategy is suitable for identifying causal road impacts on the local economy. First, the second step estimates allowed us to correct (or alleviate) measurement errors and to identify expected downward biased OLS elasticities. Second, third step regressions are critical to fixing the non-random placement of road bias even after fixing measurement error. In this case, our findings suggest that the PAC prioritizes economic development over regional balance, as its actions favor more prosperous areas, thus upward biasing "free from measurement error" second step estimates. Results remain unchanged under several robustness checks.

We found a quite stable road elasticity ranging from 0.011 to 0.017. In other words, a one percent increase in highway investments increased municipal GDP *per capita* by 0.011-0.017%. Using a more easy-to-interpret measure, results point to a return rate to highway investment of around 21.3% in Brazil, which aligns with studies worldwide. Brazil would need to invest 2.5 times more in road infrastructure from this return rate, reaching a proper road stock of 15% of the national GDP. This result proved the high rentability of highway investment in the developing country context.

While we have contributed to the empirical literature on infrastructure and economic activity in several ways, some gaps remain. First, it is hard to conclude if our second step estimates are fully corrected for measurement errors in the road variable. In addition, cost-related IVs could fix some parts of the expected omitted variable bias, and additional validations might be useful. Proposing other cost-related instruments and statistically testing them is an obvious way to confirm our identification strategy and results. Second, a similar empirical approach could be adapted for panel data models. In this case, we could include other cost types such as national demand and institutional risks, an issue we overlooked in the present work as we worked with crosssectional data. We hope to provide some new evidence on those issues in further research.

References

Adler, M. W., Pasidis, I., Levkovich, O., Lembcke, A. C., & Ahrend, R. (2020). Roads, market access and regional economic development.

Albalate, D., & Fageda, X. (2016). High-technology employment and transportation: Evidence from the European regions. *Regional studies*, 50(9), 1564-1578.

Allen, T., & Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3), 1085-1140.

Amann, E., Baer, W., Trebat, T., & Lora, J. V. (2016). Infrastructure and its role in Brazil's development process. *The Quarterly Review of Economics and Finance*, 62, 66-73.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of mathematical statistics*, 20(1), 46-63.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11, 727-753.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4), 69-85.

Aschauer, D. A. (1989). Does public capital crowd out private capital? *Journal of monetary economics*, 24(2), 171-188.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American economic review*, 110(3), 797-823.

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145, 102442.

Barzin, S., D'Costa, S., & Graham, D. J. (2018). A pseudo–panel approach to estimating dynamic effects of road infrastructure on firm performance in a developing country context. *Regional Science and Urban Economics*, 70, 20-34.

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

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

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103124.

Bird, J., & Straub, S. (2020). The Brasilia experiment: the heterogeneous impact of road access on spatial development in Brazil. *World Development*, 127, 104739.

Brasil. Exposição geral da situação econômica do Brasil – 1957. Rio de Janeiro: Conselho Nacional de Economia, 1958.

Brasil. Política de transportes e comunicações do governo do presidente Juscelino Kubitschek de Oliveira no quinquênio 1956-60. Rio de Janeiro: Departamento de Imprensa Nacional/MVOP, 1956a.

Brasil. Plano Quinquenal de Obras Rodovia rias Federais – 1956/60. Rio de Janeiro: MVOP/DNER, 1956b.

Brooks, L., & Liscow, Z. (2019). Is Infrastructure Spending Like Other Spending? In: *Economics of Infrastructure Investment*. University of Chicago Press.

Caldero n, C., & Serve n, L. (2014). Infrastructure, growth, and inequality: An overview. World Bank Policy Research Working Paper, (7034).

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.

Coşar, A. K., Demir, B., Ghose, D., & Young, N. (2022). Road capacity, domestic trade, and regional outcomes. *Journal of Economic Geography*, 22(5), 901-929.

Cosci, S., & Mirra, L. (2018). A spatial analysis of growth and convergence in Italian provinces: the role of road infrastructure. *Regional Studies*, 52(4), 516-527.

Dijkstra, E. W. (2022). A note on two problems in connexion with graphs. In Edsger Wybe Dijkstra: His Life, Work, and Legacy (pp. 287-290).

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

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79(4), 1407-1440.

Duranton, G., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. *Review of Economic Studies*, 81(2), 681-724.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81(3), 1046-1070.

Farhadi, M. (2015). Transport infrastructure and long-run economic growth in OECD countries. *Transportation Research Part A: Policy and Practice*, 74, 73-90.

Fedderke, J. W., & Bogetic , Z . (2009). Infrastructure and growth in South Africa: Direct and indirect productivity impacts of 19 infrastructure measures. *World Development*, 37(9), 1522-1539.

Felton, C., & Stewart, B. M. (2022). Handle with Care: A Sociologist's Guide to Causal Inference with Instrumental Variables.

Fernald, J. G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. *American economic review*, 89(3), 619-638.

Foster, V., Gorgulu, N., Straub, S., & Vagliasindi, M. (2023a). The Impact of Infrastructure on Development Outcomes: A Qualitative Review of Four Decades of Literature.

Foster, V., Gorgulu, N., Straub, S., & Vagliasindi, M. (2023b). The Impact of Infrastructure on Development Outcomes: A Meta-Analysis.

Frischtak, C. R., & Moura o, J. (2017). O Estoque de Capital de Infraestrutura no Brasil: uma abordagem setorial. Desafios da Nação, 133.

Frye, D. (2016). Transportation networks and the geographic concentration of employment.

Garcia-Lopez, M. A ., Holl, A., & Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, 85, 52-67.

Gertler, P. J., Gonzalez-Navarro, M., Gracner, T., & Rothenberg, A. D. (2019). Road Quality, Local Economic Activity, and Welfare: Evidence from Indonesia's Highways.

Gertler, P., Gonzalez-Navarro, M., Gracner, T., & Rothenberg, A. D. (2022). Road Maintenance and Local Economic Development: Evidence from Indonesia's Highways (No. w30454). National Bureau of Economic Research.

Ghani, E., Goswami, A. G., & Kerr, W. R. (2016). Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. *The Economic Journal*, 126(591), 317-357.

Herzog, I. (2021). National transportation networks, market access, and regional economic growth. *Journal of Urban Economics*, 122, 103316.

Holl, A. (2012). Market potential and firm-level productivity in Spain. *Journal of Economic Geography*, 12(6), 1191-1215.

Holl, A. (2016). Highways and productivity in manufacturing firms. *Journal of Urban Economics*, 93, 131-151.

Hsu, W. T., & Zhang, H. (2014). The fundamental law of highway congestion revisited: Evidence from national expressways in Japan. *Journal of Urban Economics*, 81, 6576.

Huang, Y., & Xiong, W. (2017). Geographic Distribution of Firm Productivity and Production: A ", Market Access" Approach. Working Paper.

Iasco Pereira, H. C., Romero, J. P., & Medeiros, V. (2021). Kaldor–Verdoorn's law and institutions: evidence from Brazilian municipalities. *Cambridge Journal of Economics*, 45(3), 511-536.

Imbens, G. (2014). Instrumental variables: an econometrician's perspective (No. w19983). National Bureau of Economic Research.

Jaworski, T., & Kitchens, C. T. (2019). National policy for regional development: Historical evidence from Appalachian highways. *Review of Economics and Statistics*, 101(5), 777-790.

Ke, X., & Yan, Y. (2021). Can proactive fiscal policy achieve the goal of "Beyond Keynesianism"? Evidence from China's National Highway System. *Review of Development Economics*, 25(2), 1078-1103.

Kenny, C. (2009). Measuring corruption in infrastructure: Evidence from transition and developing countries. *The Journal of Development Studies*, 45(3), 314-332.

Kornfeld, R., & Fraumeni, B. M. (2022). How Should We Measure Infrastructure? The Case of Highways and Streets (No. w30045). National Bureau of Economic Research.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1), 4850.

Lee, J. K. (2021). Transport infrastructure investment, accessibility change and firm productivity: Evidence from the Seoul region. *Journal of Transport Geography*, 96, 103182.

Li, Z., Wu, M., & Chen, B. R. (2017). Is road infrastructure investment in China excessive? Evidence from productivity of firms. *Regional Science and Urban Economics*, 65, 116-126.

Lu, H., Zhao, P., Hu, H., Zeng, L., Wu, K. S., & Lv, D. (2022). Transport infrastructure and urban-rural income disparity: A municipal-level analysis in China. *Journal of Transport Geography*, 99, 103292.

Martín-Barroso, D., Núñez-Serrano, J. A., & Velázquez, F. J. (2015). The effect of accessibility on productivity in Spanish manufacturing firms. *Journal of Regional Science*, 55(5), 708-735.

Martincus, C. V., Carballo, J., & Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. *Journal of Development Economics*, 125, 2139.

Medeiros, V., Ribeiro, R. S., & AMARAL, P. (2021a). Productivity of transportation infrastructure in Brazil: a sectoral and regional approach using dynamic panel data models. Coletânea de Trabalhos–Prêmio ABDE-BID: edição.

Medeiros, V., de Queiroz Stein, A., Ribeiro, R. S. M., & Dos Santos, U. P. (2021b). Transportation infrastructure and University-Industry collaborations: regional level evidence from Brazil. Blucher Engineering Proceedings.

Meijers, E., Hoekstra, J., Leijten, M., Louw, E., & Spaans, M. (2012). Connecting the periphery: Distributive effects of new infrastructure. *Journal of Transport Geography*, 22, 187-198.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4), 683701.

Nassif, A., Feijo , C., & Araújo, E. (2015). Structural change and economic development: is Brazil catching up or falling behind? *Cambridge Journal of Economics*, 39(5), 1307-1332.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3), 358-369.

Ottaviano, G. I. (2008). Infrastructure and economic geography: An overview of theory and evidence. EIB papers, 13(2), 8-35.

Percoco, M. (2016). Highways, local economic structure, and urban development. *Journal of Economic Geography*, 16(5), 1035-1054.

Pizer, S. D. (2016). Falsification testing of instrumental variables methods for comparative effectiveness research. *Health services research*, 51(2), 790-811.

Raiser, M., Clarke, R., Procee, P., Briceno-Garmendia, C., Kikoni, E., Kizito, J., & Viñuela, L. (2017). Back to Planning.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9, 21-58.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, 5, 1339-1398.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9, 131-165.

Rokicki, B., & Stępniak, M. (2018). Major transport infrastructure investment and regional economic development–An accessibility-based approach. *Journal of Transport Geography*, 72, 36-49.

Roberts, M., Melecky, M., Bougna, T., & Xu, Y. (2020). Transport corridors and their wider economic benefits: A quantitative review of the literature. *Journal of Regional Science*, 60(2), 207-248.

Sheard, N. (2014). Airports and urban sectoral employment. *Journal of Urban Economics*, 80, 133-152.

Souza Jr, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F., ... & Azevedo, T. (2020). Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing*, 12(17), 2735.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *The Review of economic studies*, 83(3), 1263-1295.

Straub, S. (2008). Infrastructure and growth in developing countries: recent advances and research challenges. World Bank policy research working paper, (4460).

Straub, S. (2011). Infrastructure and development: A critical appraisal of the macrolevel literature. The *Journal of Development Studies*, 47(5), 683-708.

Wang, Z., Wu, G. L., & Feng, Q. (2020). Productivity of core infrastructure investment in China: An input–output approach. *The World Economy*, 43(12), 3384-3406.

Xu, M. I. N. G., & Feng, Y. U. A. N. (2022). How transportation infrastructure affects firm Productivity? Evidence from China. *China Economic Quarterly International*, 2(1), 55-69.

Yang, Y. (2018). Transport infrastructure, city productivity growth and sectoral reallocation: Evidence from China. International Monetary Fund.

Young, A. (2022). Consistency without inference: Instrumental variables in practical application. *European Economic Review*, 147, 104112.

Zhang, X., Hu, Y., & Lin, Y. (2020). The influence of highway on local economy: Evidence from China's Yangtze River Delta region. *Journal of Transport Geography*, 82, 102600.

APPENDIX A

Table A1. Variables description

Type	Variable	Description	Source
Highway	Highway Investments	PAC Highway Investments (R$)	MINFRA
	Highway Intervention	1 if the municipality received PAC highway investments and 0 otherwise	MINFRA
	Highway Length	Highway length (km) growth between 2006 and 2018	PNLT (2007) and DNIT
IVs	Legal Protected Area	1 if the municipality is intersected by a legal protected area and 0 otherwise	MMA
	Environmental Embargos	1 if there was an environmental embargo in the municipality during this period and 0 otherwise	IBAMA
	Forest area	Forest area (km²)/Total area (km²)	
	Slope	Area with slope above 20% (which corresponds to hilly areas) (km²)/Total area (km²)	INPE
	Altitude	Average altitude (m)	INPE
	Non-plain areas	Non-plain areas/Total area (km²)	INPE
	Urban infrastructure	Building and infrastructure area (km²)/ Total area (km²)	MAPBIOMAS (Souza *et al.*, 2020)
	Population density	Population/Area (km²)	IBGE
	Land conflicts	1 if there exist a household involved in land conflicts and 0 otherwise	CEDOC - CPT
	Cost Index 1	First component of the MCA	-
	Cost Index 2	Second component of the MCA	-
	LCP-MST Starting and Ending Road Points	Distance to the nearest LCP-MST hypothetical line using starting and ending road points as hubs	MINFRA
	LCP-MST REGIC	Distance to the nearest LCP-MST hypothetical line using REGIC first three classes as hubs	REGIC
	LCP-MST Historical	Distance to the nearest LCP-MST hypothetical line using historical cities (cities with population above 50,000 in 1920) as hubs	IBGE
	Brasília Plan	Distance to the nearest Brasília Plan line (weighted by the municipality area share into the buffer zone)	Bird and Straub (2020)
	JK Road Cruise	Distance to the nearest JK Road Cruise line (weighted by the municipality area share into the buffer zone)	Brasil (1956)
	Potential Road Intervention Area	Distance to the nearest road segment classified as heavy traffic (D, E or F classification)	PNLT (2007)
	Non-Random Allocation Index	First component of the PCA	-
Dependents	GDP *per capita*	Gross Domestic Product (R$)/Population	RAIS and IBGE
	Wages	Wages (R$)	RAIS/MTE
	Firms	Number of firms	RAIS/MTE
	Employment	Number of workers	RAIS/MTE
Controls	GDP *per capita*, lagged	Gross Domestic Product (R$)/ Number of workers in 2007	IBGE
	Share of poor people (%)	Population below the poverty line/Total population	
	Area	Municipality area (km²)	IBGE
	Work force	Number of formal workers	RAIS/MTE
	Agriculture share (%)	Agriculture Value Added (R$)/ Total Value Added (R$)	IBGE
	Exports share (%)	Municipal Exports (US$) / National Exports (US$)	MDIC
	Distance to state road	Distance (km) to the nearest state road	MINFRA
	Distance to railroad	Distance (km) to the nearest railroad	MINFRA
	Distance to port	Distance (km) to the nearest federal port	MINFRA
	Railways stations in 1920	Number of railways stations in 1920	Rede Ferroviária Federal S/A
	Distance to Brasília	Distance (km) to the capital Brasília	
	Institutional Quality	Institutional quality municipal index (IQIM)	Ministry of Planning
	Human Capital (%)	Workers with master of doctoral degree/Total workers	RAIS/MTE

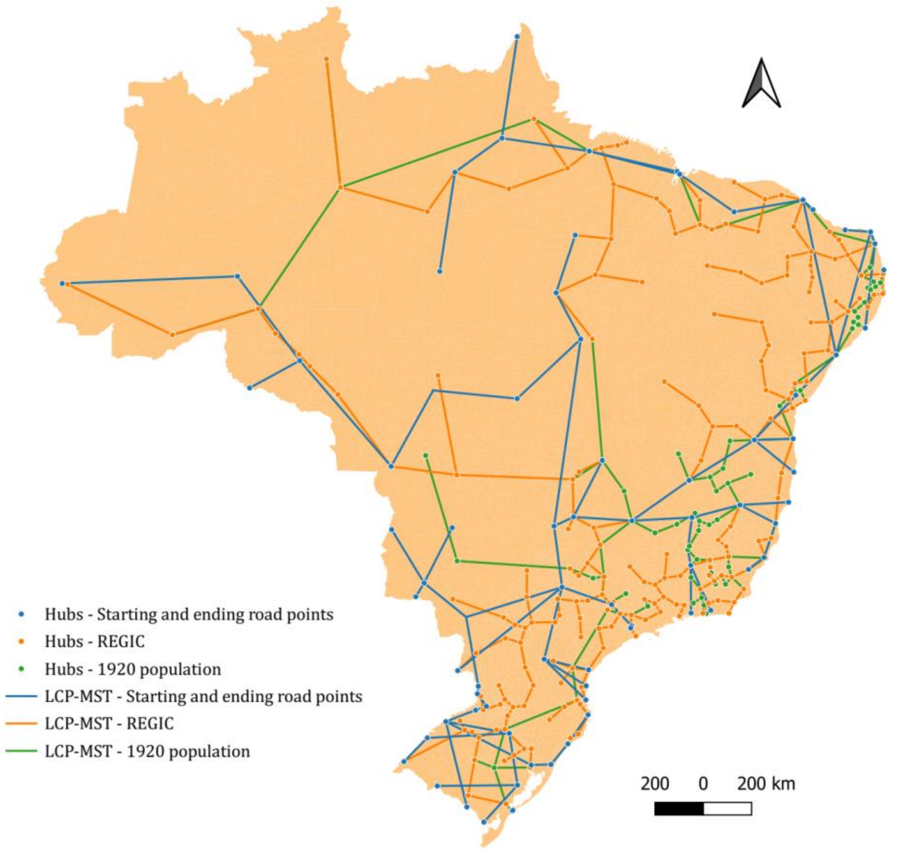
Source: Author's elaboration. Note: IBGE - Brazilian Institute of Geography and Statistics; ANA - National Water and Sanitation Agency; BCB - Central Bank of Brazil; INPE - National Institute for Space Research; IBAMA - Brazilian Institute of Environment and Renewable Natural Resources; MDIC - Ministry of Development, Industry, Commerce and Services; MINFRA - Ministry of Infrastructure; PNLT - National Transport Logistics Plan; MT – Ministry of Transport; RAIS - Annual Social Information Report; MTE - Ministry of Labor and Employment; SIM - Mortality Information System; MS – Ministry of Health.

Table A2. Descriptive statistics

Type	Variable	Obs	Mean	Std. Dev.	Min	Max
Highway	Highway Investments	5,570	1.273	3.449	0.000	14.772
	Highway Intervention	5,570	0.125	0.331	0.000	1.000
IVs	Legal Protected Area	5,570	0.424	0.494	0.000	1.000
	Environmental Embargos	5,570	0.385	0.487	0.000	1.000
	Forest area	5,568	0.377	0.261	0.000	0.994
	Slope	5,565	0.134	0.177	0.000	0.745
	Altitude	5,507	412.476	293.136	0.000	1628.000
	Non-plain areas	5,575	0.998	0.009	0.749	1.000
	Urban infrastructure	5,555	0.020	0.069	0.000	1.000
	Population density	5,565	3.134	1.417	-2.029	9.454
	Land conflicts	5,575	0.063	0.243	0.000	1.000
	Cost Index 1	5,550	0.000	1.141	-1.375	4.578
	Cost Index 2	5,550	0.000	1.020	-2.573	10.767
	LCP-MST Starting and Ending Road Points	5,565	3.706	2.110	-13.816	6.941
	LCP-MST REGIC	5,565	3.825	1.074	-0.112	6.763
	LCP-MST Historical	5,565	4.233	1.307	-1.141	7.344
	Brasília Plan	5,566	4.965	1.635	1.609	7.310
	JK Road Cruise	5,568	4.181	1.648	1.204	7.310
	Potential Road Intervention Area	5,565	4.605	1.452	-3.546	7.670
	Non-Random Allocation Index	5,565	0.000	1.634	-2.279	2.851
Dependents	GDP *per capita*	5,564	0.329	0.346	-2.130	3.083
	Wages	5,563	0.068	0.142	-0.632	7.093
	Firms	5,548	0.178	0.290	-0.301	4.170
	Employment	5,564	0.793	0.498	-2.272	6.529
Controls	GDP *per capita*, lagged	5,564	2.791	0.724	1.096	6.533
	Share of poor people (%)	5,565	41.057	22.776	0.700	90.760
	Area	5,570	6.205	1.279	1.271	11.980
	Work force	5,536	6.049	2.169	0.000	15.023
	Agriculture share (%)	5,564	0.220	0.153	0.000	0.839
	Exports share (%)	5,570	0.000	0.001	0.000	0.034
	Distance to state road	5,565	5.599	37.296	0.001	740.818
	Distance to railroad	5,565	333.165	258.873	0.353	1271.520
	Distance to port	5,565	91.916	215.006	0.031	2081.518
	Railways stations in 1920	5,570	0.536	2.877	0.000	107.000
	Distance to Brasília	5,565	1075.822	445.390	0.000	2872.215
	Institutional Quality	5,505	3.023	0.551	1.000	4.904
	Human Capital (%)	5,564	0.001	0.004	0.000	0.174

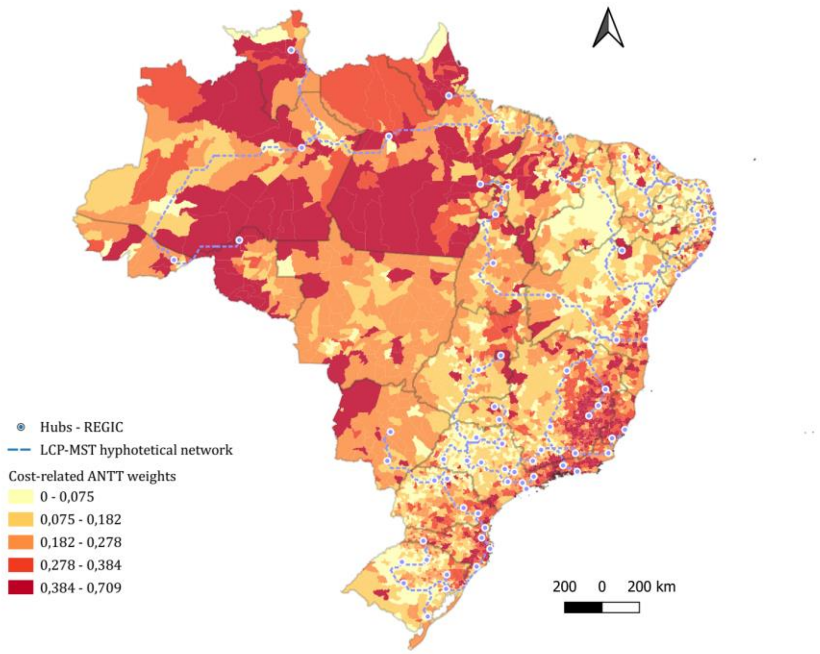
Source: Authors' elaboration.

Figure A1. LCP-MST hypothetical road networks: Euclidean distance



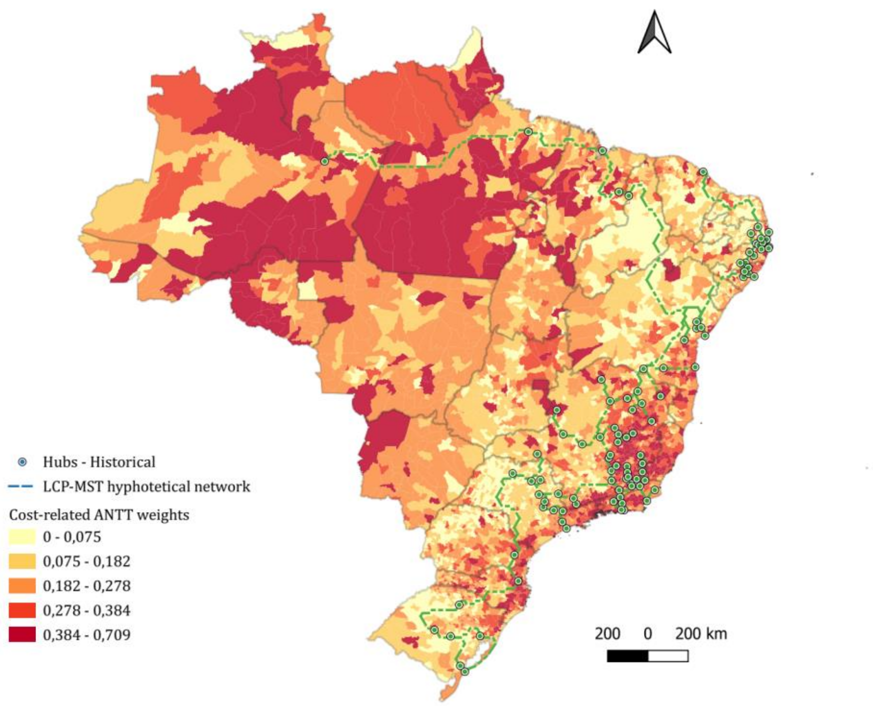
Source: authors' elaboration.

Figure A2. LCP-MST hypothetical road networks: REGIC hubs using cost-related weights in the cost minimization path



Source: authors' elaboration.

Figure A3. LCP-MST hypothetical road networks: Historical hubs using cost-related weights in the cost minimization path



Source: authors' elaboration.

APPENDIX B

Figure B1. Correlation matrix: National Highway investments and potential cost-related IVs

Variables	Infrastructure		Cost-Related IVs											Non-Random Allocation IVs					
			Environmental		Geographical				Human Physical			Aggregated		LCP-MST			Political		Intervention Areas
	Log Highways Investments	Highway Intervention	Legal Protected Areas	Environmental Embargo	Forest Area	Sloped Area	Altitude	Non-Plain Areas	Urban Infrastructure	Populational Density	Land Conflicts	Cost Index 1	Cost Index 2	LCP-MST Start/End Points	LCP-MST REGIC	LCP-MST Historical	Brasilia Plan	JK Road Cruise	Intervention Areas
Log Highways Investments	1.00																		
Highway Intervention	0.97	1.00																	
Legal Protected Areas	0.12	0.12	1.00																
Environmental Embargo	0.19	0.17	0.24	1.00															
Forest Area	0.01	0.01	0.14	0.09	1.00														
Sloped Area	-0.08	-0.07	0.11	0.05	0.08	1.00													
Altitude	-0.10	-0.08	-0.05	-0.05	-0.24	0.19	1.00												
Non-Plain Areas	-0.10	-0.08	0.01	-0.03	0.06	0.04	0.17	1.00											
Urban Infrastructure	0.15	0.14	0.08	0.06	-0.12	-0.04	-0.02	-0.16	1.00										
Populational Density	0.06	0.06	-0.06	-0.09	-0.31	0.14	0.04	-0.09	0.58	1.00									
Land Conflicts	0.10	0.09	0.14	0.12	0.09	-0.09	-0.13	0.02	0.00	-0.10	1.00								
Cost Index 1	0.18	0.17	0.76	0.71	0.13	0.36	-0.01	-0.03	0.29	0.08	0.12	1.00							
Cost Index 2	0.17	0.15	-0.03	0.08	-0.14	-0.70	-0.14	-0.15	0.74	0.31	0.07	0.00	1.00						
LCP-MST Start/End Points	-0.12	-0.10	-0.02	-0.03	0.17	0.02	0.09	0.08	-0.20	-0.26	0.00	-0.06	-0.16	1.00					
LCP-MST REGIC	-0.11	-0.11	-0.05	-0.08	0.07	-0.03	0.02	0.00	-0.16	-0.31	-0.10	-0.12	-0.10	0.12	1.00				
LCP-MST Historical	-0.08	-0.07	-0.07	-0.10	0.09	-0.11	-0.03	-0.03	-0.19	-0.29	-0.05	-0.16	-0.07	0.16	0.28	1.00			
Brasilia Plan	-0.02	-0.03	-0.06	-0.11	-0.04	0.10	-0.22	-0.09	-0.12	0.03	0.01	-0.09	-0.16	-0.11	-0.03	0.02	1.00		
JK Road Cruise	-0.05	-0.05	-0.01	-0.05	-0.09	-0.02	-0.07	0.13	-0.21	-0.19	0.04	-0.08	-0.14	0.00	0.03	0.11	0.40	1.00	
Intervention Areas	-0.08	-0.08	0.01	0.00	0.42	-0.26	-0.26	0.10	-0.25	-0.44	0.14	-0.12	-0.01	0.21	0.09	0.17	0.01	0.16	1.00

Source: authors' elaboration.

Table B2. Highway Investments and Cost-Related IVs: Naive OLS Regressions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Legal Protected Areas	0.390***									0.350***	0.434***		0.423***		
	(0.10)									(0.10)	(0.10)		(0.10)		
Environmental Embargos		0.372***								0.276**		0.374***		0.367***	
		(0.11)								(0.11)		(0.11)		(0.11)	
Forest Area			-0.037							0.325					
			(0.24)							(0.26)					
Sloped Area				-1.270***						-1.126***	-1.364***	-1.199***	-1.362***	-1.201***	-1.141***
				(0.33)						(0.35)	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
Altitude					-0.001***					-0.001***					
					(0.00)					(0.00)					
Non-plain areas						-32.862***				-22.882**					
						(9.40)				(9.28)					
Urban Infrastructure							3.981***			3.316**	3.455***	3.534***	3.454***	3.530***	3.686***
							(1.28)			(1.37)	(1.29)	(1.29)	(1.29)	(1.29)	(1.30)
Populational Density								-0.041		-0.135					
								(0.14)		(0.14)					
Land conflicts									0.561**	0.486**			0.525**	0.542**	0.557**
									(0.25)	(0.25)			(0.25)	(0.25)	(0.25)
Constant	-5.289***	-5.361***	-5.597***	-5.262***	-5.668***	26.971***	-4.451***	-5.572***	-5.553***	18.795**	-3.887***	-4.021***	-3.853***	-3.981***	-4.185***
	(1.32)	(1.31)	(1.32)	(1.31)	(1.31)	(9.44)	(1.33)	(1.31)	(1.31)	(9.28)	(1.34)	(1.33)	(1.33)	(1.32)	(1.33)
Observations	5468	5468	5466	5468	5468	5468	5456	5468	5468	5455	5456	5456	5456	5456	5456
R^2 Adjusted	0.148	0.148	0.146	0.148	0.154	0.153	0.148	0.146	0.147	0.165	0.153	0.152	0.154	0.153	0.151

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table B2. Highway Investments and Cost-Related IVs: Naive OLS Regressions without controls

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Legal Protected Areas	0.853***									0.527***	0.853***		0.765***		
	(0.10)									(0.09)	(0.10)		(0.10)		
Environmental Embargos		1.320***								1.100***		1.298***		1.228***	
		(0.10)								(0.10)		(0.10)		(0.10)	
Forest Area			0.087							-0.074					
			(0.17)							(0.16)					
Sloped Area				-1.579***						-1.473***	-1.734***	-1.651***	-1.558***	-1.503***	-1.289***
				(0.21)						(0.22)	(0.21)	(0.21)	(0.21)	(0.21)	(0.20)
Altitude					-0.001***					-0.001***					
					(0.00)					(0.00)					
Non-plain areas						-33.504***				-22.878**					
						(8.83)				(9.55)					
Urban Infrastructure							7.325***			5.424***	6.679***	6.627***	6.759***	6.684***	7.214***
							(1.17)			(1.39)	(1.15)	(1.13)	(1.15)	(1.13)	(1.17)
Populational Density								0.125***		0.064					
								(0.05)		(0.05)					
Land conflicts									1.505***	0.910***			1.159***	1.078***	1.396***
									(0.26)	(0.25)			(0.26)	(0.25)	(0.26)
Constant	0.911***	0.765***	1.241***	1.483***	1.750***	34.721***	1.130***	0.881***	1.178***	23.592**	1.013***	0.865***	0.952***	0.803***	1.216***
	(0.05)	(0.05)	(0.08)	(0.06)	(0.09)	(8.82)	(0.05)	(0.15)	(0.05)	(9.52)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Observations	5570	5570	5568	5565	5507	5570	5555	5565	5570	5493	5550	5550	5550	5550	5550
R^2 Adjusted	0.015	0.035	-0.000	0.006	0.010	0.008	0.021	0.002	0.011	0.078	0.041	0.060	0.048	0.066	0.036

Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table B3. Highway Investments and Cost-Related IVs: Naive Logit Regressions without controls

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Legal Protected Areas	0.731***									0.518***	0.750***		0.691***		
	(0.08)									(0.09)	(0.08)		(0.08)		
Environmental Embargos		1.005***								0.864***		1.010***		0.964***	
		(0.08)								(0.09)		(0.08)		(0.09)	
Forest Area			0.155							0.105					
			(0.15)							(0.16)					
Sloped Area				-1.335***						-1.453***	-1.501***	-1.424***	-1.366***	-1.303***	-1.111***
				(0.25)						(0.27)	(0.25)	(0.26)	(0.26)	(0.26)	(0.26)
Altitude					-0.001***					-0.000**					
					(0.00)					(0.00)					
Non-plain areas						-14.920***				-11.203***					
						(3.82)				(4.14)					
Urban Infrastructure							3.616***			1.923***	3.177***	3.236***	3.252***	3.296***	3.566***
							(0.52)			(0.63)	(0.49)	(0.49)	(0.50)	(0.50)	(0.54)
Populational Density								0.113***		0.101**					
								(0.04)		(0.04)					
Land conflicts									0.913***	0.507***			0.660***	0.631***	0.854***
									(0.13)	(0.14)			(0.14)	(0.14)	(0.14)
Constant	-2.301***	-2.419***	-2.000***	-1.783***	-1.618***	12.946***	-2.034***	-2.307***	-2.019***	8.422**	-2.212***	-2.335***	-2.257***	-2.382***	-1.972***
	(0.06)	(0.06)	(0.07)	(0.05)	(0.07)	(3.81)	(0.04)	(0.13)	(0.04)	(4.12)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)
Observations	5570	5570	5568	5565	5507	5570	5555	5565	5570	5493	5550	5550	5550	5550	5550
R^2 Adjusted															

Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table B4. Highway Investments and Cost-Related IVs: Naive Logit Regressions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Legal Protected Areas	0.339*** (0.10)									0.326*** (0.10)	0.399*** (0.10)		0.394*** (0.10)		
Environmental Embargos		0.201* (0.11)								0.122 (0.11)		0.205* (0.11)		0.201* (0.11)	
Forest Area			0.221 (0.26)							0.559** (0.28)					
Sloped Area				-1.328*** (0.39)						-1.478*** (0.40)	-1.505*** (0.39)	-1.278*** (0.39)	-1.506*** (0.39)	-1.282*** (0.39)	-1.252*** (0.39)
Altitude					-0.001*** (0.00)					-0.001*** (0.00)					
Non-plain areas						-17.157*** (4.98)				-11.308*** (4.28)					
Urban Infrastructure							1.767*** (0.62)			1.339* (0.70)	1.342** (0.63)	1.434** (0.62)	1.341** (0.63)	1.432** (0.63)	1.514** (0.63)
Populational Density								-0.089 (0.11)		-0.133 (0.11)					
Land conflicts									0.242 (0.17)	0.195 (0.17)			0.219 (0.17)	0.236 (0.17)	0.244 (0.17)
Constant	-6.683*** (1.09)	-6.727*** (1.09)	-6.866*** (1.10)	-6.484*** (1.11)	-6.856*** (1.09)	10.030** (5.02)	-6.276*** (1.12)	-6.766*** (1.10)	-6.847*** (1.10)	5.570 (4.33)	-5.744*** (1.12)	-5.852*** (1.12)	-5.728*** (1.12)	-5.831*** (1.12)	- 5.959*** (1.13)
Observations	5467	5467	5465	5467	5467	5467	5455	5467	5467	5454	5455	5455	5455	5455	5455
R^2 Adjusted															

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table B5. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: OLS Regressions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Log Highways Investments	0.0006 (0.00)	0.0004 (0.00)	0.0078*** (0.00)	0.0069*** (0.00)	0.0041*** (0.00)	0.0043*** (0.00)	0.0043*** (0.00)	0.0042*** (0.00)	0.0040*** (0.00)	0.0040*** (0.00)	0.0040*** (0.00)	0.0036*** (0.00)	0.0035** (0.00)	0.0035** (0.00)
Log GDP 2007			-0.2792*** (0.01)	-0.2823*** (0.01)	-0.3227*** (0.01)	-0.3370*** (0.01)	-0.3390*** (0.01)	-0.3435*** (0.01)	-0.3444*** (0.01)	-0.3446*** (0.01)	-0.3446*** (0.01)	-0.3661*** (0.02)	-0.3721*** (0.02)	-0.3721*** (0.02)
Log Municipal area				0.0258*** (0.00)	0.0136*** (0.00)	0.0043 (0.00)	0.0042 (0.00)	0.0004 (0.00)	-0.0029 (0.00)	-0.0031 (0.00)	-0.0035 (0.00)	0.0126** (0.01)	0.0134*** (0.01)	0.0134*** (0.01)
Log employment					0.0269*** (0.00)	0.0396*** (0.00)	0.0394*** (0.00)	0.0407*** (0.00)	0.0424*** (0.00)	0.0425*** (0.00)	0.0427*** (0.00)	0.0275*** (0.00)	0.0268*** (0.00)	0.0268*** (0.00)
Agriculture share (% GDP)						0.2852*** (0.04)	0.2890*** (0.04)	0.2780*** (0.04)	0.2876*** (0.04)	0.2872*** (0.04)	0.2881*** (0.04)	0.2898*** (0.04)	0.2774*** (0.04)	0.2774*** (0.04)
Exports by municipality (% total)							5.3664 (4.64)	5.4599 (4.61)	5.6395 (4.62)	5.6710 (4.63)	5.7111 (4.63)	7.2610 (4.70)	7.4979 (4.84)	7.4931 (4.84)
Distance to Brasilia								-0.0001*** (0.00)	-0.0002*** (0.00)	-0.0002*** (0.00)	-0.0002*** (0.00)	-0.0001** (0.00)	-0.0001*** (0.00)	-0.0001*** (0.00)
Distance to the nearest state road									0.0007*** (0.00)	0.0007*** (0.00)	0.0007*** (0.00)	0.0006*** (0.00)	0.0006*** (0.00)	0.0006*** (0.00)
Distance to the nearest port										0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)
Distance to the nearest railroad											0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)	0.0000 (0.00)
Share of poor people (%)												-0.0042*** (0.00)	-0.0044*** (0.00)	-0.0044*** (0.00)
Number of railway stations in 1920													-0.0013 (0.00)	-0.0013 (0.00)
Institutional Quality													0.0198** (0.01)	0.0198** (0.01)
Graduate education (% workers)														0.0396 (0.72)
Constant	0.3280*** (0.00)	0.4168*** (0.04)	1.1915*** (0.05)	0.9970*** (0.05)	1.0422*** (0.06)	1.0023*** (0.05)	1.0088*** (0.05)	1.2527*** (0.08)	1.3167*** (0.08)	1.3100*** (0.09)	1.2939*** (0.09)	1.3994*** (0.10)	1.3633*** (0.10)	1.3633*** (0.10)
Observations	5564	5564	5564	5564	5535	5535	5535	5535	5535	5535	5535	5535	5478	5478
State FE	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2 Adjusted	-0.000	0.071	0.237	0.242	0.256	0.265	0.265	0.268	0.271	0.271	0.271	0.284	0.285	0.285

Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

APPENDIX C

Table C1. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: 2SLS IV Regressions

	1	2	3	4	5	6
Second stage						
Log Highways Investments	0.0109** (0.00)	0.0109** (0.00)	0.0147*** (0.00)	0.0144*** (0.00)	0.0097* (0.01)	0.0096** (0.00)
First stage						
Legal Protected Areas	0.4795*** (0.10)		0.4877*** (0.10)		0.4387*** (0.10)	
Environmental Embargos	0.3503*** (0.11)		0.3530*** (0.11)		0.3369*** (0.11)	
Sloped Area	-1.4095*** (0.31)		-1.5196*** (0.31)		-1.6325*** (0.31)	
Cost Index 1	-0.4213*** (0.02)	-0.4253*** (0.02)				
Cost Index 2		0.2683*** (0.05)		0.2654*** (0.05)		0.2326*** (0.05)
LCP-MST REGIC		0.3602*** (0.07)		0.3676*** (0.07)		0.3668*** (0.07)
LCP-MST Historical			-0.3840*** (0.02)	-0.3869*** (0.02)		
JK Road Cruise					-0.3312*** (0.02)	-0.3312*** (0.02)
Observations	5402	5391	5402	5391	5402	5391
KP Wald F Statistic	118.735	162.994	116.089	158.400	112.011	151.930
Effective F Statistic	111.454	122.519	111.443	121.939	96.225	103.699
LIML critical value for tau=5%	19.954	21.214	19.696	20.974	20.885	22.720
R^2	0.23	0.23	0.22	0.22	0.23	0.23

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table C2. Falsification test: OLS regressions

	1	2	3	4	5	6	7	8	9	10	11	12
Legal Protected Areas	0.0256 (0.03)		0.0267 (0.03)		0.0272 (0.03)		0.0262 (0.03)		0.0250 (0.03)		0.0269 (0.03)	
Environmental Embargos	0.0349 (0.03)		0.0344 (0.03)		0.0349 (0.03)		0.0334 (0.03)		0.0351 (0.03)		0.0342 (0.03)	
Sloped Area	0.0243 (0.10)		0.0217 (0.10)		0.0198 (0.10)		0.0207 (0.10)		0.0215 (0.10)		0.0204 (0.10)	
Cost Index 1		0.0219 (0.02)		0.0222 (0.02)		0.0226 (0.02)		0.0217 (0.02)		0.0219 (0.02)		0.0222 (0.02)
Cost Index 2		0.0066 (0.03)		0.0073 (0.03)		0.0078 (0.03)		0.0072 (0.03)		0.0073 (0.03)		0.0075 (0.03)
LCP-MST Starting and Ending Road Points			-0.0040 (0.01)	-0.0040 (0.01)					0.0164 (0.02)	0.0135 (0.02)		
Potential Road Intervention Area					0.0061 (0.01)	0.0056 (0.01)			0.0112 (0.02)	0.0098 (0.02)		
Brasília Plan							-0.0054 (0.01)	-0.0049 (0.01)	-0.0074 (0.02)	-0.0062 (0.02)		
Non-random Allocation Index											-0.0081 (0.01)	-0.0075 (0.01)
Observations	403	400	403	400	403	400	403	400	403	400	403	400
R^2	0.41	0.40	0.41	0.40	0.41	0.41	0.41	0.40	0.41	0.41	0.41	0.40

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table C3. Federal Highway Investments and Municipal Outcomes Growth, 2007-2018: 2SLS IV Regressions

	1	2	3
Second stage	ΔEmployment	ΔFirms	ΔWages
Log Highways Investments	0.0168*** (0.00)	0.0214*** (0.00)	0.0697*** (0.01)
First stage			
Cost Index 1	0.2207*** (0.05)	0.2376*** (0.05)	0.2133*** (0.05)
Cost Index 2	0.3808*** (0.07)	0.4250*** (0.07)	0.3683*** (0.07)
Non-random allocation Index	-0.5069*** (0.02)	-0.5113*** (0.03)	-0.5056*** (0.02)
Observations	5141	5127	5141
KP Wald F Statistic	146.472	149.102	145.535
Effective F Statistic	114.103	117.330	112.904
2SLS critical value for tau=5%	20.686	20.826	20.609
R^2	0.12	0.47	0.20

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table C4. Federal Highway Intervention and Municipal GDP *per capita* Growth, 2007-2018: 2SLS IV Regressions

	1	2	3	4	5	6	7	8	9	10
Second stage										
Highway Intervention	0.1613***	0.1594***	0.1236**	0.1206**	0.1057**	0.1065**	0.1231**	0.1215**	0.1262**	0.1239**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
First stage										
Legal Protected Areas	0.0445***		0.0332***		0.0405***		0.0322***		0.0325***	
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)	
Environmental Embargos	0.0208**		0.0235**		0.0228**		0.0226**		0.0226**	
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)	
Sloped Area	-0.0940***		-0.1309***		-0.0997***		-0.1175***		-0.1185***	
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
Cost Index 1		0.0227***		0.0185***		0.0215***		0.0178***		0.0179***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Cost Index 2		0.0264***		0.0349***		0.0280***		0.0301***		0.0303***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
LCP-MST Starting and Ending Road Points	-0.0392***	-0.0396***					-0.0129***	-0.0129***		
	(0.00)	(0.00)					(0.00)	(0.00)		
Potential Road Intervention Area			0.0326***	0.0330***			0.0091*	0.0102**		
			(0.00)	(0.00)			(0.00)	(0.00)		
Brasília Plan					-0.0316***	-0.0317***	-0.0135***	-0.0128***		
					(0.00)	(0.00)	(0.00)	(0.00)		
Non-random Allocation Index									-0.0503***	-0.0507***
									(0.00)	(0.00)
Observations	5402	5391	5190	5178	5469	5457	5126	5115	5126	5115
KP Wald F Statistic	114.906	159.088	96.673	137.599	124.326	172.444	72.896	90.639	107.042	148.520
Effective F Statistic	104.916	128.927	96.316	118.452	107.287	131.457	74.150	85.131	105.074	125.647
2SLS critical value for tau=5%	19.156	18.280	18.468	17.721	19.684	19.487	24.055	24.703	19.174	17.705
R^2	0.22	0.22	0.24	0.24	0.23	0.23	0.24	0.24	0.23	0.24

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table C5. Federal Highway Length and Municipal GDP *per capita* Growth, 2007-2018: 2SLS IV Regressions

	1	2	3	4	5	6	7	8	9	10
Second stage										
Highway Length	0.0599***	0.0618***	0.0392*	0.0402*	0.0413*	0.0423**	0.0464**	0.0481**	0.0433**	0.0449**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
First stage										
Legal Protected Areas	0.0631**		0.0412		0.0553**		0.0407		0.0376	
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
Environmental Embargos	0.0735**		0.0816***		0.0842***		0.0747**		0.0742**	
	(0.03)		(0.03)		(0.03)		(0.03)		(0.03)	
Sloped Area	-0.0595		-0.0824		-0.0950		-0.0639		-0.0433	
	(0.08)		(0.08)		(0.08)		(0.08)		(0.08)	
Cost Index 1		0.0473***		0.0397***		0.0441***		0.0386***		0.0374***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Cost Index 2		0.0284**		0.0287**		0.0278*		0.0265*		0.0216
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
LCP-MST Starting and Ending Road Points	-0.0962***	-0.0963***					-0.0365***	-0.0371***		
	(0.01)	(0.01)					(0.01)	(0.01)		
Potential Road Intervention Area			0.0834***	0.0836***			0.0682***	0.0690***		
			(0.01)	(0.01)			(0.02)	(0.02)		
Brasília Plan					-0.0701***	-0.0700***	0.0159	0.0171*		
					(0.00)	(0.00)	(0.01)	(0.01)		
Non-random Allocation Index									-0.1205***	-0.1204***
									(0.01)	(0.01)
Observations	5402	5391	5190	5178	5469	5457	5126	5115	5126	5115
KP Wald F Statistic	78.667	105.397	67.451	90.156	79.975	106.662	49.599	59.166	73.442	97.481
Effective F Statistic	74.122	97.551	73.153	96.950	64.749	85.279	49.766	60.408	71.571	98.108
2SLS critical value for tau=5%	19.242	16.584	19.462	15.656	20.755	18.607	24.240	23.681	19.691	16.136
R^2	0.21	0.21	0.23	0.23	0.23	0.22	0.23	0.23	0.23	0.23

All regressions include the following set of control variables: GDP *per capita* in 2007; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

APPENDIX D

Table D1. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: LIML IV Regressions

	1	2	3	4	5	6	7	8	9	10
Second stage										
Log Highways Investments	0.0167*** (0.00)	0.0163*** (0.00)	0.0128** (0.01)	0.0123*** (0.00)	0.0112** (0.00)	0.0111** (0.00)	0.0127** (0.01)	0.0126*** (0.00)	0.0130*** (0.00)	0.0128*** (0.00)
First stage										
Legal Protected Areas	0.4597*** (0.10)		0.3555*** (0.10)		0.4232*** (0.10)		0.3389*** (0.10)		0.3423*** (0.10)	
Environmental Embargos	0.3312*** (0.11)		0.3478*** (0.11)		0.3547*** (0.11)		0.3383*** (0.11)		0.3384*** (0.11)	
Sloped Area	-1.3606*** (0.31)		-1.6060*** (0.31)		-1.4378*** (0.32)		-1.4601*** (0.30)		-1.4676*** (0.29)	
Cost Index 1		0.2586*** (0.05)		0.2274*** (0.05)		0.2470*** (0.05)		0.2177*** (0.05)		0.2190*** (0.05)
Cost Index 2		0.3571*** (0.07)		0.4427*** (0.07)		0.3744*** (0.07)		0.3882*** (0.07)		0.3896*** (0.07)
LCP-MST Starting and Ending Road Points	-0.3875*** (0.02)	-0.3918*** (0.02)					-0.1193*** (0.04)	-0.1204*** (0.04)		
Potential Road Intervention Area			0.3252*** (0.02)	0.3294*** (0.02)			0.0893* (0.05)	0.1025** (0.05)		
Brasília Plan					-0.3170*** (0.01)	-0.3176*** (0.01)	-0.1421*** (0.03)	-0.1314*** (0.03)		
Non-random Allocation Index									-0.4994*** (0.02)	-0.5035*** (0.02)
Observations	5402	5391	5190	5178	5469	5457	5126	5115	5126	5115
KP Wald F Statistic	113.896	156.650	97.221	137.336	123.195	169.934	72.567	89.717	106.925	147.383
Effective F Statistic	104.053	117.056	96.535	108.326	108.318	122.313	73.324	78.795	104.841	112.493
LIML critical value for tau=5%	11.784	19.258	11.535	19.848	11.675	19.872	9.119	14.261	12.006	19.770
R^2	0.22	0.22	0.23	0.24	0.23	0.23	0.23	0.23	0.23	0.23

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table D2. Federal Highway Investments and Municipal GDP *per capita* Growth, 2007-2018: 2SLS IV Regressions, Bootstrap

	1	2	3	4	5	6	7	8	9	10
Second stage										
Log Highways Investments	0.0167*** (0.00)	0.0163*** (0.00)	0.0128** (0.01)	0.0123*** (0.00)	0.0112** (0.00)	0.0111** (0.00)	0.0127* (0.01)	0.0126*** (0.00)	0.0130*** (0.00)	0.0128*** (0.00)
First stage										
Legal Protected Areas	0.4717*** (0.10)		0.3522*** (0.09)		0.5168*** (0.09)		0.2715*** (0.09)		0.3919*** (0.10)	
Environmental Embargos	0.3521*** (0.10)		0.2537** (0.11)		0.4822*** (0.11)		0.2391** (0.11)		0.4339*** (0.11)	
Sloped Area	-1.3929*** (0.31)		-1.5401*** (0.31)		-1.4923*** (0.32)		-1.6424*** (0.30)		-1.1440*** (0.29)	
Cost Index 1		0.1414*** (0.05)		0.1988*** (0.05)		0.1725*** (0.05)		0.2889*** (0.05)		0.3117*** (0.05)
Cost Index 2		0.3435*** (0.07)		0.5802*** (0.08)		0.4647*** (0.06)		0.4671*** (0.07)		0.3779*** (0.07)
LCP-MST Starting and Ending Road Points	-0.3643*** (0.02)	-0.3687*** (0.02)					-0.1954*** (0.04)	-0.1582*** (0.04)		
Potential Road Intervention Area			0.3293*** (0.02)	0.3150*** (0.02)			0.0421 (0.05)	0.0788 (0.05)		
Brasília Plan					-0.3058*** (0.01)	-0.3461*** (0.02)	-0.1296*** (0.04)	-0.1202*** (0.03)		
Non-random Allocation Index									-0.5297*** (0.03)	-0.5502*** (0.02)
Observations	5402	5391	5190	5178	5469	5457	5126	5115	5126	5115
KP Wald F Statistic	113.896	156.650	97.221	137.336	123.195	169.934	72.567	89.717	106.925	147.383
R^2	0.22	0.22	0.23	0.24	0.23	0.23	0.23	0.23	0.23	0.23

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Bootstrapped standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

Table D3. Federal Highway Investments and GDP *per capita* Growth, 2007-2018: 2SLS IV Regressions, Additional Robustness Checks

	1	2	3	4	5
Second stage					
Log Highways Investments	0.0165*** (0.00)	0.0124** (0.01)	0.0172*** (0.01)	0.0103*** (0.00)	0.0118*** (0.00)
First stage					
Cost Index 1	0.2105*** (0.05)	0.1756*** (0.05)	0.1637*** (0.05)	0.4942*** (0.10)	0.2240*** (0.05)
Cost Index 2	0.4829*** (0.08)	0.3690*** (0.07)	0.4697*** (0.08)	0.7491*** (0.15)	0.3935*** (0.07)
Non-random allocation Index	-0.5200*** (0.03)	-0.4506*** (0.03)	-0.4649*** (0.03)	-0.5613*** (0.04)	-0.4986*** (0.02)
Observations	4553	4685	4123	5115	5013
KP Wald F Statistic	139.029	113.102	104.261	63.067	144.673
Effective F Statistic	115.795	81.081	83.184	49.303	108.496
2SLS critical value for tau=5%	19.585	21.902	20.523	24.925	21.311
R^2	0.26	0.23	0.25	0.22	0.20

All regressions include the following set of control variables: GDP *per capita* in 2006; state fixed effects; municipality area; work force; agriculture share; exports share; distance to the nearest state road; distance to nearest railroad; distance to nearest port; railways stations in 1920; distance to Brasília; institutional quality; human capital. Robust standard errors reported in parentheses. * 0.1 ** 0.05 *** 0.01.

[1] The PAC included several infrastructure sectors. In this study, we focus on road transportation.
[2] An extensive literature on Economic Geography has examined the infrastructure issue. For a more detailed literature review, we recommend Straub (2008) and Ottaviano (2008).
[3] In these cases, buildings are related to lane duplications and road improvements instead of construction and paving.
[4] We cannot fully compare our data with the one from Medeiros *et al.* (2021b) as they accounted for maintenance expenses. However, they compared the road investments annual average of R$ 12.31 billion in the PAC period against an annual average of R$ 5.26 billion between 1995 and 2006.
[5] We can likely observe measurement error in the road length variable as well, as the PNLT and DNIT files are not fully comparable. In addition, there is methodological variations over the years in relation to road classifications as federal, state level and so forth. Then, this variable should be used with caution.
[6] The data are divided into six groups: Federal Full Protection and Sustainable Use Conservation Units, State Full Protection and Sustainable Use Conservation Units, and Municipal Full Protection and Sustainable Use Conservation Units.
[7] We take the 5-years sum to avoid potential outliers at using annual data.
[8] This variable is related to construction and infrastructure and is used to generated urban density areas measures.
[9] Following Faber (2014), we use the Dijkstra's (1959) optimal route algorithm to compute least costly construction paths between any bilateral pair of targeted nodes. Then, we use these bilateral cost parameters in combination with Kruskal's (1956) minimum spanning tree algorithm to identify the subset of routes that connect all targeted nodes on a single continuous graph subject to global network construction cost minimization.
[10] About 69% of the studies listed in Figure 1 used GDP, GDP per capita or some productivity measure as one of the main dependent variables.
[11] The IQIM index was constructed considering three dimensions: first, the degree of participation of the population in the public administration, considering capacity of decision, equality, deliberation, and influence in resource allocation; second, the financial capacity of each municipality, the degree to which the funds are generated by the local economic activity without federal resource transfers; and third, the management capacity of local government. The IQIM was normalised to stay between zero and one. The closer to one the index is, the better the institutions are (Iasco-Pereira, Romero, and Medeiros, 2021).
[12] In Appendix B (Table B2), we include naı ve OLS estimations without controls. We also try our highway intervention dummy variable as an additional highway investment variable (Tables B3-B4). The results remain unchanged.
[13] We use Multiple Correspondence Analysis (MCA) for mixed data to generate our composite indexes. We use the first two components as they accumulate 59% of the original data variation. We name the first component as Cost Index 1, and it is calculated by $Cost\ Index\ 1 = 0.58 * Legal\ Protected\ Area + 0.51 * Embargos + 0.13 * Sloped\ Area + 0.08 * Urban\ Infrastructure$.
The second component, which we name Cost Index 2, is calculated by $Cost\ Index\ 2 = 0.00 * Legal\ Protected\ Area + 0.01 * Embargos + 0.49 * Sloped\ Area + 0.58 * Urban\ Infrastructure$.
[14] We also try LCP-MST IVs based on REGIC and historical cities as robustness checks (Table C1 in Appendix C). Results remain stable. We also try the JK Road Cruise IV as a robustness check for historical IV.
[15] To construct our index, we use Principal Component Analysis (PCA). The Non-random Allocation Index is measured as follow: $Nonrandom\ Allocation\ Index = 0.58 * LCP|MST + 0.58 * BrasíliaPlan - 0.57 * PotentialInterventionAreas$. We use the first principal component, presenting 89% of cumulative data variation.
[16] It is important to note that those two datasets are not fully comparable, as methodological changes occurred and some measurement error is expected. Then, results need to be taken with caution.
[17] In unreported estimations, the exceptions are urban infrastructure and populational density variables, which presented significant coefficients in all specifications even controlling for past GDP, population, distance to complementary infrastructures and past infrastructure. This result seems to indicate that expropriation and interferences IVs in the urban context are likely violating exclusion restriction. On the other hand, the land conflict IV, which is related to the rural context, appears to be uncorrelated with GDP *per capita* growth in the falsification sample.