Informality and Development: The Role of Transportation Infrastructure in Reducing Self-Employment*

Nicolas Guida-Johnson†

Job Market Paper

October 10, 2022

Click here for the latest version

Abstract

I study the effects of improvements in transportation infrastructure on informality. To deal with endogeneity issues, I implement two complementary identification strategies. First, taking advantage of the staggered rollout of highways, I apply a generalized difference-in-difference regression model. Second, I conduct an instrumental variable strategy by exploiting the fact that municipalities along the route of important cities in Brazil were more likely to be connected to the federal highway system. I find that once a municipality is connected to the federal highway network, there is a reduction in informality, measured by the self-employment rate among non-agricultural, low-skilled workers. Moreover, I show that connected municipalities have higher GDP per capita, larger firms in the formal sector, and higher demand for formal employment. These results suggest that improved transportation infrastructure induces the growth and development of the formal sector.

Keywords: Informality, Transportation Infrastructure, Self-Employment, Brazil, Market Access

JEL Codes: J46, O12, O17, O18

*I am very grateful to my advisors in this project, M. Daniele Paserman, Patricia Cortés and Martin Fiszbein. For helpful feedback I thank Kevin Lang, Dilip Mookherjee, Samuel Bazzi, Yuhei Miyauchi, César Garro-Marín, Maximiliano García González, Yuheng Zhao, and seminar participants at Boston University. All remaining errors are my own.

†Boston University, Department of Economics. Contact Information: guidaj@bu.edu.

1 Introduction

The informal economy, comprising workers and firms operating at the margins of laws and regulations, is widespread in emerging and developing countries. More than two-thirds of the employed population in these countries work in the informal sector (ILO, 2018), and receive relatively low wages, have minimal job security, and experience high earnings volatility (Gomes et al., 2020). To a large extent, informality is a byproduct of poverty (La Porta and Shleifer, 2014). Due to the small size and slow growth of the modern and formal sector of the economy, a significant proportion of the labor force resort to informal activity as a survival strategy.

Developing and emerging countries also face challenges and deficiencies related to transportation infrastructure. For example, Latin American countries have elevated logistics expenses which on average account for more than half the price of delivered goods (OECD, 2018). Inadequate transportation infrastructure can considerably constrain the growth and expansion of the formal sector of the economy by limiting access to external markets and deterring investments in modern production technologies, namely in supply-chain distribution and inventory management. In such a case, the formal sector may remain small, offering relatively high-quality goods to a minority of formal workers. In contrast, a significant proportion of workers and firms could find it profitable to run small businesses in the informal economy offering low-quality goods and services.

In this paper, I analyze the effects of improved transportation infrastructure on informality. To do so, I consider the expansion of the federal highway system in Brazil between 1970 and 2010. In many ways, Brazil offers an ideal setting for this study. First, a substantial proportion of the employed population in Brazil operates in the informal sector. Using the self-employment rate as a proxy for informality, almost 40% of employed individuals were informal workers in 1970. This proportion decreased in the following 40 years but has remained above 20% as of 2010 (Figure 1). Second, road transportation has significant importance to the Brazilian economy. It was responsible for approximately 65% of the freight transportation in 2016 (CNT, 2018). Third, Brazil's federal highway network significantly expanded between 1970 and 2010. In 1970, federal highways were concentrated in the country's southeast region, with a total length of approximately 30,000 km of roads. By 2010, the road network expanded to traverse practically the entire country with 75,000 km of infrastructure.

Combining geo-spatial and census data and exploiting the staggered rollout of highways, I find that access to transportation infrastructure is associated with a reduction in informality. In this analysis, I show that in municipalities that gain access to the federal highway system, there is a reduction in the self-employment rate among non-agricultural, low-skilled workers. The magnitude of this effect is between 10% and 20%, depending on the empirical strategy. This reduction in informality is likely to be explained by economic factors associated with improvements in transportation infrastructure and the expansion of the formal economy. First, I find that municipalities with access to highways have a larger GDP per capita. Previous literature has shown that higher levels of GDP per capita are linked with lower levels of informality (Loayza and Rigolini, 2006; La Porta and Shleifer, 2014). Development and growth come from the formal sector of the economy, and its expansion shrinks the informal sector.

Second, municipalities with access to the federal highway system have a greater share of large and medium formal firms. This finding is consistent with previous literature showing that reductions in transportation costs lead to an increase in establishment size (You, 2021; Atack et al., 2008). Transportation infrastructure facilitates the implementation of supply-chain logistics, distribution, and inventory technologies. This, in turn, allows the entrance of large firms for which the possibility to coordinate the distribution of inputs and products across potentially different locations and an efficient inventory management are essential for their growth and profitability. As larger and more productive modern and formal firms start operating in a specific area, the demand for formal workers increases, and workers transition from informality into formality. Moreover, these larger formal firms displace small, inefficient, and informal businesses. This mechanism is essential in the retail sector. Close to 40% of non-agricultural self-employed workers operate as retail shop workers and street vendors (Table A3). I show that municipalities with access to highways have a larger number of large retailers (hypermarkets and supermarkets). These establishments are likely to force the exit of small, inefficient, and informal retail businesses and induce the growth of the formal sector.

Third, I find that the wage gap between formal and informal workers is higher in municipalities connected to the federal highway system. Formal workers earn relatively higher wages in municipalities with highway access than in municipalities without highway access. This finding, coupled with the reduction of informality associated with transportation infrastructure, is consistent with a relatively higher demand for formal workers and a larger size of the formal and modern economy in connected municipalities.

Finally, I show that my results on the effects of transportation on informality cannot be explained by internal migration. My main results are nearly identical after excluding internal migrants from my analysis, indicating that migration plays a small, or even null, role in explaining the association between improved access to transportation infrastructure and informality.

The key challenge in identifying the effects of highways is the potential endogenous placements of roads. A simple analysis that compares informality rates between municipalities with and without highway access is bound to be biased. The direction of the bias is ex-ante unclear. For instance, policymakers might choose to connect two or more municipalities with relatively higher economic potential. Since better economic conditions are likely to be associated with lower levels of informality, a simple comparison between municipalities would overestimate the effect of highways on informality. Alternatively, policy-makers could opt to promote relatively underdeveloped areas of the country. In such a case, the estimates would be biased toward zero. In general, policymakers could consider different (potentially unobservable) economic, social, and political factors when designing development plans to construct or expand highways. Therefore, I propose two different and complementary empirical strategies to address these empirical challenges.

First, I exploit the time variation in the expansion of the federal highway system to estimate a difference-in-difference (DiD) model in which municipalities gain access to the highway system at different periods. Both the staggered expansion of the highway system and the fact that some municipalities do not have access to highways by the end of my sample period enable me to construct a reliable counterfactual for the evolution of the main outcomes in the municipalities that gained access to highways. To address concerns raised by the recent econometric literature on staggered DiD design (Borusyak et al., 2021; Goodman-Bacon, 2021; De Chaisemartin and d'Haultfoeuille, 2020), I adopt a stacked-by-event approach (Cengiz et al., 2019; Deshpande and Li, 2019; Vannutelli, 2022), which ensures that my treatment effects are estimated solely based on comparisons between municipalities with access to highways and municipalities without access to highways (both municipalities that gain access later and municipalities that do not have access by the end of my sample period). As a robustness check, I also apply the alternative estimator proposed by De Chaisemartin and d'Haultfoeuille (2020) and obtain nearly identical estimates.

In my second empirical strategy, I exploit the relative location of municipalities to leverage conditionally exogenous variation in the probability of a municipality having access to a highway. This approach allows me to implement an instrumental variable strategy (IV). The basic intuition for this strategy is as follows: A municipality located between two important cities (such as state capitals or populated cities) might be traversed by a highway (and therefore gain access to the highway system) that connects these two cities. In this case, access to highways is not determined by the economic, political, or social conditions of the municipality but by the fact that the municipality is located between two important urban areas. This idea is formalized by constructing a hypothetical optimal network that connects state capitals, populated municipalities, and municipalities located on the country's international borders with straight lines. Then, the hypothetical network is used as an instrumental variable for actual highway placement. This strategy is similar to that used in Chandra and Thompson (2000), Banerjee et al. (2020), Faber (2014), Michaels (2008), Morten and Oliveira (2014), and Perez (2018).

Additionally, I consider two approaches to measure highway access. In the first approach, I use a binary indicator of whether a federal highway intersects the municipality. This method provides estimates of the direct impact of the expansion of the highway system. In a second approach, following Donaldson and Hornbeck (2016), I use a market access measure that considers Brazil's entire highway network and the 1970 population distribution to construct a continuous connectivity measure. The second approach has two clear advantages over the binary measure: It accounts for differences in treatment intensity by incorporating differences in connectivity that highways might provide to different municipalities as well as potential spillover effects arising from changes in connectivity in neighboring municipalities.

To address potential endogeneity concerns, I show that informality, measured by the share of non-agricultural and low-skilled employed individuals that are self-employed, exhibits parallel trends in the periods before municipalities had access to highways. This fact provides suggestive evidence in favor of the identification assumption within the DiD strategy. I also show that evidence of anticipation effects is non-existent–there is no evidence that economic agents in either connected or never-connected municipalities changed their behavior in response to the knowledge of expansion plans of the federal highway system. On the other hand, to provide suggestive evidence in favor of the IV empirical strategy, I show that, at baseline, municipalities that lie along the hypothetical network are not systematically different from those that do not. Moreover, I drop municipalities in the nodes of the hypothetical network (in these municipalities, the exclusion restriction is less likely to hold since these municipalities are potentially the target of other policies) from the analysis and include, in every specification, municipality fixed effects to control for distance to the nearest node in the hypothetical network. Additionally, I show that the hypothetical network does not represent historical travel routes between municipalities in Brazil. Therefore, the effects attributed to the highway system are due to the new highways and not to historical travel routes between cities.

This paper contributes to different strands of the literature. First, it adds to a greater understanding of the main drivers of informality and the process of formalization. Existing literature has emphasized and explored the role of government regulations (de Soto, 2001; Djankov et al., 2002; Bruhn, 2011; Kaplan et al., 2011; Piza, 2018; Rocha et al., 2018), corruption, institutional quality, and tax avoidance (Johnson et al., 1998; Loayza, 1996; Friedman et al., 2000; De Andrade et al., 2016; Levy, 2010) as key determinants of the size of the informal economy. Here, I complement this literature by highlighting the relevance of inadequate transportation infrastructure as an important constraint to the growth and expansion of the formal sector.

A related study by Zárate (2022) shows that improvements in intra-city transportation (the expansion of subway lines) in Mexico City reduce informality rates. This author argues that in cities in developing countries, workers in remote areas may prefer to work in low-paid informal jobs close to their homes rather than incurring the high cost of commuting to formal employment. Therefore, transit improvements may provide better access to formal jobs, leading to an expansion of the formal sector. My study differs from Zárate (2022) in two important aspects. First, I analyze the role of large-scale transportation infrastructure improvements at the national level. The Brazilian federal highway system impacts the transportation costs of goods and people across an entire country, while the expansion of intra-city subway lines primarily affects commuting costs. Therefore, improvements in transportation infrastructure have different economic implications. For instance, while firms may find new subway lines irrelevant in their decisions regarding the distribution of inputs and outputs across different locations, they might choose to invest in modern logistic technologies when better highways are available. Second, the main mechanisms linking better transportation infrastructure and reductions in informality rates greatly differ. Zárate (2022) highlights that the reduction in commuting costs associated with better intra-city transportation provides workers with better access to formal jobs. This paper supports the premise that improved transportation infrastructure promotes the growth and development of the formal sector by increasing access to external markets and inducing investments in modern production technology.

In this paper, I also contribute to the literature on the prevalence of micro and small enterprises in developing economies. Existing explanations for this phenomenon include limited entrepreneurial talent or managerial capital (Lucas Jr, 1978; Bloom et al., 2013), severe contracting problems for hiring outside managers (Akcigit et al., 2021), and intra-city market segmentation (You, 2021). Given that most informal firms in developing economies are micro or small enterprises (La Porta and Shleifer, 2008), the informal economy and the process of formalization are key in determining firm size in developing economies. In this paper, I show that transportation infrastructure affects firm size by promoting the growth of the formal sector. Poor transportation infrastructure leads to a small and slow-growing formal sector of the economy. Consequently, many workers and firms find it profitable to operate small and inefficient businesses in the informal economy. When better transportation infrastructure is available, larger and more efficient formal firms force these small and informal businesses to exit the market.

Finally, this paper adds to the large body of research that seeks to estimate the causal effects of transportation infrastructure in developing economies. Transportation infrastructure has been shown to raise the value of agricultural land (Donaldson and Hornbeck, 2016), increase agricultural trade and income (Donaldson, 2018), positively affect per capita GDP levels without affecting growth rates (Banerjee et al., 2020), induce internal migration (Morten and Oliveira, 2014), and accelerate urban decentralization (Baum-Snow et al., 2017). I add to this literature by providing causal estimates of the impact of large-scale transportation infrastructure on informality.

The remainder of this paper is organized as follows. Section 2 provides details on the main characteristics of the Brazilian federal highway network. Section 3 offers a simple conceptual framework to understand how transportation infrastructure affects informality. Section 4 describes the data sources and details how I measure the municipality's connectivity to the highway network and the informality rate. Section 5 presents the empirical strategies and the main results. Section 6 provides robustness checks. Section 7 explores the mechanisms that explain the main findings. Section 8 concludes.

2 Background: The Brazilian Federal Highway System

Brazil is the world's fifth largest country by geographical area (around 8.5 million km^2). This extensive territory would allow Brazil to rely relatively more on other potentially more efficient modes of transportation. However, the road system has significant importance for the Brazilian economy; road transportation was responsible for about 65% of the cargo movement in 2015 (Ministério da Infraestrutura, 2019).

As of 2010, the federal highway system in Brazil consists of approximately 75,000 kilometers of roadway. Before 1960, the highway system was negligible, with the few existing roads concentrated in the country's southeast region. The largest increase in the extension of the system occurred between 1960 and 1981, when the federal highway system expanded from 7,477 kilometers to 52,117 kilometers; an average of 2,000 kilometers per year. After 1981, there was a significant reduction in the extension's growth rate with an average of 350 kilometers added per year between 1982 and 2010 (Departamento Nacional de Infraestrutura de Transportes, 2016).

A significant event for Brazil's length and structure of the federal highway system was the establishment of Brasilia as the new capital in 1960. From that point in time, the extension of the network increased significantly, as mentioned above. The location of the new capital (center of the territory's country) allowed the projection and construction of the radial highways (*rodovias radiais*): a set of highways that run in eight different directions from Brasilia, connecting Brasilia with state capitals and municipalities located on international borders.

Besides radial highways, the federal highway system is composed of four groups of roads: 1) longitudinal highways (*rodovias longitudinais*) that run north-south; 2) crossing highways (*rodovias transversais*) that run east-west; 3) diagonal highways (*rodovias diagonais*) that run northeast-southwest and northwest-southeast directions; and 4) connecting highways (*rodovias de ligacao*) that connect relevant points of two or more federal highways, connect highways with ports, and allow access to areas of national security and strategic relevance for the country.

The National Ministry of Infrastructure is responsible for the definition of transportation policies, the promotion of technical studies, and the direction of proposals and projects for inclusion in Investment Partnerships Programs (the major effort to expand and upgrade the network of highways within a public-private partnership model). The construction, management, and maintenance of the highway system are under the control and supervision of the National Department of Transportation Infrastructure (Departamento Nacional de Infraestrutura de Transportes; DNIT), a federal authority within the National Ministry of Infrastructure.

3 Conceptual Framework

In this section, I outline a simple conceptual framework to illustrate how access to highways might affect informality. I begin by describing the dual view on informality which is key to understanding the potential role of transportation infrastructure in the informal and formal sectors of the economy. Next, I characterize how improvements in transportation infrastructure can induce the expansion and growth of the formal economy and the consequently decline in informal activity.

3.1 The Dual View of Informality

As described by La Porta and Shleifer (2014), the literature on informality has developed mainly three different approaches to informality, which differ both in terms of the causes of informality and in terms of policy recommendations for formalization. The first approach views the informal sector as a reservoir of entrepreneurial potential, repressed by government regulations (de Soto, 2001). Following this perspective, formalization stems from reducing or eliminating government regulations. A second approach highlights the potential benefits associated with operating informally. Both firms and workers voluntarily choose to operate in the informal economy to avoid taxes and regulations (Levy, 2010; Farrell, 2004). Therefore, strengthened regulations, not their elimination, induce formalization. The third perspective argues that informality is a byproduct of poverty: informality represents a survival strategy for low-skilled individuals. From this approach, that follows the development tradition of Lewis (1954) and Harris and Todaro (1970), the formal and informal sectors are fundamentally different. In this dual view, development comes from the formal economy, and its expansion eventually shrinks the informal economy.

These three approaches do not necessarily represent competing frameworks. As highlighted by Ulyssea (2018), these three perspectives can reflect the behavior and characteristics of heterogeneous firms and workers operating in developing economies. However, the evidence shows that, to a large extent, informality is the result of a survival strategy, and as such, the evidence is largely consistent with dual models of informality (La Porta and Shleifer, 2014). In these models, formal and informal sectors as largely segregated, producing different types of products with different labor and capital and serving different customers. In developing countries with significant levels of poverty and inequality, the small, formal, and modern sector of the economy employs a low fraction of the labor force. In the absence of unemployment insurance and social safety nets, those workers who cannot find employment in the formal economy resort to informal activity as a subsistence strategy. Consequently, the informal sector comprises workers and firms running small businesses that produce low-quality products for low-income customers. In contrast, the formal and modern sector of the economy is formed by productive firms that employ high-skilled workers and produce higher-quality and more expensive goods consumed by the minority of formal workers. Informal businesses do not represent a threat to formal firms and are not sufficiently productive to transition to formality.

This approach to informality encompasses several ideas and formal theories that highlight the separation of the formal and informal economies and the constraints on the expansion of the formal sector. Many economic forces, such as demand-side constraints or cost disadvantages associated with taxes and regulations, keep the two sectors separated. In this paper, I claim that inadequate transportation infrastructure might represent a relevant constraint to the growth and expansion of the formal economy. In such a case, the small and slow-growing formal sector provides limited job opportunities to the workforce. In response, many workers and firms resort to informal activity.

3.2 Transportation Infrastructure and the Informal Economy

Here I highlight two potential channels through which transportation infrastructure can facilitate the growth of the formal economy: (1) it promotes investment in supply-chain logistics and inventory technologies, and (2) it improves access to external markets.

Concerning the first channel, distribution and inventory management are, to a particular extent, key to the growth and expansion of the formal sector of the economy. Transportation infrastructure, such as a highway network, facilitates the investment in supply-chain logistics and inventory technologies that allow the efficient execution and implementation of distribution and inventory tasks. Further, these technologies allow the establishment of larger and more productive firms that force smaller and less productive firms to exit the market. The retail industry is an excellent example of these dynamics and is a significant employment sector for many informal workers.

In developing economies, a significant share of unskilled self-employed individuals run small businesses specializing in one type of product. These individuals, who mostly function on the fringe or outside the margins of the law, run small shops or operate as street vendors selling vegetables, fruits, meats, and other products. In general, they offer low-quality products to low-income consumers. As long as the formal retail sector remains small and efficient and large retailers do not enter the market, informal retail workers find it profitable to run small-scale operations. However, improved transportation infrastructure promotes the entrance and expansion of large retailers, for which supply-chain logistics and inventory technology are essential for their operations.

The technological advantages associated with efficient distribution and inventory control, coupled with economies of scale allow large retailers to significantly reduce their operating costs. As such, the cost advantages of operating at the margin of the law for informal businesses would no longer be sufficient to compete with the formal retail sector. Therefore, the growth and expansion of large retailers (promoted by improved transportation infrastructure) would force small and informal retail businesses to exit the market. At the same time, the expansion of the formal sector induces an increase in the demand for labor in the formal economy. Both processes encourage the transition of workers from informality to formality.

The second channel through which transportation infrastructure can induce the growth of the formal economy is the increase in access to external markets. With significant levels of poverty and inequality, the (limited) local demand for goods and services produced by the formal sector might represent a constraint to the growth of the formal sector. By reducing transportation costs, transportation infrastructure might expand the aggregate demand for goods produced locally by the formal sector and induce its expansion. The increase in the demand for formal goods expands the formal sector, and so does the formal sector's demand for labor. A greater labor demand from the formal economy implies a higher transition of workers from the informal to the formal sector. However, it can be argued that improved access to external markets could also expand the demand for goods locally produced by the informal sector. In such a case, the transition of workers from informality to formality might be offset by an increase in the demand for labor of the informal economy. Nonetheless, this offsetting effect is unlikely to dominate, given the characteristics of the informal sector. For most of the informal sector, selling their products in external markets would not represent a profitable activity.

Additionally, a reduction in transportation costs typically lowers input prices. This, in turn, would increase production in the formal sector and expand the demand for labor, inducing a higher transition of workers from the informal to the formal sector. Analogously, it could be argued that lower input prices would also benefit the informal sector. However, the effect of lower input prices is likely to be relatively more significant for formal and modern firms. Given the productivity differences between sectors, the growth in output due to lower input prices is likely to be higher in the formal economy. Therefore, the potential offsetting effect on informality, if any, would not dominate the growth and expansion of the formal sector.

There are, of course, many other ways transportation infrastructure can affect informality. Reducing transportation costs can increase growth rates in local economic activity (Ghani et al., 2016; Storeygard, 2016), raising income and inducing an increase in the demand for modern, high-quality products produced by the formal sector. Enhanced intercity connectivity may bring formal job opportunities in other locations to previously informal workers (Zárate, 2022). Alternatively, improved access to capital could raise investment in productive activities (Duan et al., 2021).

Finally, it is important to highlight that the transition from informality to formality might depend on individual characteristics such as age and gender. First, younger workers have less sector-specific experience in informality (lower opportunity cost of transition to formality) and lower search costs (due to factors such as the absence of children). Furthermore, informal activity might provide flexibility in allocating time between housework, child-raising, and work outside the household. Therefore, attitudes towards women's allocation of time between these activities might diminish the transition of women from informality to formality.

4 Data & Measurement

To analyze the impact of municipality-level access to highways, I combine data from official maps on the evolution of the federal highway system with census data containing information at the individual level on variables of interest associated with employment status, class of worker, education level and demographic characteristics. Moreover, I use different surveys to obtain municipality-level measures such as GDP per capita and the number and size of registered firms. In this section, I provide details concerning these data sources.

Additionally, I describe the methods used to measure the two crucial variables in my analysis: the municipality's connectivity to the highway network and informality rate. The former is measured using two approaches. First, I use a binary indicator that takes value one if a municipality is intersected by a highway, and zero otherwise. Second, I implement a market access approach that measures the municipality's connectivity considering the entire network of highways and population distribution. For informality rate, I consider self-employment rate as a proxy for informality. In the following subsections, I provide details on their construction, their relevance and their benefits and potential disadvantages.

4.1 Federal Highway System

I use geo-spatial data on the federal highway system compiled by the Ministério dos Transportes, Portos e Aviação Civil (Ministry of Transport, Ports and Civil Aviation) in Brazil. This information comes in a set of maps that shows the extension and structure of the federal highway system for each decade between 1970 to 2010. The information available from these maps allows me to identify the geo-location of each highway and to determine through which municipalities a new or existing highway runs.

This set of maps is shown in Figure 2. Between 1970 and 2010, there is substantial variation in the extension and structure of the federal highway system. At the beginning of this period, highways were mostly concentrated in the southeast region of the country. The total extension of the network was approximately 39,000 km. Over time, the network extended initially to the northeast and eventually, albeit to a lesser extent, to the northwest region. In 2010, the network extension more than doubled the 1970 extension, reaching approximately 98,000 km. It is clear from these maps that the Amazon rainforest, which spans the western and central parts of the country, constitutes a natural barrier to the extension and density of the network. Nevertheless, by 2010 the northwest region was connected to the remainder of the country through the highway network.

4.1.1 Measuring the Municipality's Connectivity to the Highway Network

To measure the municipality's connectivity to the highway network, I propose two different approaches. First, I define highway access at the municipality level as a federal highway intersecting the geographical area of the municipality. Consequently, this measure of connectivity takes the form of a binary indicator that takes value one if the municipality is intersected by a highway, and zero otherwise. Table A1 displays the change over time in the number of municipalities with highway access (i.e. municipalities intersected by a highway). In 1970, 51.1% of the municipalities already had access to a highway. Each decade shows an increase in the number of municipalities intersected by a highway; the largest change occurred between 1980 and 1990, when the number of municipalities with highway access increased by 28.3%.

Using a binary indicator has two main advantages: estimates of the causal effects of highways are easy to interpret, and the analysis of identification assumptions in different empirical strategies is straightforward. However, the binary indicator does not incorporate potential differences in connectivity level that highways might provide to different municipalities. Highways are not homogeneously distributed throughout Brazil. For instance, the southeast region has a denser network of highways than other regions in the country. Therefore, a municipality located in the southeast that is intersected by a new highway has an increase in connectivity higher than a municipality in another region. Furthermore, the binary indicator treats all municipalities that are not intersected by a highway as untreated; however, the municipality's connectivity can increase if a neighboring municipality is traversed by a new highway. This type of spillover effect might be a relevant source of underestimation of the causal effect of highways if indirectly treated municipalities are considered as controls.

To address these issues, I propose to use a market access approach to measure the municipality's connectivity. In particular, I use a simplified version of Donaldson and Hornbeck (2016) approach which takes the following form,

$$MarketAccess_{kt} = \sum_j \tau_{kjt}^{-\theta} Pop_j \qquad (1)$$

where τ_{kj} is the transportation cost from municipality k to municipality j in period t, θ is the trade elasticity and, Pop_j is the population of municipality j in 1970. Therefore, this market access measure represents the connectivity level of a municipality considering all possible destinations and the population size of them, weighted by the trade elasticity.

To construct this measure, I first calculate the cost of transportation from each municipality to all possible destinations in the following manner. First, for each year, I construct a link from every municipality's centroid to the nearest highway within 50 kilometers (in robustness exercises, I relax the cutoff to 100 and 200 kilometers). Then, I compute the total minimum distance of traveling through the highway network from each municipality to all potential destination municipalities. For example, a municipality for which the nearest highway in a certain period is farther than 50 kilometers will not have access to any municipality and for the remainder of municipalities, it will not be a potential destination.

Given data limitations, in particular the lack of information on other types of roadways and the quality of federal highways, the transportation cost between two municipalities is calculated as the sum of the distance from the municipalities' centroids to the nearest highways and the traveling distance through the highway network. Implicitly, by computing the transportation cost in this manner, I am assuming that (1) the per kilometer cost of transport from the centroid of the municipality to the highway is the same as the per kilometer cost of traveling through the highway network, (2) the per kilometer cost of transport from the centroid of the municipality to the highway is the same for every municipality, and (3) the per kilometer cost of traveling through the highway network is the same everywhere (i.e. no differences exist in highway quality within the entire network).

Finally, using the transportation costs, I compute the market access measure as in equation (1) using the 1970 population size of each municipality. I consider the baseline population data since this outcome is potentially endogenous to extensions of the highway network. Following Allen and Atkin (2016), I set the trade elasticity at 1.5 ($\theta = 1.5$) based on the mean gravity coefficient for developing country samples in Head and Mayer (2014). In robustness exercises, I show that the results do not change for different values of the parameter θ.

An underlying assumption on the measures of the municipality's connectivity outlined above is that in every municipality where a highway is located there exists entry and exit ways. This assumption does not seem strong given that most highways in the network are ground-level roads and therefore the construction of entry and exit ways is relatively inexpensive. Additionally, neither the maps nor supplemental data sources provide road-level information such as class, surface type, and pavement conditions. Consequently, it is assumed that the quality and type of highways are uniform throughout the entire network.

4.2 Census Data

I use decennial census data between 1970 and 2010, obtained from the Integrated Public Use Micro Samples (IPUMS). The 1970 and 1980 samples include 25% of the Brazilian population, while the 1991, 2000, and 2010 samples include 10% of the population. Approximately five million households and twenty million individuals are included in each sample.

My sample of workers consists of individuals between ages 18 and 64 who were employed during the year of the survey. Residents of institutional group quarters such as prisons and other institutions are excluded along with unpaid family workers. The census microdata allows me to identify the class of worker (i.e., employed, self-employed, or wage/salary worker), the occupation of the worker, and the industry of employment. Furthermore, the municipality in which the worker resides can be identified from the census microdata. Therefore, all outcomes and control variables used in this paper are constructed at the municipality level which enables me to link the information on highway access with variables constructed from census data.

4.2.1 Measuring Informality at the Municipality Level

Measuring informality is inherently difficult as by definition informal activities are not officially registered. In this paper, I use self-employment as a measure of informality. This measure has been extensively proposed as a proxy for informality[1] and it is strongly correlated with other measures of the informal economy.

Several methods have been proposed to measure the size and the evolution of the informal economy. These include the percentage of the GDP produced by the informal sector, the rate of tax evasion, measures inferred from aggregate electricity consumption, the number of registered firms per 1,000 inhabitants, the share of the employed population that is self-employed and the proportion of salary workers that declare not having a legal contract with their employer. All these measures are highly correlated in developing countries (La Porta and Shleifer, 2014). Among them, self-employment rate and the share of salary workers without legal contracts are the two most used measures of informality. Figure A2 shows a strong correlation between these two measures: countries with a large proportion of salary workers without legal contracts have also a

[1]The literature on informality has extensively used self-employment as either a proxy for informality or part of the informal sector. Some examples are: Antón et al. (2012), Azuara and Marinescu (2013), Almeida and Carneiro (2012), Chen et al. (2006), Falco et al. (2011), Fiess et al. (2010), Funkhouser (1996), Gasparini and Tornarolli (2009), Gong and Van Soest (2002), Loayza and Rigolini (2011), Magnac (1991), Maloney (2004), among others.

large proportion of self-employment. This strong correlation reinforces self-employment as a good proxy for informality.

Ideally, my measure of informality would include wage or salaried workers without legal contracts; however, only in the 2000 and 2010 census surveys were salary workers asked whether they held a legal contract with their employer. Therefore, in 1970, 1980 and 1991, the available data does not allow me to identify salary workers without legal contracts.

Moreover, I choose to focus on non-agricultural workers. The main reason for this is that the transition out of agricultural self-employment might represent mainly a process of structural transformation, and not a process of formalization. Therefore, to avoid confounding structural transformation with formalization I restrict the sample to non-agricultural workers.

Given these facts, the two main outcome variables of this paper are: (1) the share of non-agricultural employed individuals that are self-employed and (2) the share of non-agricultural, non-college (low-skilled) employed individuals that are self-employed. Both measures are computed at the municipality level. By restricting the sample to self-employed individuals without a college education, I exclude the group of skilled self-employed workers comprised of professionals and technicians with general high productivity and who are incorporated in the modern economy. Examples of skilled self-employed workers include lawyers and medical doctors who own their own practice. It is highly unlikely that these workers operate in the informal economy.

Figure 1 shows the evolution of self-employment in Brazil. In 1970, the percentage of non-agricultural self-employment was 36%. The whole period under analysis is characterized by a reduction of the self-employment rate. By 2010, self-employment was approximately 22%; a 39% reduction from its level in 1970. It is also relevant to note that the two measures of informality are close in terms of levels and follow a similar pattern of evolution during the period of analysis. Only after 1990 can one observe a minor divergence in the two series.

The similarity between these two measures indicates that most non-agricultural self-employed individuals do not have a college education and that among college-educated individuals, self-employment is not a common employment status. This is confirmed in Table 1: in 1970, only 10% of high-skilled (college educated) individuals are self-employed while close to 40% of low-skilled (non-college educated) individuals participate in the labor market as self-employed. Moreover, in the baseline year self-employment was more prevalent among relatively older individuals (among workers aged 35-64, self-employment was close to 50%, while among workers aged 18-24, self-employment was around 20%) and among male workers (41% of non-agricultural male workers were self-employed, while 26% of non-agricultural female workers were self-employed).

Additionally, Table 1 shows substantial variation across municipalities in each of the statistics presented.

For instance, the top 10% of municipalities in terms of non-agricultural self-employment had at least a 55% self-employment rate, while the bottom 10% had less than an 18% rate. The remaining statistics show similar variability across municipalities.

4.2.2 Additional Municipality Characteristics

The census data allows me to construct the following set of variables at the municipality level: the share of the adult population that is illiterate, the share of population with college degree, the share of population living in rural conditions, the share of married individuals, the share of individuals employed in manufacturing, labor force participation rate, unemployment rate, the share of immigrants, mean wage of formal workers, mean wage of informal workers and the share of internal migrants. Panel A of Table A2 shows the 1970 summary statistics for this set of variables. On average, the share of illiterate individuals is relatively high at approximately 35% of the adult population and there is a low share of college-educated individuals and immigrants. Almost 45% of the population lived in rural conditions, consistent with a low share of manufacturing employment. As in the case of self-employment, all these variables present substantial variation across municipalities.

4.3 Additional Surveys

From the Cadastro Central de Empresas (Central Registry of Companies), I obtained information at the municipality level concerning the number of registered firms and the number of registered firms by size. This survey covers the years between 2019 and 2006 and includes the entire population of firms registered by the central authority. In my analysis, I use the 2010 data. Additionally, from Contas Regionais e Nacionais do Brasil (Regional and National Accounts of Brazil), I obtained data on the GDP at the municipality level. This data is available for the year 2010 and can be disaggregated by economic activity. Panel B of Table A2 shows summary statistics for this set of variables. Again, there is substantial variation across municipalities in terms of GDP per capita and the number of registered firms. On average, most firms in a municipality are small (i.e., less than 20 employees) and the average firm had 2.7 employees. In the case of firm size, there is less variability across municipalities–in 90% of the municipalities, at least 96% of the firms were small organizations.

5 Empirical Strategy & Main Results

The identification of the casual effect of infrastructure developments, such as the construction and extension of highways, is challenging. In principle, the ideal experiment would involve randomly selecting the location of highways. However, the decision of building a highway and its location is made following development plans designed by policy makers. These policy makers might choose to connect two or more municipalities with relatively higher economic potential. Additionally, they might choose to build or extend highways following political favoritism. This is especially true for infrastructure, such as highways, that might cover areas of the country under the political control of different administrations. Alternatively, policy makers could choose to benefit relatively underdeveloped areas of the country. For these reasons, municipalities connected to the federal highways are likely to be different, both in terms of observable and unobservable characteristics, from municipalities not connected. A simple analysis that compares outcome variables between these two groups of municipalities is bound to be biased.

To address these empirical challenges, I propose two different, yet complementary empirical strategies. First, I exploit the time variation in the expansion of the federal highway system. This characteristic allows me to estimate a difference-in-difference (DiD) model in which municipalities gain access to the highway system in different periods of time. Both the staggered expansion of the highway system and the fact that many municipalities do not have access to highways by the end of my sample period, allows me to construct a reliable counterfactual for the evolution of the main outcomes in the municipalities that gained access to highways, within the context of a DiD model. Second, I exploit the relative location of municipalities to leverage conditionally exogenous variation in the probability of a municipality having access to a highway. This allows me to implement an instrumental variable strategy. The basic intuition for this strategy is as follows: a municipality located between two important cities (such as state capitals or populated cities) might be traversed by a highway (and therefore gaining access to the highway system) that connects these two cities. In this case, gaining access to highways is not determined by the economic, political, or social conditions of the municipality but only by the fact that the municipality is located between two important urban areas.

In the following two subsections, I describe the specific characteristics of these two strategies, their identification assumptions, and describe the main results obtained using these empirical approaches.

5.1 Difference-in-Difference: OLS

The staggered expansion of the Brazilian federal highway system between 1970 and 2010 allows me to implement a difference-in-difference (DiD) strategy. Using the binary indicator for highway access, the analysis compares municipalities that gain access to a federal highway (treated municipalities) to municipalities that have access to highways in a later period (not-yet-treated municipalities) and municipalities that by 2010 are not connected to the highway system (never-treated municipalities). In this context, the two-way fixed effects model to estimate the casual effect of highway access would be the following,

$$y_{kt} = \beta Highway_{kt} + \alpha X_{kt} + \delta_k + \gamma_t + \epsilon_{kt} \tag{2}$$

where y_{kt} is the share of employed individuals that are self-employed at time t in municipality k. $Highway_{kt}$ is a dummy variable that equals one if the municipality is intersected by a federal highway, and zero otherwise, X_{kt} represent a set of control variables at the municipality level, δ_k are municipality fixed effects and γ_t are time fixed effects.

A recent and growing econometric literature highlights the potential pitfalls of two-way fixed effects estimators with staggered adoption (Borusyak et al., 2021; Goodman-Bacon, 2021; De Chaisemartin and d'Haultfoeuille, 2020). In particular, the estimation of β in equation (2) would be biased in the presence of heterogeneous effects across time or units. The estimated coefficient is a weighted average of all the possible 2x2 comparisons, including the comparisons between already-treated units and not-yet-treated units. One way to overcome this problem is to focus only on the comparisons where not-yet-treated units and never-treated units serve as controls. This approach has been proposed and implemented by Deshpande and Li (2019), Sun and Abraham (2021) and Vannutelli (2022), among others. For example, for the year 1990 in my setting, treated units are those municipalities that gained access to a federal highway between 1980 and 1990 (i.e. a highway was constructed intersecting the area of the municipality between 1980 and 1990), not-yet-treated units are those that gain access in a later period and, never-treated units are municipalities that did not have access to a federal highway by the end of my sample period (2010).

To focus on these comparisons, the dataset must be modified in the following way: First, separate datasets for the years 1980, 1990, 2000 and 2010 are created. In each of these datasets, municipalities that gain access to highways in that year are considered treated, while municipalities with access in later years (not-yet-treated units) and municipalities without access (never-treated units) serve as controls. Second, event-time dummies relative to the year of treatment are created in every dataset. Finally, these datasets are appended together. The resulting dataset allows me to estimate the following specification,

$$y_{kt} = \beta Highway_{kc} + \beta_{DD} Highway_{kc} \times Post_{kt} + \sum_{j=-4}^{j=3} \beta_j D^j + \alpha X_k \times t + \delta_k + \gamma_t + \epsilon_{kt} \tag{3}$$

where $Highway_{kc}$ is a dummy variable that takes value 1 if the municipality k is a treated municipality (i.e., municipality intersected by a highway) in cohort c. Notice that, given the structure of the data, one municipality can appear multiple times both as treated and as control. Therefore, $Highway_{kc}$ is not collinear with municipality fixed effects. $Post_{kt}$ is a dummy equal to 1 once a municipality has access to a federal highway. To control for event-time trends that are not captured by year fixed effects, I include D^j; a set of relative event-time dummies. They take value 1 if year t is j periods after (or before) treatment. Standard errors are clustered at the municipality level, allowing for serial correlation over time (Bertrand et al., 2004).

The vector X_k contains 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural areas, share of individuals employed in manufacturing, and log of population) and geographical variables (distance to the border, distance to the capital, distance to the coast and area). Each of these variables is interacted with a linear time trend to control for differential trends in informality that might be correlated with highway access.

Table 2 presents the results of estimating equation (3). Columns (1) and (2) show the effect of highways on the share of self-employment among non-agricultural employed individuals. After including control variables, municipality fixed effects, and year fixed effects, municipalities with highway access show an average reduction of 3.1 percentage points in the proportion of non-agricultural self-employed individuals. Relative to the baseline mean of the outcome variable, this effect corresponds to an 8.5% reduction in informality. Once the sample is restricted to individuals without college education (columns (3) and (4) of Table 2), estimated coefficients are similar. When control variables are included, highway access is associated with a reduction of 3.4 percentage points in the proportion of non-agricultural, low-skilled self-employed individuals; a reduction of approximately 9% relative to the mean in 1970. Finally, columns (5) and (6) show that highways do not affect self-employment among high-skilled individuals. The effect of highways on informality, measured by self-employment, is concentrated exclusively on low-skilled individuals.

To acquire a better understanding of the economic relevance of these effects, I consider the change in the self-employment rate over time (Figure 1). On average, Brazil experienced a decennial reduction of 3.5 percentage points in the self-employment rate over the period under analysis. Equivalently, between 1970 and 2010, the self-employment rate reduced by an average of 12% every decade. These figures demonstrate the economic relevance of the effects found in Table 2; the effects of highways on informality are almost three quarters the average decennial reduction of informality in Brazil over the period under analysis.

The identifying assumption for the estimated effects in Table 2 to be causal is that the evolution of informality in municipalities that do not have access to federal highways (either because they gain access later or because they never have access) is a good counterfactual for the evolution of informality in municipalities that gain access to federal highways. To analyze how reasonable this assumption is, one can investigate the presence of differences in pre-trends. This analysis can be performed estimating a modified version of specification (3),

$$y_{kt} = \beta Highway_{kc} + \sum_{j=-4}^{j=3} \alpha_j D^j \times Highway_{kc} + \sum_{j=-4}^{j=3} \beta_j D^j + \Delta X_k \times t + \delta_k + \gamma_t + \epsilon_{it} \quad (4)$$

where the main difference relative to specification (3) is that $Highway_{kc}$ is now interacted with a set of relative event-time indicators, D^j. Consequently, the parameters α_j measure the change in informality of the treated municipalities j years after (before) treatment, relative to the pre-treatment year, and relative to the change in informality in control municipalities. My data allows me to estimate treatment effects up to four periods from the onset. Notice that I use decennial data therefore, for example, α_2 would represent the effect of highways 20 years after the municipality has been connected to the federal highway system.

Results of estimating specification (4) are presented in Figures 4 and 5. First, Figure 4 plots the estimated effects of highways before and after municipalities gained access to highways on the share of non-agricultural self-employed individuals. Even though the coefficients 40, 30 and 20 years before the introduction of highways are not statistically different from zero, their point estimates are not close to zero and therefore, the results in Table 2 for all non-agricultural self-employed should be taken with caution. However, when I focus on non-agricultural, non-college educated self-employed individuals (Figure 5), the evidence is much stronger in favor of the lack of differential pre-trends. In effect, the estimated coefficient for each decade before the introduction of highways is close to zero and not significant.

Although, as previously mentioned, self-employment was not a common working status for college educated individuals (in 1970, 10% of college educated and non-agricultural workers were self-employed), it is possible that the evolution and share of high-skilled self-employed individuals might be the reason for the differences between Figures 4 and 5. In particular, in the process of expanding the highway network, it might have been possible that policy makers targeted municipalities with relatively higher economic conditions. These better economic conditions are likely to have been positively correlated with a larger share of high-skilled self-employment, such as medical doctors, lawyers, engineers, and other professionals with their own practice. However, when I focus on low-skilled self-employment, municipalities targeted to gain access to highways do not differ in terms of this outcome before the introduction of highways. This fact reinforces the causal interpretation of the effect on low-skilled self-employment rates and reduces the concern that the estimated coefficients are driven by pre-existing differences.

Moreover, Figures 4 and 5 allows me to analyze the dynamic effects of highways. Ten years after the construction of a highway, a reduction in informality is observed in municipalities with highway access. This reduction is shown to increase over time, and in every decade, municipalities with highway access experienced a lower informality rate. It is possible that initially, only economic activities sensitive to transportation infrastructure led to the growth and expansion of the formal economy, and consequently, other economic activities may have expanded and induced further growth of the formal economy.

Besides differential pre-trends, an additional source of concern is the presence of anticipation effects. It is possible that years before a highway is open to traffic in a certain municipality, the population (firms, workers, and local authorities) knew it would be connected to the federal highway system. If economic agents changed their behavior based on this knowledge, my treatment effects could be biased upward or downward depending on the type of anticipatory behavior. For example, in expectation of a new highway firms could choose to invest in a new location for their operations, increasing the demand for formal labor and potentially decreasing informality in the municipality. This type of behavior would lead to an underestimation of my treatment effects; economic activity in municipalities that serve as controls (not-yet-treated municipalities) would start changing as it does in treated municipalities. On the other hand, as economic agents learn that a municipality will have access to the federal highway system, investment in municipalities without access to highways (and in which there are no future plans to build a highway) could decline; firms could choose to close locations, and in general economic activity could be reduced, leading to an increase in informality. In this case, my estimations would suffer from upward bias; the treatment in treated municipalities has a negative effect on control municipalities.

To provide evidence against these two types of anticipatory behavior, I can use specification (4) and look at the β_j coefficients and inspect whether the control units change their behavior before becoming treated. Notice that these coefficients are event-time fixed effects estimated off of the control group. Moreover, I can restrict the sample for which this specification is estimated to separately analyze the two types of behavior mentioned above. By restricting the sample to treated and not-yet treated municipalities, I can look for relative changes in outcomes before treatment indicative of economic agents in not-yet treated municipalities behaving as treated municipalities before receiving treatment. Additionally, by restricting the sample to treated and never-treated municipalities, I can identify relative changes in outcomes before treatment indicative of economic agents in non-treated municipalities changing their behavior as a response to treatment in other municipalities.

The results of these exercises are presented in Appendix Figures A5 and A6. Figure A5 shows the estimated coefficients when the sample is restricted to treated and not-yet treated municipalities. There is no evidence of the presence of relative changes in outcome around treatment that would be indicative of economic agents in not-yet treated municipalities behaving as in treated municipalities before receiving treatment. Moreover, Appendix Figure A6 shows the estimated β_j coefficients when the sample is restricted to treated and never-treated municipalities. Similarly, this figure shows no evidence that economic agents in never treated municipalities changed their behavior as a response of treatment in other municipalities.

An additional potential concern is the presence of differential time-varying shocks that simultaneously occurred during the expansion of Brazilian highway system. To the best of my knowledge, there are no other factors correlated with the expansion of the highway system across the country and over the period under analysis. It is important to highlight that for these shocks to be troublesome in my estimation they would need to affect systematically different treated and untreated municipalities. Given both the temporal and geographical amplitude of my sample, it is unlikely that such shocks be present in my analysis.

To corroborate my analysis using specifications (3) and (4), I apply the methods and alternative estimator recently proposed by De Chaisemartin and d'Haultfoeuille (2020), which is robust to heterogeneous treatment effects across groups or over time. I start by estimating the "naive" two-way fixed effects model in equation (2) and compute the weights attached to the regressions. Panel A of Appendix Table A20 presents the results. First, it is important to highlight that estimates are all very similar to the ones in Table 2. Second, the coefficients on the binary indicator for highway access is obtained as a weighted average of 1,164 ATTs, of which only 10% receive a negative weight. Furthermore, negative weights sum up to only -0.04, suggesting that the relative importance of ATTs receiving negative weights is limited[2]. Additionally, results using the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020), presented in Panel B of Table A20, are similar to the ones obtained in Panel A and to the ones shown in Table 2[3]. Overall, this evidence suggests that the potential problem arising from heterogeneous treatment effects is limited in this setting.

On the other hand, to estimate the effect of highway access using the market access measure as main independent variable, I can estimate the following two-way fixed effects model,

$$y_{kt} = \beta MarketAccess_{kt} + \alpha X_k \times t + \delta_k + \gamma_t + \epsilon_{kt} \tag{5}$$

[2] To further analyze whether negative weighting represents a problem, one can look at the two diagnostic measures to assess the robustness of the estimated coefficient on highway access, $\hat{\beta}_{fe}$, to treatment effect heterogeneity. The first one, $\underline{\sigma}_{fe}$, corresponds to the ratio between the $\hat{\beta}_{fe}$ and the standard deviation of the weights, which is a proxy of the degree of heterogeneity in ATEs across treated groups and time periods, and reflects the minimal value under which it would be possible to have a β_{fe} which is of the opposite sign of the true ATT. The second one, $\underline{\underline{\sigma}}_{fe}$, is a proxy of the minimal amount of treatment effect heterogeneity under which it would be possible to obtain a β_{fe} which is of the opposite sign of all the ATEs. These diagnostic measures are presented in the last two rows of Panel A of Appendix Table A20. Reassuringly, across all outcomes and specifications, $\underline{\sigma}_{fe}$ is as large as the $\hat{\beta}_{fe}$ estimate per se, thus implying that a substantial amount of treatment effect heterogeneity across municipality-period cells would be required to invalidate the "naive" estimates.

[3] One can also use the De Chaisemartin and d'Haultfoeuille (2020)'s methods to compute placebo estimators looking at outcomes' evolution in pre-treatment periods, as well as the different dynamic treatment effects over time. I present results for the full dynamic specifications in Appendix Figures A3 and A4. Results look very similar to the ones in Figures 4 and 5.

where, as previously described, y_{kt} is the share of employed individuals that are self-employed at time t in municipality k. $MarketAccess_{kt}$ is the market access (continuous variable) of municipality k and time t, estimated using equation (1). X_k is a set of municipality characteristics interacted with a linear time trend, δ_k are municipality fixed effects and γ_t are time fixed effects.

As shown in De Chaisemartin and d'Haultfoeuille (2020), even with a continuous treatment variable as $MarketAccess_{kt}$, the two-way fixed effect estimator can be biased in the presence of heterogeneous treatment effects across groups or over time. However, as discussed previously, the potential issues associated with heterogeneous treatment effects in my setting appears to be limited (see Appendix Table A20). Consequently, the estimated coefficients from specification 5 are unlikely to suffer from a bias as a result of heterogeneous treatment effects.

Table 3 shows the results of estimating specification (5)[4]. The market access variable is normalized, therefore, the corresponding estimated coefficients indicate the change in the share of self-employment for one standard deviation increase in market access. Consequently, an increase of one standard deviation in market access reduces the share of non-agricultural self-employed individuals by 2.9 percentage points. Consistent with previous results, this effect is concentrated on low-skilled self-employed individuals; the reduction for this group is 3.1 percentage points, while the reduction for high-skilled individuals is close to zero and not significant. Relative to the mean of the dependent variable in 1970, a one standard deviation in market access is associated with a reduction of 8% in the share of non-agricultural, low-skilled self-employed workers.

To compare the results in Table 3 with the DiD results presented in Table 2, it is useful to compute by how much the market access of a municipality is increased when a highway intersects the territory of the municipality. This change in market access can be considered as the direct effect of highways on municipalities' connectivity levels. Indirect effects are associated with the increase in connectivity in a certain municipality when a neighboring municipality gain access to a new highway. Table A15 shows that when a municipality is intersected by a federal highway its market access increases by 0.879 standard deviations. Therefore, if I take the 3.1 percentage points decrease in informality estimated for a one standard deviation increase in market access, the direct effect of highways is approximately a 2.7 percentage points reduction in informality. Comparing this estimate with the DiD estimates in Table 2, the magnitude of the effect using the market access approach is slightly smaller than that obtained using a binary measure of highway access.

Even though the estimates from equation (5) may not suffer from a bias because of heterogeneous treat-

[4] To make estimates comparable to those obtained using the binary indicator, specification (5) is estimated on the sample of municipalities that in 1970 were not intersected by a highway (i.e., municipalities that in 1970 were intersected by a highway are dropped from the analysis). Notice that, by focusing only on the comparisons between treated and not-yet-treated units and treated and never-treated units, specifications (3) and (4) were estimated on the sample of municipalities that in 1970 were not intersected by a highway.

ment effects, it is likely that $MarketAccess_{kt}$ is an endogenous variable in regression model (5). This variable is based on the highway network and as discussed above, the location and availability of highways are likely to be the result of policy making decisions that consider characteristics of the municipality. These characteristics could be both observable and unobservable time-varying features, leading to an omitted variable bias in the estimation of the main coefficient in equation (5). Therefore, in the following section I address potential endogeneity concerns by implementing an instrumental variables strategy.

5.2 Instrumental Variable: Euclidean Network

In this second empirical strategy, I instrument the municipality's connectivity to the highway network with its connectivity to a hypothetical optimal network. When capital municipalities, populated municipalities, and border municipalities are connected by federal highways, intermediate municipalities not explicitly targeted need to be connected to the highway system. This instrumental variable strategy exploits the fact that municipalities conveniently located between two or more important municipalities are more likely to gained access to the federal highways system. This strategy is similar to that used in Chandra and Thompson (2000), Banerjee et al. (2020), Faber (2014), Michaels (2008), Morten and Oliveira (2014) and Perez (2018).

To build this hypothetical optimal network, I implement the following procedure. First, based on the government's goal to connect Brasilia (the national capital) to the state capitals, I construct a minimum path between Brasilia and all state capitals in eight different directions. This strategy is similar to the one used in Morten and Oliveira (2014), resulting in an identical network. Second, I identify the set of border cities: municipalities bordering Uruguay, Argentina, Paraguay, Peru, Colombia, Venezuela, and Guyana. I label these municipalities as border municipalities. I also identify the set of the top three municipalities in terms of population in 1970 in each state, excluding the state capital. Municipalities in this set are labeled as populated cities. Finally, I include these two sets of municipalities in the network by connecting each of them to a state capital, a border municipality, and/or a populated city. These connections are constructed minimizing the total length of the entire network. The result is a hypothetical network that connects, through straight lines, the following targeted municipalities: the national capital, state capitals, populated cities in 1970 and border cities.

Figure 3 shows the resulting hypothetical network. Straight lines connect targeted municipalities. For illustrative purposes, Figure A1 focuses on the northwest region of the state of Paraná and the southern part of the state of São Paulo. The figure shows the hypothetical network along the actual highway system in 1970. The hypothetical network correctly predicts a connection between the state capital of Paraná, Curitiba, and São Paulo. Municipalities located between Curitiba and São Paulo, along the hypothetical network, are predicted to have access to a highway and they actually receive a highway. However, municipalities that do not lie along the straight line connecting these two municipalities are not predicted to have access to a highway and in fact did not have access in 1970. Similarly, the straight line connecting Ponta Grossa and Londrina, two populated municipalities in the state of Paraná, predicts highway placement, and municipalities located along this line are predicted to have access to a highway. Notice also that the hypothetical network does not predict the entire system of highways. For instance, from São Paulo there are highways heading east (to the left), that are not predicted by the hypothetical network.

Based on the hypothetical network, I can construct two different instrumental variables. The first one is a binary indicator that takes on the value one if the municipality is intersected by the hypothetical network and zero otherwise. The second represents the predicted market access based on the hypothetical network,

$$PredictedMarketAccess_k = \sum_j \nu_{kj}^{-\theta} Pop_j \tag{6}$$

where ν_{kj} is the transportation cost from municipality k to municipality j but now calculated through the hypothetical euclidean network. The procedure used to calculate the predicted measure of market access in equation (6) is similar to the one described above–the only difference is that rather than using the actual network of highways, the hypothetical network is used to compute transportation costs.

The key identification assumption is that, conditional on distance to the nearest targeted municipality, location along the hypothetical network influences economic outcomes only through its effect on highway placement. In other words, the effect of the location of an intermediate municipality located between, for example, a state capital and a border city, on informality (and other economic outcomes) is realized only through the placement of a federal highway. Below, I discuss the potential threats to identification.

First, targeted municipalities in the hypothetical network are more likely to be also the target for actual highways and other infrastructure developments. Therefore, the exclusion restriction is less likely to hold in targeted municipalities (national capital, state capitals, populated cities in 1970 and border cities). To address this concern, I restrict the analysis to municipalities not targeted by the hypothetical network.

Second, municipalities that are located closer to targeted municipalities are mechanically more likely to lie along the hypothetical network. If proximity to targeted municipalities has an independent effect on informality, the exclusion restriction would be invalid. Therefore, in all regression estimations I always control for distance to the nearest targeted municipality.

Third, it might be possible that municipalities that lie along the hypothetical network are different from municipalities outside this network in terms of unobserved characteristics. To provide evidence against this potential concern, I compared in-network municipalities to out-of-network municipalities in a set of observed characteristics at the baseline period. Table A5 shows this comparison. Overall, there does not seem to be systematic and significant differences between these two sets of municipalities, which adds support to the assumption that these municipalities do not differ in terms of unobservable characteristics. Although not statistically significant, connected municipalities have fewer college educated individuals, fewer immigrants, more married individuals, more males, a lower average total income, more illiterates, a higher labor force participation rate and, worse conditions in terms of access to sewage, drinking water and electricity. The two significant differences are in terms of the share of the population living in rural conditions and population. For this reason, these characteristics are always included as controls in the remainder of the analysis.

Finally, since the hypothetical network connects important municipalities, the network could reflect pre-existing historical travel routes between cities. If this is the case, effects attributed to the highway system may instead be due to the effects of the historical travel routes and not to the new highways. To provide evidence against this potential concern, I analyze whether the hypothetical network predicts population growth in the period 1940-1950; i.e., before the construction of federal highway system began. Table A6 shows the result of regressions of population growth between 1940 and 1950 on an indicator for network status. This analysis reveals that municipalities that lie along the hypothetical network do not exhibit higher population growth in the period before highways were constructed.

The specification to be used to implement this instrumental variable strategy has the following form,

$$y_{kt} = \beta Connectivity_{kt} + \alpha X_k \times t + \delta_k + \gamma_t + \epsilon_{kt} \tag{7}$$

where, as previously stated, y_{kt} is the self-employment in municipality k in period t, X_k are control variables at the municipality level interacted with linear time trends, δ_k are municipality fixed effects and γ_t are time fixed effects. The main independent variable, $Connectivity_{kt}$, is either the binary indicator when a municipality is intersected by a highway (instrumented with the binary indicator that takes on the value one if the municipality is intersected by the hypothetical network) or the continuous measure on connectivity based on the market access approach (instrumented with the predicted market access measure from equation (6))

It is important to highlight that both instrumental variables are fixed over time. Therefore, the approach taken to estimate specification (7) consists of instrumenting the endogenous variable with the interaction terms between the time-invariant instrumental variable and time fixed effects.

Table 4 presents the first stage regression results of the binary indicator for actual highway access at the municipality level on the hypothetical optimal network. As shown in column (1), the instrument is strongly correlated with the endogenous variable: municipalities that lie along the hypothetical network are more likely to be intersected by an actual highway. This significant correlation remains once time fixed effects are included in column (2). As previously mentioned, the hypothetical network is fixed over time. Therefore, to use it as an instrument in a panel setting with municipality fixed effects, I need to resort to its interactions with time fixed effects. In column (3) of Table 4, I first show the correlation of the interaction without including municipality fixed effects. The instrument and interactions are still significant and positive. Finally, in column (4), along with control variables, I include municipality fixed effects. The interactions are all positive and significant and the F-statistic equals 45.9, indicating a strong first stage.

Additionally, Table 5 shows the first stage regression results of actual market access on predicted market access. In every specification, the predicted market access is strongly correlated with the actual market access measure. When controls, municipality fixed effects and year fixed effects are included, the interactions between the predicted market access and year fixed effects are significant, and the F-statistic equals 82.2, indicating a strong first stage.

Table 6 presents the 2SLS estimates on the effects of highways on self-employment using the binary indicator as main independent variable. Columns (1) and (2) display the effect on the share of self-employment among non-agricultural employed workers. The results indicate that highway access is associated with a decrease in self-employment of approximately 8.7 percentage points. As shown previously with the DiD strategy, the effects are exclusively concentrated on low-skilled self-employment. Relative to 1970, highway access is associated with a reduction of 22% in informality, measured by the share of non-agricultural and low-skilled workers that are self-employed.

On the other hand, Table 7 shows the 2SLS results when the municipality's connectivity is measured using the market access approach. An increase of one standard deviation in market access reduces the share of non-agricultural employed individuals that are self-employed by 8.2 percentage points. Consistent with previous results, the effect is concentrated in low-skill self-employed individuals; the reduction for this group is 7.8 percentage points while the reduction for high-skill individuals is close to zero and not significant. Relative to the mean of the dependent variable in 1970, a one standard deviation in market access is associated with a reduction of 20.6% in the share of non-agricultural and low-skilled workers that are self-employed.

An alternative approach to specification (7) consists of using the hypothetical network to predict the change in the municipality's connectivity between 1970 and 2010, and analyze the effect of the change in the municipality's connectivity on the change in informality. In particular, this alternative specification takes the following form,

$$\Delta y_{kr} = \beta \Delta Connectivity_k + \alpha X_k + \delta_r + \epsilon_{kr} \tag{8}$$

where Δy_{kr} is the change in self-employment in municipality k located in region r between 1970 and 2010, $\Delta Connectivity_k$ is the change in connectivity in municipality k between 1970 and 2010. As before, X_k are control variables measured in 1970. δ_r are region fixed effects. Notice that, when using the binary indicator for the municipality's connectivity, $\Delta Connectivity_k$ takes the value one if municipality k was not intersected by a highway in 1970, but it was in 2010. And it takes value zero if the municipality remained without highway access by 2010. Similarly, when using the continuous measure of market access, $\Delta Connectivity_k$ measures the change in market access for municipality k between 1970 and 2010. In both cases, $\Delta Connectivity_k$ is instrumented with the instrumental variables derived from the hypothetical network, which predicts highway placement and market access, respectively.

Appendix Tables A11 and A12 show the 2SLS results of estimating equation (8)[5]. First, Table A11 presents the results using the binary indicator for the municipality's connectivity. Estimates indicate that municipalities that gained access to a highway between 1970 and 2010 (i.e., the municipality was not intersected by a highway in 1970, but it was in 2010) experienced a reduction of 7.5 percentage points in the share of non-agricultural self-employed individuals. As in the previous results, this effect is concentrated on low-skilled self-employed individuals: better highway access reduces self-employment among non-agricultural, low-skilled workers by 6.9 percentage points. Furthermore, Table A12 shows that an increase in one standard deviation in market access between 1970 and 2010 is associated with a reduction of 6.8 percentage points in self-employment among non-agricultural workers, and a reduction of 6.6 percentage points in self-employment among non-agricultural, low-skilled workers. Overall, these results are very similar to the ones previously describe based on specification (7), providing support to the instrumental variable strategy.

Finally, it is worth noticing that the IV estimates (either those based on specification (7) or specification (8)) are larger, in absolute value, than the corresponding OLS estimates[6] . There are at least two potential reasons for this. First, the IV estimates identify a local average treatment effect among the set of compliers. In this case, compliers are municipalities that gained access to the highway system because of their location

[5]Appendix Tables A7 and A8 present the OLS results of estimating specification (8). Both set of results are very similar to those obtained in the previous section. Additionally, Appendix Tables A9 and A10 show the first stage regressions of the change in municipality's connectivity (measure by either the binary indicator or the continuous market access measure) on predicted highway access and predicted market access, respectively. In both cases, the instruments are strongly correlated with the independent variable and the F-statistics suggest strong first stage regressions.

[6]It is worth noticing that these differences are not the consequence of dropping targeted municipalities (nodes) in the instrumental variable strategy. Appendix Tables A13 and A14 reproduce the OLS estimates dropping municipalities located at the nodes of the hypothetical optimal network. Results are almost identical to those obtained when these municipalities are not excluded from the analysis (Tables 2 and 3).

but would not have gained access otherwise. It is possible that the economic returns of highways are higher in these municipalities than in always-taker units. In my setting, always-taker units are those municipalities that regardless of their relative location, a highway would be constructed intersecting its territory. A potential reason for this may be political favoritism: local authorities politically connected with the federal government may push for the construction of infrastructure developments, even when the economic returns of these developments is relatively low. Alternative, always-takers units may be municipalities where national strategic infrastructures (for instance, ports, international airports, and military bases) are located. These strategic infrastructures may induce better overall economic conditions in these municipalities, leading to already lower levels of informality, and therefore a limited potential effect of improved transportation on informality. For these reasons, it might be the case that the economic return of highways in terms of their potential to affect informality is relatively higher in complier municipalities.

A second reason for the difference between IV and OLS estimates is the presence of classical measurement error in the measures of the municipality's connectivity. In the presence of classical measurement error, OLS estimates will tend to be biased towards zero. However, the instrumental variable strategy would correct this potential bias.

5.2.1 Heterogeneous Effects by Age and Gender

To further investigate the effects of highways on informality, I examine whether the effects of improved transportation infrastructure differ by age and gender.

Using specification (7), Tables 8 and 9 show the results by age group and gender, respectively. Although not statistically different from each other, the point estimates in Table 8 suggest that the effect of improved transportation infrastructure is higher for older workers. However, relative to the baseline year, the estimated effect is larger for younger workers. In 1970, the average municipality had a self-employment rate of 20% among non-agricultural and non-college educated individuals aged 18-24; then, the estimated effect corresponds to a 39% reduction in informality. On the other hand, for workers aged 25-34 and 35-64 the self-employment rate was 37.4% and 49.9%, respectively. Therefore, the effect of highways corresponds to a reduction of 26% for the 25-34 group and a reduction of 20% for the 35-64 group. As previously mentioned, transition out of informality might depend on individual characteristics such as age. Younger workers have less sector-specific experience in informality (lower opportunity cost of transition to formality) and lower search costs (due to factors such as the absence of children) that would make them more likely to transition from informality to formality.

From Table 9, the effects are largely concentrated on male workers. Self-employment among non-agricultural, non-college educated and male workers reduced in municipalities that gained access to highways by 8.7 percentage points. In the case of self-employed women, the effect is -2.1 percentage points, but not statistically significant. Relative to level of the outcome variable in the baseline year, the effect for men represents a reduction of 20% in informality. For women, this effect corresponds to a reduction of 7%. These results are consistent with differential attitudes towards women's allocation of time between housework, child raising and work outside the household. These results are expected since informal activity might provide more flexibility in allocating time between these activities and women devote relatively more time to housework and child raising. Consequently, transition out of informality is expected to be lower for women than for men.

Qualitatively similar results by age and gender are obtained with a DiD specification (Appendix Tables A16 and A17) and using the market access as a measure for the municipality's connectivity (Appendix Tables A18 and A19).

6 Robustness Checks

Thus far, the results consistently show that access to transportation infrastructure has a negative effect on informality. In this section, I explore the robustness of the main findings by executing two different empirical exercises. First, I extend the definition of informality to include salaried workers without legal contracts. Next, I present results using different cut-offs and trade elasticity values in the construction of the market access variable.

6.1 Alternative Definition of Informality

In the main analysis, due to a lack of individual-level data on the legal status of employed salaried workers in the 1970, 1980 and 1991 censuses, self-employment rates were used as proxies for informality. However, from the 2000 and 2010 censuses one can identify whether salaried workers had legal contracts with their employers. Therefore, using these two data samples, I extend the measure of informality by including salaried workers without legal contracts. Appendix Table A21 shows the results of the effect of highway access and market access on this extended measure of informality, which is constructed as the share of non-agricultural, low-skilled individuals who are either self-employed or salaried workers without a legal contracts. It is worth noting that between 2000 and 2010 only 35 municipalities gained access to a federal highway. Therefore, the implied low variability in both the binary indicator and the market access measure leads to more imprecise estimates. Nevertheless, results in Appendix Table A21 are consistent with the main findings of this paper: highway access or an increase in market access leads to a reduction in informality, measured by this extended definition.

6.2 Alternative Definitions of Market Access

In the first step of the construction of the market access measure, defined in equation (1), I linked every municipality's centroid to the nearest highway within 50 kilometers. According to this assumed cut-off, municipalities for which the closest highway is farther than 50 kilometers did not have access to any other municipality and were not considered as potential destination for other municipalities. In such a case, the value of market access for these municipalities would be zero. Additionally, the parameter θ (trade elasticity) in equation (1) was set at 1.5. In this subsection, I show that results are not subject to these values. Appendix Table A22 shows that results are robust to different values of the cut-off (50, 100 and 200 kilometers) and to varying θ from 1 to 26.83, considered in the literature (Donaldson, 2018; Head and Mayer, 2014).

7 Mechanisms

The key and robust finding so far is that municipalities with improved transportation infrastructure show in response a lower rate of informality. In this section, I investigate a set of underlying economic forces that can explain the effects of transportation infrastructure on informality. In particular, I show that municipalities with highway access have higher GDP per capita, a higher number medium and large firms (including a larger number of large retailers; hypermarkets and supermarkets) and a higher wage gap between formal and informal workers. Additionally, I show that my main findings on informality cannot be explained by internal migration. Given the available data, I focus on the year 2010 and combine different surveys with information at the municipality level with the geo-spatial information on the federal highway system in that year.

Real GDP per Capita. Column (1) of Table 10 shows that municipalities connected to the federal highway system have a GDP per capita 16% higher than municipalities not connected. Similarly, as presented in column (2), a one standard deviation increase in market access is associated with an increase of 17% in GDP per capita. These results are consistent with previous findings in the literature. In particular, Banerjee et al. (2020) show that Chinese counties with better transportation infrastructure have a higher GDP per capita.

In poor municipalities, with low levels of GDP per capita, informality provides subsistence income for a significant share of workers who are unable to find formal employment. The small formal economy in these municipalities provides employment, income, and goods to a reduced proportion of the population. To expand and grow, the formal sector requires to invest in modern production technologies. However, when most of the population do not buy modern manufactured goods produced by the formal economy, these investments are not profitable and the transition to formality is slow.

The process of economic development, partially induced by improved infrastructure, raises income and, consequently, inefficient informal activities are replaced by formal efficient production. From the demand side, higher income shifts the demand from low-quality and cheap goods produced by the informal economy towards higher-quality and more expensive goods produced by the formal economy. This in turn facilitates the expansion and growth of the formal sector. As the formal sector grows, its demand for labor increases and more workers transition from informal to formal employment.

Firm Size and Number of Firms. Columns (3), (5) and (7) of Table 10 show that in municipalities connected to the federal highway system the share of small firms (less than 20 employees) is smaller while the share of medium firms (20 to 50 employees) and large firms (more than 50 employees) is larger than in municipalities not connected. Highway access implies a reduction of 0.6% in the proportion of small firms and an increase of 33% and 28% in the share of medium and large firms, respectively. Similarly, columns (4), (6) and (8) of Table 10 show that one standard deviation increase in market access is associated with a reduction of 0.7% in the proportion of small firms and with an increase of 20% and 29% in the proportion of medium and large firms, respectively. This shift in the composition of firms within connected municipalities is accompanied by an increase in the total number of registered firms: column (9) shows that highway access is associated with an increase of 43% in the number of registered establishments while a one standard increase in market access implies a 44% increase in the number of registered establishments. Overall, this evidence indicates that transportation infrastructure induces the growth in the number of large and medium firms while decreases the total number of small firms.

Transportation infrastructure facilitates the implementation of supply-chain logistics, distribution, and inventory technologies. This in turn allows the formation of larger firm for which a quick and easy availability of inputs, the possibility to coordinate the distribution of inputs and products across potentially different locations and an efficient inventory management are key for their growth. Additionally, these firms operate in the formal sector: they are productive enough to survive in the formal sector and their size prevents them from operating informally (as such, they are more likely to be inspected by the government). As larger and more productive firms start operating in a certain location, not only the demand for formal workers increases (and workers transition from informality into formality), but also, they displace small and informal businesses.

A good example of these dynamics is found in the retail sector. Large retailers (hypermarkets and supermarkets) have a significant advantage over small, single-store retailers. These advantages mainly lie in their technology edge (logistics, distribution, and inventory control) and scale. In particular, Basker and Pham (2008) argue that Wal-Mart's better technology allowed it to grow, and this expansion lowered its operating costs through economies of scale. The emergence and establishment of a more-efficient firm, such as Wal-Mart, will necessarily tend to affect less-efficient local firms. Accordingly, Jia (2008) argues that Wal-Mart's expansion alone can explain around 50 to 70% of the net exit of small retailers between 1988 and 1997 in US.

In developing economies, the retail sector is mainly composed of small and informal businesses specialized in selling one group of products; for example, shops that sell only fruits and vegetables or shops that sell only meat and derivatives. The cost advantages associated with operating informally of these small businesses are no longer sufficient to compete with the formal retail sector when large retailers enter the market and take advantage of their technological edge and scale. Therefore, these large retailers force the exit of small, inefficient, and informal retail businesses and induce the growth of the formal sector.

The retail sector is not only a good illustration of the link between transportation infrastructure and formalization but also an economically relevant sector for Brazilian informal workers. Column (1) of Table A3 shows the distribution of non-agricultural self-employed workers by industry in 1970. Non-agricultural self-employment is highly concentrated in wholesale and retail trade; approximately 36% of these workers operate in this sector. In comparison, 12.2% of non-agricultural salaried workers operate in Wholesale and Retail (column (2) Table A3). Self-employed individuals in the retail sector work as sales shop workers, street vendors, commercial representatives, and other retail-specific occupations (Table A4). Given these facts, it is expected that a large proportion of low-skilled self-employed workers will be displaced by the formal retail sector when transportation infrastructure becomes available.

Further evidence to support that transportation infrastructure induces the formation of large retailers is presented in Table 11. Column (1) shows that municipalities with highway access have an average of 1.5 more large retailers (hypermarkets and supermarkets) than municipalities without highway access. This effect corresponds to an 18% increase in the number of large retailers. Similarly, as presented in column (2), a one standard deviation increase in market access is associated with an increase of 1.8 in the total number of large retailers (22% increase). As explained above, transportation infrastructure allows the investment in logistic and distribution technologies. These technologies are key to the grow and development of large firms in the retail sector. In turn, large retailers force the exit of small, inefficient, and informal retail businesses, and induce the growth of the formal sector and expand the demand for formal employment.

The Wage Gap Between Formal and Informal Workers. On average, formal workers earn 6% more than informal workers. However, as presented in columns (11) and (12), this gap is higher in municipalities with access to transportation infrastructure. Column (11) of Table 10 shows that in municipalities with highway access the gap is 4.5% higher than in municipalities not connected. Similarly, column (12) indicates that a one standard deviation increase in market access is associated with a 5% increase in the wage gap between formal and informal workers.

The interpretation of the evidence presented above on the positive effects of transportation on real GDP per capita, firm size, and number of firms as potential factors for the growth of the formal sector and the reduction of informality relies on a relatively higher demand for labor of the formal sector in connected municipalities. The finding that the relative wage of formal workers is higher in municipalities with highway access coupled with the reduction of informality represents suggestive evidence that in fact the demand for formal employment is higher in this group of municipalities. As discussed above, this higher demand for formal employment is the result of the economic development, induced by infrastructure, which raises the overall level of income and increases the demand for modern goods produced by the formal economy. Economic development also promotes the establishment of larger and more efficient firms that displace inefficient informal activity and increase the demand for formal employment.

Internal Migration. One could argue that the employment prospects and potential earnings for self-employed individuals are relatively enhanced in municipalities without highway access. In response, self-employed individuals living in municipalities with access to a federal highway may choose to migrate to municipalities without highway access. Similarly, one could expect a flow of salaried workers in the opposite direction if job opportunities and potential wages for salaried workers are relatively better in municipalities with highway access. Therefore, it is possible that the results found on the association between transportation infrastructure and the reduction of informality can be explained solely by compositional changes induced by internal migration.

To investigate the role of internal migration, Table 12 presents the effect of transportation infrastructure on informality excluding internal migrants. Internal migrants are defined as workers that reside in a different municipality than where they were born. For reference, columns (1) and (2) show the effect of highway access and market access on informality (measured by the share of low-skilled, non-agricultural employed individuals that are self-employed) for all workers. In columns (3) and (4), I restrict the sample by excluding internal migrants. The coefficients on both the binary indicator for highway access and the continuous measure of market access are virtually identical to those in column (1) and (2), indicating a small, or even null, role of internal migration.

It is worth noticing that the previous exercise does not rule out the presence of internal migration.

The potential flows of self-employed and salaried workers across municipalities affect informality rates in both origin and destination municipalities. Excluding internal migrants adjusts informality rates only in destination municipalities. As such, results in Table 12 provide suggestive evidence that the effects of highways on informality cannot be fully explained by internal migration. If internal migration played a major role in explaining the main results of this paper, we should observe a significant reduction in the estimated coefficients once internal migrants are excluded.

8 Conclusion

Exploiting the expansion of the Brazilian federal highway system, this paper shows that improvements in transportation infrastructure reduce informality. An increase in the municipality's connectivity to the federal highway system (measured either by a binary indicator for highway access or by a continuous measure of market access) reduces the share of self-employment among non-agricultural, low-skilled workers; a proxy for informality rate.

This central result is found using two different, yet complementary empirical strategies: difference-in-differences (that takes advantage of the staggered expansion of the highway system) and instrumental variables (which exploits the relative location of municipalities to leverage conditionally exogenous variation in the probability of a municipality having access to a highway). Depending on the empirical strategy, I find that improved transportation infrastructure is associated with a reduction between 10% and 20% in informality.

Furthermore, I find that reductions in informality are mostly concentrated among younger workers. For these workers, the transition to formality can be relatively less costly due to lower sector-specific experience in informality and due to lower search costs. Additionally, results suggest that male workers are more likely to transition to formality than female workers. This result is consistent with the fact that Informal activity provides certain flexibility in allocating time between housework, child raising, and work outside the household. Attitudes towards women's allocation of time can diminish the transition of women from informality to formality.

Transportation infrastructure plays a central role in the development of the formal sector. Accordingly, this paper shows that municipalities with access to highways have a higher GDP per capita, larger firms in the formal sector, and higher demand for formal employment. Of particular relevance is the link between transportation infrastructure and firm size in the retail sector: improved transportation infrastructure facilitates the entrance of large retailers in a given local market. Empirically, I find that municipalities with highway access have a larger number of supermarkets and hypermarkets. The emergence and establishment of more-efficient retailers will necessarily tend to affect less-efficient local businesses. In developing countries, small-scale retail businesses typically operate in the informal economy; a significant proportion of informal workers are retail shop workers or street vendors. Therefore, these large retailers force the exit of small and informal retail businesses, inducing the growth of the formal sector.

Better and more roads have the potential to induce the growth and development of the formal sector. As the formal economy expands, the demand for formal employment increases, and more workers transition from informal to formal employment. At the same time, a more developed formal sector forces the exit of small, inefficient, and informal businesses.

References

Akcigit, Ufuk, Harun Alp, and Michael Peters, "Lack of selection and limits to delegation: firm dynamics in developing countries," *American Economic Review*, 2021, *111* (1), 231–75.

Allen, Treb and David Atkin, "Volatility and the Gains from Trade," Technical Report, National Bureau of Economic Research 2016.

Almeida, Rita and Pedro Carneiro, "Enforcement of labor regulation and informality," *American Economic Journal: Applied Economics*, 2012, *4* (3), 64–89.

Andrade, Gustavo Henrique De, Miriam Bruhn, and David McKenzie, "A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms," *The World Bank Economic Review*, 2016, *30* (1), 24–54.

Antón, Arturo, Fausto Hernández Trillo, and Santiago Levy, *The end of informality in México?: fiscal reform for universal social insurance*, Vol. 1300, Inter-American Development Bank Washington, DC, 2012.

Atack, Jeremy, Michael R Haines, and Robert A Margo, "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70," Technical Report, National Bureau of Economic Research 2008.

Azuara, Oliver and Ioana Marinescu, "Informality and the expansion of social protection programs: Evidence from Mexico," *Journal of health economics*, 2013, *32* (5), 938–950.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 2020, *145*, 102442.

Basker, Emek and Van H Pham, "Wal-Mart as catalyst to US-China trade," *Available at SSRN 987583*, 2008.

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

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust differences-in-differences estimates?," *The Quarterly journal of economics*, 2004, *119* (1), 249–275.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does management matter? Evidence from India," *The Quarterly journal of economics*, 2013, *128* (1), 1–51.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, 2021.

Bruhn, Miriam, "License to sell: the effect of business registration reform on entrepreneurial activity in Mexico," *The Review of Economics and Statistics*, 2011, *93* (1), 382–386.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional science and urban economics*, 2000, *30* (4), 457–490.

Chen, Martha, Joann Vanek, and James Heintz, "Informality, gender and poverty: A global picture," *Economic and Political Weekly*, 2006, pp. 2131–2139.

CNT, *Somente 12,4% da malha rodoviária brasileira é pavimentada.*, Confederação Nacional do Transporte, Brasil, 2018.

de Soto, H., *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, Black Swan, 2001.

Deshpande, Manasi and Yue Li, "Who is screened out? Application costs and the targeting of disability programs," *American Economic Journal: Economic Policy*, 2019, *11* (4), 213–48.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The regulation of entry," *The Quarterly Journal of Economics*, 2002, *117* (1), 1–37.

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

_ **and Richard Hornbeck**, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Duan, Liaoliao, Dongxiao Niu, Weizeng Sun, and Siqi Zheng, "Transportation infrastructure and capital mobility: evidence from China's high-speed railways," *The Annals of Regional Science*, 2021, *67* (3), 617–648.

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

Falco, Paolo, Andrew Kerr, Neil Rankin, Justin Sandefur, and Francis Teal, "The returns to formality and informality in urban Africa," *Labour Economics*, 2011, *18*, S23–S31.

Farrell, Diana, "The Hidden Dangers of the Informal Economy," *The McKinsey Quarterly*, 2004, *3*, 26.

Fiess, Norbert M, Marco Fugazza, and William F Maloney, "Informal self-employment and macroeconomic fluctuations," *Journal of Development Economics*, 2010, *91* (2), 211–226.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton, "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries," *Journal of public economics*, 2000, *76* (3), 459–493.

Funkhouser, Edward, "The urban informal sector in Central America: Household survey evidence," *World development*, 1996, *24* (11), 1737–1751.

Gasparini, Leonardo and Leopoldo Tornarolli, "Labor informality in Latin America and the Caribbean: Patterns and trends from household survey microdata," *Revista Desarrollo y Sociedad*, 2009, (63), 13–80.

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr, "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing," *The Economic Journal*, 2016, *126* (591), 317–357.

Gomes, Diego BP, Felipe S Iachan, and Cezar Santos, "Labor earnings dynamics in a developing economy with a large informal sector," *Journal of Economic Dynamics and Control*, 2020, *113*, 103854.

Gong, Xiaodong and Arthur Van Soest, "Wage differentials and mobility in the urban labour market: a panel data analysis for Mexico," *Labour economics*, 2002, *9* (4), 513–529.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Harris, John R and Michael P Todaro, "Migration, unemployment and development: a two-sector analysis," *The American economic review*, 1970, *60* (1), 126–142.

Head, Keith and Thierry Mayer, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

ILO, *Women and Men in the Informal Economy: A Statistical Picture*, ILO, 2018.

Jia, Panle, "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Johnson, Simon, Daniel Kaufmann, and Pablo Zoido-Lobaton, "Regulatory discretion and the unofficial economy," *The American economic review*, 1998, *88* (2), 387–392.

Jr, Robert E Lucas, "On the size distribution of business firms," *The Bell Journal of Economics*, 1978, pp. 508–523.

Kaplan, David S, Eduardo Piedra, and Enrique Seira, "Entry regulation and business start-ups: Evidence from Mexico," *Journal of Public Economics*, 2011, *95* (11-12), 1501–1515.

Levy, Santiago, *Good intentions, bad outcomes: Social policy, informality, and economic growth in Mexico*, Brookings Institution Press, 2010.

Lewis, William Arthur, "Economic development with unlimited supplies of labour," *Manchester School of Economic and Social Studies*, 1954, *22* (2), 139–191.

Loayza, Norman and Jamele Rigolini, *Informality trends and cycles*, Vol. 4078, World Bank Publications, 2006.

Loayza, Norman V, "The economics of the informal sector: a simple model and some empirical evidence from Latin America," in "Carnegie-Rochester conference series on public policy," Vol. 45 Elsevier 1996, pp. 129–162.

_ **and Jamele Rigolini**, "Informal employment: safety net or growth engine?," *World Development*, 2011, *39* (9), 1503–1515.

Magnac, Th, "Segmented or competitive labor markets," *Econometrica: journal of the Econometric Society*, 1991, pp. 165–187.

Maloney, William F, "Informality revisited," *World development*, 2004, *32* (7), 1159–1178.

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

Morten, Melanie and Jaqueline Oliveira, "Migration, roads and labor market integration: Evidence from a planned capital city," *Unpublished manuscript*, 2014.

OECD, *Enhancing Connectivity through Transport Infrastructure: The Role of Official Development Finance and Private Investment*, Organisation for Economic Cooperation and Development (OECD), Paris., 2018.

Perez, Santiago, "Railroads and the rural to urban transition: Evidence from 19th-century argentina," *Manuscript, Univ. California, Davis*, 2018.

Piza, Caio, "Out of the Shadows? Revisiting the impact of the Brazilian SIMPLES program on firms' formalization rates," *Journal of Development Economics*, 2018, *134*, 125–132.

Porta, Rafael La and Andrei Shleifer, "The unofficial economy and economic development," *Brookings Papers on Economic Activity*, 2008, *Fall*, 275–352.

_ and _ , "Informality and development," *Journal of economic perspectives*, 2014, *28* (3), 109–26.

Rocha, Rudi, Gabriel Ulyssea, and Laísa Rachter, "Do lower taxes reduce informality? Evidence from Brazil," *Journal of development economics*, 2018, *134*, 28–49.

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

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Ulyssea, Gabriel, "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 2018, *108* (8), 2015–47.

Vannutelli, S, "From Lapdogs to Watchdogs: Random Auditor Assignment and Municipal Fiscal Performance," *Working Paper*, 2022.

You, Wei, "The economics of speed: the electrification of the streetcar system and the decline of Mom-and-Pop stores in Boston, 1885–1905," *American Economic Journal: Applied Economics*, 2021, *13* (4), 285–324.

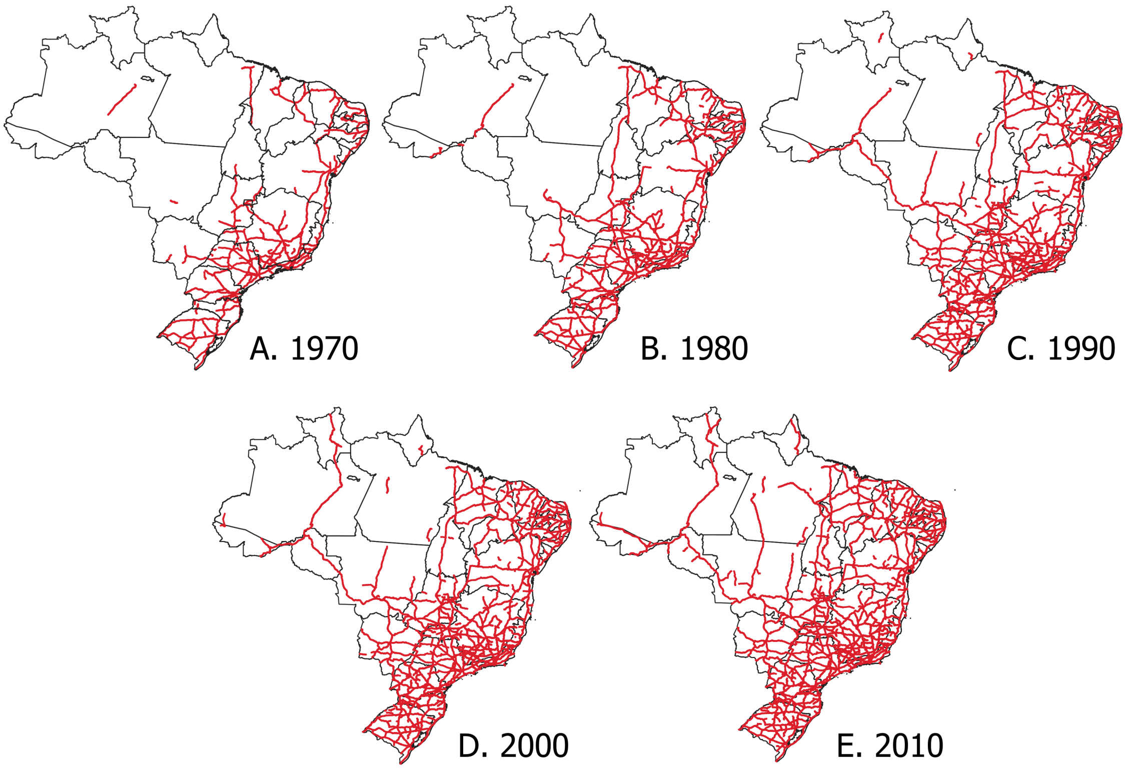
Zárate, Román D, "Spatial Misallocation, Informality, and Transit Improvements," *Policy Research Working Paper*, 2022, *WPS 9990.*

9 Tables and Figures

Figure 1: Evolution of Self-Employment in Brazil, 1970-2010

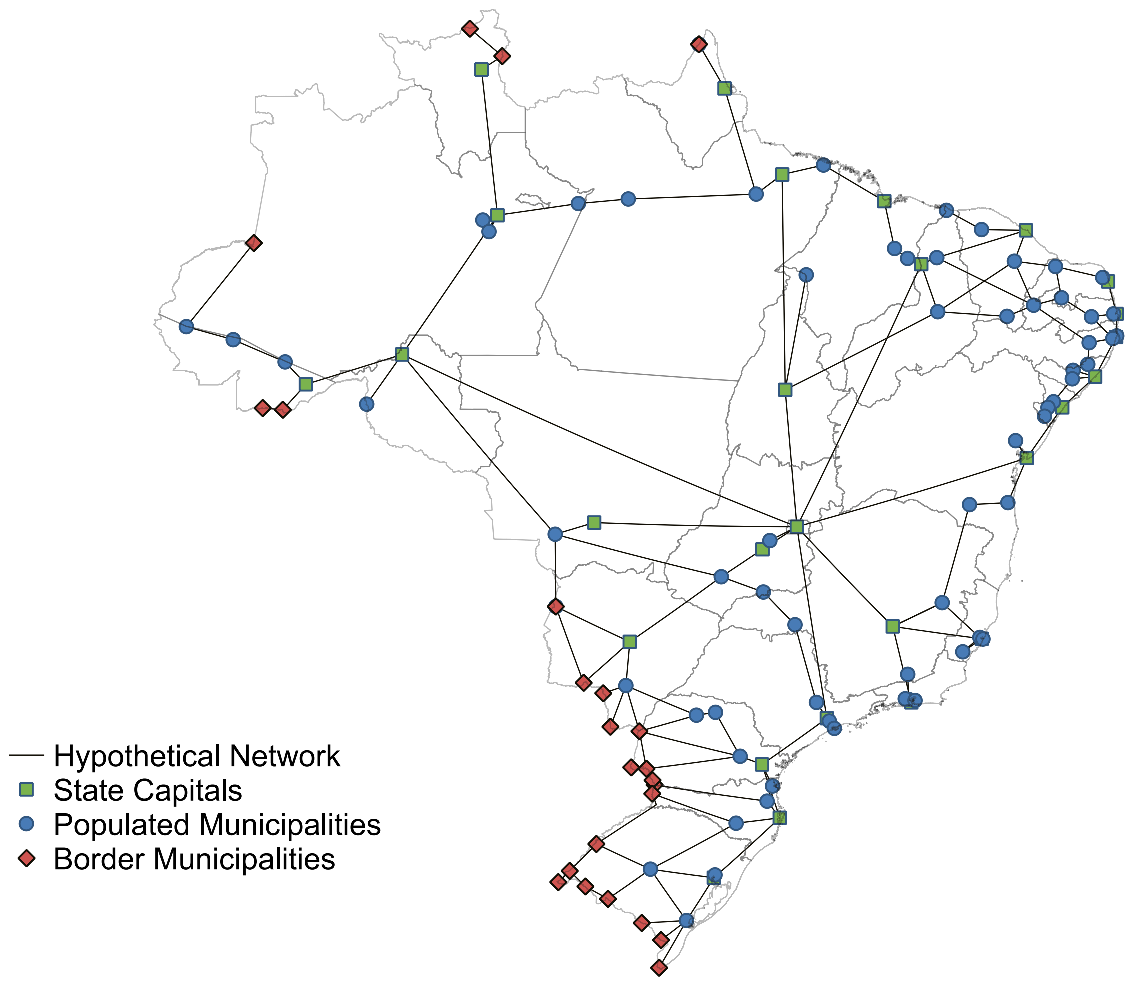
Notes: This figure shows the evolution of self-employment among non-agricultural and among non-agricultural and non-college workers in Brazil between 1970 and 2010. Each point in the series is the weighted average of the corresponding variable among municipalities (weights are the 1970 municipality share of the national population).

Figure 2: Evolution of Federal Highways in Brazil, 1970-2010



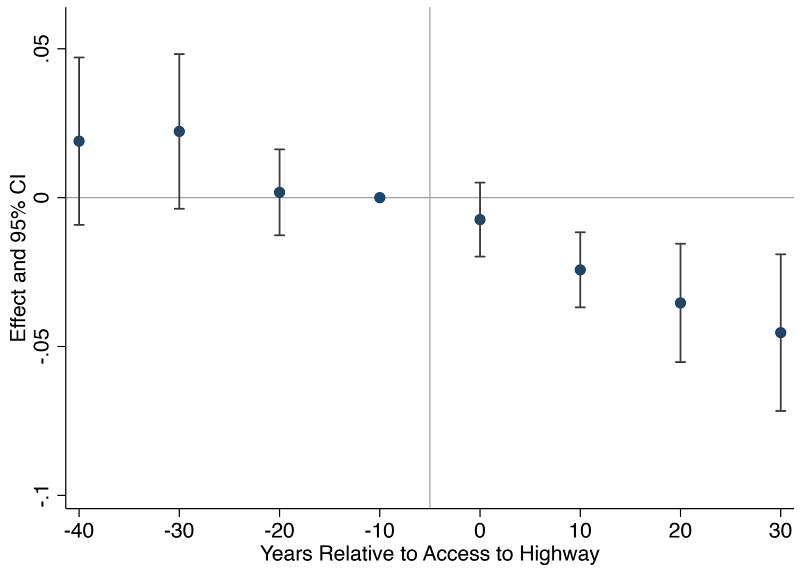
Notes: This figure shows the state of the federal highway system in Brazil in each year for the period 1970 to 2010. Highways are represented with red lines.
Maps available at: https://www.gov.br/infraestrutura/pt-br/centrais-de-conteudo/rodo-evolucao-pdf (*last access Feb 2022*)

Figure 3: Hypothetical Optimal Network



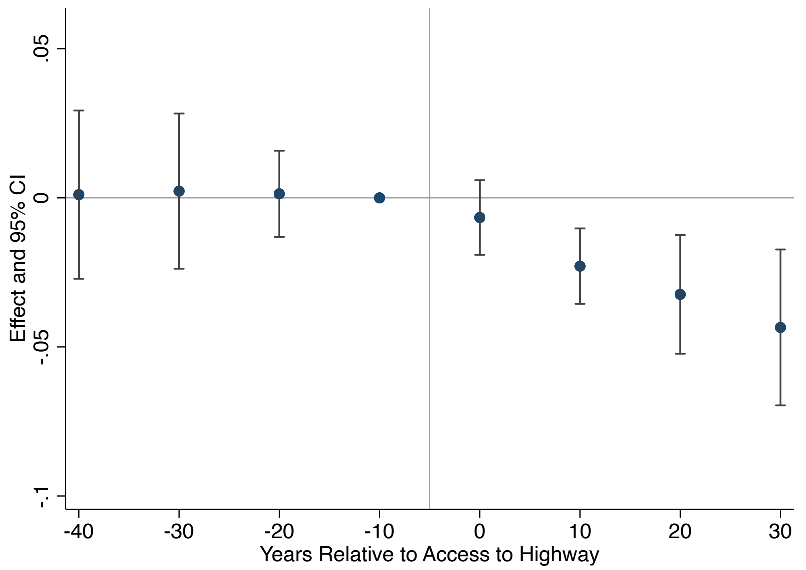
Notes: This figure shows the optimal hypothetical network. The construction of this network is described in Section 5.2. Straight lines connect state capitals, populated municipalities and border municipalities (municipalities located at international border).

Figure 4: Event Study Results: Share of Non-Agriculture Employed Individuals that are Self-Employed



Notes: This figure reports coefficients and 95% confidence intervals estimated according to specification (4). The estimation includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Figure 5: Event Study Results: Share of Non-Agriculture and Low-Skilled Employed Individuals that are Self-Employed



Notes: This figure reports coefficients and 95% confidence intervals estimated according to specification (4). The estimation includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Table 1: Summary Statistics: Main Outcomes in 1970

	Mean	Std. Dev	P10	P25	P50	P75	P90
Share of Self-Employment among:							
Non-Agricultural Workers	0.366	0.139	0.175	0.262	0.362	0.462	0.545
Non-Agricultural and Low-Skilled Workers	0.379	0.143	0.187	0.277	0.372	0.480	0.564
Non-Agricultural and High-Skilled Workers	0.106	0.109	0.000	0.029	0.090	0.151	0.206
Non-Agricultural Workers Aged 18-24	0.191	0.125	0.062	0.099	0.165	0.252	0.354
Non-Agricultural Workers Aged 25-34	0.354	0.150	0.158	0.239	0.348	0.451	0.546
Non-Agricultural Workers Aged 35-64	0.490	0.154	0.260	0.383	0.503	0.606	0.675
Non-Agricultural and Low-Skilled Workers Aged 18-24	0.199	0.129	0.064	0.106	0.171	0.264	0.379
Non-Agricultural and Low-Skilled Workers Aged 25-34	0.374	0.155	0.166	0.260	0.370	0.476	0.570
Non-Agricultural and Low-Skilled Workers Aged 35-64	0.499	0.156	0.266	0.391	0.507	0.614	0.684
Non-Agricultural Male Workers	0.408	0.153	0.189	0.295	0.417	0.511	0.605
Non-Agricultural Female Workers	0.263	0.164	0.084	0.132	0.230	0.364	0.479
Non-Agricultural and Low-Skilled Male Workers	0.411	0.155	0.191	0.295	0.420	0.523	0.608
Non-Agricultural and Low-Skilled Female Workers	0.292	0.172	0.093	0.158	0.273	0.409	0.513

Notes: This table shows statistics on self-employment rate at the municipality level for different groups of the population in 1970. The sample consists of individuals who were between age 18 and 64 and who were working in the year of the survey. Residents of institutional group quarters such as prisons and other institutions are dropped along with unpaid family workers. Low-Skilled workers are those individuals without college education. Skilled-workers are those with college education complete.

Table 2: The Effect of Highway Access on Self-Employment: DiD Estimates

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Highway	0.002	0.002	0.003	0.002	0.003	0.002
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Highway x Post	-0.048***	-0.031***	-0.051***	-0.034***	-0.002	-0.001
	(0.009)	(0.006)	(0.009)	(0.006)	(0.006)	(0.005)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	9361	9361	9361	9361	9361	9361
R2	0.757	0.802	0.760	0.805	0.458	0.494
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the OLS estimates on the effects of highways on self-employment using equation (3). Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 3: The Effect of Market Access on Self-Employment: OLS Estimates

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.036***	-0.029***	-0.049***	-0.031***	0.003	0.002
	(0.011)	(0.010)	(0.011)	(0.010)	(0.015)	(0.016)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3515	3515	3515	3515	3515	3515
R2	0.527	0.543	0.587	0.544	0.472	0.501
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the OLS estimates on the effects of market access on self-employment using equation (5). Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 4: First Stage Regressions

Dependent Variable: Dummy variable $HighwayAccess_{it}$

	(1)	(2)	(3)	(4)
In-Network	0.232***	0.232***	0.146***	
	(0.031)	(0.031)	(0.030)	
In-Network X 1980				0.198***
				(0.050)
In-Network X 1990			0.098**	0.250***
			(0.033)	(0.041)
In-Network X 2000			0.109**	0.239***
			(0.036)	(0.041)
In-Network X 2010			0.134***	0.255***
			(0.037)	(0.043)
Constant	0.302***	0.080***	0.118***	0.132
	(0.020)	(0.019)	(0.017)	(0.187)
Municipality FE	NO	NO	NO	YES
Year FE	NO	YES	YES	YES
Controls	NO	NO	NO	YES
Obs	3390	3390	3390	3390
R2	0.055	0.129	0.132	0.713
F-Statistic	56.191	116.375	69.574	45.892

Notes: This table reports the first stage regressions for instrumenting highway placement with the hypothetical optimal network. The dependent variable is a dummy that takes value one if the municipality is intersected by a federal highway. In-network is a dummy variable that takes value one if the municipality is intersected by the hypothetical optimal network. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 5: First Stage Regressions: Predicted Market Access and Actual Market Access

Dependent Variable: Market Access

	(1)	(2)	(3)
Predicted Market Access	0.548^{***}		
	(0.029)		
Predicted Market Access X 1980		0.017^{***}	0.018^{***}
		(0.005)	(0.006)
Predicted Market Access X 1990		0.018^{***}	0.020^{***}
		(0.006)	(0.007)
Predicted Market Access X 2000		0.023^{***}	0.024^{***}
		(0.007)	(0.007)
Predicted Market Access X 2010		0.033^{***}	0.034^{***}
		(0.008)	(0.008)
Municipality FE	NO	YES	YES
Year FE	NO	YES	YES
Controls	NO	NO	YES
Obs	3390	3390	3390
R2	0.075	0.566	0.654
F-Statistic	65.256	80.426	82.211

Notes: This table reports the first stage regressions for instrumenting actual market access with predicted market access. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Predicted market access is calculated using equation (6) based on the hypothetical optimal network and the distribution of population in 1970. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 6: The Effect of Highway Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Highway	-0.082***	-0.087***	-0.077***	-0.084***	0.002	0.002
	(0.028)	(0.030)	(0.028)	(0.030)	(0.027)	(0.030)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3390	3390	3390	3390	3390	3390
R2	0.412	0.481	0.389	0.455	0.432	0.433
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the 2SLS estimates on the effects of highways on self-employment using using the hypothetical optimal network as an instrument for highway placement. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 7: The Effect of Market Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.078***	-0.082***	-0.074***	-0.078***	-0.001	0.002
	(0.017)	(0.016)	(0.018)	(0.017)	(0.019)	(0.019)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3390	3390	3390	3390	3390	3390
R2	0.568	0.593	0.547	0.598	0.431	0.433
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the 2SLS estimates on the effects of market access on self-employment using equation (7) and instrumenting market access with predicted market access. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table 8: The Effect of Highway Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed, by Age Group

	18-24		25-34		35-64	
	(1)	(2)	(3)	(4)	(5)	(6)
Highway	-0.050*	-0.078**	-0.093***	-0.097***	-0.114***	-0.098***
	(0.027)	(0.030)	(0.032)	(0.035)	(0.032)	(0.034)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3390	3390	3390	3390	3390	3390
R2	0.225	0.241	0.404	0.438	0.474	0.543
Mean Dep Var	0.164		0.289		0.409	
Mean Dep Var in 1970	0.199		0.374		0.499	

Notes: This table reports the 2SLS estimates on the effects of highways on self-employment using using the hypothetical optimal network as an instrument for highway placement for different age groups. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table 9: The Effect of Highway Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed by Gender

	Male		Female	
	(1)	(2)	(3)	(4)
Highway	-0.094***	-0.087***	-0.004	-0.021
	(0.030)	(0.033)	(0.038)	(0.037)
Year Fixed Effects	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES
Controls	NO	YES	NO	YES
Obs	3390	3390	3390	3390
R2	0.276	0.360	0.207	0.332
Mean Dep Var	0.354		0.241	
Mean Dep Var in 1970	0.411		0.292	

Notes: This table reports the 2SLS estimates on the effects of highways on self-employment using using the hypothetical optimal network as an instrument for highway placement for male and female workers separately. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table 10: The Effect of Highway Access on Real GDP, Firm Size, Number of Firms and Wage Gap. 2010. 2SLS

	Log(Real GDP per Capita)		Share of Firms' Establishments: Large		Share of Firms' Establishments: Medium		Share of Firms' Establishments: Small		Number of Registered Firms' Establishments		Log Wage Gap (Formal/Informal)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Highway Access	0.160†		0.002†		0.005†		-0.006†		244.3†		0.045†	
	(0.013)		(0.000)		(0.000)		(0.001)		(39.815)		(0.006)	
Market Access		0.155†		0.002†		0.003†		-0.007†		240.2†		0.040†
		(0.012)		(0.000)		(0.000)		(0.001)		(39.815)		(0.005)
Region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
R2	0.538	0.538	0.096	0.104	0.142	0.149	0.160	0.169	0.118	0.119	0.444	0.451
Obs	5451		5451		5451		5451		5451		5451	
Mean Dep Var	2.528		0.007		0.015		0.978		564.117		0.060	

Notes: This table reports the 2SLS estimates on the effects of highways and market access on real GDP per capita, the share of small, medium and large firm establishments, the number of registered firms and the log of the ratio between formal and informal wage, in 2010. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and region fixed effects. Controls include 2000 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area). Standard errors clustered at the region level in parenthesis. All regressions are weighted by 2000 municipality share of national population. Significance levels: † $p<0.01$.

Table 11: The Effect of Highway Access on Number of Large Retailers (Hypermarkets and Supermarkets). 2SLS. 2010

Dependent Variable: Number of Large Retailers (Hypermarkets and Supermarkets)		
	(1)	(2)
Highway Access	1.521***	
	(0.166)	
Market Access		1.322***
		(0.149)
Region FE	YES	YES
Controls	YES	YES
R2	0.345	0.446
Obs	5451	
Mean Dep Var	8.369	

Notes: This table reports the 2SLS estimates on the effects of highways and market access on the number of large retailers (hypermarkets and supermarkets) in 2010. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and region fixed effects. Controls include 2000 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area). Standard errors clustered at the region level in parenthesis. All regressions are weighted by 2000 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table 12: The Effect of Highway Access and Market Access on Self-Employment. 2SLS

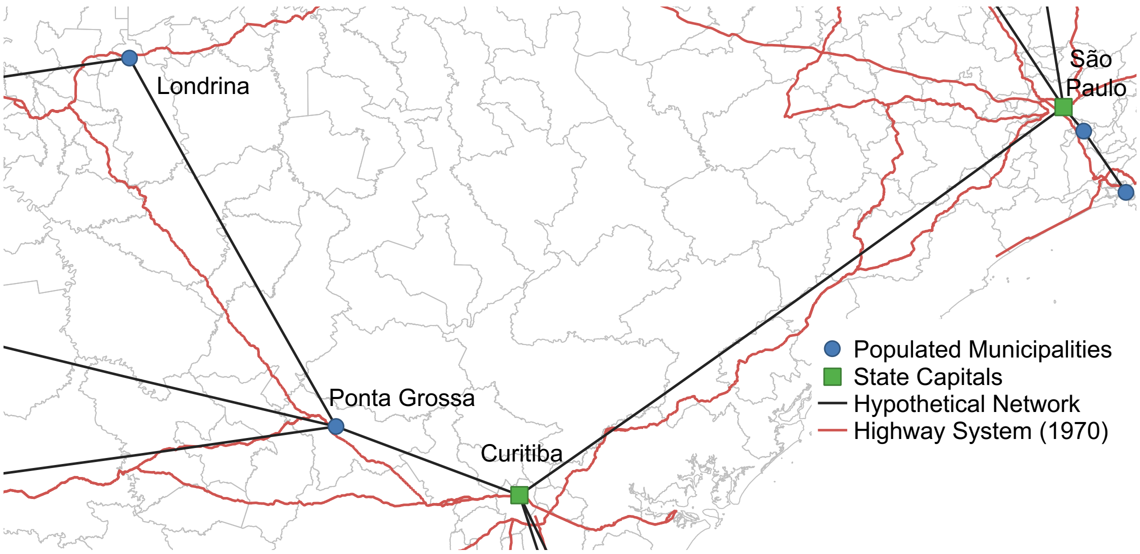
Dependent Variable: Share of Low Skilled and Non-Agriculture Employed Individuals that are Self-Employed

	All		Excluding Internal Migrants	
	(1)	(2)	(3)	(4)
Highways	-0.084***		-0.083***	
	(0.030)		(0.030)	
Market Access		-0.078***		-0.076***
		(0.017)		(0.018)
Year Fixed Effects	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Obs	3390	3390	3390	3390
R2	0.455	0.501	0.465	0.521

Notes: This table reports the 2SLS estimates on the effects of highways and market access on the share of low-skilled and non-agricultural employed individuals that are self-employed. Internal migrants are defined as those individuals that do not live in the municipality where they were born. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in the share of internal migrants for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

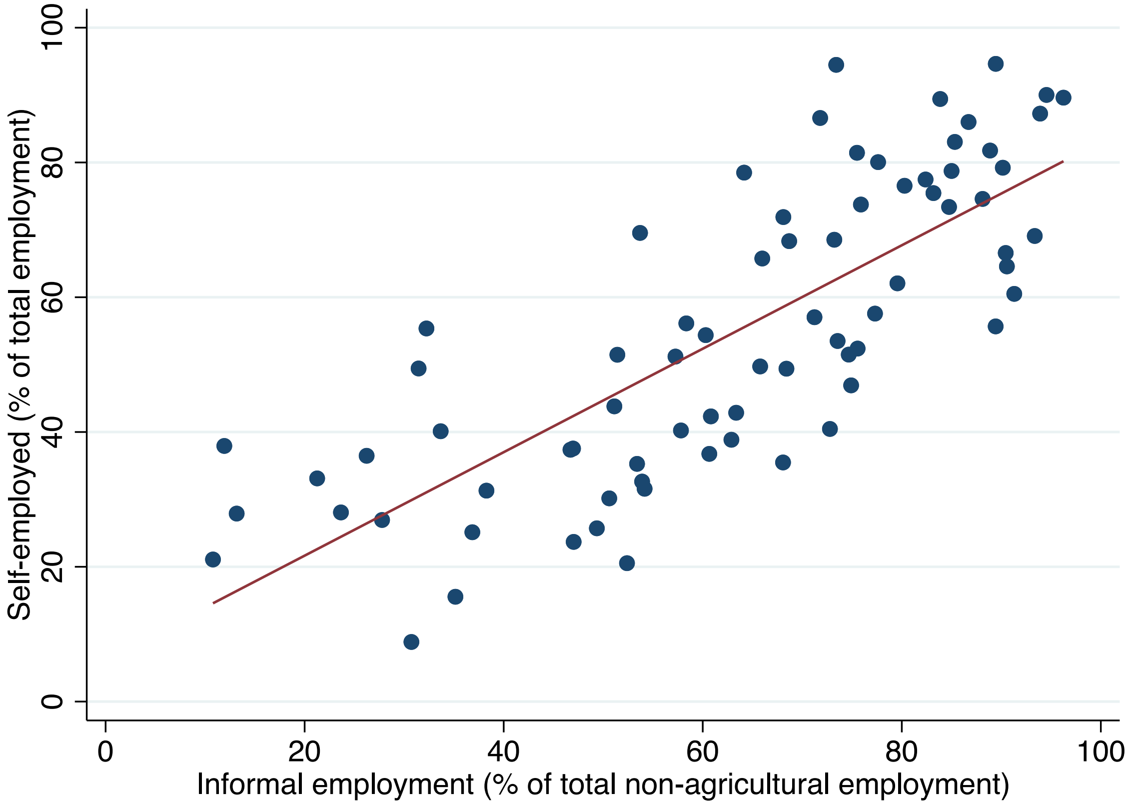
A Appendix: Tables and Figures

Figure A1: Hypothetical Optimal Network and Highway System, 1970. Northwest of Paraná and south of São Paulo.



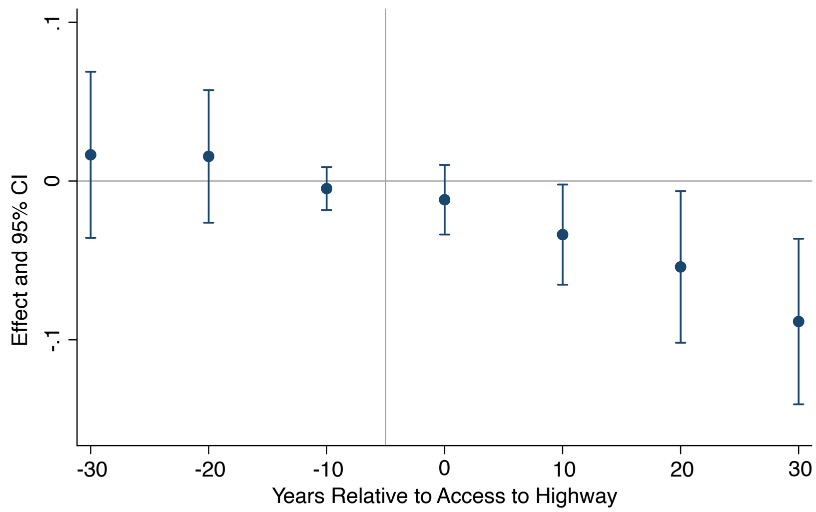
Notes: This figure shows the optimal hypothetical network and the highway system in the northwest of the state of Paraná and south of the state of São Paulo. The construction of the network is described in Section 5.2. Straight lines connect state capitals, populated municipalities and border municipalities (municipalities located at international border). Red lines represent the highway system in 1970.

Figure A2: Cross-Country Correlation Between Informal Salary Employment and Self-Employment



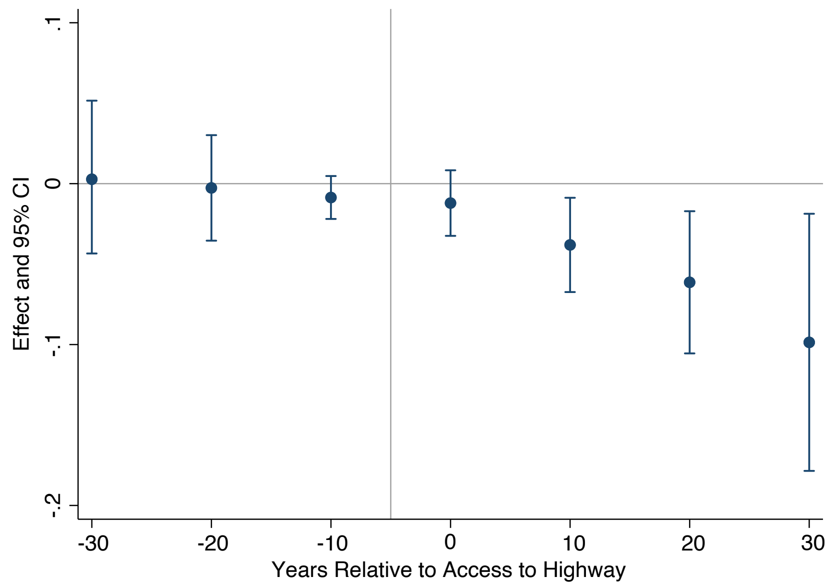
Notes: This figure shows the level of informal employment (percentage of the total non-agricultural employment salary wage that is unregistered) and the level of self-employment (percentage of total employment) for a set of 75 countries at the latest available year for each year. The set of countries includes high income countries, middle income countries and low income countries. The red line represents the linear relationship between the two variables derived from the data. Source: World Bank Database

Figure A3: Event Study Results: Share of Non-Agriculture Employed Individuals that are Self-Employed. De Chaisemartin and d'Haultfoeuille (2020) Estimator.



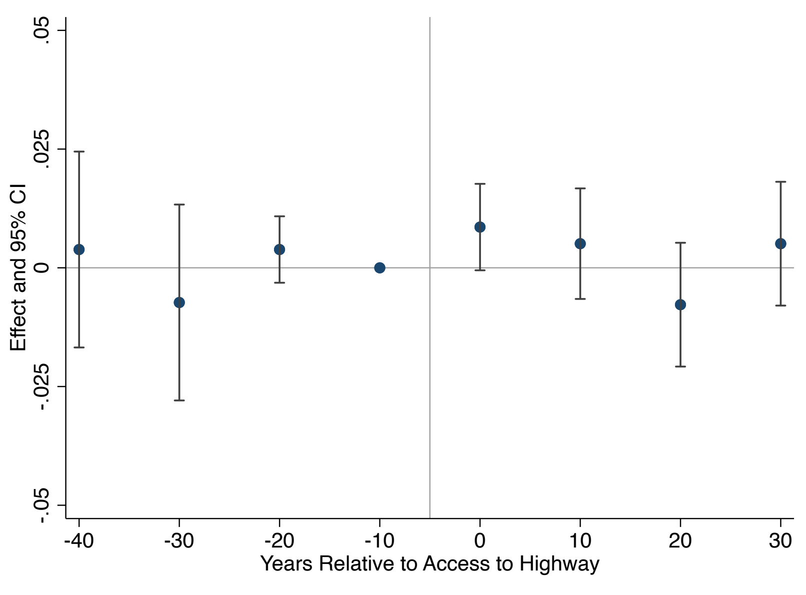
Notes: This figure reports coefficients and 95% confidence intervals estimated according to De Chaisemartin and d'Haultfoeuille (2020) estimator. The estimation includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Figure A4: Event Study Results: Share of Non-Agriculture and Low-Skilled Employed Individuals that are Self-Employed. De Chaisemartin and d'Haultfoeuille (2020) Estimator



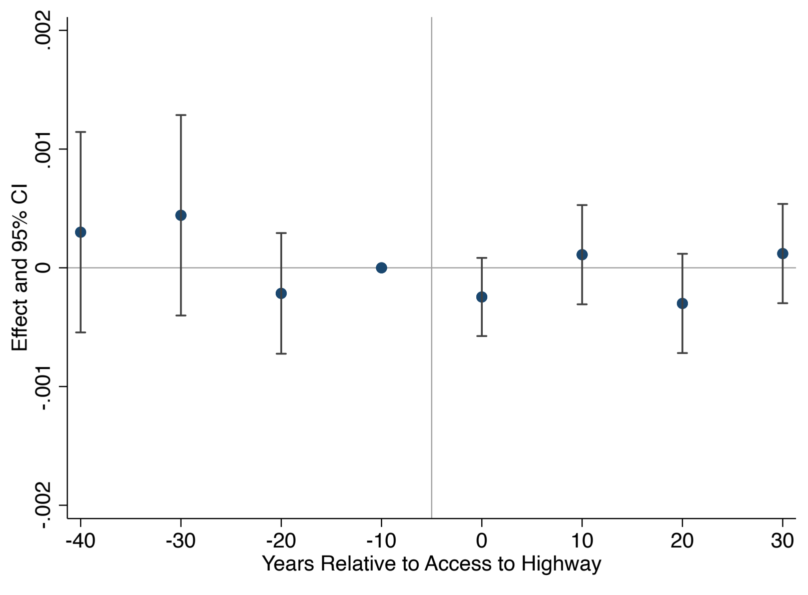
Notes: This figure reports coefficients and 95% confidence intervals estimated according to De Chaisemartin and d'Haultfoeuille (2020) estimator. The estimation includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Figure A5: Event Study Results: Share of Non-Agriculture and Low-Skilled Employed Individuals that are Self-Employed. Anticipation Effects in Control Group (Not-Yet-Treated Municipalities)



Notes: This figure reports estimates and 95% confidence intervals of coefficients β_j in specification (4). The estimation sample includes treated and not-yet-treated municipalities. Regression includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Figure A6: Event Study Results: Share of Non-Agriculture and Low-Skilled Employed Individuals that are Self-Employed. Anticipation Effects in Control Group (Never Treated Municipalities)



Notes: This figure reports estimates and 95% confidence intervals of coefficients β_j in specification (4). The estimation sample includes treated and never treated municipalities. Regression includes year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Regression is weighted by 1970 municipality share of national population.

Table A1: Municipalities with Highway Access. 1970-2010

	Municipalities with Access to Highway		
Year	Number	Percentage (%)	Variation (%)
1970	736	51.1	
1980	870	60.5	10.1
1990	1039	72.2	28.3
2000	1082	75.2	4.1
2010	1117	77.6	3.2

Notes: This table shows the number, percentage and the decennial variation in the number of municipalities with highway access. Highway access means that a highway intersects the municipality. The total number of municipalities is 1439. By the end of 2010, 322 municipalities did not have access to a highway.

Table A2: Summary Statistics at Municipality Level

Panel A: Summary Statistics in 1970	Mean	Std. Dev	P10	P25	P50	P75	P90
Share of Illiterate	0.343	0.199	0.123	0.164	0.301	0.503	0.651
Share of College Educated	0.01	0.012	0.000	0.001	0.003	0.015	0.029
Share of Immigrants	0.02	0.031	0.000	0.001	0.005	0.024	0.075
Share Rural	0.426	0.331	0.004	0.043	0.472	0.737	0.834
Share of Married	0.646	0.047	0.585	0.617	0.645	0.679	0.705
Lab. Force Participation Rate	0.541	0.044	0.487	0.515	0.539	0.562	0.587
Unemployment Rate	0.009	0.008	0.000	0.002	0.006	0.015	0.019
Share Manufacturing	0.125	0.123	0.018	0.032	0.076	0.183	0.342
Panel B: Summary Statistics in 2010	Mean	Std. Dev	P10	P25	P50	P75	P90
GDP per capita (Reais)	12520	13416	4218	5695	9846	14965	22044
# Registered Firms	162	463	26	46	81	163	316
Average Size Firm	2.715	2.963	0.846	1.275	2.018	3.273	5.044
Share of Large Firms (≥50 employees)	0.005	0.009	0.000	0.000	0.000	0.007	0.017
Share of Medium Firms (20 to 49 employees)	0.012	0.018	0.000	0.000	0.005	0.019	0.032
Share of Small Firms (< 20 employees)	0.983	0.023	0.957	0.974	0.99	1.000	1.000

Notes: Panel A of this table shows statistics on variables at the municipality level in 1970. The sample consists of individuals who were between age 18 and 64 and who were working in the year of the survey. Residents of institutional group quarters such as prisons and other institutions are dropped along with unpaid family workers. Panel B shows statistics on GDP per capita, number of registered firms and share of firms by size at the municipality level in 2010.

Table A3: Distribution of Non-Agricultural Workers by Industry and by Occupation, 1970

	Self-Employed, (%)	Salaried, (%)
Industry	(1)	(2)
Mining and Extraction	1.5	1.2
Manufacturing	8.4	26
Utilities	0.1	2.8
Construction	11.8	12.6
Wholesale and Retail Trade	36.1	12.2
Hotels and Restaurants	0.2	0.3
Transportation, Storage, and Communications	9.2	9.6
Financial Services and Insurance	0.2	3.5
Business Services and Real Estate	2.4	1.6
Education	0.9	8.8
Health and Social Work	1.5	2.8
Other Services	9.8	4.8
Private Household Services	17.9	13.8

Notes: This table shows the distribution of non-agricultural workers by industry and class of worker in 1970.

Table A4: Distribution of Non-Agricultural Self-Employed by Occupation (Top Occupations), 1970

Occupation	(%)
Retail shop workers	25.36
Tailors and seamstresses	8.91
Street vendors	8.16
Drivers	7.75
Laundresses and ironers	6.91
Stone masons	6.27
Carpenters	2.72
Barbers and hairdressers	2.65
Painters and whitewashers	1.99
Mechanics for int. combust. Engines	1.73
Cabinet makers and joiners	1.54
Shoemakers	1.53
Teamsters (oxen, horse, etc.)	1.13
Gold miners	0.99
Straw hat makers	0.94
Electricians	0.88
Commercial representatives	0.81

Notes: This table shows the distribution of non-agricultural self-employment by occupation. The occupations displayed comprise 80% of all non-agricultural self-employed individuals.

Table A5: Summary Statistics for Municipalities in 1970 by Optimal Network Status

	In Network		Not In Network		Difference	
	Mean	Std. Dev.	Mean	Std. Dev		Std. Error
Share College Individuals	0.002	0.003	0.003	0.004	-0.001	(0.000)
Share Immigrant	0.006	0.014	0.007	0.013	-0.001	(0.002)
Share Rural	0.671	0.219	0.600	0.258	0.071*	(0.027)
Share Married	0.659	0.044	0.656	0.042	0.003	(0.005)
Share Male	0.502	0.028	0.494	0.026	0.008	(0.004)
Mean Total Income	1098.650	536.003	1172.804	594.438	-74.154	(61.303)
Share Illiterate	0.441	0.186	0.440	0.191	0.002	(0.022)
Lab. Force Participation Rate (%)	54.530	5.244	53.991	5.145	0.539	(0.459)
Unemployment Rate (%)	0.372	0.600	0.456	0.574	-0.083	(0.056)
Share Manufacturing	0.076	0.100	0.087	0.107	-0.012	(0.006)
Share of HH with Sewage	0.109	0.154	0.158	0.194	-0.049	(0.032)
Share of HH with Drinking Water	0.205	0.215	0.264	0.259	-0.059	(0.032)
Share of HH with Electricity	0.262	0.249	0.321	0.276	-0.059	(0.038)
Working Age Population	6369.424	7834.278	4790.975	8189.782	1578.450*	(618.203)
Municipalities	603		732			

Notes: This table shows means and standard deviations of various characteristics at the municipality level for municipalities intersected (in-network) and not intersected (not in network) by the hypothetical optimal network. Municipalities in the nodes of the network are excluded. All these characteristics are measure at the baseline year. The difference and standard error is compute based on regression of each characteristic on an indicator variable that takes value one if the municipality is intersected by the hypothetical optimal network. Significance levels: * p<0.05 ** p<0.01 *** p<0.001

Table A6: Placebo Test.

Dependent Variable: Population Growth in 1940-1950		
Connected	0.031	0.025
	(0.020)	(0.033)
State Fixed Effects	No	Yes
Obs	950	950
R2	0.006	0.012
Mean Dep. Var	0.12	0.12

Notes: This table reports the OLS estimates of the regression of population growth between 1940 and 1950 on a dummy variable that takes value one if the municipality is intersected by the hypothetical optimal network. Due to available data the number of municipalities is lower than in the rest of the analysis. Standard errors clustered at the state level in parenthesis. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A7: The Effect of Highway Access on Self-Employment: OLS Estimates

Dependent Variable: Change in Share of Non-Agriculture Employed Individuals that are Self-Employed, 1970-2010

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Highway Access	-0.059***	-0.062***	-0.058***	-0.061***	-0.001	0.001
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Region Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	703	703	703	703	703	703
R2	0.043	0.048	0.044	0.050	0.048	0.048
Mean Dep Var	-0.1420		-0.1349		-0.0368	

Notes: This table reports the OLS estimates on the effects of change in highway access (binary indicator) on the change in self-employment. Δ Highway Access is a dummy variable that takes value one if the municipality is intersected by a federal highway in 2010, but not in 1970. It takes value zero for all other cases. All regressions include region fixed effects. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A8: The Effect of Market Access on Self-Employment: OLS Estimates

Dependent Variable: Change in Share of Non-Agriculture Employed Individuals that are Self-Employed, 1970-2010

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Market Access	-0.050***	-0.058***	-0.051***	-0.055***	-0.001	0.001
	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)	(0.011)
Region Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	703	703	703	703	703	703
R2	0.052	0.054	0.055	0.055	0.049	0.050
Mean Dep Var	-0.1420		-0.1349		-0.0368	

Notes: This table reports the OLS estimates on the effects of change in market access (continuous measure of connectivity) on the change in self-employment. Δ Market Access is the change in municipality's market access between 1970 and 2010. All regressions include region fixed effects. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A9: First Stage Regressions

Dependent Variable: Δ $HighwayAccess_{it}$ *(binary indicator), 1970-2010*

	(1)	(2)
In-Network	0.156***	0.158***
	(0.024)	(0.025)
Region Fixed Effects	YES	YES
Controls	NO	YES
Obs	678	678
R2	0.059	0.062
F-Statistic	52.187	50.198

Notes: This table reports the first stage regressions for instrumenting the change in highway access (binary indicator) between 1970 and 2010 with the hypothetical optimal network. In-network is a dummy variable that takes value one if the municipality is intersected by the hypothetical optimal network. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A10: First Stage Regressions

Dependent Variable: Δ *MarketAccess*$_{it}$ *(continuous measure), 1970-2010*

	(1)	(2)
Predicted Market Access	0.257***	0.298***
	(0.021)	(0.020)
Region Fixed Effects	YES	YES
Controls	NO	YES
Obs	678	678
R2	0.079	0.081
F-Statistic	66.187	64.148

Notes: This table reports the first stage regressions for instrumenting the change in market access (binary indicator) between 1970 and 2010 with the hypothetical optimal network. Predicted market access is the market access derived from the hypothetical optimal network. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A11: The Effect of Highway Access on Self-Employment: 2SLS Estimates

Dependent Variable: Change in Share of Non-Agriculture Employed Individuals that are Self-Employed, 1970-2010

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Highway Access	-0.065*	-0.075*	-0.059*	-0.069*	-0.001	-0.001
	(0.039)	(0.044)	(0.034)	(0.040)	(0.030)	(0.030)
Region Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	678	678	678	678	678	678
R2	0.043	0.046	0.043	0.053	0.043	0.043
Mean Dep Var	-0.1420		-0.1349		-0.0368	

Notes: This table reports the 2SLS estimates on the effects of change in highway access (binary indicator) on the change in self-employment using the hypothetical optimal network as an instrument. Δ Highway Access is a dummy variable that takes value one if the municipality is intersected by a federal highway in 2010, but not in 1970. It takes value zero for all other cases. All regressions include region fixed effects. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A12: The Effect of Market Access on Self-Employment: 2SLS Estimates

Dependent Variable: Change in Share of Non-Agriculture Employed Individuals that are Self-Employed, 1970-2010

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Market Access	-0.062*	-0.068*	-0.060*	-0.066*	-0.001	-0.001
	(0.035)	(0.038)	(0.034)	(0.039)	(0.032)	(0.0303)
Region Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	678	678	678	678	678	678
R2	0.055	0.058	0.056	0.057	0.051	0.052
Mean Dep Var	-0.1420		-0.1349		-0.0368	

Notes: This table reports the 2SLS estimates on the effects of change in market access (continuous measure of connectivity) on the change in self-employment using the predicted market access derived from the hypothetical optimal network as an instrument. Δ Market Access is the change in municipality's market access between 1970 and 2010. All regressions include region fixed effects. Controls include 1970 municipality variables: share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A13: The Effect of Highway Access on Self-Employment: DiD Estimates (excluding municipalities at nodes of hypothetical network)

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Highway x Post	-0.043***	-0.032***	-0.046***	-0.035***	-0.001	-0.001
	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	9107	9107	9107	9107	9107	9107
R2	0.757	0.802	0.760	0.805	0.458	0.494
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the OLS estimates on the effects of highways on self-employment using equation (3). Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A14: The Effect of Market Access on Self-Employment: OLS Estimates (excluding municipalities at nodes of hypothetical network)

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.037***	-0.031***	-0.047***	-0.036***	0.002	0.001
	(0.011)	(0.010)	(0.012)	(0.011)	(0.015)	(0.016)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3390	3390	3390	3390	3390	3390
R2	0.527	0.543	0.587	0.544	0.472	0.501
Mean Dep Var	0.292		0.317		0.072	
Mean Dep Var in 1970	0.366		0.379		0.106	

Notes: This table reports the OLS estimates on the effects of market access on self-employment using equation (5). Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A15: Correlation Between Binary Indicator of Highway Access and Market Access

Dependent Variable: Market Access

	(1)	(2)	(3)	(4)
Highway	0.901***	0.902***	0.877***	0.879***
	(0.012)	(0.013)	(0.011)	(0.009)
Municipality FE	NO	NO	YES	YES
Year FE	NO	YES	YES	YES
Controls	NO	NO	NO	YES
Obs	3515	3515	3515	3515
R2	0.323	0.324	0.441	0.544

Notes: This table reports the regression of market access on a dummy variable that takes value one if the municipality is intersected by a highway. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A16: The Effect of Highway Access on Self-Employment: DiD Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed, by Age Group

	18-24		25-34		35-64	
	(1)	(2)	(3)	(4)	(5)	(6)
Highway	-0.000	-0.001	0.003	0.002	0.005*	0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Highway x Post	-0.034***	-0.024***	-0.054***	-0.038***	-0.057***	-0.037***
	(0.008)	(0.006)	(0.009)	(0.008)	(0.010)	(0.007)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	9361	9361	9361	9361	9361	9361
R2	0.647	0.682	0.744	0.774	0.796	0.828
Mean Dep Var	0.164		0.289		0.409	
Mean Dep Var in 1970	0.199		0.374		0.499	

Notes: This table reports the OLS estimates on the effects of highways on self-employment using equation (3) for different age groups. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A17: The Effect of Highway Access on Self-Employment: DiD Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed by Gender

	Male		Female	
	(1)	(2)	(3)	(4)
Highway	0.004	0.002	-0.001	0.000
	(0.003)	(0.002)	(0.002)	(0.002)
Highway x Post	-0.051***	-0.031***	-0.033***	-0.021***
	(0.009)	(0.007)	(0.009)	(0.007)
Year Fixed Effects	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES
Controls	NO	YES	NO	YES
Obs	9361	9361	9361	9361
R2	0.792	0.822	0.625	0.692
Mean Dep Var	0.354		0.241	
Mean Dep Var in 1970	0.411		0.292	

Notes: This table reports the OLS estimates on the effects of highways on self-employment using equation (3) for male and female workers separately. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A18: The Effect of Market Access Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed, by Age Group

	18-24		25-34		35-64	
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	-0.052***	-0.064***	-0.067***	-0.074***	-0.090***	-0.094***
	(0.013)	(0.012)	(0.007)	(0.008)	(0.012)	(0.015)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs	3390	3390	3390	3390	3390	3390
R2	0.366	0.399	0.459	0.573	0.471	0.603
Mean Dep Var	0.164		0.289		0.409	
Mean Dep Var in 1970	0.199		0.374		0.499	

Notes: This table reports the 2SLS estimates on the effects of market access on self-employment using equation (7) and instrumenting market access with predicted market access, for different age groups. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * p<0.10 ** p<0.05 *** p<0.01.

Table A19: The Effect of Market Access on Self-Employment: 2SLS Estimates

Dependent Variable: Share of Non-Agriculture Low-Skilled Employed Individuals that are Self-Employed by Gender

	Male		Female	
	(1)	(2)	(3)	(4)
Market Access	-0.091***	-0.093***	-0.003	-0.004
	(0.016)	(0.016)	(0.020)	(0.013)
Year Fixed Effects	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES
Controls	NO	YES	NO	YES
Obs	3390	3390	3390	3390
R2	0.436	0.472	0.355	0.411
Mean Dep Var	0.354		0.241	
Mean Dep Var in 1970	0.411		0.292	

Notes: This table reports the 2SLS estimates on the effects of market access on self-employment using equation (7) and instrumenting market access with predicted market access, for male and female workers separately. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A20: The Effect of Highway Access on Self-Employment: "Naive" and Robust Estimators

Dependent Variable: Share of Non-Agriculture Employed Individuals that are Self-Employed

	All		Low Skilled (Non-College)		High Skilled (College)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: "Naive" specification						
Highway	-0.040	-0.028	-0.042	-0.029	-0.001	-0.001
	(0.008)	(0.006)	(0.008)	(0.006)	(0.005)	(0.005)
Obs	3515	3515	3515	3515	3515	3515
R2	0.748	0.791	0.748	0.791	0.445	0.490
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
% ATTs with negative weights	11.51	9.62	11.51	9.62	11.51	9.62
Sum of negative weights	-0.006	-0.0382	-0.006	-0.0382	-0.006	-0.0382
$\underline{\sigma}_{fe}$	0.038	0.026	0.042	0.028	0.019	0.013
$\underline{\underline{\sigma}}_{fe}$	1.422	0.112	1.584	0.121	0.723	0.057
Panel B: De Chaisemartin and d'Haultfoeuille (2020) estimator						
Highway	-0.052	-0.04	-0.055	-0.044	-0.002	-0.002
	(0.012)	(0.015)	(0.011)	(0.017)	(0.014)	(0.014)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Municipality Fixed Effects	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES
Obs[(a)]	4892	4892	4892	4892	4892	4892

Notes: Panel A reports the OLS estimates on the effects of highways on self-employment using equation (2). Panel B reports the estimates using De Chaisemartin and d'Haultfoeuille (2020) estimator. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. All regressions include year fixed effects and municipality fixed effects. Controls include 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. (a) Number of municipality-period observations used to estimate each effect. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population.

Table A21: The Effect of Highway Access on Self-Employment and Non-Registered Salaried Employment

	Dependent Variable: Share of Non-Agriculture and Non-College Employed Individuals that are Self-Employed or Salaried without legal contract			
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Highway	-0.026**	-0.032**		
	(0.012)	(0.015)		
Market Access			-0.023*	-0.029*
			(0.013)	(0.016)
Municipality Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Obs	714	700	714	700
R2	0.976	0.978	0.893	0.903

Notes: This table reports the OLS and 2SLS estimates on the effects of highways on informality. Informality is defined as the share of non-agricultural and non-college employed individuals that are either self-employed or salaried workers without legal contract. Highway is a dummy variable that takes value one if the municipality is intersected by a federal highway. Market access is calculated using equation (1) for each year based on the full highway network in that year and the distribution of population across municipalities in 1970. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. Controls include 2000 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 2000 municipality share of national population. Significance levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A22: Robustness of The Market Access Based Results to Varying the Connection Cut-Offs and Trade Elasticity (θ) Value

	$\theta = 1.5$	$\theta = 1$	$\theta = 1.1$	$\theta = 3.6$	$\theta = 3.73$	$\theta = 3.80$	$\theta = 6.74$	$\theta = 12.86$	$\theta = 26.83$
Cut-Off	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
50 km	-0.078	-0.080	-0.080	-0.079	-0.078	-0.078	-0.078	-0.077	-0.075
	(0.017)	(0.017)	(0.017)	(0.017)	(0.015)	(0.017)	(0.017)	(0.019)	(0.017)
100 km	-0.074	-0.074	-0.074	-0.073	-0.072	-0.072	-0.071	-0.071	-0.065
	(0.017)	(0.016)	(0.017)	(0.016)	(0.016)	(0.018)	(0.016)	(0.016)	(0.016)
200 km	-0.066	-0.066	-0.066	-0.065	-0.065	-0.064	-0.063	-0.063	-0.057
	(0.017)	(0.018)	(0.019)	(0.017)	(0.017)	(0.017)	(0.018)	(0.019)	(0.017)

Notes: This table reports the 2SLS estimates on the effects of market access on self-employment for different values of θ and different values of the connecting cut-off. Each coefficient is based on a separate 2SLS regression with 3,515 observations. The dependent variable is the share of non-agricultural and low-skilled individuals that are self-employed. Market access is normalized so that the corresponding coefficient indicates the change in self-employment for a one standard deviation increase in market access. All regressions include year fixed effects and municipality fixed effects. All regressions include the following controls: 1970 municipality variables (share of the adult population that is illiterate, share of population living in rural conditions, share of employed individuals in manufacturing, and log of population) and geographical municipality variables (distance to the border, distance to the capital, distance to the coast and area), each of them interacted with a linear time trend. Standard errors clustered at the municipality level in parenthesis. All regressions are weighted by 1970 municipality share of national population.