Quantifying the causal effects of high speed rail: Evidence from the Madrid-Barcelona Corridor

Jose M. Carbo*, Daniel J. Graham, Anupriya, Daniel Casas Bofarull

Railway and Transport Strategy Centre
Department of Civil Engineering
Imperial College London

Extended abstract submitted for
hEART 2018
21st February, 2018

---

1. Background and Research Objective

Transport investments play a key role in the economy by providing direct benefits of the form of travel time savings for passengers and freight, as well as inducing wider benefits via agglomeration economies (Graham, 2007). Given the scale of investments, quantification of economic impacts is crucial to understand the returns to substantial public funds. High-speed rail (HSR) represents a very large scale investment. Modeling the causal impact of HSR on economy is empirically challenging because (1) allocation of HSR infrastructure is endogenous to localized characteristics such as economic performance, and (2) the link between such investments and productivity is simultaneously determined by other confounding factors. The literature estimating this causal effect is limited.

In this study, we contribute to the growing literature in this field by estimating the causal impact of HSR on Spain's regional economy, focusing on the effect of the Madrid-Barcelona corridor, which covers 621 km between the two cities and was built between 2003 and 2008. The HSR network in Spain is the largest in Europe and the second largest in the world after China. But its economic impact remains understudied. The objective of this paper is to use causal econometric methods to fill this gap in the literature. We use the plausibly random treatment received by the intermediate cities on the line: Lleida and Tarragona.

---

*Corresponding author. Email address:j.carbo-martinez@imperial.ac.uk

2. Data and Methodology

We exploit quasi-experimental conditions to study the causal effect of HSR on Spain's regional productivity. We adopt two different methodologies. First, we use a difference-in-differences (DID) estimation that provides a valid framework to control for temporal and confounding biases, by making comparisons between control and treated groups in the pre-treatment and post-treatment period. Second, we extend our research with a synthetic control method analysis of the impact of HSR on Lleida and Tarragona, two provinces with stops along the Madrid-Barcelona HSR line. Synthetic control methods have been used in comparative case studies since the work of Abadie and Gardeazabal (2003). A key assumption for establishing casual inference from synthetic control methods is that the treatment is exogenous to the variable of interest. We use the fact that Lleida and Tarragona received treatment (i.e. got HSR service) due to their geographical location rather than their economic performance.

Our response variables of interest are: (1) GVA, (2) GVA per employee, (3) number of companies registered, and (4) employment rate. We use an extensive panel dataset comprising economic information on all Spanish provinces for the years 1992-2015, obtained from the INE, Spain website [1]. We also add a set of covariates [2] to control for variation in the impact of HSR.

2.1. Difference-in-Difference Method

To quantify the impact of HSR, our estimate of interest is the Average Treatment Effect (ATE) defined as the difference in mean response under treated and control status:

$$\tau = E[Y_{it}(1)] - E[Y_{it}(0)] \tag{1}$$

where $Y_{it}(1)$ and $Y_{it}(0)$ are responses under treatment and control status respectively.

We estimate this ATE using the following difference-in-difference specification:

$$\begin{aligned}\ln(Y_{it}) = \beta_0 + \tau(DID) + \sum_{n}^{N} \beta_n X_{it} + \sum_{m}^{Y} \beta_m \gamma_t \\ + \sum_{j}^{G} \beta_j \alpha_i + u_{it}\end{aligned} \tag{2}$$

where, β_is are parameters to be estimated and $X_i t$ is the set of covariates mentioned previously. α_is and γ_ts are region and year specific dummy variables that capture regional heterogeneity and yearly shocks respectively. The

[1] INE, Spain: http://www.ine.es/

[2] Covariates: gross capital formation over GVA (named as investment), average years of education, average compensation, and the weight in GDP of the agricultural, industrial, construction and services sectors

term $DID \epsilon \{0,1\}$ equals 1 for treated regions in post-treatment stage, and 0 otherwise.

For each response variable, non-treated Spanish provinces which satisfy the parallel trend assumption are chosen as control group.

2.2. Synthetic Control Method

We use the same procedure as in Abadie and Gardeazabal (2003) and Abadie et al. (2015). The objective is to create a synthetic control region as a weighted average of the available control regions. We built the synthetic region so that it resembles relevant economic characteristics of the treated regions before receiving treatment[3][4]. These control regions are Spanish provinces that do not have access to HSR stations during the period 1992-2015. Once we have created the synthetic control region, we compare its economic evolution to the one of the treated region.

Let J be the number of available control regions and let K be the number of economic characteristics. We are interested in assigning a weight to each control region. These weights are chosen so that the synthetic region better approximates the economic characteristics of the treated region. The optimal $Jx1$ vector of weights $W = (w_1, .., w_J)$ is obtained from the following minimization problem:

$$\underset{W}{\text{Minimize}} \quad (X_1 - X_0 W)' V (X_1 - X_0 W) \qquad (1)$$

such that

$$\sum_{i=1}^{J} w_i = 1$$

where X_1 is a $Kx1$ matrix containing all the economic characteristics of the treated region and X_0 is the KxJ vector containing the economic characteristics of the control regions. The solution to this problem is a vector of optimal weights W*(V), which should add up to one. This optimal vector depends on V, a KxK diagonal matrix that assigns weights to each economic characteristic.

We choose the matrix V so that the variable of interest in the treated region is best replicated by its synthetic counterpart before treatment. We focus on two variables of interest: Number of firms and GVA per employee. Let T_p be the number of years observed previous to the treatment. The problem of choosing V can be specified as follows:

$$\underset{V}{\text{Minimize}} \quad (Z_1 - Z_0 \mathrm{W}^*(\mathrm{V}))'(Z_1 - Z_0 \mathrm{W}^*(\mathrm{V})) \qquad (2)$$

[3]Among the economic characteristics, we consider we consider the set of covariates mentioned previously.

[4]Lleida's HSR station was built in 2003 and Tarragona's in 2007

where Z_1 is a $T_p x1$ vector with a time series of the variable of interest in the treated region, and Z_0 is a $T_p xJ$ vector with a time series of the variable of interest in the synthetic control region

Once we have built our synthetic control region, we compare its evolution with respect to the treated region during the years following the treatment.

3. Results and further work

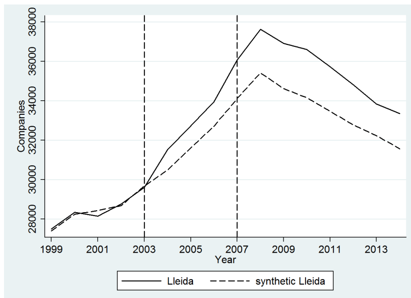
The key results that emerge from our analysis are as follows:

1. From our DID analysis, we find statistically significant effects from HSR investment (at the 95% level) on labour productivity measured as GVA per employee, which increases by 1.084% in the treated provinces along the HSR line relative to the control group. This could be attributed to expansion in GVA which shows a statistically significant increase of 2.272%. There is no statistically significant impact on employment. The number of companies are found to increase by 3.292%. When Barcelona and Madrid are excluded from the analysis, we get a 2.061% increase in labour productivity i.e. GVA per employee. We again find an expansion in GVA of 2.409% but no statistically significant impact on employment. The number of companies are also found to increase by 2.808%.

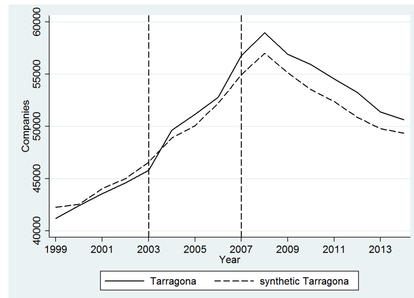
2. Figure 1a shows the evolution of number of firms in synthetic Lleida and Lleida. Two red vertical dotted lines mark year 2003 and year 2007, the years in which Lleida and Tarragona got treatment. Synthetic Lleida provides an excellent fit for number of firms in Lleida prior to 2003. Immediately after 2003, the number of firms in Lleida grows faster than in its synthetic counterpart. This divergence gets bigger after 2007, when the HSR line was enhanced. We find that, by 2009, the number of firms in Lleida is 7 percent higher than its synthetic control region. The impact of HSR on the number of firms in Tarragona is similar, figure 1b. Synthetic Tarragona also provides a fine fit for the number of firms in Tarragona before 2003. After 2003, number of firms in Tarragona starts to diverge from its synthetic region, although the HSR service will not be available in Tarragona until the end of 2006. This might indicate that there was an anticipation effect, with firms moving or being created in the province before the station was opened. The number of firms in Tarragona was 5 percent higher than its synthetic control region by 2010.

3. This presents evidence that the opening of the HSR line created incentives for new firms in the region. We find that this lead to an increase in the economic activity in the areas, since the effects of high speed rail are also significant in terms of GVA per employee, figures 2a and 2b. To test the robustness of our results, we will check the significance of our

findings applying the same method to all provinces. We will also evaluate to what extent our results are driven by particular weighted provinces.

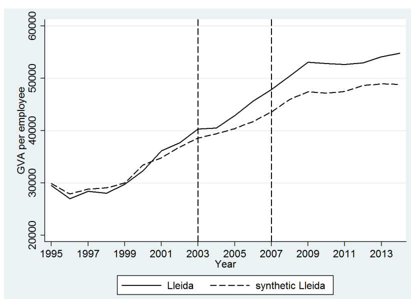


(a) Synthetic Lleida and Lleida

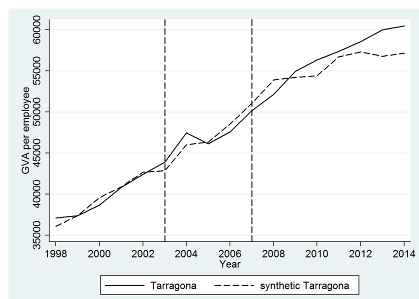


(b) Synthetic Tarragona and Tarragona

Figure 1: Number of companies



(a) Synthetic Lleida and Lleida



(b) Synthetic Tarragona and Tarragona

Figure 2: GVA per employee

References

Abadie, A., Diamond, A. and Hainmueller, J. (2015), ‘Comparative politics and the synthetic control method’, *American Journal of Political Science* **59**(2), 495–510.

Abadie, A. and Gardeazabal, J. (2003), ‘The economic costs of conflict: A case study of the basque country’, *The American Economic Review* **93**(1), 113–132.

Graham, D. J. (2007), ‘Agglomeration, productivity and transport investment’, *Journal of transport economics and policy (JTEP)* **41**(3), 317–343.