Assessing the Impact of High-speed Rail on Population Growth: the case of Spain

Daniel Ruiz Palomo

February 24, 2025

PRELIMINARY AND INCOMPLETE

Abstract

Over the past three decades, Spain has built the second largest high-speed rail network in the world, massively reducing travel times across the country. Did the construction of the high-speed rail network contribute to the concentration of population and employment in the main metropolitan areas, or did it contribute to the relocation of both population and employment to outlying regions? This paper proposes an empirical methodology to answer this question. To address the endogeneity between growth and route allocation, the paper uses an inconsequential unit approach based on a hypothetical minimum construction cost network. Preliminary results suggest that more exposed regions experienced greater population growth.

1 Introduction

Over the last three decades, Spain invested a huge amount of resources in building one of the largest high-speed rail networks in the world. According to the International Union of Railways (UIC), the Spanish high-speed rail network is currently 3.661 km long, being the second longest network in the world. Moreover, there is a strong commitment to further extend it with more than a 1.000 kilometers under construction.

This articles studies the impact that the construction of the high-speed rail network had on population growth of Spanish municipalities between 2001 and 2021. To conduct my analysis, I derive an equation that relates population growth to the change in accessibility that a municipality experienced after the construction of the high-speed rail network, a variable that I call relative average time change. To measure it, I use the 2001 Spanish census to build a national commuting matrix. Additionally, I create GTFS files for all long-distance and regional train services operating in 2001 from hard-copy timetables to measure the changes in travel time from 2001 to 2021.

The major challenge in estimating my empirical equation is the likely correlation between the exposure to the high-speed rail network and other unobserved variables that also explain population growth. To address this concern, I rely on an empirical strategy based on the inconsequential unit approach. As in Faber 2014, I build a hypothetical minimum cost high-speed rail network connecting all provincial capitals in Spain taking into account exclusively the topography of the terrain, an unarguably exogenous feature. The distance to this hypothetical network is then used as an instrument for the relative average time change (RATC). This strategy seeks to identify the conveniently located municipalities due to exogenous features and use this source of exogenous variation to estimate the effect that my measure of exposure to the high-speed rail network has on population growth. The identifying assumption is that being close to the hypothetical minimum construction cost network has an effect on population growth only through the changes in travel times due to the construction of the high-speed rail network, conditional on a set of observable variables.

The results suggest that the construction of the high-speed rail network had a positive yet small effect on population growth. In my preferred specification, a one minute change in relative average time leads to a 3 percentage points increase in population growth over the 20 years of analysis.

This research contributes to the literature on the impact of transport infrastructure on the organization of economic activity within a country, summarized in Redding and Turner 2015.

This section serves as an introduction to the article. Section 2 describes the Spanish high-speed rail network and the available data. In section 3, I derive the empirical equation used to assess the impact that the construction of the high-speed rail network had on population growth. Section 4 discusses the empirical strategy used to estimate the empirical equation. Section 5 presents the estimation results and section 6 concludes.

2 Background and Data

2.1 Spain's High-speed Railway Network

The origins of the Spanish high-speed rail network can be traced back to the eighties, when the government decided to upgraded the railway connection between Andalusia, the southernmost region of the country, and Madrid. The project, initially thought as a new connection relying on existing infrastructure, progressively evolved to a complete new infrastructure constructed in international gauge that allowed the circulation of trains at 300 km/h.

In 1992, this new railway corridor was opened and it had a huge impact on the Spanish society. The AVE, the commercial name given to operate the new high-speed services, was perceived as a new mode of transport: as comfortable and convenient as a regular train and as quick as a plane[1]. Soon, having a high-speed station in a region became a symbol of modernity and prosperity and the political battle for pushing the government to plan and construct completely new lines started.

Figure 1: High-speed rail network as planned in the 2000 National Transport Infrastructure Plan

In 2000, the Spanish government published a national transport infrastructure plan, which was the seed for the currently existing high-speed rail network. The plan emphasized the connection between Madrid and *all* mainland Spanish province capitals, regardless of the construction cost or potential travel demand. It was best summarized by the Spanish prime minister José María Aznar when

[1] In fact, AVE in Spanish means bird, a name that tried to emphasize that this new train service was closer to a plane than to a regular train.

he said that "every province capital would be connected in less than four hours and a half to the centre of the Iberian Peninsula". Since then, every new version of the transport infrastructure plan has maintained this core principle while adding the unlikely construction of some transversal corridor.

The corridor connecting Madrid and Seville was followed by the construction of the corridors connecting Madrid and Barcelona in 2008 and Madrid and Valencia in 2010. The Great Recession, which in Spain was exacerbated by the following national debt crisis, slowed down the deployment of the network but new sections have progressively been inaugurated since then. Currently, the network is more than 3,500 km long, being the second largest high-speed rail network in the world only after China's, a country almost 20 times bigger.

2.2 Data

This section describes the data used in the estimation. Municipality level population data are taken from the 2001 and 2021 Spanish census. Additionally, the 2001 Spanish census has detailed information about the commuting patterns of the entire country since it asks to each respondent its municipality of residence and its municipality of work. I use this data to construct a commuting matrix between every municipality in 2001.

The map containing the geo-referenced Spanish municipalities was taken from the Spanish National Geographic Institute (IGC). Additionally, I created a map of centroids for each municipality, used to compute the travel time between them, using the latitude and longitud provided by the IGC. Finally, I also used the map of slopes, a 2m pixel grid that covers Spain containing information about the average slope in every pixel.

To compute travel times between municipalities in 2021, I use the GTFS for all long distance and regional services operated by RENFE, the Spanish national railway operator. These files are regularly updated by the Spanish Ministry of Transport and they can be found in multiple webpages[2]. For 2001, I created myself the GTFS for all long distance and regional services operated by RENFE. All the travel time calculations were made using Pereira et al. 2021, a free package in R that implements an efficient routing algorithm that allows modal choice. I used the road network downloaded from OSM as a mode for accessing the rail mode.

[2] In particular, I used https://nap.transportes.gob.es/

3 Empirical equation

This section derives the empirical equation used to test the impact that the construction of the high-speed rail network had on population growth of Spanish municipalities. This is done in two steps. In the first step, I derive a linear equation that relates population growth to changes in travel times. In the second step, I use a residential choice model based on a nested logit model to aggregate all changes in travel times into a single variable that I call relative average time change.

Consider a country divided into N municipalities indexed by $i = 1, \ldots, N$. Population of municipality i in year τ can be written as[3]

$$R_i^\tau = \pi_i^\tau R^\tau \tag{1}$$

where R^τ is total country population in year τ and π_i^τ is the proportion of citizens who live in municipality i in year τ.

The proportion of citizens living in a municipality can be seen as a non-linear function of variables such as amenities, productivity, land availability and travel times to other municipalities. With a little abuse of notation, this relationship can be written as

$$\pi_i^\tau = \pi_i(t_{is}^\tau, \ldots, t_{rs}^\tau) \quad \forall s, r \tag{2}$$

where t_{rs}^τ is the travel time between any two municipalities r and s in year τ. I purposely omit the dependence on other variables to simplify the notation and to clarify the exposition of the arguments to follow.

As argued in Baum-Snow and Ferreira 2015, a non-linear relationship like (2) can be approximated with a linear one through implicit differentiation with respect to time, measuring each explanatory variable in first-differences and using the partial derivatives of the function π_i as coefficients. In my case, this means that I can write the *change* in the proportion of residents living in a municipality as a linear function of the *difference* in travel time induced by the construction of the high-speed rail network.

Specifically, assume that between year τ and year $\tau + T$ there is an improvement in the transport infrastructure that induces a change in travel times between some municipalities. Using a first-order Taylor expansion around the proportion of residents who live in municipality i in year τ, I can find an equation for the proportion of residents who live in region i in year $\tau + T$ as

$$\ln \pi_i^{\tau+T} \approx \ln \pi_i^\tau + \sum_r \sum_s \frac{\partial \ln \pi_i}{\partial t_{rs}^\tau} \Delta t_{rs} \tag{3}$$

[3]Throughout the article, subscripts will denote spatial variables and superscripts will denote temporal variables.

where Δt_{rs} is the change in travel time between municipalities r and s from year τ to year $\tau + T$.

Combining equations (1) and (3), I get a linear equation that relates local population growth to changes in travel time, national population growth and unobserved variables

$$\ln R_i^{\tau+T} - \ln R_i^{\tau} = \ln R^{\tau+T} - \ln R^{\tau} + \sum_r \sum_s \frac{\partial \ln \pi_i}{\partial t_{rs}^{\tau}} \Delta t_{rs} + +u_i \tag{4}$$

In equation (4), the term Δt_{rs} is an explanatory variable while the term $\frac{\partial \ln \pi_i(\tau)}{\partial t_{rs}^{\tau}}$ is the *heterogeneous* effect that the change in travel time between municipalities r and s has on population growth in municipality i.

While this equation can in principle be brought to the data, there are at least two issues that must be addressed before. On the one hand, it might be the case that there are more improved origin-destinations pairs than municipalities, meaning that there are more explanatory variables than observations. On the other hand, even in the case when there are enough observations, the effect of the same explanatory variable may have opposite signs on different municipalities.

To see this, consider a country with three municipalities where only one origin-destination pair improves its travel time after the construction of a high-speed railway line. Due to this, the newly connected municipalities have better accessibility with respect to the remaining one. Some residents will thus migrate there and the share of citizens living in the benefited regions will increase out of the share of citizens living in the remaining one. In this case, there is a reshuffle of population because of a change in relative accessibility and the same explanatory variable will thus induce opposite effects on different municipalities.

A solution to these problems is to theoretically model the residential choice of residents and its dependence with travel time to other regions. In this way, the sum of heterogeneous effects in equation (4) can be transformed into a single homogeneous effect that captures the total impact that the construction of the high-speed rail network has on population growth. The key idea will be to model residential choice using a nested logit model. In the upper layer, residents will choose their residence municipality while in the lower layer residents will choose their workplace municipality taking into account the time it takes to commute from their residence. This will allow me to have a closed-form equation for the residential choice probabilities as well as for its derivative with respect to travel times.

Let me assume that (indirect) utility of a citizen w who lives in i and travels to j has the following additive form

$$U_{ij}(w) = V_i + V_{ij} + \varepsilon_i(w) + \varepsilon_{ij}(w)$$

where V_i is a common component of utility that depends only on features of i, V_{ij} is a common component of utility that depends on features of both i and j,

$\varepsilon_i(w)$ is an idiosyncratic components of utility that depends only on features of i and $\varepsilon_{ij}(w)$ is an idiosyncratic components of utility that depends on features of both i and j.

The term V_{ij} captures the disutility of travelling from i to j due to monetary costs or travel time as well as features of the destination. For simplicity, I will assume that the marginal disutility of travel time is constant and homogeneous which means that $\frac{\partial V_{ij}}{\partial t_{ij}} = -\beta$. This assumption is not innocuous since it imposes two important restrictions. On the one hand, it assumes that every extra minute of travel time creates the same disutility. On the other hand, it assumes that the disutility is homogeneous throughout all possible origin and destinations pairs. While the first restriction can be easily relaxed imposing a more sophisticated functional form [4], the second restriction turns out to be crucial for the derivation of my empirical equation.

As derived in the appendix, the proportion of residents who choose to live in region i is

$$\pi_i = P[\max_j U_{ij} > \max_j U_{rj}, \forall r \neq i] = \frac{e^{\theta V_i} \mathbb{A}_i^{\theta/\mu_i}}{\sum_r e^{\theta V_r} \mathbb{A}_r^{\theta/\mu_r}} \tag{5}$$

where the term $\mathbb{A}_i$ is the inclusive value of nest i, whose functional form is

$$\mathbb{A}_i = \sum_j e^{\mu_i V_{ij}}$$

In this context, the inclusive value can be interpreted as a measure of accessibility since it adds all the disutilities of travelling from i to any destination s. Notice that when the sum of all travelling disutilities is small (large), the accessibility will be high (low).

Equation (5) states that the relative population living in i depends both on its local features (captured by V_i) and accessibility (captured by $\mathbb{A}_i$) relative to the local features and accessibility of other municipalities. Intuitively, a high proportion of residents will choose to live in i when living there is good *compared to* living in other municipalities. As I will discuss next, the construction of the high-speed rail network has an effect on population growth precisely through the change in the relative accessibility between municipalities.

The next step consists in computing the impact that a change in travel time has on the relative population living in a municipality i. Firstly, consider a change in travel time that directly affects the municipality

$$\frac{\partial \ln \pi_i}{\partial t_{is}^{\tau}} = \frac{\theta}{\mu_i} \frac{(1-\pi_i)}{\mathbb{A}_i} \frac{\partial \mathbb{A}_i}{\partial t_{is}^{\tau}} = -\theta\beta\pi_{s|i} + \theta\beta\pi_{is} < 0 \tag{6}$$

[4] A possible different specification for the disutility of travel time is $V_{ij} = -t_{ij}^{\alpha}$. The convexity of this function guarantees that each extra minute of travel time creates more disutility than the previous one.

where π_{is} is the percentage of residents commuting i to s in year τ and $\pi_{s|i}$ is the percentage of residents commuting to s conditional on living in i.

Next, consider a change in travel time between any two other municipalities

$$\frac{\partial \ln \pi_i}{\partial t^{\tau}_{rs}} = -\frac{\theta}{\mu_r}\frac{\pi_r}{\mathbb{A}_r}\frac{\partial \mathbb{A}_r}{\partial t^{\tau}_{rs}} = \theta\beta\pi_{rs} > 0 \tag{7}$$

where π_{rs} is the percentage of residents commuting r to s in year τ.

As equation (6) and (7) state, a change in travel time that directly benefits a municipality i will attract population while a change in travel time that benefits other municipalities will expel population from municipality i. The overall effect of the construction of the high-speed rail network on population growth will thus be the sum of these individual effects

$$\sum_r \sum_s \frac{\partial \ln \pi_i}{\partial t_{rs}} \Delta t_{rs} = -\theta\beta \left(\sum_s \pi_{s|i}\Delta t_{is} - \sum_r \pi_r \sum_s \pi_{s|r}\Delta t_{rs} \right) \tag{8}$$

whose value will be positive or negative depending on what effect dominates.

The term in the right-hand side of equation (8) can be interpreted as a variable that measures the relative average time change (RATC) in each municipality due to the construction of the high-speed rail network. While the first term is a weighted average of changes of travel times from municipality i to any other municipality the second term is a weighted average of the average of changes of travel times. Hence, its difference captures the average time change of a municipality i with respect to the national average time change.

To gain intuition about what the variable RATC captures, it is interesting to consider the simple case of two municipalities $i = \{1, 2\}$. In this case, the value of RATC for the first municipality simplifies to

$$RATC_1 = -(1 - \pi_1)\Delta t_{12}(\pi_{2|1} - \pi_{1|2})$$

which implies that the sign of the effect on population growth depends exclusively on the relative proportion of resident who commute between the municipalities.

When $\pi_{2|1} > \pi_{1|2}$, the value for $RATC_1$ is positive[5] and population in municipality 1 increases at the expenses of municipality 2. Although the travel time change is the same for both municipalities, it is relatively more beneficial for residents of municipality 1 since a greater share of them uses the improved connection. For the same reason, when $\pi_{1|2} > \pi_{2|1}$, residents from municipality 2 are comparatively more benefited and population in municipality 2 grows at the expense of municipality 1. In the particular case that $\pi_{2|1} = \pi_{1|2}$, there is no population growth in any municipality even though there is a change in travel time. The reason is that the residents of both municipalities are equally affected and the relative average travel time does not change.

[5] I consider that travel time between 1 and 2 decreases which means that $\Delta t_{12} < 0$

Combining equations (4) and (8), I get the equation

$$g_i = g + \theta\beta RATC_i + u_i \tag{9}$$

where I define $g_i = \ln R_i^{\tau+T} - \ln R_i^{\tau}$, $g = \ln R^{\tau+T} - \ln R^{\tau}$ and $RATC_i = -\left(\sum_s \pi_{s|i}\Delta t_{is} - \sum_r \pi_r \sum_s \pi_{s|r}\Delta t_{rs}\right)$.

Equation (9) is the empirical equation used to assess the impact that the construction of high-speed rail had on population growth. It is a a simple linear equation with a single explanatory variable whose value can be measured with the data I describe in Section 2 and it states that municipalities that experienced a higher relative average time change should grow more, conditional on national population growth. Next section explains the empirical strategy to estimate the value of coefficient $\theta\beta$ while section 5 reports the results for such estimation.

4 Empirical Strategy

The theoretical framework explained in the previous section is used to estimate the effect that the construction of the high-speed rail network had on population growth of Spanish municipalities between 2001 and 2021. The baseline specification is a linear equation of the form

$$g_i = \alpha + \beta RATC_i + \gamma X_i + u_i \tag{10}$$

where g_i is population growth of municipality i between 2001 and 2021, $RATC_i$ is the relative average time saved in municipality i between 2001 and 2021 and X_i is a vector of exogenous control variables described later in the text.

As explained in the previous section, the value of $RATC_i$ is a weighted average of the travel times changes induced by the construction of the high-speed rail network

$$RATC_i = -\left(\sum_s \pi_{s|i}^{2001}(t_{ij}^{2021} - t_{ij}^{2001}) - \sum_r \pi_r^{2001} \sum_s \pi_{s|r}^{2001}(t_{ij}^{2021} - t_{ij}^{2001})\right)$$

where the weights $\pi_{s|i}^{2001}$ is the percentage of residents living in municipality i who were commuting to municipality s in 2001 and $\pi_{s|i}^{2001}$ is the percentage of residents living in municipality i.

Table 1 summarizes the values for the relative average time saved (RATC). Approximately half of Spanish municipalities experienced a positive RATC between 2001 and 2021, a fact in line with the great extension of the high-speed rail network. Despite this, the largest gains are concentrated in a minority of municipalities since less than 10% of them actually experienced an value larger than 5 minutes. Regarding the negative values, most of the municipalities experienced values between 0 and -1 min while only less than 1% of them experienced a value smaller than -10 min.

Figure 2 shows that the spatial distribution of the RATC is not homogeneous. The positive vales are concentrated around the northwestern regions, the areas around Calatayud, Cuenca and Granada, in the south of Spain. All these areas have in common the fact that a high-speed station was built and a sizeable share of their population was commuting out of the region in 2001. The negative values are concentrated around the mountainous regions in the north, Extremadura and the area North of Madrid. These regions did not receive rail transport improvements over the period of analysis and, in fact, some of them lost some train services.

The OLS estimator can be used to obtain an unbiased estimate for β provided that ATS and the unobserved variables that also explain population growth are uncorrelated. Unfortunately, there are reasons to believe that this is not the case since the Spanish government systematically prioritized the connection of the provincial capitals to the high-speed rail network. These municipalities

	RATC (min)
Observations	8104
Mean	0.66
Standard deviation	5.06
Minimum	-140.42
1st percentile	-8.19
10th percentile	-1.28
25th percentile	-0.85
50th percentile	-0.40
75th percentile	1.15
90th percentile	4.19
99th percentile	17.80
Maximum	92.16

Table 1: Summary statistics for the relative average time change (RATC)

and their neighbors are economically different than the rest of the country and they might be affected by other unobserved variables that also affect population growth, confounding the effect of my explanatory variable.

As an alternative, I use an empirical strategy based on the inconsequential unit approach (Redding and Turner 2015). While some municipalities were indeed targeted to be part of the network due to its population, economic output or political salience, some others experimented a change in travel times because they happen to be conveniently located.

To leverage this fact, I follow Faber 2014 and I construct a hypothetical minimum construction cost (HMCC) high-speed rail network connecting all provincial capitals in Spain taking into account exclusively the topography of the terrain, an unarguably exogenous variable (figure 3). The distance from each municipality to this hypothetical network is then used as an instrument for the average time saved.

The hypothetical minimum construction cost high-speed rail network is constructed in two-steps. Firstly, I identify the minimum construction cost paths connecting any pair of provincial capitals. To do this, I take into account the fact that high-speed rail construction costs heavily depend on the slope of the terrain it crosses. While flat areas only require the construction of the track and catenary, more steep areas also require the construction of specific infrastructure such as viaducts or tunnels, which makes the cost per kilometer significantly higher [a reference is needed here]. Using historical construction costs for the Spanish network, I estimate that the cost of building in flat areas (those with a slope lower than 3%) is 9 M€ /km while the cost of building is more steep areas is 39 M€ /km[reference is needed]. In the second step, I select the minimum number of paths that connects all province capitals using Kruskal's minimum spanning tree algorithm. The algorithm selects among all possible paths the smallest number of them that guarantees that every node is connected to the network.

The estimation of β using as an instrument the distance to the hypothetical minimum construction cost network requires the instrument to be uncorrelated with the unobserved variables that also explain population growth. As in Faber 2014, there are at least two threads that may compromise this assumption.

Since the hypothetical network is built choosing the paths with the minimum possible slope, which are generally river valleys or mountains passes, the instrument might be picking features of this geographical areas that historically attracted population for other reasons[6]. In other words, the municipalities closer to the hypothetical network could be places more attractive to live precisely because they are conveniently located. To address this concern, I add as controls a set of variables that reflect the existing economic conditions before the construction of the high-speed rail network.

Additionally, by construction of the instrument, municipalities closer to the provincial capitals are systematically closer to the hypothetical minimum construction cost network. A thread to the exclusion restriction arises if being close to the provincial capitals has also a direct effect on population growth. To overcome this problem, I include as control the log-distance between a municipality and the closest provincial capital.

In conclusion, the identifying assumption of this paper is that being close to the hypothetical minimum construction cost network only affects population growth through its impact on average time saved conditional on the log-distance to the closest provincial capital and pre-existing economic conditions before the construction of the high-speed rail network.

[6]For instance, a river valley is an attractive place to live due to its water and land availability or due to its endowment of other transport infrastructure.

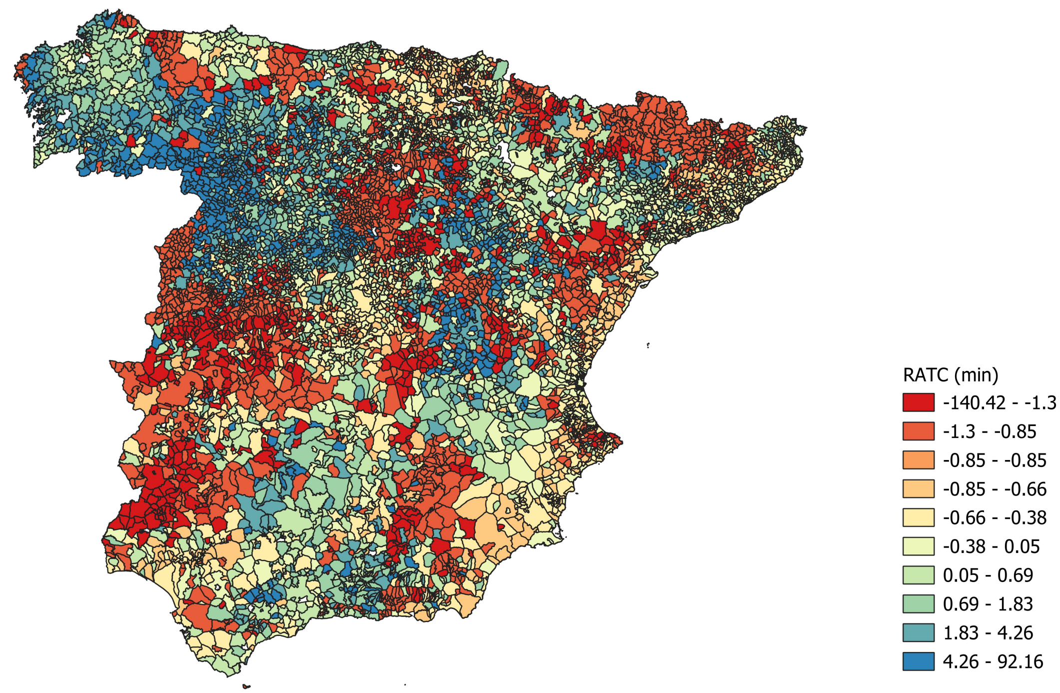

Figure 2: Spatial distribution of the relative average time saved (RATC)

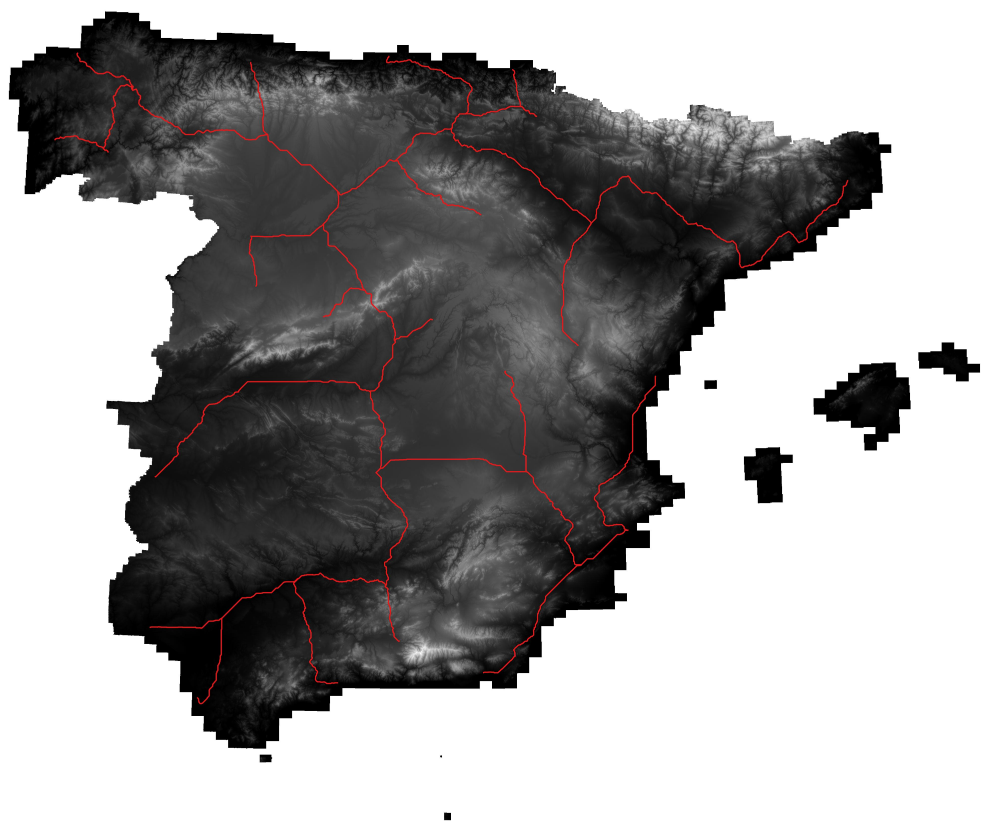

Figure 3: Hypothetical minimum cost network connecting all provincial capitals.

5 Estimation Results

This section reports estimation results for equation (10). Table 2 reports the first stage result for the distance to the minimum construction cost network with and without additional controls. The distance to the minimum construction cost network remains statistically significant within a province conditional on log distance to the nearest provincial capital and the log population in 2001. The relative average time change is more likely to be higher for municipalities with higher distance to the provincial capital in smaller pre-existing municipalities.

	(1) RATC	(2) RATC
Distance to MCCN	-0.0158*** (0.0023)	-0.0205*** (0.0025)
log Population 2001		-0.0724* (0.0410)
log Distance to Capital		0.2619*** (0.0701)
Province FE	Yes	Yes
Observations	8057	8057
R	0.2086	0.2094
First stage F-Stat	46.166	70.455

Table 2: First stage regressions

Table 3 presents OLS and IV results regressing population growth between 2001 and 2021 on the relative average time change before and after including the log distance to the nearest provincial capital and the log population in 2001. The IV estimate of the effect of the relative time change in population growth is positive and statistically significant.

The OLS estimates are negative and statistically different than zero, regardless of the inclusion of control variables. The difference between these estimates and the IV estimates suggests a negative correlation between my explanatory variable and unobservable features that also affected population growth over the period of analysis.

The inclusion of controls in the IV estimation leads to smaller effects. As argued in the empirical strategy, there is a reasons to believe that the exclusion restriction suggesting a positive correlation between my instrument and unobserved variable that also affect population growth.

	(1) OLS No controls	(2) OLS With Controls	(3) IV No controls	(4) IV With controls
Relative Average Time Saved (min)	-0.0014*	-0.0020***	0.2626***	0.0307***
	(0.001)	(0.001)	(0.0403)	(0.0101)
log Population 2001		0.0433***		0.0544***
		(0.003)		(0.0029)
log Distance to Capital		-0.1930***		-0.1655***
		(0.007)		(0.0078)
Province FE	Yes	Yes	Yes	Yes
Observations	8057	8057	8057	8057
R	0.325	0.463		

Table 3: Effect of RATC on population growth between 2001 and 2021

6 Conclusions

The construction of large-scale transport infrastructure has general equilibrium effect other than the reduction in travel times. The changes in travel times that these projects induce create new opportunities for residents and firms to relocate affecting the spatial distribution of economic activity. To evaluate such changes, this article explores that impact that the construction of the Spanish high-speed rail network had on population growth at the municipality level over the 2001 to 2021 period.

Using a reduced-form equation derived from an equilibrium condition of a theoretical model, this article shows that the construction of the high-speed rail network contributed to the population growth of the relatively more exposed municipalities. In particular, a relative average time changed induced to the construction of the high-speed rail network contributed to an increase of 3 percentage points in population growth over the 2001 to 2021 period.

References

Baum-Snow, N. and Ferreira, F. (2015). "Causal inference in urban and regional economics". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 3–68.

Faber, B. (2014). "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System". In: *Review of Economic Studies* 81.3, pp. 1046–1070.

Pereira, R. H. M. et al. (2021). "r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R". In: *Findings*. DOI: 10.32866/001c.21262. URL: https://doi.org/10.32866/001c.21262.

Redding, S. J. and Turner, M. A. (2015). "Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity". In: *Handbook of Regional and Urban Economics*. Ed. by G. Duranton, J. V. Henderson, and W. C. Strange. Vol. 5. Handbook of Regional and Urban Economics. Elsevier, pp. 1339–1398. DOI: https://doi.org/10.1016/B978-0-444-59531-7.00020-X. URL: https://www.sciencedirect.com/science/article/pii/B978044459531700020X.