Road Endpoints and City Sizes*

Bruno Barsanetti†

June 2021

Abstract

The completion of transportation infrastructure frequently takes many years and occurs gradually. How does the gradual construction of transportation infrastructure affect the distribution of economic activity across the sites it serves? I examine the long-run effects of the timing of railroad construction on city sizes. I first present a model that predicts that towns that are railroad endpoints for longer become persistently larger. I then show that, in a sample of Brazilian railroad towns, time as endpoint strongly predicts town size: each additional year that a town was a railroad endpoint in the past is associated with a town population 0.112 log point larger in 2010. Additional testable implications, historical evidence, and instrumental variable estimates suggest that such association reflects the causal effects predicted by the model.

JEL Classification Codes: R12, O18, R40.
Keywords: railroads, city sizes, second nature, path dependence.

---

*I appreciate comments from Mario Cannella, José Carreño, Alípio Ferreira, Seema Jayachandran, Cynthia Kinnan, and Chris Udry, seminar audiences at Insper, FEA-USP, and EPGE-FGV, and research funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. All errors are mine.

†EPGE Brazilian School of Economics and Finance. Praia de Botafogo 190, Botafogo, Rio de Janeiro, RJ, 22250-900, Brazil. Phone: +55 (21) 3799-5832. E-mail: bruno.barsanetti@fgv.br

1 Introduction

Transportation infrastructure projects frequently leave important and persistent effects on economic activity, as shown, among others, by Duranton and Turner (2012), Jedwab and Moradi (2016), and Bird and Straub (2020). It is also common for the completion of these projects to take many years, so there is often a gradual opening of the transportation structure to use.[1] Still, most research on the effects of transportation infrastructure abstracts from this aspect: theoretical models are often static, empirical findings mostly focus on access to transportation. My point in this paper is that the timing of infrastructure completion plays a major role in the way transportation infrastructure shapes economic activity around it.

When road construction is gradual, there is a succession of endpoint sites, a fact I exploit to examine if and how the timing of road opening affects the long-run distribution of economic activity along the road. I begin by considering this question within a simple theoretical framework, whose main prediction is that towns that spend more years as road endpoints become persistently larger, even after they no longer are endpoints. I then present empirical evidence which suggests this effect is substantial. In a sample of railroad towns in Brazil, a long historical time as railroad endpoint predicts a large city size. Additional testable implications, historical evidence, and instrumental variable estimates indicate such positive relationship is causal.

Why does the timing of gradual opening of a road matter for economic activity along the road? At each site to which a road has just arrived, the economic benefits of the road depend on how productive other sites on the road are. Such dependence is the result of agglomeration shadows (Fujita *et al* 1999, Cuberes *et al* 2021): if nearby sites are very productive, firms producing at the newly arrived site might not be able to compete with firms from these nearby locations. The intuition for the results in this paper is that, with a slower (faster) timing of road construction, local productivity growth implies that the strenght of these agglomeration shadows would be stronger (weaker). The timing of road construction then determines the distribution of economic activity

[1] Examples of gradual completion and opening include the São Paulo subway Line 2, that opened in different stages from 1991 to 2010, the Golden Quadrilateral highway upgrading project in India, which was completed in different waves from 2001 to 2013, and the Second Avenue subway in New York, which was partially opened in 2017 but may take more than a decade for completion.

along the road.

I formalize this intuition in a stylized model of an economy in which a road is built over time. The road connects interior locations with a gateway through which this economy trades with the rest of the world. Towns provide urban services to a rural hinterland that produces a tradable agricultural good. The road reduces the costs of moving the agricultural good to the gateway and provides urban firms with access to inputs for the production of the services they provide. New towns emerge on the road as it gets constructed. Over time, towns become more productive in the provision of urban services.

The main prediction of the model is that a town that was a road endpoint for more years will be both larger and farther away from the next town. To understand the result, note that neighboring towns compete for hinterland markets and that new towns emerge on newly constructed segments of the road. If a town spends a longer time as an endpoint, it is more productive in providing urban services at the time the road reaches new sites farther ahead. Therefore, the town casts a larger urban shadow over these sites, preventing their urban development.[2] The first new town will be pushed farther down the road, and the endpoint town will persistently serve a larger hinterland. Since hinterland size is associated with town size, a town that was an endpoint for longer will be larger even after it is no longer a road endpoint.

As evidence for the theoretical predictions, I document the associations between time as endpoint and city size for a sample of railroad towns that surround the city of São Paulo, Brazil. São Paulo is the center of a radial railroad network that was built toward frontier regions in the west. Since the railroads promoted the settlement of the frontier, many new towns emerged around the railroads, which makes the empirical setting suitable for a test of the model predictions. In fact, classic studies in geography already highlight the importance of railroad endpoints for city growth in the region (Deffontaines 1938, Monbeig 1952). As a case study, I show that the model predictions are consistent with the historical development of São José do Rio Preto, a town that was a railroad

[2]Cuberes *et al* (2021) have shown for the United States that proximity to urban centers may hurt town growth due to increased competition. This is particularly true for the period of 1840-1920, which is precisely the time period when most of the railroad construction I examine took place. Other papers have also documented urban shadows; see, e.g., Bosker and Buringh (2017) and Hodgson (2018).

endpoint for 21 years due to plausibly exogenous reasons, including a long legal dispute over the railroad concession.

I show that time as endpoint is a strong predictor of city size. Each additional year that a town was a railroad endpoint in the past is associated with a municipality urban population that is 0.112 log point larger and an urban GDP that is 0.124 log point larger in 2010. Time as endpoint also predicts distance to the next municipality farther down the railroad: each additional year as endpoint is associated with a distance to the next town that is 0.037 log point greater. Such associations are consistent with the model predictions. They are also robust to alternative specifications, to different inference procedures, and to the inclusion of a large set of geographic fundamentals as control variables. When assessing the bias due to omitted variables according to the approach proposed by Altonji *et al* (2005) and Oster (2019), it seems unlikely that the estimated coefficients are due to omitted geographic fundamentals. And although the baseline estimates use urban population data from the 2010 census, data from previous census years produce almost identical estimates, suggesting that the association between past time as endpoint and city sizes is persistent.

I test for an asymmetry in the theoretical predictions: time as endpoint increases distance to the *next* town but does not affect distance to the *previous* town. The intuition for this asymmetry lies in the fact that urban shadows in the model matter for city emergence precisely on sites that were just connected by the railroad. In fact, the data shows no association between time as endpoint and distance to the previous municipality. The estimated coefficient (-0.007) is statistically indistinguishable from zero, precisely estimated, and five times smaller than the coefficient for the association with distance to the next municipality. Such asymmetric pattern brings credibility to the interpretation that the positive correlation between time as endpoint and town size reflects the model predictions.

An instrumental variable approach confirms the empirical findings. I instrument time as endpoint with the logarithm of distance to the closest already-incorporated municipality farther down the railroad line. The location of incorporated municipalities is a proxy for the location of already established towns. The instrument thus exploits the position of a site with respect to important urban centers that would eventually be connected by the railroad: when one such center was closer, the railroad company had more incentives to build rapidly in order to connect it earlier and profit from its demand for transportation. This incentive is plausibly unrelated to the growth potential of the site. As expected, a larger distance to the next already incorporated town increases time as endpoint.

The two-stage least squares estimates point on the same direction as the ordinary least squares coefficients. The point estimates imply that one additional year as endpoint increases urban population by 0.228 log point. As before, the results indicate positive effects on distance to the next town but not on distance to the previous town. Moreover, two robustness exercises suggest the validity of the exclusion restriction. First, I account for the threat that the instrument could correlate with geographic fundamentals by showing that the estimates are robust to the inclusion of a large set of control variables. Second, I exploit distance to the closest already-incorporated municipality farther up (instead of farther down, as used to construct the instrument) to show that the instrumental variable results are not due to remoteness at the time of railroad arrival. In sum, the instrumental variable estimates suggest that the positive association between time as endpoint and town size reflects a causal effect of the timing of railroad provision.

The rest of the introduction discusses the contribution to the literature. Section 2 presents the model, its main predictions, and model extensions. Section 3 presents the historical setting, the empirical framework, the data, and a historical case study. Section 4 presents the associations between time as endpoint and city sizes, the robustness of the results, the instrumenal variable findings, and a discussion of the mechanisms. Section 5 concludes with implications for infrastructure policy and for future research.

1.1 Related Literature

This paper contributes to the empirical literature on the effects of transportation infrastructure on economic activity; see Redding and Turner (2015) for a review. Railroads have been shown to impact economic activity in developing economies; for instance, see Jedwab and Moradi (2016), Qin (2017), Donaldson (2018), Banerjee *et al* (2020), Forero *et al* (2020), or Américo (2021). In general, papers in the literature consider the effect of proximity to railroad lines or of changes in the overall transportation infrastructure that lead to increased market access. Since I examine how the gradual opening of the railroad matters through the time each town spends as an endpoint, I contribute to the literature by investigating the time dimension of investments in transportation infrastructure. To the extent of my knowledge, the timing of railroad construction has been used in economics only as a source of variation to examine different questions, such as place-based industrial policies by Fan and Zou (2019) or the long-run effects of immigration by Sequeira *et al* (2020), but not as an economic phenomenon of interest itself.

My findings relate to the literature on persistence in spatial economics; see Lin and Rauch (2020) for a review. The literature contains evidence that historical legacy factors may permanently affect the spatial distribution of economic activity. Historical transportation infrastructure is one such legacy factor; see, for instance, Jedwab *et al* (2017), Wahl (2017), Brooks and Lutz (2019), Barsanetti (2021), and Flückiger *et al* (2021). If the past arrival of a railroad to a site affects urban development, it can also shape the later spatial development of the region, explaining why time as endpoint could affect city size.

This paper also adds to a theoretical literature on how transportation infrastructure determines city sizes; see Sasaki (1992), Fujita and Mori (1996), Mun (1997), and Fajgenbaum and Schaal (2020). The models in these papers are either static or assume that the transportation network does not change, while in my model the road is built over time. Two recent papers consider transportation improvements in a dynamic setting: Trew (2020) models an economy in which the transportation sector may change endogenously over time, and Balboni (2019) estimates a dynamic spatial quantitative model to evaluate the welfare effects of road building in Vietnam. These papers do not focus on the effects of gradual transportation opening, focusing instead on aggregate economic growth or welfare.

2 Model

2.1 Environment

Consider a railroad that is built along a route.[3] The different sites y in this economy are located on the half-line $Y \equiv [0, \infty)$. Site $y = 0$ is a gateway (a port town) through which this economy trades with the rest of the world (RoW).[4] Time is also continuous: $t \in T \equiv [0, \infty)$. Starting from the initial period $t = 0$, a railroad is built along the half-line Y. The railroad construction schedule is exogenous. At time t, the railroad has been built up to site $y^r(t)$. Assume $y^r(.)$ is absolutely continuous, so it can be written as $y^r(t) = \int_0^t g^r(\theta)d\theta$; $g^r(t)$ is the expansion rate of the railroad at time t. For each site y, define the time the railroad reaches it as $t^r(y) \equiv min\{t \in T | y^r(t) \geq y\}$.

There are three goods in this economy:

1. an agricultural good a that is exported to the RoW through the gateway town $y = 0$. Denote its price on site y and at time t as $p(y, t)$. The price it is transacted with the RoW is $p(0, t) = p$ for all $t \in T$. The agricultural good can be costlessly transported by rail. When it is transported overland through a transportation mode other than rail, there is an iceberg cost: $e^{\tau_a d}$ units have to be shipped for one unit to arrive at a destination that is at a distance d.

2. an intermediate good m that is imported from the RoW at the gateway town $y = 0$. Normalize the price of the intermediate good at $y = 0$ to 1. It can be costlessly transported by rail, but the cost of transporting it overland outside of the railroad is prohibitive ($\tau_m = \infty$).

3. urban services s that cannot be traded with the RoW and are transported overland at an iceberg cost τ_s regardless of the transportation mode. One can think of s as a non-tradable good that does not benefit from the railroad to reach a larger market. This is a simplifying assumption that captures some features of non-tradable sectors. A variety of services that are focused on serving local needs rely on tradable inputs. For instance, an auto repair shop

[3]There is nothing specific about railroads in the model. I use the term railroad because it connects to the empirical setting examined later on the paper. The model predictions should be valid for other types of roads.

[4]Other papers that consider transport infrastructure on a half-line include Mori (2012) and Coşar and Fajgelbaum (2016).

provides its services within a limited ratio, to customers who can bring their vehicles to the shop. A railroad reduces the cost the shop has in purchasing components but may not reduce the shop's ability to reach customers. Let $q(y,t)$ be the price of this good in site y at time t.

There are two types of agents in this economy. First, there are immobile farmers, uniformly distributed across the half-line Y. The presence of uniformly distributed farmers is a common assumption in spatial economics models; see Fujita *et al* (1999) and Proost and Thisse (2019). Normalize the mass of farmers in each site y to 1. Each farmer produces one unit of the agricultural good a per period. They consume both the agricultural good a and the urban services s, that they value according to the utility function $u(a,s) = a^{\beta} s^{1-\beta}$.

There are also urban firms. They use one unit of the intermediate good m to produce $e^{A(y,t)}$ units of urban services s. Note that $A(y,t)$ measures the logarithm of the site-specific productivity at period t, which is the same across all firms that are located at that site y. In each period, there is free entry of urban firms, that decide on which site to locate. If they enter a site y, they produce with the same productivity as other firms that were in that site. That is, they immediately learn from the experience of firms that were in that site in the past. I say that, at time t, there is a town on site y if there is positive production of urban services at that site: $s(y,t) > 0$. I assume that, when firms are indifferent between entering or not into a site, they enter; this assumption guarantees uniqueness of the equilibrium.

Initially, log-productivity is $A(0,0) = A_0 \geq 0$ at the port town. The port town is the only town at the initial period, as the required intermediate good is not available at other sites. For the other sites $y > 0$, assume that initial log-productivity is $A(y,0) = 0$.

I assume that, over time, site-specific productivity increases in towns but not in non-town sites. This is a simplifying assumption that captures the idea that there is learning in cities. It implies intertemporal local spillovers: the presence of firms in a site in the past implies that, in the future, that site will be more productive. The assumption is that productivity growth follows:

$$\dot{A}(y,t) = \begin{cases} 0 & \text{if } s(y,t) = 0 \\ g(A(y,t)) & \text{if } s(y,t) > 0 \end{cases}$$

where $g(.)$ is a Lipschitz continuous, non-negative, and non-increasing function. I assume that the productivity growth of a newly formed town is sufficiently fast when compared to the expansion rate of the railroad. This assumption guarantees that, at any point in time, there is a finite number of cities in the economy.

Assumption 1. $\tau_s \cdot sup\ g^r(t) < g(0)$

Productivity growth in cities acts as a force favoring the agglomeration of urban services. Dispersion forces exist due to the transport costs. Assumption 1 is a statement comparing such agglomeration and dispersion forces.

2.2 Equilibrium

I first examine the prices and the distribution of urban firms across space at time t. At time t, the railroad has been built up to $y^r(t)$. Using the transportation costs for the agricultural good, the price of the agricultural good is given by:

$$log\ p(y,t) = \begin{cases} log(p) & \text{if } y \leq y^r(t) \\ log(p) - \tau_a[y - y^r(t)] & \text{if } y > y^r(t) \end{cases} \tag{1}$$

Given these prices, the expenditure from site y farmers on urban services will be given by $(1-\beta)p(y,t)$. Since the prices of the agricultural good affect farmers' income, the demand for urban services will depend on the agricultural good prices.

What about the prices of urban services? Urban services are supplied at towns that exist at time t. Denote the set of all such towns as $\mathbb{Y}(t) \equiv \{y \in Y | s(y,t) > 0\}$. Since the production of urban services requires the use of the intermediate good, all towns will be on sites that have been reached by the railroad: $\mathbb{Y}(t) \subset [0, y^r(t)]$. Moreover, the free entry condition implies that urban services prices at city $y \in \mathbb{Y}(t)$ satisfy $log\ q(y,t) = -A(y,t)$. Accounting for transportation costs, the price of urban services at site y is:

$$log\ q(y,t) = min_{\hat{y} \in \mathbb{Y}(t)} \Big(-A(\hat{y},t) + \tau_s |y - \hat{y}| \Big) \tag{2}$$

Over time, the site-specific log-productivity $A(y,t)$ of each town increases. As a consequence, urban services prices must fall everywhere. This fact has an important implication: any site that, at a time t, **(a)** had been reached by the railroad ($y \leq y^r(t)$) and **(b)** was not a town on that period ($y \notin \mathbb{Y}(t)$) will never be a town in a future period. To understand this implication, note that a site y can become a town only if its productivity is high enough so urban services can be produced as cheap as it takes to buy them from somewhere else: $log\ q(y,t) \geq -A(y,t)$. When a site is not a town, the right-hand side of inequality 2 does not change over time, while the left-hand side goes down.

It is also possible to show that, once a town emerges on a site y, it will always exist. This occurs for two reasons. First, relative to towns that will emerge later, the town at y will always be at least weakly more productive.[5] As a consequence, local firms will always be able to sell locally at a lower price than firms from younger towns. Second, relative to towns that already existed when y emerged, productivity growth at y will always be larger, since $g(.)$ is non-increasing. Hence, if local firms can sell at a lower price than firms from existing towns at time t, they would also be able to do so in a future time $t' > t$.

Because of these two properties, there will be a single equilibrium in this economy. There is always a finite number of towns, that expand over time. The location of the towns can be solved recursively, starting from the first town (the gateway town). The proposition below formalizes this solution.

Proposition 1. *There is an unique equilibrium. The set of towns $\mathbb{Y}(t)$ is finite and non-decreasing over time. At the initial period, $\mathbb{Y}(0)$ consists only of $y_0 = 0$. A recursive procedure defines $\mathbb{Y}(t)$:*

1. *Let $\{y_0, y_1, ..., y_N\}$ be the set of towns at a given time. Define $t_N = t^r(y_N)$.*

2. *For all $t \in [t_N, t_{N+1})$, $\mathbb{Y}(t) = \{y_0, y_1, ..., y_N\}$. A next town is founded only in period t_{N+1}, that satisfies:*

[5] This result is a consequence on the assumptions on productivity growth. It ensures the equilibrium will be unique. I later discuss these assumptions.

$$t_{N+1} = min\left\{ t > t_{N+1} \;\middle|\; \tau_s[y^r(t) - y_N] = A(y_N, t) \right\}$$

3. *At time* t_{N+1}*,* $\mathbb{Y}(t_{N+1}) = \{y_0, y_1, ..., y_N, y_{N+1}\}$*, where* $y_{N+1} = y^r(t_{N+1})$*.*

The proof of the proposition is in Appendix A. In what follows, I provide a sketch of the proof. The argument in the preceding two paragraphs implies that, at each period t, there is a finite number of towns, and that the set of towns grows with t. The proof uses this fact to construct a sequence of dynamical systems. In each dynamical system, the set of towns is fixed and productivity, city sizes and prices change according to the equilibrium condition. The solution to each of these dynamical systems is unique. Eventually, it becomes profitable for an urban firm to enter at a new railroad endpoint $y^r(t)$. This happens precisely at the site at which distance to the previous endpoint town y_N is sufficiently large so that a new firm can produce at a lower cost to local customers at $y^{N+1} \equiv y^r(t)$. This occurs if $A(y^{N+1}, t) = 0 \geq A(y_N, t) - \tau_s[y^r(t) - y_N]$. When this happens, the set of towns and the dynamical system may be redefined to include y^{N+1}. The argument follows by induction.

According to Proposition 1, the next town to emerge will be on the first site where, due to transportation costs, the price of urban services provided by the last existing town on the railroad is high enough so that it becomes profitable for a firm to establish itself at that site. Assumption 1 is an upper bound to the railroad expansion rate that implies the site of the next town will not be reached immediately.

Proposition 1 characterizes, at any time t, the set of towns in equilibrium. To completely characterize the equilibrium, we need to find prices $q(y, t)$ and town sizes $s(y, t)$. Prices are given by equations (1) and (2). And, given the prices, it is possible to define which sites are served by each town (town hinterland). If $\mathbb{Y}(t) = \{y_0, y_1, ..., y_N\}$, denote the hinterland served by town y_n as $[d_n(t), d_{n+1}(t)]$. The hinterland border $d_n(t)$ is the site where farmers are indifferent between consuming from y_{n-1} and y_n.[6] It is given by:

[6]For completeness, let $d_0(t) = 0$ and $d_{N+1}(t) = \infty$.

$$d_n(t) = \frac{y_n + y_{n-1}}{2} - \frac{A(y_n, t) - A(y_{n-1}, t)}{2\tau_s} \tag{3}$$

The total production of urban services in town y_n is given by the demand from its rural hinterland $[d_n(t), d_{n+1}(t)]$. Consumption of urban services from farmers in site $y \in [d_n(t), d_{n+1}(t)]$ will be $\frac{(1-\beta)p(y,t)}{q(y,t)}$. Accounting for the iceberg transportation costs, using the expression for the prices of the agricultural good and the urban services, total production of a non-endpoint town y_n (i.e. a town that was not the last one to emerge along the road, $y_n < y_N$) is:

$$s(y_n, t) = p(1-\beta)e^{A(y_n,t)}\Big[d_{n+1}(t) - d_n(t)\Big] \tag{4}$$

For the endpoint town y_N, total production of urban services is:

$$s(y_N, t) = p(1-\beta)e^{A(y_n,t)}\left[y^r(t) - d_N(t) + \frac{1}{\tau_a}\right] \tag{5}$$

These equations, jointly with Proposition 1, characterize the equilibrium.

2.3 Predictions

Given the characterization of the equilibrium, what are the key predictions of the model? In particular, what are the persistent effects of being an endpoint on town size?

I compare towns that were railroad endpoints for different time intervals. That is, I compare towns according to the time it took the railroad to reach the next town to emerge on it. Denote a town site by y_n. From Proposition 1, the next town to emerge will be at the closest site y_{n+1} that satisfies the following equation:

$$\tau_s\big(y_{n+1} - y_n\big) = A\big(y_n, t^r(y_{n+1})\big) \tag{6}$$

The equation above can be illustrated by Figure 1. When the railroad is built slowly, the last endpoint town will be larger and more productive at the time the railroad reaches a site y farther down the road. Urban firms will have less incentives to settle there. Hence, the next town to emerge will be farther away.

Prediction 1. *As a town is a railroad endpoint for longer, the distance to the next town farther down the railroad increases.*

Note that towns that are farther up the endpoint town have already emerged, and so their positions have already been determined. Hence, the model predictions with respect to distance to the closest town are asymmetric. The model predicts that time as endpoint increases distance to the next (farther down) town, but it makes no prediction with respect to distance to the previous (farther up) town. I exploit this asymmetry in the empirical part of the paper.

There is also a model prediction on town sizes. If a town is an endpoint for longer, it will be persistently larger. This occurs for two reasons. First, the next town on the railroad will be farther away, directly increasing town size according to equation (3). Second, the next town will have had emerged later, so it will be (at least weakly) less productive at any future time. According to equation (3), the productivity difference also increases hinterland size. As a consequence, the town will be larger.

Prediction 2. *As a town is a railroad endpoint for longer, its size permanently increases.*

2.4 Discussion

The model presented above is a stylized model. Its purpose is to unveil the basic economics of how time as a road endpoint can persistently impact the spatial distribution of economic activity along the road. The intuition to understand the model predictions is that the emergence of a town on a site depends on whether new firms on that site can compete with firms from the closest town. If the closest town is an endpoint for longer, its firms will be more competitive, preventing nearby city emergence. As a result, towns that spend more time as endpoints become persistently larger as they serve larger markets.

A virtue of the model solution is that it is unique and has an analytical characterization, which implies the predictions are general and do not depend on particular parameterizations of the model. However, tractability is achieved through some restrictive assumptions. The model may no longer be tractable when some of these simplifications are relaxed. In the rest of this section, I discuss the role played by different assumptions and argue that the basic economic considerations the model highlights are robust to these changes. I also discuss how these extensions may affect the mechanisms for why time as endpoint increases city sizes. Online Appendix B.1 contains formal details of the extensions.

Dynamic agglomeration economies. There are some simplifications on the functional form taken by agglomeration economies in the model. In particular, the model allows for agglomeration economies as new firms learn from previous firms that were on the same site, so all of them end up with the same productivity. These are intertemporal spillovers, as new firms benefit from the location decisions of previous firms.[7] However, one could think of more complicated assumptions for productivity growth. For instance, productivity growth could depend positively on the size of the urban service sector in the town. Indeed, there are some empirical findings showing that the concentration of economic activity may lead to faster city growth, a fact that has been called dynamic agglomeration economies; see Glaeser *et al.* (1992), Henderson *et al* (1995), Henderson (2003), Cingano and Schivardi (2004) and Hanlon and Miscio (2017).[8]

These dynamic agglomeration economies may be incorporated in the model by extending the functional form of productivity growth to $\dot{A}(y,t) = g(A(y,t), m(y,t))$, where $m(y,t)$ indicates the quantity of intermediate goods used by urban service firms on site y at time t; by definition, $s(y,t) = e^{A(y,t)}m(y,t)$. In this case, the function $g(.)$ will be non-increasing in $A(y,t)$ and non-decreasing in $m(y,t)$. Note that, since endpoint towns serve all the half-line that extends to their right, their urban sector will be larger, and they will benefit from an additional boost to productivity growth as a consequence of the larger urban sector.

Do dynamic agglomeration economies introduce a new effect of time as endpoint? Depending

[7]Other models have similar assumptions. For instance, Assumption 3 in Rauch (1993) implies that productivity in a site equals the highest it has ever been, so local knowledge does not depreciate.

[8]Glaeser *et al.* (1992) and Hanlon and Miscio (2017) find no evidence for dynamic within-industry agglomeration economies, although their findings support the existence of inter-industry growth externalities. Evidence for dynamic within-industry agglomeration economies are found by Henderson *et al* (1995), Henderson (2003), and Cingano and Schivardi (2004). Note that, although there is a single urban sector in the model, cities often concentrate many urban activities; hence, it is not clear whether one should expect the dynamic agglomeration economies in the setting of this paper to be driven by within-industry or inter-industry externalities.

on the relation between productivity growth and town size, time as endpoint could lead productivity in a town towards a persistently higher level. The model is no longer tractable with this alternative assumption, so in Online Appendix B.1.1 I present an example that shows how a longer time as endpoint may lead to a higher steady-state productivity if there are dynamic agglomeration economies. In the baseline model, towns that were endpoints for longer are larger because they serve a larger hinterland. When the model allows for dynamic agglomeration economies, then higher competitiveness in the long run is a second reason why towns that were endpoints for longer become larger. The intuition is that the productivity growth associated with time as endpoint may allow towns to overcome a “productivity trap" by growing faster while an endpoint.

Static agglomeration economies. A larger town size could directly increase the contemporary level of firm productivity. The inclusion of such static agglomeration economies complicates the model, as town sizes and productivities must now be determined jointly. In Online Appendix B.1.2, I present an example of how static agglomeration economies may be incoporated to the model. In the example, the same qualitative predictions as in the baseline model hold: a larger time as endpoint increases distance to the next town and town size. Unlike in the baseline model and as in the extension with dynamic agglomeration economies, the static agglomeration economies also imply that more time as endpoint persistently increases the town productivity in producing urban services. This occurs precisely as a consequence of the larger town size.

Farmer mobility. The model assumes that farmers are immobile and uniformly distributed across the space. However, transportation infrastructure is often associated with the opening of agricultural frontiers (Jedwab and Moradi 2016, Monbeig 1952), so it seems important to understand whether the main model predictions are robust to this assumption. Other city emergence models, such as the ones in Fujita *et al* (1999), assume that farmers are mobile. In Online Appendix B.1.3, I show that farmer mobility may be included into the model without changing its qualitative predictions.

Mobile workers. In the baseline model, intermediate goods are the only input in the production of urban services. In Online Appendix B.1.4, I show how the model may be extended so labor is also used by urban firms. I assume workers are mobile and choose where they work, so wages adjust to equalize worker utility across towns. The inclusion of labor implies a weaker relationship between productivity and prices, but the qualitative predictions of the model remain the same.

Other considerations. Other than the considerations brought by the extensions above, three other characteristics of the model are worth discussing.

First, there are simplifying assumptions on transportation costs. These assumptions bring tractability to the model. For instance, since the costs to transport the intermediate goods are prohibitive ($\tau_m = \infty$), no towns exist beyond the railroad endpoint. If this assumption is relaxed, some towns would exist beyond the railroad endpoint even at $t = 0$. The pre-existing towns would continue existing and would be larger than new towns. However, if $\tau_m > \tau_s$, then the railroad expansion could lead to the emergence of new towns as it benefits the entry of firms on sites newly reached by the railroad.

Second, firms in two economic sectors could directly benefit from locating on an endpoint town. The first sector is the railroad itself. Railroad endpoints often have railroad yards and concentrate maintenance personnel. They can also serve as a base for the railroad construction team. The second sector consists of inter-modal transportation activities, such as the portage activities that influenced city location in the United States (Bleakley and Lin 2012). If there are linkages from either of these sectors into urban services, productivity growth in the endpoint town could be faster. A town that spends more time as endpoint is also a town that benefits for longer from these linkages, and as a result it may become persistently larger.

Third, there is a single urban good in the model. Hence, the model does not account for any potential effect on the city hierarchy.[9] Still, time as endpoint may affect not only city size, but the number of urban services provided by the town. In this case, the relevant hinterland should be defined not only by the distance to the previous and next towns along the railroad line, but -for some goods -by the distance to the previous and next town of the same or larger hierarchy.

[9]City hierarchy is associated with Central Place Theory; see Hsu (2012) for a recent model with many urban goods and an urban hierarchy.

3 Empirical Approach

3.1 Background

I empirically test the model predictions by investigating the relationship between time as endpoint and city size for a sample of towns along historical railroad lines built to connect the city of São Paulo with frontier regions to the west. Two reasons motivate the choice of this sample for the analysis. First, the setting shares similarities with the economy I model in Section 2. Second, qualitative research in geography already highlights the importance of endpoint towns in the region. This sub-section presents background information on the sample region.

The city of São Paulo was first connected by rail in 1867, when the opening of the São Paulo Railway connected it to the nearby seaport in Santos. With the subsequent construction of new railroads, the railroad network in the state took a radial pattern, with São Paulo as the main transportation hub. Many of the new lines were built toward low population density regions in the west of the state of São Paulo or in neighboring states.

Before the railroad era, overland transportation costs in the region were high due to the hilly and forested terrain and the absence of navigable rivers (Leff 1982, Summerhill 2003). The main alternative transportation mode was the mule. Railroad building integrated the west of the state with markets near the coast, effectively incorporating it into a fast-growing economy. Coffee production -the main export industry in the region -expanded along the railroads. At the same time, a myriad of new towns emerged along the rail lines. This expansion coincided with a period of unprecedented economic growth, industrialization, and mass migration from Europe, during which São Paulo became the dynamic center of the Brazilian economy (Monbeig 1952, Luna and Klein 2018). Railroads arguably explain some of the regional development in the period: in a recent paper, Américo (2021) shows that proximity to railroad lines increased employment and the number of establishments in manufacturing.

Note there are many similarities with the model economy. Railroad lines were built from a gateway towards the interior of the country. The main economic activity of the region was the production of cash crops for export, in particular coffee. Railroad construction reduced transportation costs. Finally, new towns emerged along the railroad. In these towns, merchants and other businesses imported goods from the city of São Paulo or from the rest of the world to sell them to local farmers.

The geography of the state of São Paulo was deeply influenced by the railroads that were built in the century following 1867, to the extent that sub-regions in the state were named for the railroads that connected them (Menucci Giesbrecht 2001, Luna and Klein 2018). Contemporary geographers highlighted the importance of railroad endpoints for the emergence and growth of towns in this setting. The seminal references in the literature are Deffontaines (1938) and Monbeig (1949, 1952). Railroad endpoint towns were described as *bocas do sertão* - "gateways to the frontier" -, the last outposts of the modern and industrializing economy that expanded from the east of the state. In these towns, a variety of merchants and firms provided goods and services to farmers who were expanding the agricultural frontier. Monbeig (1949, 1952) argued that, even after the railroad line was extended beyond the *boca do sertão* town and it no longer was an endpoint, the town kept its economic importance. This occurred due to the strength of the businesses established at the town; as Monbeig (1949, p. 64) noted: "The former "boca do sertão" kept, thanks to its well-known traders and bankers, their hospitals, and their schools, its influence over the former frontier."[10]

3.2 Empirical Framework

The model and the geography literature suggest that there is a positive association between the number of years a city was a railroad endpoint and its size. I document this association by estimating the following equation:

$$y_i = \beta EndpointYears_i + X_i'\theta + u_i \tag{7}$$

where i indicates a municipality, y_i is a dependent variable such as the logarithm of total urban population in 2010, $EndpointYears_i$ is the number of years that the municipality was a railroad

[10]Unlike Monbeig, Deffontaines (1938) expected many of the economic benefits that towns experienced as endpoints to disappear after the railroad expanded, and so he was skeptical that past endpoint status would persistently affect town sizes. The empirical results presented below favor the views of Monbeig, as opposed to Deffontaines'.

endpoint in the past, and X_i is a vector of control variables (see below).

For most of my analysis, I estimate equation 7 by ordinary-least squares. In Section 4, I present an instrumental variable estimation as support for a causal interpretation of β. I perform inference using heteroskedasticity-robust estimates of the standard errors. However, the results are robust to the use of an inference procedure that allows for cross-sectional dependence between municipalities that are geographically close; see Appendix Table B.6.

In the baseline specification, control variables are the year of railroad arrival, the logarithm distance to São Paulo, and a dummy equal to one if the municipality was already incorporated by the time of railroad arrival. All control variables are likely predictors of city size. The year of railroad arrival is associated with how long a municipality has been affected by railroad access. The distance to São Paulo is a measure of access to the largest metropolis in the region. Finally, if a municipality was already incorporated by the time of railroad arrival, then it was likely an already relatively sizeable urban settlement.[11] Only 30% of the sample municipalities were incorporated before the railroad arrival. Most of the sample thus consists of towns that did not exist before the railroad and that developed around a station that served a rural area, or of towns that were small by the time of railroad arrival.[12] In alternative specifications, I consider larger sets of control variables.

Data. Information on the expansion of the São Paulo railroad network comes from two sources. The railroad lines and the location of each station are from shapefiles released by the *Agência Nacional de Transportes Terrestres* (ANTT), the Brazilian federal agency responsible for regulating railway and road transportation. The list of railroad stations includes both active stations and stations no longer in use. I match each station in the ANTT dataset to station opening dates from *Estações Ferroviárias do Brasil* (EFB), a comprehensive website with information on railroad stations.[13] Using this information, I construct for each station the year of arrival of the railroad line. Finally,

[11]The incorporation year refers to the year the town became the seat of a *vila* (in the colonial or imperial, pre-1889 period) or of a *município* (in the republican, post-1889 period). The information was collected from IBGE; see Appendix B.3.

[12]Santa Gertrudes is an example of a settlement founded after the railroad arrival; originally, it was only a railroad stop to serve a farm in the area. Presidente Bernardes, on the other hand, is an example of an unincorporated small village that grew after the railroad arrival.

[13]See also Menucci Giesbrecht (2001), a book by the website curator.

I use the station level information to construct the year of railroad arrival for each municipality. In this way, for each municipality I define the total number of years the municipality was an endpoint as the difference between the railroad arrival year for the next municipality on the railroad line and the municipality own arrival year.[14]

The sample consists of municipalities along the railroad lines that were built toward the west of the state of São Paulo; some lines extend into the neighboring states of Paraná and Minas Gerais.[15] The sample broadly coincides with the geograhic region studied by Monbeig (1952). I exclude municipalities that are endpoints today. All railroad lines either started in the city of São Paulo or in some other railroad line that start in it. In this way, for each municipality I can calculate its distance, along the railroad, to the city of São Paulo. The city of São Paulo does not belong to the sample. There are 203 sample municipalities, with a combined urban population in 2010 of almost 13 million inhabitants. The railroad arrival dates span the period from 1867 to 1967.

As measures of city size, I use urban population from the 2010 Brazilian Population Census and estimates of urban GDP for the same year from the *Instituto Brasileiro de Geografia e Estatística* (IBGE). Brazilian municipalities are administratively split into urban and rural regions (IBGE 2017); urban population is the total population living in the urban regions. The main results are robust to the use of urban population from previous census years. I define urban GDP as the combined GDP in the manufacturing, services, and government sectors.

Auxiliary data sources are described in Online Appendix B.3.

Summary statistics. The map in Figure 2 shows the sample municipalities, the railroad lines, and the number of years each municipality was a railroad endpoint. Of the 203 sample municipalities, 49 were railroad endpoints for at least 2 years. The summary statistics are shown in Table 1.[16]

[14]In some cases, there is a bifurcation of the railroad lines and there are two or three next municipalities; I define years as endpoint as the difference to the first next municipality to be reached by a railroad.

[15]Therefore, I exclude two railroads that were built to connect São Paulo to other state capitals: the one connecting São Paulo to Rio de Janeiro, the federal capital, and the one connecting it to the southern capitals of Curitiba and Porto Alegre. I also exclude the lines connecting São Paulo to the nearby port of Santos.

[16]Zipf's law does not hold for this sample. Instead, there is a concave relationship between log rank and log city population; see Appendix Figure B.1. Previous studies have documented Zipf's law for the universe of Brazilian cities; for instance, see Chauvin *et al* (2017).

3.3 Historical Example

As noted by Monbeig (1952), unexpected events often caused some towns to be railroad endpoints for longer. In what follows, I present a historical case study of a sample town that remained for many years a railroad endpoint due to a plausibly exogenous reason. The case of this town illustrates the model predictions and supports a causal interpretation for the empirical findings presented in Section 4.

No other sample municipality was a railroad endpoint for as long as São José do Rio Preto. The town is located on a railroad line built by the *Estrada de Ferro Araraquara* (EFA), a private Brazilian company that started its activities in 1896. Its first railroad segment was completed by 1898. As already noted by Monbeig (1952), the setbacks this company would later face provide an ideal experiment to investigate the effects of time as railroad endpoint.

The EFA railroad was a promising enterprise (Silva and Tosi 2014). It would serve the expanding agricultural frontier in the northwest of the São Paulo state. In 1900, the federal government awarded to the company a concession that allowed it to build until the Mato Grosso capital of Cuiabá. The concession granted tax benefits to the company. As a consequence, the company issued new debt to finance railroad building.

However, the company soon faced a variety of setbacks. In 1906, the state government imposed a tax on new coffee plantations, which negatively affected the movement of the railroad and reduced the returns on the investment (Silva and Tosi 2014). Moreover, the concession to Cuiabá was unexpectedly reviewed and canceled. With its financial conditions aggravated by the mismanagement of an administration that took over the company in 1909, the railroad line reached São José do Rio Preto in 1912.

EFA was then acquired by the São Paulo Northern Railroad Company, a company controlled by L. Behrens and Söhne, a German bank that held many EFA bonds. This new company, however, had no investment capacity to expand or even maintain the railroad, particularly after the start of World War I affected the German bank. After overdue salaries led to a strike by the railroad workers in 1919, the state government canceled the concession and took over the railroad. A long court litigation followed the government takeover, postponing new investment on the railroad. As a consequence, São José do Rio Preto was a railroad endpoint for 21 years. The railroad line was extended towards the west only in 1933.

The period as endpoint is considered a period of fast growth for the city. An excerpt from the municipality entry in the Brazilian Encyclopedia of Municipalities (IBGE 1958, v. 30, p.189) reads: "The arrival of the EFA railroad in 1912 was the beginning of a 'golden age' in the development of São José do Rio Preto, which became a center of commerce for the region due to its position on the railroad endpoint." As illustrative of the fast growth of the city, in 1925 the municipality received 46% of the 2,345 newly arrived international immigrants that moved to municipalities served by the railroad.[17]

Figure 3 maps the gradual opening of the *Estrada de Ferro Araraquara*. The figure also shows each municipality urban population in 2010. In relation to the other municipalities, São José do Rio Preto was a railroad endpoint for much longer: it was a railroad endpoint for 21 years, while the mean time as endpoint among other EFA municipalities is 1.62 year (the standard deviation is 2.35 years).

Today, São José do Rio Preto is the largest town along the railroad. Its urban population in 2010, of 383 thousand inhabitants, was sixteen times larger than the mean city size among other EFA municipalities (24 thousands inhabitants), and much larger than the second-largest urban population (112 thousand inhabitants). The distance to the next municipality (24 km), Mirassol, is also larger than the counterpart for any of the other 28 EFA municipalities, and twice the mean distance to the next municipality (12 km). That is, just as the number of years as endpoint, the urban population and the distance to the next municipality of São José do Rio Preto are outliers relative to the distribution of these variables among other EFA municipalities. The urban history of São José do Rio Preto is then illustrative of the model predictions, a relevant historical example given its long time as endpoint is plausibly unrelated to its growth potential.

[17] See *Secretaria da Agricultura, Comércio e Obras Públicas* (1925), pages 103-107. In that year, the state of São Paulo received 54,678 immigrants.

4 Empirical Findings

4.1 Main Findings

Main results. The ordinary least squares estimates of equation 7 are shown in Table 2. In Panel A, the dependent variable is the logarithm of urban population in 2010. In Panel B, it is the logarithm of urban GDP in 2010. And in Panel C, it is the logarithm of distance to the next municipality along the railroad. In column (1), I report the coefficients on years as endpoint when there are no control variables. Scatterplots of these relationships are shown in Figure 4. In column (2), I report the coefficients after the inclusion of the baseline controls. The results indicate that one additional year as railroad endpoint is associated with a larger urban population in 2010 by 0.112 log point, a larger urban GDP by 0.124 log point, and a larger distance to the next municipality by 0.037 log point. All these coefficients are statistically different than zero at a 1% significance level. Hence, the correlations in the data are in line with the two model predictions.

The same associations are documented when the dependent variables in equation 7 are in levels; see column (3) of Table 2. One additional year as endpoint is associated with 9,351 additional urban residents, an urban GDP that is R$ 237,220 larger, and a next municipality that is 0.675 km farther away along the railroad line.

Omitted geographical fundamentals. A challenge in interpreting the estimated coefficients as the effect of time as endpoint on city sizes is that time as endpoint may correlate with geographic factors that determine city sizes. To address this concern, I examine whether the results are robust to the inclusion of a large set of geographic fundamentals as controls. To select the fundamentals, I have drawn on the economic history of the region. The fundamentals include variables such as altitude and the terrain ruggedness index (TRI), the logarithm of the potential yields for coffee (the main export crop at the time of the railroad expansion), maize (the main staple crop) and sugarcane (a historical crop that remains relevant today), the logarithm of average days a year that support malaria transmission, the share of the municipality covered by different soil types, and a dummy equals one if a main river crosses the municipality. Before the malaria eradication programs that began in 1950, malaria was considered a barrier to the development of cities (Kohlhepp *et al*, 2014).

The soil variables include the shares of the municipality covered by latosols and acrisols, the two most common soil types in the sample region, as well as the share covered by *terra roxa*, a productive type of latosol that was sought after by farmers during the expansion of the agricultural frontier. In Online Appendix Table B.1, I present the correlations between time as endpoint and these fundamentals. Although some of the fundamentals correlate with some of the baseline controls, particularly with distance to São Paulo, there is no association with time as endpoint. These results suggest that omitted fundamentals are unlikely to be a threat.

In fact, the main results are robust to the inclusion of this large set of geographic fundamentals as controls; see column (4) of Table 2.[18] Column (5) also includes a quadratic polynomial of latitude and longitude, which partially accounts for omitted fundamentals that vary smoothly over space. In comparison with column (2), the new coefficients in column (4) and (5) barely change. Therefore, even when comparing geographically similar municipalities, time as endpoint is positively associated with city sizes.

Under the assumption that selection on observable geographic fundamentals is informative about selection on unobservable fundamentals, I can use the changes in coefficients and in the explanatory power of the model after the inclusion of the controls to examine such omitted variable bias (Altonji *et al* 2005, Oster 2019). The coefficients with all the controls in column (5) are slightly smaller than the baseline coefficients in column (2), but the R-squareds are substantially larger. Following Oster (2019), I calculate the lower bound for the coefficient under the assumption that selection on unobservable geographic fundamentals is no larger than selection on observables fundamentals. The bounds are displayed in columns (4) of Table 2, at the bottom of each panel. For the calculation of the bounds, I follow Oster's (2019) recommendation and assume that the inclusion of the unobservables would lead to a model with an R-squared 30% larger than the R-squared of the model that includes all observables. The bounds are positive and only slightly lower than the estimated coefficients, further suggesting that the main results are not due to omitted geographic fundamentals.

Distance to the previous municipality. As additional support to the model predictions, I exam-

[18] I include both the level and the square of altitude and terrain ruggedness.

ine the association between time as endpoint and distance to the *previous* town along the railroad. According to the model, more time as endpoint leads to a larger distance to the next town along the railroad but not to a larger distance to the previous municipality. This occurs because, at the time the railroad reaches an endpoint site, the locations of previous towns have already been determined and are independent of later expansion rates of railroad construction.[19] This empirical test is useful since it is not clear why other explanations for the correlations documented in Table 2 would imply a null association between time as endpoint with distance to the previous town but a positive association with distance to the next town.

In Panel D of Table 2, I show the estimated coefficients of equation 7 when the distance to the previous municipality is the dependent variable. The correlation is graphed in Panel (d) of Figure 4. The coefficients (-0.007 in the specification with the baseline controls) are not statistically significant, often negative, and of a magnitude that is more than five times smaller than the coefficients on distance to the next municipality in Panel C. Moreover, the results are robust to the different specifications discussed above. Indeed, there seems to be no association between time as endpoint and distance to the previous municipality.

Predictive power. It is important to point out that time as endpoint is a strong predictor of modern city sizes. Column (1) of Table 2 shows that the R-squared of a simple regression of log city sizes on years as endpoint is 0.100. Appendix Table B.2 expands on this result by reporting the Shapley values of different groups of predictors. I consider four groups: **(i)** years as endpoint; **(ii)** the baseline controls (year of railroad arrival, whether a municipality was incorporated before the railroad arrival, and the log distance to São Paulo); **(iii)** the geographic fundamentals included in the specification of column (4) of Table 2; and **(iv)** the quadratic polynomial of latitude and altitude that is also included in column (5) of Table 2. The Shapley values are the average marginal contributions to the R-squared across all specifications that include different subsets of these groups.[20]

[19]When dynamic agglomeration economies are accounted for, the town that precedes the endpoint may be abandoned; see case 2 of Proposition B.1. In this case, time as endpoint could affect distance to the previous town. Note however that the prediction is still asymmetric: time as endpoint always affects distance to the next town but only in some cases it affects the distance to the previous town. To the extent of my knowledge, there was no abandonment of towns in the historical setting I examine.

[20]See Henderson *et al* (2018) for a paper that uses the Shapley value to measure the relative importance of different sets of geographic fundamentals.

Column (1) of Appendix Table B.2 displays the Shapley values when the dependent variable is the logarithm of urban population. Although the predictive power of time as endpoint is not as large as the (joint) predictive power of the baseline controls, geographic fundamentals, or coordinates polynomial, its marginal effect on the R-squared is between 50 and 80% of the marginal effect of each group. These estimates suggest that time as railroad endpoint is an important second-nature geographic fundamental in determining the distribution of population within municipalities along the railroad network. Similar results hold when the dependent variable is the logarithm of urban GDP or of distance to the next municipality; see columns (2) and (3). In contrast, column (4) shows that time as endpoint is a very weak predictor of distance to the previous municipality.

Persistence. The results above indicate a strong association between time as endpoint in the past and city sizes in 2010. The model also predicts such association to be persistent. As evidence for persistence, I show that the association between time as endpoint and the logarithm of urban population in 2010 is similar to associations with urban population in previous census years. Appendix Table B.3 reports the estimates of the coefficients of equation 7 when the dependent variable is the logarithm of urban population in 1970, 1980, 1991, and 2000.[21] The coefficients are almost identical across all years, although for past census years the estimated coefficients are more precise.

Further robustness. The associations documented above are robust to a variety of concerns.

For the baseline results, the measure of time as endpoint is linear in the years a municipality was a railroad endpoint. In Appendix Table B.4, I consider different measures. First, I use a dummy variable equal to one if years as endpoint exceeds a threshold. I report the estimated coefficients for thresholds of 1, 2, 5 and 10 years. The results are not only robust to these alternative measures, but the estimated coefficients are usually increasing in the threshold. Second, I use the inverse hyperbolic sine of years as endpoint as an alternative measure. The inverse hyperbolic sine is a concave function, which reduces the influence of observations with more years as endpoint. Results are again robust to the alternative measure. Third, I include both linear and quadratic terms of time as endpoint. The coefficient on the square of years as endpoint is always close to zero and

[21] For the censuses of 1970, 1980, and 1991, I aggregate the variables to minimum comparable areas due to the incoporation of new municipalities; see Reis *et al.* (2008). The most recent year of railroad arrival in the sample is 1967.

statistically insignificant, suggesting the associations are linear.[22]

Appendix Table B.5 presents the coefficients of equation 7 under different specifications. In column (1), I control for the Brazilian GDP growth in the year of railroad arrival.[23] In column (2), I flexibly control for the year of railroad arrival by including a third-degree polynomial of the year. These estimates suggest that the main results are not driven by economic conditions at the time of the railroad arrival. In column (3), I include railroad company fixed effects to equation 7. The estimates suggest that the main results are not due to differences in historical city growth among different railroad lines.

As the railroads promoted the settlement of the region and most of the towns emerged only after railroad arrival, there is little information on pre-railroad economic conditions that could affect later town sizes. Column (4) of Appendix Table B.5 addresses this issue by using information from the 1872 Census, the first national census in Brazil. As control variables, I include the following characteristics of the original 1872 municipality from which each sample municipality originated: the logarithm of population density, the shares of slaves and foreigners in the population, the literacy rate, the share of free school-aged children in school, and the share of employment in non-agriculture. To avoid including endogenous variables as controls, the estimates use only the sub-sample of towns reached by the railroad after 1872. The empirical findings are again robust to the inclusion of the additional control variables.

Finally, the main results are robust to alternative inference methods. In Appendix Table B.6, I allow for cross-sectional dependence of the errors due to geographic proximity as in Conley (1999). I consider two different distance cutoffs (50 and 75 km) for this dependence to exist. The estimated standard errors for the coefficients of interest are often smaller than the ones in the baseline estimation, so the use of robust standard errors that ignore cross-sectional dependence is a conservative inference procedure.

[22]Naturally, the effects of time as endpoint are not unbounded and so decreasing returns should eventually apply, but such decreasing returns do not appear within the range of time as endpoint in the sample.

[23]Estimates of past GDP growth are from Araújo *et al* (2008).

4.2 Instrumental Variable Findings

The previous subsection shows positive associations between years as endpoint and urban population, urban GDP, or distance to the next municipality. These findings are consistent with the model predictions. There is no association between years as endpoint and distance to the previous municipality, a pattern that is also consistent with the model. The associations are robust to different specifications and seem unlikely to be explained by unobserved geographic fundamentals. However, since time as endpoint is possibly endogenous to future city size, such associations might not reflect the causal effect of years as endpoint. I now present instrumental-variables estimates in support of a causal interpretation for the empirical findings.

Identification strategy. The instrument I use exploits the incentives of the railroad company to build faster so it reaches an urban center farther down the road. I construct the instrumental variable as follows. For each sample municipality, I identify the closest town that satisfies two conditions: **(i)** it is farther down the same railroad line as the first municipality;[24] **(ii)** at the year the railroad reached the first municipality, the latter town was already an incorporated municipality.[25] I then compute the distance between the two municipalities along the railroad line, a measure I refer to as the distance to the next incorporated municipality. For 76 sample municipalities, there was no already-incorporated municipality farther down the railroad line, so I assign the maximum distance in the sample (154.25 km) for these municipalities.[26] The excluded instrument is the logarithm of this distance.

The main identification threat for a causal interpretation of the ordinary least squares estimates is that a site with a greater potential of becoming a large town in the long run might be a site the railroad company had an interest in connecting, but less interest in expanding the railroad after it. As a result, that site would remain a railroad endpoint for longer. The use of distance to the next incorporated municipality as an instrument overcomes this threat by leveraging variation on

[24]Note this condition also restricts the closest town to be on a railroad built by the same company as the first municipality. Such restriction makes sense as the instrument should affect how fast that company built the railroad.

[25]This closest municipality is restricted to be always different than the own municipality.

[26]The results are robust to assigning larger values to these municipalities; these estimates are reported on Online Appendix Table B.8.

the incentives of the railroad companies in building farther down the site. Since towns became incorporated municipalities only after achieving some economic and populational importance, they were also sites railroad companies were interested to connect.[27] Proximity to one such already incorporated municipality would then lead to faster railroad construction and, as a consequence, less time as endpoint. This incentive associated with distance to the next incorporated municipality would plausibly be at play regardless of the potential of a site in becoming a large town.

Note the instrumental variable approach I use is similar to a common identification strategy for the effects of transportation that exploits a site position with respect to cities that will eventually be connected by the road; see, e.g., Faber (2014), Banerjee *et al* (2020), and Bird and Straub (2020).[28] The assumption behind this identification strategy is that, although such position influences whether a site gets access to the road, the position is *incidental* to characteristics of a site that would cause city growth in the future. The identification strategy I use is different because, instead of inducing variation on access to transportation infrastructure, the instrument induces variation on time as a railroad endpoint (conditional on receiving the railroad). The instrumental variable satisfies the exclusion restriction if distance to the next incorporated municipality affects city sizes only through its impact on time as endpoint. Later on this subsection, I present two robustness exercises in support of this identification assumption.

IV results. An instrumental variable must affect time as endpoint. It will be the case if proximity to an already incorporated municipality incentivized the railroad company to build faster. In fact, there is a positive relationship between log distance to the next incorporated municipality and years as endpoint. Column (1) of Panel A of Table 3 displays the first-stage estimates. An increase of one logarithm point in such distance increases time as endpoint by 0.773 year. This is a moderately strong first stage, with a robust F-statistic on the excluded instrument of 19.28.

The two-stage least squares estimates are shown in columns (2) to (4) of Panel A of Table 3.[29]

[27] Américo (2021) uses least-cost paths between initially incorporated municipalities and seaports or state capitals to instrument for access to railroads in Brazil. He finds a strong positive relationship between proximity to these least-cost paths and to railways, showing that the position of already incorporated municipalities affected railroad placement.

[28] Redding and Turner (2015) refer to this identification strategy as the inconsequential unit approach.

[29] The reduced form estimates are shown on Appendix Table B.7.

The estimates confirm the findings from the ordinary least squares by suggesting positive effects of time as endpoint on urban population, urban GDP, and distance to the next municipality, but not on distance to the previous municipality. The two-stage least squares estimates for the effects on city sizes or distance to the next municipality are statistically significant at a 5% significance level. Inference is robust to the use of an Anderson-Rubin test; see the bottom of the panel. Note that the two-stage least squares estimates are larger than the ordinary-least squares coefficients: the estimates indicate that one additional year as endpoint increases urban population by 0.228 log point.[30]

Robustness. I now present two robustness tests to assess the validity of the instrumental variable approach. A first identification threat is that distance to the next incorporated municipality depends on the geographic position of the municipality and, as a consequence, could be associated with geographic fundamentals that affect city sizes. I address this concern by including all the observable geographic fundamentals used in column (5) of Table 2 (including the coordinates polynomial) as included instruments. The main results are robust to the inclusion of the geographic fundamentals; see Panel B of Table 3 for the estimates.

A second threat to identification is that a larger distance to the next incorporated municipality may indicate that a site was relatively more *remote*. Redding and Sturm (2008) have shown that remoteness affects city growth, so initial remoteness could affect long-run city size.[31] To address this concern, I exploit the fact that the instrument is constructed based only on distance to the closest incorporated municipality farther *down* the railroad line, not farther *up*. I exploit this fact in two ways.

First, I show in Appendix Table B.9 that the logarithm of distance the closest already incorporated municipality farther up the railroad line does not seem correlated with years as endpoint or

[30]A plausible reason for a downward bias to the OLS is that a fast-growing endpoint town would attract revenues to the railroad company. Railroads were costly investments, so railroad companies often used their own revenues to finance construction. As a consequence, a faster growing town would be an endpoint for a shorter period. Since it would likely be larger in the long run, this relation would introduce the downward bias to the OLS.

[31]The findings of Redding and Sturm (2008) indicate negative effects of remoteness, which would introduce a negative bias to the two-stage least squares estimates for the effects on city size.

with city sizes. Such exercise may be interpreted as the use of a “placebo instrumental variable", which should not affect time as endpoint but that may still be correlated with remoteness. The results support the validity of the instrumental variable strategy.

Second, I directly include a proxy of the remoteness of a municipality at the time of railroad arrival as an included instrument. I define such proxy as the logarithm of distance to the closest incorporated municipality on the railroad line, *either up or down the line*. In this way, the estimates exploit only variation in proximity to next incorporated municipality that is not associated with variation to proximity with incorporated municipalities anywhere on the railroad. The estimates are shown on Panel C of Table 3. Despite the more restrictive conditions on the variation used for identification, the instrument remains a strong predictor of time as endpoint, with a robust F-statistic of 15.06. The proxy of remoteness is actually negatively associated with time as endpoint (conditional on the instrumental variable). It is not associated with city sizes. The two-stage least squares are robust to this alternative specification, suggesting that the instrument is not picking up the effects of past remoteness.

Summing up, the instrumental variables estimates are in line with the model predictions and the associations documented in the previous subsection. The estimates are again robust to the inclusion of a large set of observable fundamentals and do not seem a consequence of past remoteness. In this way, the instrumental variables approach suggests the plausibility of a causal interpretation for the positive association between time as endpoint and city sizes.

4.3 Decomposition of the Effects

In this sub-section, I use the theoretical model as the basis for a brief discussion about which channels explain the positive effect of time as endpoint on city size. The model predicts that the logarithm of town sizes (measured by the total production of urban services) is a linear function of the logarithm of urban sector productivity and of the logarithm of the town's hinterland; see equation 4. By taking the logarithm of equation 4 and then the derivative with respect to time as endpoint T_{end}, the effect on city size may be decomposed into:

$$\frac{d\,log\big(\text{ City Size }\big)}{d\,T_{\text{end}}} = \frac{d\,log\big(\text{ Productivity }\big)}{d\,T_{\text{end}}} + \frac{d\,log\big(\text{ Hinterland Size }\big)}{d\,T_{\text{end}}} \tag{8}$$

The equation above highlights that a larger city size is a consequence of a larger hinterland or of a higher productivity. Note that, in the baseline model, long-run productivity is not affected by time as endpoint, so the effects on hinterland size should be equal to the effects on city size. However, when there are dynamic agglomeration economies, time as endpoint may persistently increase urban services productivity; see the example in Online Appendix B.1.1. Static agglomeration economies also predict a larger productivity, due to the productivity advantages of larger cities; see the example in Online Appendix B.1.2

The left-hand side of equation 8 is the effect of years as endpoint on city sizes estimated in the previous sub-sections. With urban GDP as a measure of city size, the left-hand side is between 0.124 (according to the baseline OLS estimates) and 0.213 (according to the instrumental variable estimates).

The right-hand side, on the other hand, consists of the sum of two effects I do not observe. Consider productivity first: it refers to the inverse marginal cost for urban services, for which there is no data.[32] Hinterland size is also not directly observable: it consists of all the rural sites that consume from the town. However, there are some proxies for it. In what follows, I use the effects on these proxies of hinterland sizes to assess each of the channels on the left-hand side of equation 8.

A first proxy variable for hinterland size is the distance to the next municipality. Hinterland size is likely larger than this distance, depending also on distance to the previous town and productivity differences between the neighboring towns; see equation 3. As time as endpoint does not affect distance to the previous town and assuming there are no effects in productivity, as in the baseline

[32]It is tempting to use nominal wages as a measure of labor productivity, just as in many papers that estimate agglomeration economies; see Combes and Gobillon (2015). However, since urban services are non-tradable goods, wages would also depend on the prices and would thus be poor proxies for productivity. Online Appendix B.1.4 extends the model to include mobile labor. In that case, wages are negatively associated with productivity. This occurs because of workers' spatial arbitrage condition. Note the extension does not contain any disutility costs associated with density, such as scarce land for housing, which might reverse this association.

model, then the fact that distance to the next municipality is lower than hinterland size implies that the effects on the logarithm of the former are an upper bound for the effects on the logarithm of the latter. That is, if there were no effects on productivity, then the effects on log hinterland size should be no greater than the effects on log distance to the next municipality. The ordinary least squares coefficients in Table 2 suggest that the effects on hinterland size are smaller than the effects on city sizes. This is not the case for the two-stage least squares coefficients, which are relatively similar for both variables; see Table 3.

A second proxy for the logarithm of hinterland size is the logarithm of municipality area. Most municipalities in Brazil include both an urban district and a surrounding rural region, so the municipality area is a crude but reasonable indicator of hinterland size. The ordinary-least squares coefficient on the logarithm of municipality area is 0.022, which is lower than the coefficient for log urban GDP. However, the two-stage least squares coefficient is 0.263, which is close to the effect of 0.213 on log urban GDP. See Online Appendix Table B.10 for the estimates.

In sum, the instrumental variable findings suggest that the effects on city sizes can be attributed mostly to effects on hinterland size, while the ordinary least squares suggest effects on productivity may play a role. This discrepancy and the lack of direct data on hinterland sizes and productivity prevent the exercise above from providing unequivocal evidence on the relative importance of the two channels. Moreover, the decomposition is based on a stylized model. The validity of the decomposition thus depends on simplified model assumptions.[33] Still, the discussion above is useful as it provides a more nuanced view of how past time as endpoint affects city size.

4.4 Frontier Hypothesis

The empirical results are consistent with the model predictions, so I interpret the effects of time as endpoint according to the mechanisms described by the model. However, endpoint towns were considered as gateways to the agricultural and settlement frontier, so time as endpoint could correlate with the duration of historical exposure (proximity) to the frontier. Bazzi *et al* (2020) show

[33] For instance, the log linear decomposition in equation 8 does not hold when farmers are mobile; see equation B.9 in Online Appendix B.1.3.

that, in the United States, historical exposure to the frontier is associated with cultural traits such as individualism and distrust of the government. Their findings support an established hypothesis from American historians that frontier experience shaped local culture. A similar hypothesis also exists for Brazil: Monbeig (1952) extensively discuss how the frontier attracted entrepreneurial individuals whose mindset was shaped by a worldview highlighting individual effort. If these cultural traits favor urban activity, then they could explain some of the effects of time as endpoint on city sizes.

I argue that this alternative mechanism is unlikely to be relevant. First, it does not explain the effects on distance to the next municipality. Second, it is not clear why individualism and distrust for the government would favor the emergence of large towns. Third, although I do not have the rich data on values that Bazzi *et al* (2020) use to support their hypothesis, I follow their analysis and use the results of an election in which these frontier values were noticeable as a proxy for culture. In 2018, presidential candidate Jair Bolsonaro ran on a platform of small government and support for small businesses and entrepreneurs. In fact, Bolsonaro received strong support from historical frontier regions, such as from the west of the country. Bolsonaro's support was also high in the sample municipalities: the average vote shares for Bolsonaro in the first and second rounds of the 2018 presidential elections were, respectively, 59% and 73%, well above the national shares of 46% and 55% of voters. However, neither the ordinary-least squares estimates nor the instrumental variable findings suggest that time as railroad endpoint affected the Bolsonaro vote share; see Appendix Table B.11. These null results suggest that time as endpoint is either uncorrelated with past frontier exposure or that frontier exposure did not influence cultural traits within the sample.

5 Conclusion

This paper makes the point that the timing of gradual construction of transportation infrastructure persistently affects the spatial distribution of economic activity. I present a stylized model that makes explicit how this effect occurs. The model has two key predictions: long-run town size and distance to the next town along the railroad are increasing in the time a town spends as a railroad endpoint. I test these predictions in a sample of railroad towns in Brazil, building on classic work in geography that highlighted the importance of railroad endpoints for urban growth in this setting. I quantify the positive associations between time as endpoint and town size and distance to the next town. I also test an additional model prediction, present instrumental variable estimates, and discuss a historical case study as support for the interpretation that the positive associations reflect the causal effects predicted by the model.

The question of how the gradual construction of transportation infrastructure impacts the spatial distribution of economic activity is relevant to policy makers, particularly in developing economies where funds for investment in infrastructure are scarce. The timing of road construction is often a policy choice. The theoretical argument and empirical findings in this paper suggest it is a consequential one. There is a need for richer models to understand the optimal timing of road building when funds are scarce, quantify the welfare importance of the timing of road opening, and identify the role that can be played by complementary policies such as value capture taxes or road use pricing.

Further empirical research is needed to identify in which cases the gradual opening of transportation infrastructure affects economic activity. This paper focuses on frontier city emergence along railroads, but the economic forces behind the persistent effects of time as endpoint are more general. The gradual opening of transportation infrastructure matters because a transportation structure increases the economic density around it, which in turn casts agglomeration shadows on sites that still have not been connected by the transportation structure. Time as endpoint may affect the strength of this agglomeration shadow if firms already served by the transportation infrastructure become more productive over time. The explanation put forward in this paper thus relies on the interaction between positive effects of access to transportation infrastructure, agglomeration shadows, and local productivity growth. Such economic considerations plausibly apply to other settings and transportation structures. The broader point from this paper is that changes in access to transportation may not be sufficient to understand the effects of transportation. The history behind such access also matters.

References

[1] Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), 151-184.

[2] Américo, P. (2021). The industrialization paths: railroads and structural transformation in Brazil, 1872-1950. *Working paper.*

[3] Araújo, E., Carpena, L., & Cunha, A. B. (2008). Brazilian business cycles and growth from 1850 to 2000. *Estudos Econômicos*, 38(3), 557-581.

[4] Balboni, C. (2019). In harm's way: infrastructure investments and the persistence of coastal cities. *Working paper*.

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

[6] Barsanetti, B. (2021). Cities on pre-Columbian paths. *Journal of Urban Economics*, 122 (2), 103317.

[7] Bazzi, S., Fiszbein, M., & Gebresilasse, M. (2020). Frontier culture: the roots and persistence of "rugged individualism" in the United States. *Econometrica*, 88(6), 2329-2368.

[8] Bleakley, H., & Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, 127(2), 587-644.

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

[10] Bosker, M. & Buringh, E. (2017). City seeds: geography and the origins of the European city system. *Journal of Urban Economics*, 98, 139-157.

[11] Brooks, L., & Lutz, B. (2019). Vestiges of transit: urban persistence at a microscale. *Review of Economics and Statistics*, 101(3), 385-399.

[12] Chauvin, J. P., Glaeser, E., Ma, Y., & Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98, 17-49.

[13] Cingano, F., & Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2(4), 720-742.

[14] Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, vol. 5, Elsevier.

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

[16] Coşar, A. K., & Fajgelbaum, P. D. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, 8(1), 24-56.

[17] Cuberes, D., Desmet, K., & Rappaport, J. (2021). Urban growth shadows. *Jorunal of Urban Economics*, 122(3), 103334.

[18] Deffontaines, P. (1938). The origin and growth of the Brazilian network of towns. *Geographical Review*, 28(3), 379-399.

[19] Donaldson, D. (2018). Railroads of the Raj: estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934.

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

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

[22] Fajgelbaum, P. D., & Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, 88(4), 1411-1452.

[23] Fan, J., & Zou, B. (2019). Industrialization from scratch: the Construction of the Third Front and local economic development in hinterland China. *Working paper*.

[24] Flückiger, M., Hornung, E., Larch, M., Ludwig, M., & Mees, A. (2021). Roman transport network connectivity and economic integration. *Working paper*.

[25] Forero, A., Gallego, F. A., González, F., & Tapia, M. (2020). Railroads, specialization, and population growth: evidence from the first globalization. *Journal of Population Economics*, 1-46.

[26] Fujita, M., & Mori, T. (1996). The role of ports in the making of major cities: self-agglomeration and hub-effect. *Journal of Development Economics*, 49(1), 93-120.

[27] Fujita, M., Krugman, P., & Venables, A. (1999). *The spatial economy: cities, regions, and international trade.* Cambridge: MIT Press.

[28] Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126-1152.

[29] Hanlon, W. W., & Miscio, A. (2017). Agglomeration: a long-run panel data approach. *Journal of Urban Economics*, 99, 1-14.

[30] Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53(1), 1-28.

[31] Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5), 1067-1090.

[32] Henderson, J. V., Squires, T., Storeygard, A., & Weil, D. (2018). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1), 357-406.

[33] Hodgson, C. (2018). The effect of transport infrastructure on the location of economic activity: railroads and post offices in the American West. *Journal of Urban Economics*, 104, 59-76.

[34] Hsu, W. T. (2012). Central place theory and city size distribution. *The Economic Journal*, 122(563), 903-932.

[35] IBGE (2017). *Classifcação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação*. Rio de Janeiro: IBGE.

[36] IBGE (1958). *Enciclopédia dos municípios brasileiros*. Rio de Janeiro: IBGE.

[37] Jedwab, R., Kerby, E., & Moradi, A. (2017). History, path dependence and development: evidence from colonial railways, settlers and cities in Kenya. *The Economic Journal*, 127(603), 1467-1494.

[38] Jedwab, R., & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: evidence from Africa. *Review of Economics and Statistics*, 98(2), 268-284.

[39] Kohlhepp, G., Soethe, P.A., Martineschen, D., da Costa Pereira, C.H., Mathias, D., Antoniuk, E., Boechat, F.B., Fullgraf, F., da Silva, N.P., Paulino,S. & Neubauer, S.N. (2014). *Colonização agrária no Norte do Paraná: processos geoeconômicos e sociogeográficos de desenvolvimento de uma zona subtropical do Brasil sob a influência da plantação de café*. Eduem, Maringá, Portuguese translation of the German original from 1975.

[40] Leff, N. (1982). *Underdevelopment and development in Brazil*. Winchester: Allen & Unwin.

[41] Luna, F.V., & Klein, H. (2018). *An economic and demographic history of São Paulo, 1850-1950*. Stanford: Stanford University Press.

[42] Lin, J., & Rauch, F. (2020). What future for history dependence in spatial economics? *Regional Science and Urban Economics*, 103628.

[43] Menucci Giesbrecht, R. (2001). *Um dia o trem passou por aqui.* São Paulo: Studio4.

[44] Monbeig, P. (1949). A divisão regional do estado de São Paulo. *Anais do Geógrafos Brasileiros*, 1, 19-36.

[45] Monbeig, P. (1952). *Pionniers et planteurs de São Paulo*. Paris: A. Colin.

[46] Mori, T. (2012). Increasing returns in transportation and the formation of hubs. *Journal of Economic Geography*, 12, 877-897.

[47] Mun, S. I. (1997). Transport network and system of cities. *Journal of Urban Economics*, 42(2), 205-221.

[48] Proost, S., & Thisse, J.F. (2019). What can be learned from spatial economics? *Journal of Economic Literature*, 57(3), 575-643.

[49] Oster, E. (2019). Unobservable selection and coefficient stability: theory and evidence. *Journal of Business and Economics Statistics*, 37(2), 187-204.

[50] Qin, Y. (2017). No county left behind? The distributional impact of high-speed rail upgrades in China. *Journal of Economic Geography*, 17(3), 489-520.

[51] Rauch, J.E. (1993). Does history matter only when it matters little? The case of city-industry location. *The Quarterly Journal of Economics*, 108(3), 843-867.

[52] Redding, S.J., & Sturm, D.M. (2008). The costs of remoteness: evidence from German division and reunification. *American Economic Review*, 98(5), pp.1766-97.

[53] Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In *Handbook of regional and urban economics*, vol. 5, Elsevier.

[54] Reis, E., Pimentel, M., Alvarenga, A.I. & Santos, M.C.H. (2008). Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. *Rio de Janeiro, Ipea/Dimac.*.

[55] Secretaria da Agricultura, Comércio e Obras Públicas. (1925) *Relatório da Agricultura*. São Paulo.

[56] Sasaki, K. (1992). Trade and migration in a two-city model of transportation investments. *The Annals of Regional Science*, 26(4), 305-317.

[57] Sequeira, S., Nunn, N., & Qian, N. (2020). Immigrants and the making of America. *The Review of Economic Studies*, 87(1), 382-419.

[58] Silva, A. D., & Tosi, P. G. S. (2014). Considerações sobre entrelaçamento de circuitos e produções na órbita do complexo cafeeiro: o caso da Companhia Estrada de Ferro Araraquara (1896 a 1909). *Heera-Revista de História Econômica & Economia Regional Aplicada*, 10(16), 55-71.

[59] Trew, A. (2020). Endogenous infrastructure development and spatial takeoff in the First Industrial Revolution. *American Economic Journal: Macroeconomics*, 12(2): 44-93.

[60] Summerhill, W. R. (2003). *Order against progress: government, foreign investment, and railroads in Brazil, 1854-1913.* Stanford: Stanford University Press.

[61] Wahl, F. (2017). Does European development have Roman roots? Evidence from the German Limes. *Journal of Economic Growth*, 22(3), 313-349.

A Proof of Proposition 1

The proof has three steps.

STEP 1: I show that, if $y \notin \mathbb{Y}(t^r(y))$, then $y \notin \mathbb{Y}(t)$ for any $t > t^r(y)$. That is, I show that all existing cities emerge at the time the railroad reaches their site.

To see this is the case, note that for any $y \notin \mathbb{Y}(t) \cap [0, y^r(t)]$, it must be the case that $log\ q(y,t) < -A(y,t)$. In particular, if $y \notin \mathbb{Y}(t^r(y))$, then $log\ q(y, t^r(y)) < 0$. In this way, there is some $t' > t^r(y)$ such that:

$$log\ q(y,t') < log\ q(y, t^r(y)) \leq -g(0)[t' - t^r(y)] < -A(y,t')$$

where the first inequality comes from the fact that prices are decreasing, the second inequality comes from the fact that $t' - t^r(y)$ is sufficiently small, and the third inequality from the fact that $g(0)[t' - t^r(y)]$ is an upper bound on productivity growth. Hence, for any $t \in [t^r(y), t']$, $y \notin \mathbb{Y}(t)$. As a consequence, $A(y,t') = 0$. We can then iteratively repeat this same argument to show that, for any $t \in [t', t' + k(t' - t^r(y))]$, $y \notin \mathbb{Y}(t)$, where k is an arbitrary integer. Hence, $y \notin \mathbb{Y}(t)$ for any $t > t^r(y)$.

STEP 2: define $\tilde{\mathbb{Y}}(t) = \bigcup_{t' \leq t} \mathbb{Y}(t)$. This is the set of all sites that are or ever were a town. I show that, for any t, $\tilde{\mathbb{Y}}(t)$ is a finite set. As a consequence, $\mathbb{Y}(t)$ must also be finite, as it is a subset of $\tilde{\mathbb{Y}}(t)$.

First, note that as $\tilde{\mathbb{Y}}(t) \subset [0, y^r(t)]$, then $\tilde{\mathbb{Y}}(0) = \mathbb{Y}(t) = \{y_0 = 0\}$. Let $A(0,0)$ be the initial productivity at the port town. The log price of urban services from y_0 at site $y^r(t)$ will be no higher then: $log\ p(y^r(t), t) \leq -A(0,0) + \tau_s y^r(t)$. Therefore, define:

$$\underline{t_1} = \frac{A(0,0)}{\tau_s\ sup\ g^r(t)}$$

as the least time it takes for the railroad to reach a site where it is profitable for a firm to compete with firms at the port town, assuming no productivity growth in the later. Hence, for $t \leq \underline{t_1}$, it must be that $\tilde{\mathbb{Y}} = \{0\}$.

Now consider a time t such that $y^r(t) \in \mathbb{Y}(t)$. Define $\underline{\Delta t}$ as the positive number that satisfies:

$$A(y^r(t), t + \underline{\Delta t}) = \Big(\tau_s \; sup \; g^r(t)\Big)\underline{\Delta t}$$

Since $A(y^r(t), t + \underline{\Delta t})$ is concave in $\underline{\Delta t}$ and $\frac{\partial A(y^r(t),t)}{\partial \underline{\Delta t}} = g(0) > \tau_s \; sup \; g^r(t)$ (by Assumption 1), then $\underline{\Delta t}$ exists and is uniquely defined. Note that for any $t' \in (t, t + \underline{\Delta t})$, $log \; q(y^r(t'), t') < 0$, so no town will emerge in that interval. As a consequence, we can bound the cardinality of $\tilde{\mathbb{Y}}(t)$ by:

$$\#\tilde{\mathbb{Y}}(t) \leq \begin{cases} 1 & \text{, if } t < \underline{t_1} \\ k & \text{, if } t \in \big[\underline{t_1} + (k-1)\underline{\Delta t}, \; \underline{t_1} + k\underline{\Delta t}\big] \end{cases}$$

STEP 3: now I define a sequence of finite-dimensional dynamical systems that have each a unique solution and that, jointly with an updating rule for the set of existing towns, characterizes the unique equilibrium.

Dynamical system 1: it is a one dimensional dynamical system with $\mathbf{A}(t) = \big(A_0(t)\big)$. It is characterized by the following differential equation and the initial value:

$$\dot{A}_0(t) = g(A_0(t))$$

$$A_0(0) = A(0,0)$$

The equation above defines an initial vale problem (call it $\mathbf{P_0}$) which is Lipschitz continuous in $\mathbf{A}$ and continuous in t. Hence, by the Picard-Lindelöf Theorem, $\mathbf{P_0}$ has a unique and continuous solution $A_0(t)$.

Now define $t_1 \equiv min\big\{t > 0 | \tau_s y^r(t) \geq A_0(t)\big\}$. If no such t_1 exists, then there will never exist any town other than y_0. Otherwise, then $\mathbb{Y}(t_1) = \{y_0, y_1\}$, where $y_1 = y^r(t_1)$. Moreover, since $A_0(t)$ is continuous, so is $\tau_s y^r(t) - A_0(t)$. Hence, $\tau_s y^r(t_1) \geq A_0(t_1)$.

Dynamical system N: at t_N, let $\mathbb{Y}(t_N) = \{y_0, y_1, ..., y_m\}$, where $y_m = y^r(t_N)$. Note that, since $\mathbb{Y}(t) \subset \tilde{\mathbb{Y}}(t)$, then $m \leq N$. I will later show that $\mathbb{Y}(t) = \tilde{\mathbb{Y}}(t)$. Define $\mathbf{A}(t) = \big(A_0(t), ..., A_m(t)\big)$ and the initial value problem $\mathbf{P_N}$ (with initial time t_N) as:

$$\dot{A}_i(t) = g(A_i(t)) \quad i = 0, 1, ..., m$$
$$A_i(t_N) = A(y_i, t_N) \quad i = 0, 1, ..., m$$

which, by the Picard-Lindelöf Theorem, has a unique and continuous solution $\mathbf{A}(t)$.

Now I show that, along $\mathbf{A}(t)$, all towns in $\mathbb{Y}(t_N)$ will also be a town at a future time (**Claim 1**). This occurs if, for any towns y_i and y_j, the following is true:

$$-A(y_i, t) \leq -A(y_j, t) + \tau_s |y_i - y_j|$$

The equation above holds for t_N. Suppose $A(y_j, t_N) \leq A(y_i, t_N)$, then it must be the case at any future time $t > t_N$; hence, the inequality above will hold as well. Now suppose that $A(y_j, t_N) > A(y_i, t_N)$, then $g\big(A(y_j, t)\big) \leq g\big(A(y_i, t)\big)$, so the left-hand side of the equation above decreases faster than the righ-hold side; hence, the inequality is also satisfied at any $t > t_N$. This proves the claim.

Now define:

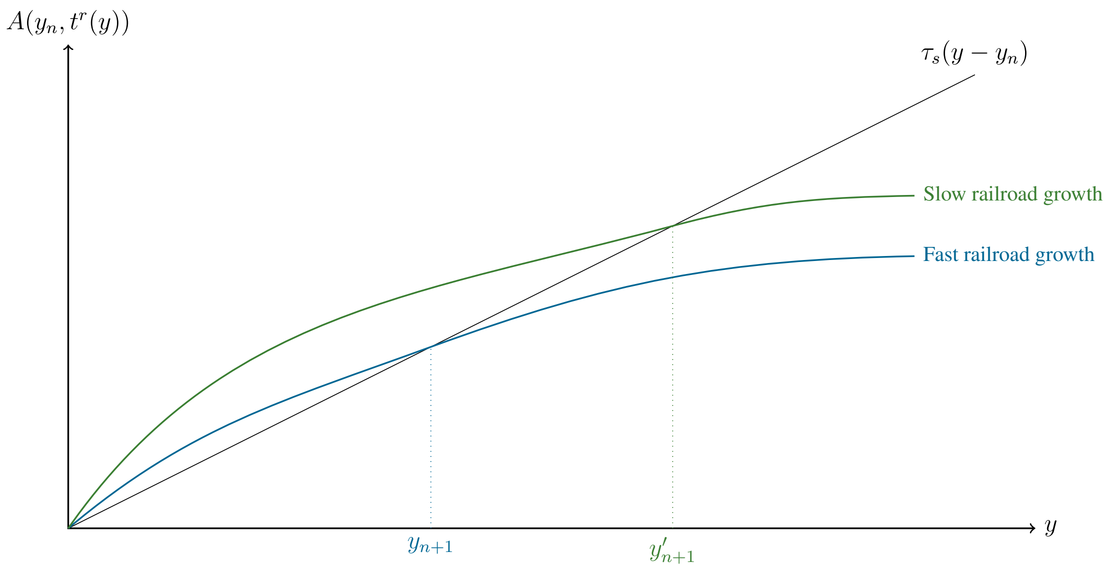
$$t_{N+1} = min\big\{t > t_{N+1} | \tau_s(y^r(t) - y_m) \geq A_m(t)\big\}$$

which will actually hold in equality, since $A_m(.)$ is continuous.

This completes the proof of uniqueness of the equilibrium. As a consequence of **Claim 1**, $\mathbb{Y}(t) = \tilde{\mathbb{Y}}(t)$. Moreover, the recursive procedure to identify the sequence $\{t_1, t_2, ...\}$ is the same recursive procedure explained in the statement of the proposition.

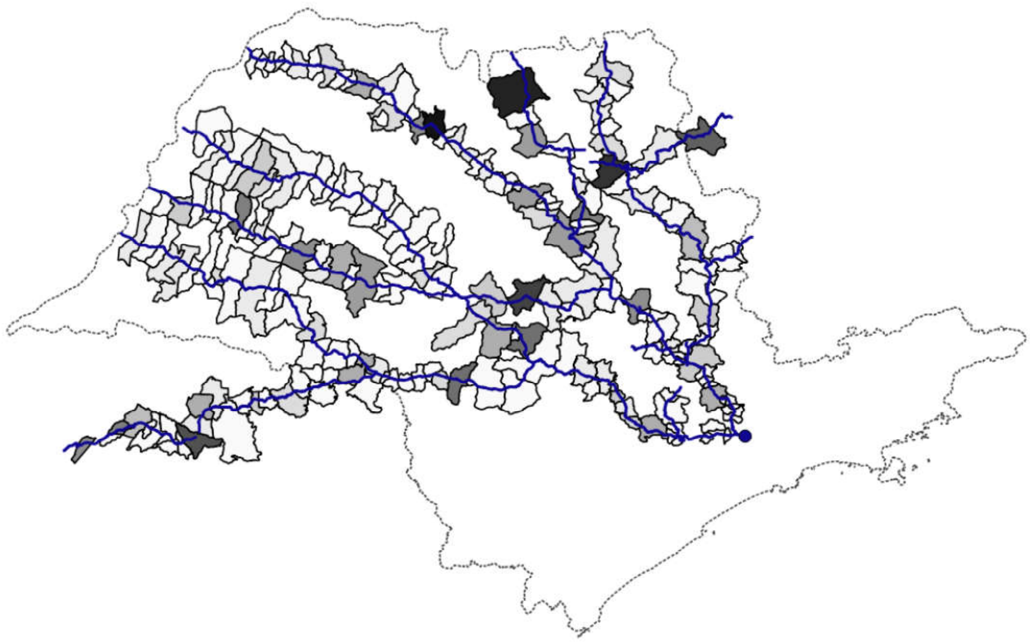
Figures and Tables

Figure 1: Distance to the next town



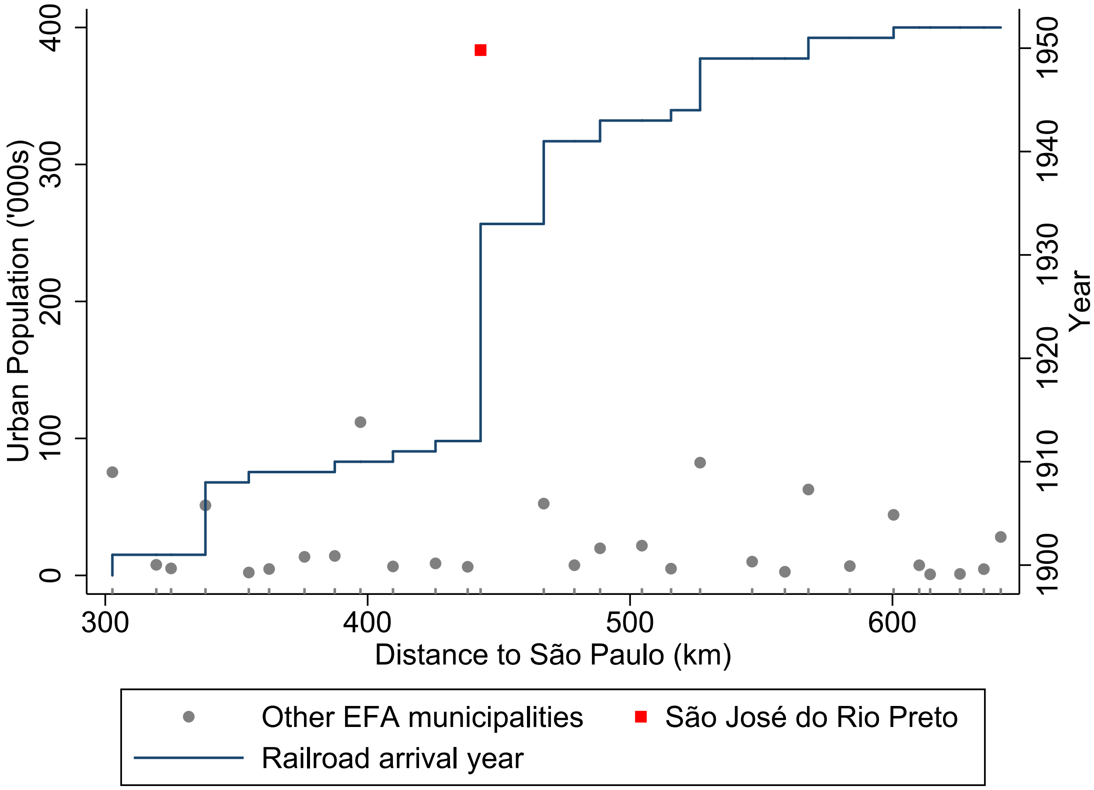
Note: The green line indicates $A(y_n, t^r(y))$ under slow railroad growth, while the navy blue indicates it under fast railroad growth.

Figure 2: Sample municipalities, railroad lines, and years as endpoint



Note: Railroad lines in blue; *Estação da Luz*, the main station in the city of São Paulo, is indicated by the blue circle. The black lines delineate sample municipalities, while the dotted gray line delineates the state of São Paulo. Years as endpoint are represented in shades of gray, with darker shades indicating more years.

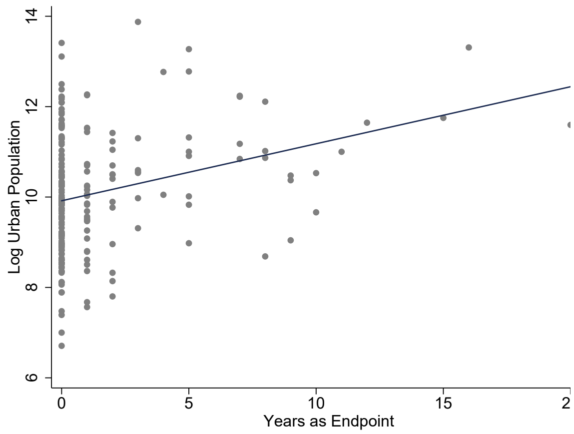
Figure 3: Railroad arrival dates and town sizes along the *E.F. Araraquara*



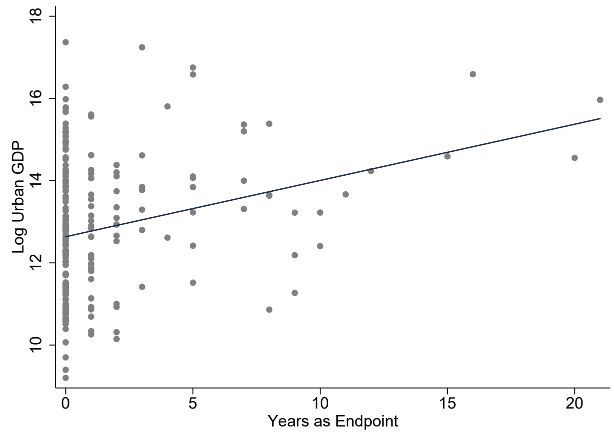
Note: The gray ticks on the horizontal axis indicate the distance of each municipality to São Paulo. The circles indicate the total urban population in 2010; the urban population of São José do Rio Preto is represented by a red square. The navy line indicates the railroad expansion: for each municipality, the year of railroad arrival is the minimum value of the line at that point, while the maximum value indicates the year of rairoad arrival to the next municipality. Hence, the vertical distance on the navy line represents years as endpoint.

Figure 4: Years as endpoint and town size

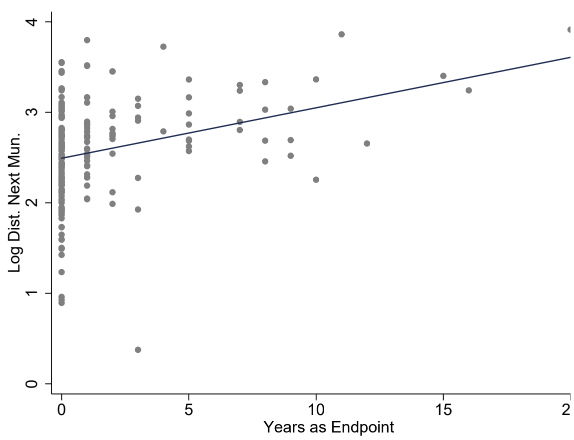
(a) Urban population



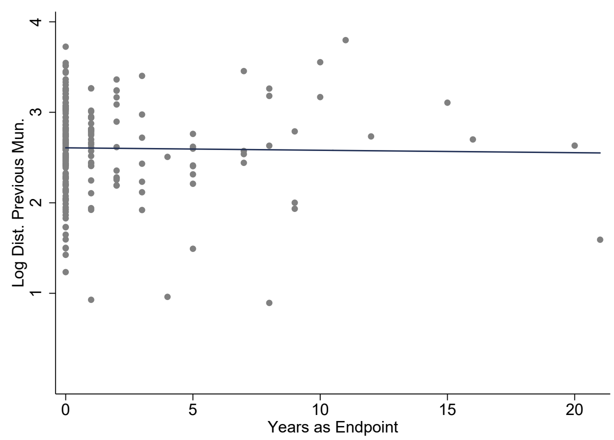
(b) Urban GDP



(c) Distance to next municipality



(d) Distance to previous municipality



Note: Each circle is a municipality. The line indicates the linear fit.

Table 1: Summary statistics

	Obs.	Mean	Std. Dev.
Years as endpoint	203	1.61	3.40
Years as endpoint ≥ 2	203	0.24	0.43
Population ('000s)	203	66.49	123.77
Population density	203	276.96	849.78
Urban population ('000s)	203	63.92	122.01
Urbanization rate	203	0.90	0.10
Per capita GDP (2010 PPP US$)	203	17152.37	15161.97
Distance to São Paulo (km)	203	382.67	181.12
Distance to next municipality (km)	203	15.22	8.17
Year of railroad arrival	203	1910.78	27.28
Year of incorporation	203	1924.98	51.47
Incorporated before railroad arrival	203	0.30	0.46
Altitude (m)	203	549.89	136.25
Terrain ruggedness index (m)	203	7.69	3.42
Main river dummy	203	0.43	0.50
Malaria suitability (days a year)	203	345.71	31.65
Potential yield (ton/ha) - coffee	203	0.69	0.22
Potential yield (ton/ha) - maize	203	2.54	1.22
Potential yield (ton/ha) - sugarcane	203	3.85	0.95
Share of *terra roxa*	203	0.34	0.38
Share of latosols	203	0.42	0.41
Share of acrisols	203	0.45	0.42
Latitude	203	-22.10	0.97
Longitude	203	-49.28	1.61

Note: See the Online Appendix for the sources and definitions of each variable.

Table 2: Years as endpoint, town sizes, and distance to next municipality

	(1)	(2)	(3)	(4)	(5)
Panel A: urban population					
Years as endpoint	0.126***	0.112***	9.351***	0.109***	0.103***
	(0.021)	(0.021)	(3.129)	(0.021)	(0.021)
Railroad arrival year		-0.006	0.210	-0.018**	-0.023**
		(0.005)	(0.639)	(0.009)	(0.010)
Log distance to São Paulo		-0.681***	-74.560**	-0.252	-1.680***
		(0.167)	(30.969)	(0.441)	(0.560)
Incorporated municipality		0.648***	49.779**	0.752***	0.721***
		(0.178)	(24.223)	(0.189)	(0.192)
R^2	0.100	0.349	0.279	0.431	0.474
Lower bound with prop. selection					0.090
Panel B: urban GDP					
Years as endpoint	0.137***	0.124***	237.220***	0.124***	0.118***
	(0.026)	(0.025)	(87.527)	(0.024)	(0.025)
R^2	0.078	0.390	0.301	0.486	0.522
Lower bound with prop. selection					0.111
Panel C: distance to the next municipality					
Years as endpoint	0.056***	0.037***	0.675***	0.038***	0.037***
	(0.009)	(0.009)	(0.214)	(0.009)	(0.009)
R^2	0.117	0.217	0.314	0.308	0.319
Lower bound with prop. selection					0.037
Panel D: distance to the previous municipality					
Years as endpoint	-0.003	-0.007	0.008	-0.006	-0.007
	(0.013)	(0.013)	(0.202)	(0.013)	(0.013)
R^2	0.000	0.042	0.051	0.180	0.184
Dependent variable in:	log	log	level	log	log
Baseline controls?		Y	Y	Y	Y
Geographic fundamentals?				Y	Y
Geographic coordinates?					Y
Observations	203	203	203	203	203

Note: Coefficients from OLS. The dependent variables are in log, except in column (3), in which they are in levels. Level population is in thousands of inhabitants, level GDP is in thousands of reais, and level distance to next or previous municipalities are in kilometers. Baseline controls are year of railroad arrival, log distance to São Paulo, and whether a municipality was incorporated before the railroad arrival. Geographic fundamentals are mean and squared altitude, mean and squared terrain ruggedness, the logarithms of the potential yields of coffee, maize and sugarcane, the logarithm of days a year that support malaria transmission, soil dummies, and a river dummy. Geographic coordinates are included as a quadratic polynomial. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table 3: Two-stage least squares estimates

	First stage (1)	Log urban population (2)	Log urban GDP (3)	Log distance next mun. (4)	Log distance previous mun. (5)
Panel A: baseline results					
Years as endpoint		0.228**	0.213**	0.217***	-0.054
		(0.092)	(0.103)	(0.077)	(0.046)
Log distance to next incorporated mun.	0.773***				
	(0.176)				
Robust F-statistic of the excluded IV	19.284				
Anderson-Rubin statistic		5.69	3.84	8.72	1.39
p-value of Anderson-Rubin test		0.017	0.050	0.003	0.238
Panel B: includes geographic fundamentals as included instruments					
Years as endpoint		0.216***	0.210**	0.218***	-0.048
		(0.076)	(0.089)	(0.078)	(0.044)
Log distance to next incorporated mun.	0.942***				
	(0.222)				
Robust F-statistic of the excluded IV	17.927				
Anderson-Rubin statistic		7.33	4.99	8.22	1.22
p-value of Anderson-Rubin test		0.007	0.026	0.004	0.269
Panel C: includes a measure of remoteness as an included instrument					
Years as endpoint		0.194**	0.213**	0.099***	-0.052
		(0.092)	(0.106)	(0.033)	(0.040)
Log distance to next incorporated mun.	1.389***				
	(0.358)				
Log distance to closest incorporated mun.	-0.969**	0.042	0.000	0.144***	-0.002
	(0.400)	(0.093)	(0.107)	(0.053)	(0.041)
Robust F-statistic of the excluded IV	15.056				
Anderson-Rubin statistic		4.33	3.87	8.55	1.96
p-value of Anderson-Rubin test		0.038	0.049	0.004	0.162
Observations	203	203	203	203	203

Note: Column (1) displays OLS estimates when the dependent variable is years as endpoint. Columns (2) to (5) display two-stage least squares estimates of the coefficient on years as endpoints. The excluded instrument is the logarithm of distance to the closest incorporated municipality farther down the railroad line. Baseline controls are included instruments in all the panels. Geographic fundamentals are mean and squared altitude, mean and squared terrain ruggedness, the logarithms of the potential yields of coffee, maize and sugarcane, the logarithm of days a year that support malaria transmission, soil dummies, a river dummy, and a quadratic polynomial of latitude and longitude. Robust standard errors in parentheses. Statistical significance (based on the Wald statistic) denoted by: * 10%, ** 5%, *** 1%

Online Appendix for

Road Endpoints and City Sizes

B.1 Model Extensions

B.1.1 Dynamic Agglomeration Economies

I present an example with three towns to illustrate how dynamic agglomeration modify the model. At any time t, at most three towns exist in this economy: the gateway town $y_0 = 0$, a temporary endpoint y_1, and a final endpoint town y_2; the locations y_1 and y_2 are endogenous. Suppose that the railroad is built at a linear speed $\hat{g}^r$ until it reaches site y_2.

I assume that productivity growth in each town is a function both of $A(y,t)$ and $m(y,t)$ according to the function:

$$g\big(A(y,t), m(y,t)\big) = \begin{cases} \hat{g}_H & \text{, if } A(y,t) \leq A_1 \\ \hat{g}_L & \text{, if } A(y,t) \in (A_1, A_2] \text{ and } m(y,t) \geq \frac{p(1-\beta)}{\tau_a} \\ 0 & \text{, otherwise} \end{cases}$$

where $A_2 > A_1$ and $\hat{g}_H > \hat{g}_L > \frac{\hat{g}_H}{2}$. Assume also that $\tau_s \hat{g}^r \in (\hat{g}_L, \hat{g}_H)$.

Initial productivity on y_0 is A_2; the gateway town already has the highest possible productivity and so its productivity does not grow. This assumption is useful because it implies the location of the first town, y_1, does not depend on the expansion rate of the railroad, $\hat{g}^r$; the latter will determine only the time y_1 was an endpoint.

The next town will appear on the first site y_1 in which transport costs make it profitable for urban service firms to enter: $y_1 = \frac{A_2}{\tau_s}$. Now consider the emergence of the final endpoint town y_2. Since $\hat{g}_H > \hat{g}^r$, y_2 only emerges after $A(y_1, t) > A_1$. Denote by t_E the time that town y_1 was a railroad endpoint. From equation 6, it must be the case that:

$$\tau_s(y_2 - y_1) = \tau_s \hat{g}^r t_E = A_1 + \big(t_E - \frac{A_1}{\hat{g}_H}\big)\hat{g}_L = A(y_1, t^r(y_1) + t^E), \tag{B.1}$$

which implies that there is a 1-to-1 correspondence between railroad construction speed and time as endpoint:

$$t_E = \frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H} \frac{A_1}{\tau_s \hat{g}^r - \hat{g}_L} \tag{B.2}$$

Note that $t_E \in \big(\frac{A_1}{\hat{g}_H}, \frac{A_1}{\hat{g}_H} + \frac{A_2 - A_1}{\hat{g}_L}\big)$ and that the distance between y_2 and y_1 is a linear function of t_E, $y_2 - y_1 = \frac{1}{\tau_s}\big[\frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H} A_1 + \hat{g}_L t_E\big]$.

The endpoint town y_2 will always serve the hallf-line that extends to its right, so equation 5 implies that $m(y_2, t) \geq \frac{p(1\beta)}{\tau_a}$. Hence, productivity in y_2 will eventually reach A_2. Now consider the former endpoint town y_1. Equation 5 and some simple algebra shows that, for $t \geq t^r(y_2)$:

$$m(y_1, t) = p(1-\beta)\left[\frac{(\hat{g}_H - \hat{g}_L)}{2\tau s \hat{g}_H} A_1 + \frac{\hat{g}_L t_E}{2\tau_s} + \frac{A(y_1, t)}{\tau_s} - \frac{A(y_2, t)}{2\tau_s}\right] \tag{B.3}$$

In particular, note that intermediate good use in town y_1 at the time y_2 emerges is an increasing function of t_E: $m(y_1, t^r(y_2)) = p(1-\beta)\big(\frac{1+2\tau_s}{2\tau_s}\big)\big[\frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H} A_1 + \hat{g}_L t_E\big]$.

There are two cases to consider. If $m(y_1, t^r(y_2)) \geq \frac{p(1-\beta)}{\tau_a}$, then town y_1 is sufficiently large to benefit from the dynamic agglomeration economies and so productivity growth in y_1 continues after y_2 emerges. Eventually, $A(y_1, t)$ reaches A_2. But if $m(y_1, t^r(y_2)) < \frac{p(1-\beta)}{\tau_a}$, then the town is too small and its productivity stagnates on $A(y_1, t^r(y_2))$. With town y_2 becoming more productive over time, town y_1 becomes smaller. If the distance between y_2 and y_1 is sufficiently small, town y_1 will eventually be abandoned. Otherwise, it will keep existing but its steady-state productivity will be strictly lower than A_2. Since $m(y_1, t^r(y_2))$ is increasing in t_E, the steady-state productivity of y_1 is an increasing function of time as endpoint t_E. The proposition below summarizes this result.

Proposition B.1. *Consider the environment above. There are $\bar{t}$ and $\underline{t}$ such that:*

1. *If $t_E \geq \bar{t}$, then $y_1 \in \mathbb{Y}(t)$ for all $t \geq t^r(y_1)$ and $lim_{t\to\infty} A(y_1, t) = A_2$.*
2. *If $t_E < \bar{t}$ and $t_E < \underline{t}$, then there is t such that $y_1 \notin \mathbb{Y}(t')$ for all $t' \geq t$.*
3. *If $t_E < \bar{t}$ and $t_E \geq \underline{t}$, then $y_1 \in \mathbb{Y}(t)$ for all $t \geq t^r(y_1)$ and $lim_{t\to\infty} A(y_1, t) < A_2$. Moreover, $lim_{t\to\infty} A(y_1, t)$ is increasing in t_E within this interval.*

Proof. Define $\bar{t}$ as the value of t_E that satisfies

$$p(1-\beta)\big(\frac{1+2\tau_s}{2\tau_s}\big)\big[\frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H} A_L + \hat{g}_L t^E\big] = \frac{p(1-\beta)}{\tau_a}.$$

If $t_E \geq \bar{t}$, then $A(y_1, t^r(y_1) + t_E) \geq \frac{p(1-\beta)}{\tau_a}$. Remember that, for $t \geq t^r(y_1) + t_E$,

$$m(y_1, t) = p(1-\beta)\left[\frac{(\hat{g}_H - \hat{g}_L)}{2\tau s \hat{g}_H}A_1 + \frac{\hat{g}_L t_E}{2\tau_s} + \frac{A(y_1, t)}{\tau_s} - \frac{A(y_2, t)}{2\tau_s}\right].$$

Therefore, $\dot{m}(y_1, t) \geq \frac{p(1-\beta)}{\tau_s}(\hat{g}_L - \frac{\hat{g}_H}{2}) > 0$ as long as $A(y_1, t) < A_2$. Hence, $A(y_1, t) \geq \frac{p(1-\beta)}{\tau_a}$ for all $t \geq t^r(y_1) + t_E$, so $\dot{A}(y_1, t) = \hat{g}_L$ as long as $A(y_1, t) < A_2$. Therefore, $lim_{t\to\infty} A(y_1, t) = A_2$. This proves case (1).

Now consider the case in which $t_E < \bar{t}$. In this case, $A(y_1, t^r(y_1)+t_E) < \frac{p(1-\beta)}{\tau_a}$, so $\dot{A}(y_1, t) = 0$ and $\dot{m}(y_1, t) \leq 0$. For a sufficiently large t, $A(y_2, t) = A_2$. Hence, it will be unprofitable for a firm to establish on site y_1 at time t if $A_2 - \tau_s(y_2 - y_1) < A(y_1, t) = A(y_1, t^r(y_2) + t^E)$. Since $y_2 - y_1 = \frac{1}{\tau_s}\left[\frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H}A_1 + \hat{g}_L t_E\right]$, then it will be eventually unprofitable only if $A_2 \geq \left[\frac{(\hat{g}_H - \hat{g}_L)}{\hat{g}_H}A_1 + \hat{g}_L t_E\right]$. Straightforward algebra shows that this condition is equivalent to $t_E < \underline{t} \equiv \frac{A_2}{2\hat{g}_L} - \frac{A_1(\hat{g}_H - \hat{g}_L)}{\hat{g}_H \hat{g}_L}$. This proves case (2).

If $t_E < \bar{t}$ and $t_E \geq \underline{t}$, then $y_1 \in \mathbb{Y}(t)$ for all $t \geq t^r(y_1)$. But since $m(y_1, t) < \frac{p(1-\beta)}{\tau_a}$ for all $t \geq t^r(y_B)$, then $\dot{A}(y_1, t) = 0$. Hence, $lim_{t\to\infty} A(y_1, t) = A(y_1, t^r(y_1) + t_E)$, which is an increasing function of t_E. This proves case (3). □

B.1.2 Static Agglomeration Economies

I present a three town example as in Subsection B.1.1: the railroad is built at linear speed $\hat{g}^r$ until it reaches an endogenous location y_2. A first interior city emerges at $y_1 < y_2$. Now, instead of intermediate good use $m(y, t)$ influencing the growth rate of productivity, I assume it affects the contemporary productivity. The log-productivity of a town y consists of two terms:

$$A^*(y, t) = A(y, t) + \phi m(y, t) \tag{B.4}$$

Note that $\phi > 0$ is a parameter that determines the strength of the agglomeration economies. I assume that (for towns) $A(y, t)$ increases at a linear speed $\hat{g}$ until productivity reaches $\overline{A} > 0$.

At $t = 0$, there is a single town at the gateway site $y_0 = 0$. I assume that $A(0, 0) = \overline{A}$. In this case, as long as no other town emerges, the town y_0 has log-productivity $A^*(y_0, t) = \overline{A} + \phi p(1 - \beta)\left(y^r(t) + \frac{1}{\tau_a}\right)$. Note that, as the railroad expands, the town productivity increases. This occurs because the higher prices for the agricultural good increase the demand for urban services from the gateway town, which results in higher productivity due to the agglomeration economies. If this increase in productivity is sufficiently strong, then urban services prices at the endpoint site $y^r(t)$ could fall as the railroad expands, preventing new towns from emerging. I introduce the following assumption to rule this out:

Assumption B.1. $\phi p(1-\beta) < \tau_s$

Due to Assumption B.1, eventually there is a site for which an *individual* firm benefits from entering. Denote this site by y_1; it is at a point at which the gateway town provides urban services with a price equal to 1. Some simple algebra shows this site is:

$$y_1 = \frac{\overline{A} + \frac{\phi p(1-\beta)}{\tau_a}}{\tau_s - \phi p(1-\beta)} \tag{B.5}$$

I assume that a town emerges at y_1. This assumption implies that there is no coordination that allows a *continuum* of urban firms to establish themselves at an earlier site, benefitting from the agglomeration economies to compete with firms at the gateway y_0. By doing so, I rule out a multiple equilibra situation that often appears in economies with static agglomeration economies. A similar assumption is used by Fujita *et al* (1999) when modeling the emergence of new cities.

Note that, as in the example in Subsection B.1.1, y_1 does not depend on the growth rate of the railroad, $\hat{g}^r$. The latter parameter then only affects the time y_1 spends as an endpoint.

After y_1 emerges, there is now a two-town economy. Town y_1 serves all the hinterland to its right, and so the productivity at y_1 immediately jumps to a positive value due to the agglomeration economies.

I now introduce one more assumption to the example. This assumption rules out that town y_1 becomes so large to the point of leading the gateway town y_0 to disappear. Note that the emergence of a third town y_2 limits the size of y_1. In fact, by using an upper bound for the productivity of y_1, then town y_2 must emerge at a site that is no farther away from the origin as the site $\overline{y}_2 = [\tau_s - \phi p(1-\beta)]^{-1}[\overline{A} + \frac{\phi p(1-\beta)}{\tau_s} + \tau_s y_1]$.

Assumption B.2. $\phi p(1-\beta)\big(\overline{y}_2 + \frac{1}{\tau_a}\big) < \tau_s y_1$

Assumption B.2 implies there is a site $d_1(t) > 0$ which defines the hinterland border between the two towns. Given $d_1(t)$, log-productivity in the two towns satisfy:

$$
\begin{aligned}
A^*(y_0,t) &= \overline{A} + \phi p(1-\beta)d_1(t) \\
A^*(y_1,t) &= min\left\{\overline{A}, \frac{\hat{g}}{\hat{g}^r}\big(y^r(t)-y_1\big)\right\} + \phi p(1-\beta)\big(y^r(t)-d_1(t)\big) + \frac{\phi p(1-\beta)}{\tau_a}
\end{aligned}
$$

The hinterland border $d_1(t)$ is defined by the site at which firms from both towns can provide urban services at the same price. Using the expressions for the logarithm of productivity above, I derive the following expression for $d_1(t)$:

$$
d_1(t) = \frac{y_1}{2} + \frac{\overline{A} - min\left\{\overline{A}, \frac{\hat{g}}{\hat{g}^r}\big(y^r(t)-y_1\big)\right\} - \phi p(1-\beta)\big(y^r(t)-y_1\big) - \frac{\phi p(1-\beta)}{\tau_a}}{2\big[\tau_s - \phi p(1-\beta)\big]} \tag{B.6}
$$

I now consider the emergence of a third town, at a site $y_2 > y_1$. As for the case of the emergence of y_1, I assume the next town emerges at the first site at which an *individual* firm can profitably enter. The following result shows that time as endpoint increases pushes y_2 to the right:

Lemma B.1. *The distance from the site y_2 to the origin (and to y_1) is non-increasing in $\hat{g}^r$.*

Proof. By using the expression for $d_1(t)$ in equation B.6, the log-productivity of town y_1 as a function of $y^r(t)$ may be written as:

$$
\begin{aligned}
F(y^r(t)) \equiv A^*(y_1,t) = min\big\{\overline{A}, \frac{\hat{g}}{\hat{g}^r}\big(y^r(t)-y_1\big)\big\} + \frac{\phi p(1-\beta)}{\tau_a} + \phi p(1-\beta)\Bigg[y_r(t) - \frac{y_1}{2} \\
- \frac{\overline{A} - min\big\{\overline{A}, \frac{\hat{g}}{\hat{g}^r}\big(y^r(t)-y_1\big)\big\} - \phi p(1-\beta)\big(y^r(t)-y_1\big) - \frac{\phi p(1-\beta)}{\tau_a}}{2\big[\tau_s - \phi p(1-\beta)\big]}\Bigg]
\end{aligned}
$$

Note that $F(y^r(t))$ is strictly increasing and concave in $y^r(t)$. For a given $y^r(t)$, $A^*(y_1,t)$ is also weakly decreasing in $\hat{g}^r$. The location y_2 is given by the value of $y^r(t)$ for which $F(y^r(t))$ intersects with $\tau_s[y^r(t)-y_1]$.

Note that $F(y_1) > 0 = \tau_s(y_1 - y_1)$. Note also that, from Assumption B.2, $d_1(t) > 0$ and so $A^*(y_1,t) < \overline{A} + \phi p(1-\beta)\big(y^r(t) - \frac{1}{\tau_a}\big)$. Hence, $F(\overline{y}_2) < \tau_s(\hat{y}_2 - y_1)$. Due to the concavity of

$F(.)$, there is a value y_1 for which $F(y_2) = \tau_s(y_2 - y_1)$. Since an increase in $\hat{g}^r$ weakly decreases $F(.)$, then it also weakly decreases y_2. $\square$

The previous lemma shows that distance to the next town decreases with the speed of railroad construction. Therefore, it increases with time as endpoint. But eventually a new town y_2 emerges. The following proposition derives the steady-state town sizes and productivities as a function from y_2. Using the previous lemma, the proposition below shows that town sizes and productivity increase with time as endpoint.

Proposition B.2. *The steady-state town size and productivity of y_1 is non-increasing in $\hat{g}^r$.*

Proof. In what follows, I ignore the time indicators t for ease of notation (as the focus is on steady-state values). In the steady state, $A(y_0) = A(y_1) = A(y_2) = \overline{A}$. Note that the intermediate good use in each town is a function of the two hinterland borders, d_1 and d_2:

$$
\begin{aligned}
m(y_0) &= p(1-\beta)d_1 \\
m(y_1) &= p(1-\beta)(d_2 - d_1) \\
m(y_2) &= p(1-\beta)(y_2 - d_2 + \frac{1}{\tau_a})
\end{aligned}
$$

By using the price conditions that determine d_1 amd d_2, the following system of equations appear:

$$
\begin{pmatrix} 2\big[\phi p(1-\beta) - \tau_s\big] & \phi p(1-\beta) \\ \phi p(1-\beta) & 2\big[\phi p(1-\beta) - \tau_s\big] \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} \tau_s y_1 \\ \big[\phi p(1-\beta) - \tau_s\big] y_2 - \tau_s y_1 - \frac{\phi p(1-\beta)}{\tau_a} \end{pmatrix}
$$

By solving the system above, I write d_1 and d_2 as a function of parameters and y_2 as:

$$
\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \kappa \begin{pmatrix} 2\big[\phi p(1-\beta) - \tau_s\big] & -\phi p(1-\beta) \\ -\phi p(1-\beta) & 2\big[\phi p(1-\beta) - \tau_s\big] \end{pmatrix} \begin{pmatrix} \tau_s y_1 \\ \big[\phi p(1-\beta) - \tau_s\big] y_2 - \tau_s y_1 - \frac{\phi p(1-\beta)}{\tau_a} \end{pmatrix}
$$

where $\kappa = \frac{1}{4\big[\phi p(1-\beta) - \tau_s\big]^2 - \phi^2 p^2 (1-\beta)^2}$. From the solution above, note that d_1 is decreasing in y_2 and d_2 is increasing in y_2. Hence, the hinterland size of town y_1 is increasing in y_2. Since, from Lemma B.1, y_2 is weakly decreasing in the speed of railroad construction $\hat{g}^r$, then the hinterland size of town y_1 is weakly decreasing in $\hat{g}^r$. Therefore, intermediate good use $m(y_1)$, log-productivity $A^*(y_1)$ and town size $s(y_1)$ are also weakly decreasing in $\hat{g}^r$. This concludes the proof. $\square$

B.1.3 Farmer Mobility

Assume that each site may be farmed by up to a mass one of farmers. Farmers on site y at period t must pay a rent to use the land; denote these land rental prices as $\ell(y,t)$. Suppose landowners reside outside the economy, so demand for the urban services comes only from farmers. A farmer has disposable income $p(y,t)-\ell(y,t)$ if she farms site y. According to farmers' preferences, the farmer that farms site y at time t consumes $\beta\frac{p(y,t)-\ell(y,t)}{p(y,t)}$ of the agricultural good and $(1-\beta)\frac{p(y,t)-\ell(y,t)}{q(y,t)}$ of urban services.

Suppose that farmers are perfectly mobile and that there is free entry of farmers into the economy. Let $v_0 > 0$ be the utility associated with farmers' outside option. As a consequence, the indirect utility at any site y that is farmed at period t is equal to v_0, which allows land prices to be written as a function of product prices:

$$\ell(y,t) = max\{0,\ p(y,t) - \frac{v_0}{\beta^\beta(1-\beta)^{1-\beta}}p(y,t)^\beta q(y,t)^{1-\beta}\} \tag{B.7}$$

The following assumption states that farmers' indirect utility on sites where the railroad has just arrived is higher than their outside option.

Assumption B.3. $v_0 < \beta^\beta(1-\beta)^{1-\beta}p^{1-\beta}$

Since product prices are (weakly) lower at $y < y^r(t)$ than at $y^r(t)$, Assumption B.3 implies that all sites served by the railroad are farmed. Hence, the incentives for urban firms to establish in a site that was just reached by the railroad are no different than in the case in which farmers were immobile. Urban firms enter a site if, and only if, they may provide urban services to $y^r(t)$ at a lower price than urban firms in the closest town. Therefore, Proposition 1 remains valid. City emergence and productivity growth in cities are thus the same as before.

All sites up to site $y^F(t) > y^r(t)$ are farmed. No sites farther away than $y^F(t)$ are farmed. I refer to $y^F(t)$ as the agricultural frontier. It is the farthest site that may be profitably farmed. At $y^F(t)$, land prices are zero and the farmer indirect utility is v_0. Using the expressions for product prices from equations 1 and 2, after some algebra $y^F(t)$ may be written as a function of the location of the railroad endpoint and the location and productivity of the endpoint town as:

$$y^F(t) = \frac{\tau_a}{\tau_a+\tau_s}y^r(t) + \frac{\tau_s}{\tau_a+\tau_s}y^N + A(y_N,t) + \frac{1}{1-\beta}log\left(\frac{\beta^\beta(1-\beta)^\beta p^{1-\beta}}{v_0}\right)$$

Note that the inclusion of mobile farmers into the model does not distort firm entry decisions, so it does not affect town locations. Still, it complicates the model in two ways. First, the model now includes an agricultural frontier that is endogenous. Second, land prices reduce the disposable income of farmers, which will change the equations that define town sizes. However, as in the baseline model, town size will be an increasing function of hinterland size and town productivity, and so increasing in time as endpoint too. Farmer mobility does not change the qualitative predictions of the model.

To complete the characterization, I consider towns sizes. The demand for urban services from the rural hinterland served by a town determine its size. Total production of a non-endpoint town y_n is:

$$s(y_n,t) = \frac{p^\beta u_0}{\beta^\beta(1-\beta)^{1-\beta}\tau_s}e^{A(y_n,t)}e^{(1-\beta)\tau_s[d_{n+1}(t)-d_n(t)]} \tag{B.8}$$

The hinterland served by the endpoint town y_N extends up to the agricultural frontier $y^F(t)$. Total production of urban services at the endpoint town is:

$$s(y_N,t) = \frac{p^\beta u_0(1-\beta)^\beta}{\beta^\beta}e^{A(y_n,t)}\left(\frac{e^{(1-\beta)\tau_s[y^r(t)-d_N(t)]}}{(1-\beta)\tau_s} + \frac{e^{[(1-\beta)\tau_s-\beta\tau_a][y^F(t)-y^r(t)]}}{(1-\beta)\tau_s-\beta\tau_a}\right) \tag{B.9}$$

B.1.4 Urban Workers

The following extension consists in adding labor as an input to the production of urban services. Assume that workers have the same preference as farmers, but they decide on where to locate according to an arbitrage condition. Let the utility of their outside option be u_0; assume u_0 is exogenous and workers enter freely in the economy. This assumption is consistent with the high levels of internal and domestic migration to São Paulo and surrounding regions at the time of railroad expansion. I include labor in the production function by assuming the production of urban services requires both labor l and intermediate goods m according to a Cobb-Douglas production function. Let $\alpha \in (0,1)$ be the expenditure share in the intermediate good. As before, $e^{A(y,t)}$ is the total factor productivity on site y at period t.

In each town $y \in \mathbb{Y}(t)$, the indirect utility of workers will be equal to their outside option, which will imply a log-linear relationship between wages and urban services prices:

$$log\ q(y,t) = log\left[\frac{(1-\beta)\beta^{\frac{\beta}{1-\beta}}}{p^{\frac{\beta}{1-\beta}} u_0^{\frac{1}{1-\beta}}}\right] + \frac{1}{1-\beta} log\ w(y,t) \tag{B.10}$$

Moreover, from the cost minimization problem of urban firms, there is a second log-linear relationship between prices and wages:

$$log\ q(y,t) = A(y,t) - log\left[(1-\alpha)^{1-\alpha}\alpha^{\alpha}\right] + (1-\alpha) log\ w(y,t) \tag{B.11}$$

By combining the two conditions, note that both wages and urban service prices are negative log-linear functions of TFP:

$$log\ w(y,t) = c_w - \frac{1-\beta}{\beta + (1-\beta)\alpha} A(y,t) \tag{B.12}$$

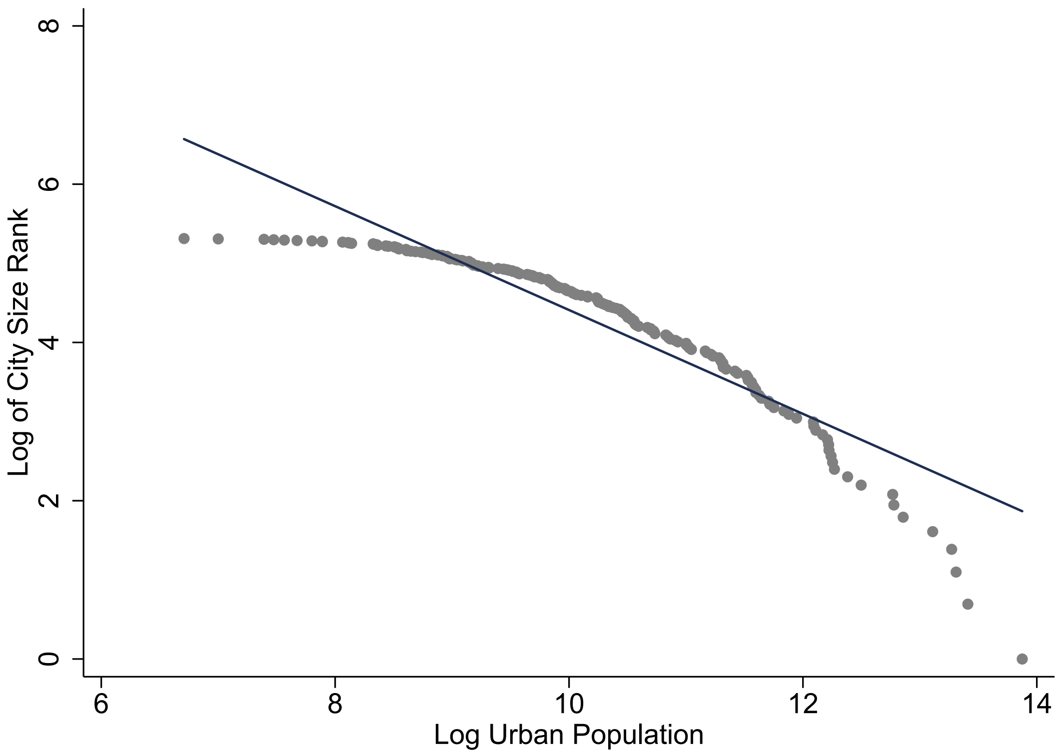
$$log\ q(y,t) = c_q - \frac{1}{\beta + (1-\beta)\alpha} A(y,t) \tag{B.13}$$

where c_w and c_q are functions of the utility and production parameters and of the outside option u_0. Without labor, the relation between prices and $A(y,t)$ is linear with a coefficient of one.

Note that the inclusion of labor implies that the demand for urban services comes not only from the rural hinterland, but also from local urban workers. However, the local workers will always consume a share $(1-\alpha)$ of the town production. Firm entry decisions and the decisions of farmers from which town to buy urban services are similar to the decisions in the baseline model. Therefore, the inclusion of labor does not lead to a qualitative change to the model and its main predictions hold.

B.2 Appendix Figures and Tables

Figure B.1: Relationship between city sizes and city rank



Note: Each circle is a sample municipality. The line indicates the linear fit

Table B.1: Balance of geographic fundamentals

Panel A:	Altitude (m)	TRI (m)	Log potential yield: coffee	maize	sugarcane
	(1)	(2)	(3)	(4)	(5)
Years as endpoint	1.949	0.016	-0.005	-0.010	-0.006
	(2.294)	(0.048)	(0.008)	(0.012)	(0.006)
Railroad arrival year	-1.809**	0.067**	-0.002	0.006**	-0.002
	(0.429)	(0.015)	(0.002)	(0.002)	(0.002)
Log distance to São Paulo	-76.194**	-5.079**	0.162**	0.038	0.209**
	(13.065)	(0.700)	(0.046)	(0.063)	(0.062)
Incorporated municipality	25.341+	-0.148	0.048	-0.041	0.019
	(14.769)	(0.401)	(0.058)	(0.082)	(0.045)
R^2	0.544	0.550	0.052	0.142	0.179

Panel B:	River (dummy)	Log malaria days	Soil dummies: terra roxa	latosol	acrisol
	(1)	(2)	(3)	(4)	(5)
Years as endpoint	0.003	-0.001	0.006	0.006	-0.004
	(0.011)	(0.001)	(0.008)	(0.008)	(0.010)
Railroad arrival year	-0.006***	-0.002***	-0.012***	-0.013***	0.008***
	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)
Log distance to São Paulo	0.135	0.172***	0.389***	0.298***	-0.255***
	(0.089)	(0.015)	(0.051)	(0.055)	(0.060)
Incorporated municipality	0.165**	0.011	0.018	0.039	-0.063
	(0.079)	(0.009)	(0.052)	(0.052)	(0.064)
R^2	0.051	0.723	0.240	0.275	0.103

Note: Coefficients from OLS. The sample contains 203 municipalities. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.2: Predictive power of different sets of variables

	Log urban population (1)	Log urban GDP (2)	Log distance next mun. (3)	Log distance previous mun. (4)
Years as endpoint	0.080	0.067	0.070	0.002
Baseline controls	0.156	0.172	0.083	0.018
Geographic fundamentals	0.137	0.166	0.132	0.132
Coordinates polynomial	0.101	0.117	0.033	0.032
R^2	0.474	0.522	0.319	0.184

Note: The table displays the Shapley values for the R^2. Dependent variables vary by column; see the labels on the top of the table. The equations are estimated by OLS. Baseline controls are year of railroad arrival, log distance to São Paulo, and whether a municipality was emancipated before the railroad arrival. Geographic fundamentals are mean and squared altitude, mean and squared terrain ruggedness, the logarithms of the potential yields of coffee, maize and sugarcane, the logarithm of days a year that support malaria transmission, soil dummies, and a river dummy. The coordinates polynomial is quadratic and includes the interaction between longitude and latitude.

Table B.3: Association with the logarithm of urban population in different census years

	2010 (1)	2000 (2)	1991 (3)	1980 (4)	1970 (5)
Years as endpoint	0.112***	0.112***	0.114***	0.111***	0.112***
	(0.021)	(0.021)	(0.018)	(0.018)	(0.017)
R^2	0.349	0.334	0.337	0.299	0.279
Observations	203	203	196	196	195

Note: Coefficients from OLS. Robust standard errors in parentheses. The equations includes baseline controls: year of railroad arrival, incorporated municipality dummy, and log distance to São Paulo. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.4: Robustness to alternative measures of time as endpoint

	Log urban population (1)	Log urban GDP (2)	Log distance next mun. (3)	Log distance previous mun. (4)
Time as endpoint $\geq$ 1 year	0.449**	0.539**	0.233***	-0.056
	(0.180)	(0.210)	(0.066)	(0.075)
Time as endpoint $\geq$ 2 years	0.666***	0.754***	0.228***	-0.053
	(0.211)	(0.249)	(0.084)	(0.090)
Time as endpoint $\geq$ 5 years	0.923***	1.018***	0.334***	-0.032
	(0.240)	(0.284)	(0.088)	(0.124)
Time as endpoint $\geq$ 10 years	1.233***	1.350***	0.420**	0.250
	(0.424)	(0.489)	(0.164)	(0.231)
IHS of years as endpoint	0.352***	0.398***	0.133***	-0.028
	(0.086)	(0.101)	(0.031)	(0.042)
Years as endpoint	0.122**	0.132**	0.055**	0.006
	(0.055)	(0.066)	(0.021)	(0.032)
Square of years as endpoint	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.004)	(0.001)	(0.002)
Observations	203	203	203	203

Note: Coefficients from OLS. Each cell indicates the estimated coefficients of a single equation, except for cells in the bottom two rows, which reports coefficient from an equation that includes both time as endpoint and the square of time as endpoint. Baseline controls are included. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.5: Further robustness

	Control for GDP growth (1)	Year polynomial (2)	Company FEs (3)	1872 Census controls (4)
Panel A: urban population				
Years as endpoint	0.112***	0.108***	0.106***	0.106***
	(0.021)	(0.019)	(0.022)	(0.022)
R^2	0.349	0.409	0.397	0.319
Panel B: urban GDP				
Years as endpoint	0.123***	0.119***	0.122***	0.123***
	(0.025)	(0.021)	(0.026)	(0.026)
R^2	0.390	0.460	0.429	0.358
Panel C: distance to next municipality				
Years as endpoint	0.037***	0.037***	0.042***	0.035***
	(0.009)	(0.009)	(0.009)	(0.009)
R^2	0.217	0.241	0.303	0.253
Panel D: distance to previous municipality				
Years as endpoint	-0.007	-0.007	-0.004	-0.011
	(0.013)	(0.013)	(0.012)	(0.011)
R^2	0.045	0.112	0.123	0.185
Observations	203	203	203	187

Note: Coefficients from OLS. The dependent variables are in log. Robust standard errors in parentheses. All columns include the baseline controls (year of railroad arrival, incorporated municipality dummy, and log distance to São Paulo). Column (1) includes national GDP growth. Column (2) includes a 3rd-degree polynomial of the year of railroad arrival. Column (3) includes railroad company fixed effects; when more than a railroad company served a municipality, I consider the first company to arrive at the municipality. Column (4) includes the following information from the origin municipality in the 1872 census as control variables: the logarithm of population density, the share of slaves in the population, the share of foreigners (born outside Brazil), the literacy rate, the share of free children 6-15 year old in school, and the share of employment outside of agriculture. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.6: Inference robust to cross-sectional dependence of the errors

	Log urban population (1)	Log urban GDP (2)	Log distance next mun. (3)	Log distance previous mun. (4)
Years as endpoint	0.112	0.124	0.037	-0.007
cutoff: 50 km	(0.017)	(0.020)	(0.010)	(0.012)
cutoff: 75 km	(0.018)	(0.021)	(0.009)	(0.012)
Observations	203	203	203	203

Note: Coefficients from OLS. Baseline controls are included. Robust standard errors in parentheses allow for cross-sectional dependence as in Conley (1999), with an uniform kernel and distance cutoffs of 50 and 75 km.

Table B.7: Reduced form estimates

	Log urban population (1)	Log urban GDP (2)	Log distance next mun. (3)	Log distance previous mun. (4)
Panel A: baseline results				
Log distance to next incorporated mun.	0.176**	0.165*	0.168***	-0.041
	(0.075)	(0.086)	(0.045)	(0.035)
R^2	0.300	0.346	0.275	0.048
Panel B: includes geographic fundamentals as included instruments				
Log distance to next incorporated mun.	0.204***	0.198**	0.205***	-0.045
	(0.074)	(0.092)	(0.056)	(0.041)
R^2	0.441	0.488	0.395	0.189
Panel C: includes a measure of remoteness as an included instrument				
Log distance to next incorporated mun.	0.270**	0.295**	0.137***	-0.072
	(0.128)	(0.147)	(0.044)	(0.050)
Log distance to closest incorporated mun.	-0.147	-0.206	0.049	0.048
	(0.159)	(0.183)	(0.053)	(0.056)
R^2	0.304	0.352	0.277	0.051
Observations	203	203	203	203

Note: Coefficients from OLS. Baseline controls are included in all the panels. Geographic fundamentals are mean and squared altitude, mean and squared terrain ruggedness, the logarithms of the potential yields of coffee, maize and sugarcane, the logarithm of days a year that support malaria transmission, soil dummies, a river dummy, and a quadratic polynomial of latitude and longitude. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.8: Robustness to instrument construction

	First stage (1)	Log urban population (2)	Log urban GDP (3)	Log distance next mun. (4)	Log distance previous mun. (5)
Panel A: additional 50 km for observations with no next already-incorporated municipality					
Years as endpoint		0.243**	0.234**	0.207***	-0.056
		(0.094)	(0.105)	(0.075)	(0.046)
Log distance to next incorporated mun.	0.703***				
	(0.166)				
Robust F-statistic of the excluded IV	17.851				
Panel B: additional 100 km for observations with no next already-incorporated municipality					
Years as endpoint		0.252***	0.248**	0.201***	-0.058
		(0.096)	(0.108)	(0.074)	(0.047)
Log distance to next incorporated mun.	0.653***				
	(0.159)				
Robust F-statistic of the excluded IV	16.834				
Panel C: additional 150 km for observations with no next already-incorporated municipality					
Years as endpoint		0.260***	0.258**	0.196***	-0.060
		(0.098)	(0.110)	(0.073)	(0.047)
Log distance to next incorporated mun.	0.616***				
	(0.154)				
Robust F-statistic of the excluded IV	16.076				
Panel D: additional 200 km for observations with no next already-incorporated municipality					
Years as endpoint		0.265***	0.266**	0.192***	-0.061
		(0.100)	(0.111)	(0.072)	(0.047)
Log distance to next incorporated mun.	0.587***				
	(0.149)				
Robust F-statistic of the excluded IV	15.488				
Observations	203	203	203	203	203

Note: Column (1) displays OLS estimates when the dependent variable is years as endpoint. Columns (2) to (5) display two-stage least squares estimates of the coefficient on years as endpoints. The excluded instrument is the logarithm of distance to the closest incorporated municipality farther down the railroad line. If there are no already-incorporated municipality farther down the line, I define the instrument as the logarithm of maximum value of such distance (154.25 km) plus the distance indicated on the panel title. In the baseline estimates shown on Table 3, I assign 154.25 km for these observations. Baseline controls are included instruments in all the panels. Robust standard errors in parentheses. Statistical significance (based on the Wald statistic) denoted by: * 10%, ** 5%, *** 1%

Table B.9: Placebo instrument

	Years as endpoint (1)	Log urban population (2)	Log urban GDP (3)	Log distance next mun. (4)	Log distance previous mun. (5)
Log distance to previous incorporated mun.	-0.293	0.063	0.089	0.063	0.013
	(0.244)	(0.103)	(0.123)	(0.041)	(0.045)
Observations	203	203	203	203	203

Note: Coefficients from OLS. Baseline controls are included in all the panels. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.10: Effects on the logarithm of municipality area

	(1)	(2)	(3)	(4)
Years as endpoint	0.063***	0.022*	0.024*	0.263***
	(0.013)	(0.013)	(0.013)	(0.071)
R^2	0.073	0.373	0.541	
Anderson-Rubin statistic				23.84
p-value of Anderson-Rubin test				0.000
Baseline controls?		Y	Y	Y
Geographic fundamentals?			Y	
Estimation method:	OLS	OLS	OLS	IV
Observations	203	203	203	203

Note: Coefficients from OLS in colums (1) to (3) and two-stage least squares in column (4). Baseline controls are included in all equations. Geographic fundamentals are mean and squared altitude, mean and squared terrain ruggedness, the logarithms of the potential yields of coffee, maize and sugarcane, the logarithm of days a year that support malaria transmission, soil dummies, a river dummy, and a quadratic polynomial of latitude and longitude. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

Table B.11: Effects on the Bolsonaro vote share

	First round vote		Second round vote	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Years as endpoint	0.001	0.004	0.001	0.002
	(0.001)	(0.005)	(0.001)	(0.005)
Observations	203	203	203	203

Note: Coefficients from ordinary or two-stage least squares; see the top of the table. Baseline controls are included in all the panels. Robust standard errors in parentheses. Statistical significance denoted by: * 10%, ** 5%, *** 1%

B.3 Data Appendix

Railroad data. The coordinates of each railroad station are from the shapefile by the *Agência Nacional e Transportes Terrestres* (ANTT) released with the 2017 *Declaração de Rede*. The shapefile contains information on the municipality of each station, the current operator, the station code, and the station name. I manually match each station by name with information from the website *Estações Ferroviárias do Brasil* (EFB); see http://www.estacoesferroviarias.com.br. For each station, I assign the year it started operating from EFB.[34]. I then define the year of railroad arrival as the minimum between the year of operation start of that station or of any other station that follows from it. All railroads in the sample were constructed from São Paulo towards the west, so this procedure indeed identifies the year of railroad arrival at each site. For each station, I identify the station that precedes it; for each station, one can go back and reach the São Paulo main station, *Estação da Luz*, after some iterations. I define the distance to the previous station as the Euclidean distance between the station and the previous station. The station distance to the city of São Paulo is the sum of all such distances until reaching the *Estação da Luz*.

For each municipality, railroad arrival year and distance to São Paulo refer to the first station in the municipality. As Figure 2 shows, almost all municipalities whose territory is crossed by a railroad have a station on it. Time as endpoint is the difference between railroad arrival year in that municipality and the lowest railroad arrival year in a next municipality. Distance to the next municipality is the difference between distance to São Paulo in that municipality and the lowest distance to São Paulo in a next municipality.

There are eight railroad companies in the sample: São Paulo Railway, Companhia Paulista de Estradas de Ferro (includes the Rio-Douradense), Estrada de Ferro Sorocabana (includes the Ituana), Estrada de Ferro São Paulo-Minas, Companhia Mogiana de Estradas de Ferro, Estrada de Ferro Araraquara, Estrada de Ferro Noroeste do Brasil, Estrada de Ferro São Paulo-Paraná. I include lines in the neighboring states of Paraná and Minas Gerais that were a continuation of the

[34] To assess the accuracy of these data, I confirmed the opening dates of some stations from EFB to opening dates from a statistical report on 1918 railroads; see Ministério da Viação e Obras Públicas (1924), *Estatística das estradas de ferro da União e das fiscalizadas pela União relativa ao ano de 1918,* Rio de Janeiro, Imprensa Nacional.

São Paulo rail network. I include only the railroad lines that are present in both the ANTT shapefile and *Estações Ferroviárias do Brasil*. As a consequence, I do not have the abandoned segment of the Noroeste after Araçatuba; this segment was replaced by an alternative route that I have in my sample. I also cannot trace back the evolution of the Mogiana in the state of Minas Gerais, nor of some smaller branches of the railroads. Finally, I do not include railroad segments built by FEPASA after the 1970s; the purpose of these segments was to connect some of these lines after they were all taken over by the government.

Sample selection. I include all municipalities along the sample railroads, with the exception of: **(a)** São Paulo, which is the origin of the network; **(b)** Bauru, the only municipality where there is a crossing of railroads (the Paulista, Sorocabana and Noroeste all go through the city); **(c)** the railroad endpoints today.

Population and GDP. Urban and rural populations are from the 2010 Brazilian census. Municipality GDP estimates are from IBGE.

Municipality coordinates indicate the centroid of the municipality territory.

Municipality incorporation year was manually collected at IBGE Cidades; see https://cidades.ibge.gov.br.

Agricultural suitability. Coffee, maize and sugarcane potential yields are from the Global Agro-ecological Zones (GAEZ) dataset from FAO. In all cases, I use the potential yield under rainfed, intermediate-input agriculture. I calculate averages over the municipality territory.

Digital elevation model. The altitude and ruggedness variable is created using as source the Japan Aerospace Exploration Agency (JAXA) ALOS Global Digital Surface Model. The terrain ruggedness index (TRI) is the square root of the average squared differences in altitude between each pixel and its eight surrounding pixels. The mean of the variables is taken over the municipality area.

Main rivers. The location of the main rivers is from the Brazilian *Agência Nacional das Águas* (ANA). I use version 1.3 of the *Base Hidrográfica Ottocodificada*. To identify the main rivers, I use the ANA map of the Paraná river basin and select the main rivers as all those highlighted in the map. In the sample area, these rivers are the Paraná, Ivaí, Paranapanema, Itararé, Tibaji, Itapetininga, Pardo, do Peixe, Tietê, Sorocaba, Grande, Turvo, Sapucaí, Piracicaba, and Moji-Guaçú.

Malaria. Malaria transmission suitability is from the Malaria Atlas. I use the mean between the days per year that could support *Plasmodium falciparum* transmission and the days per year that could support *Plasmodium vivax* transmission. Each of these is a 1 arcsec $\times$ 1 arcsec raster.

Soil. Soil information is from Embrapa. Soil variables include the share of a municipality that is covered by acrisols, by latosols, and by *terra roxa*, a type of red latosol that was sought for by coffee farmers. *Terra roxa* refers to red latosols, which is any soil with the following codes: LVdf1-11, LVef1-3, LVed1-23, and LVe1-2.

Historical GDP series. The estimates of Brazilian GDP are from Araújo *et al.* (2008).

1872 Census. The data is available at the municipality level from Cedeplar. Since many new municipalities emerged since 1872, I use the average minimum comparable areas of Reis *et al.* (2008). I use the following variables from the 1872 census: total population, slave population, population born abroad, literate population, number of free children aged 6-15 in school and out of school, population employed in different activities. When defining the non-agricultural employment share, I exclude the non-active population and the individuals occuppied in domestic labor from the calculations. Non-agriculture refers to any activity other than *lavradores* (farmers) and *criadores* (ranchers). The 1872 census is available at http://www.nphed.cedeplar.ufmg.br/pop-72-brasil/.

2018 elections. The Bolsonaro vote share is from electoral results released by the *Tribunal Superior Eleitoral* (TSE), the election authority of Brazil. The data is freely available at the TSE website; see https://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1.