*Extended abtract*

# EXTENDED ABSTRACT

**Title: Evaluating the impact of HSR stations on the creation of firms**

**Authors and e-mail of all: Anna Matas (anna.matas@uab.cat), Josep Lluis Raymond (josep.raymond@uab.cat) and Josep Lluis Roig (josepLluis.roig@uab.cat)**

**Department: Economia Aplicada**

**University: Universitat Autonoma de Barcelona**

**Subject area:** *(please, indicate the subject area which corresponds to the paper)*
**Localización de la actividad económica y análisis cluster**

**Abstract:** *(minimum1500 words)*

## 1. Introduction

The High Speed Rail (HSR) is a fast and comfortable mode of transport that has known a significant implementation in some countries. In the European Union, Spain leads the group of countries with the highest investment in high-speed as France, Germany and  Italy. With 2,531 kilometres and 37 stations, Spain has the most extensive network of HSR in Europe and the second in the world with. Until 2018, investment in HSR amounted to 51,775 million euros, 21% of which have been founded by the European Union. Following the network extension, demand has risen from 1.29 billion of passenger-km in 1995 to 11.31 billion in 2018. Figure 1 shows the high-speed lines and stations at the end of 2018.

Figure 1. High-speed network. 2018


In contrast with the network size, the rate of utilization is far below the European levels. In Spain, HSR carries 4.9 million of passenger-km per kilometre of network, whereas the equivalent figures for France and Germany are 23.1 and 17.1, respectively. The low levels of utilization can be explained by the fact that decisions to build the high-speed rail network have been based more on political grounds than on economic criteria, the underlying idea being that the connexion to the high-speed network will have a positive economic effect on the region affected. Additionally, equity reasons have been put forward to justify the investment. In this sense, the commitment of the different governments has been that nine out of ten citizens should have a HSR station no further than 30 kilometres away. So, investment continues and currently there are 904 kilometres under construction.

The advantages of high-speed in terms of travel time and comfort have its counterpart on the high investment and maintenance costs. Given the opportunity costs of public funds, it is essential to quantify the expected benefits of investing in the high-speed network.

Taking advantage of the investment policy followed in Spain, this paper evaluates the impacts of the opening of a high-speed rail station on the level of economic activity of the geographical area affected. Economic activity is measured as the number of new firms created in a given year[1]. Since the HSR only provides passenger services we expect a higher effect on those sectors that do not depend on freight traffic. In order to account for this fact, we examine four categories of activities: service sector, tourism-related activities, knowledge-intensive activities and the manufacturing sector. Our hypothesis is that the impact will be higher on the first three groups. We estimate different coefficients for each high-speed station.

The econometric strategy consists on estimating a fixed effect model using panel data. Using panel data makes it possible to account for individual and temporal specific effects and, at the same time, testing for the existence of a dynamic structure in the data. We use panel data at provincial level over the

---

[1] Additionally, we analysed the impacts on changes in residential location through two variables: changes in population and changes in house prices. However, we didn't find any significant effect for any of the two variables.

1996-2017 period. Overall, the estimation results show that, the effect of a new rail station was relatively small, if existing at all. An additional contribution of our paper consists on developing a methodological procedure to improve the accuracy of the estimation of small but positive effects.

The rest of the paper is organised as follows. In section 2 we briefly describe the related literature, and in section 3 we show the data. Section 4 outlines the econometric strategy and section 5 offers the estimation results. In section 6 we compute the distribution functions for the estimated coefficients. Finally, section 7 provides the conclusions.

## 2. Related literature

Several papers have quantified the economic impacts of HSR from different perspectives, but a consensus has not been reached yet. Whereas some papers find a positive effect, others show no effect at all (see Albalate and Bel, 2012 for a review).

## 3. Data

We use data from the Central Business Register (CBR) compiled by the National Institute of Statistics (INE). Specifically, we evaluate the impact of HSR on the number of new registered companies at a province level (NUTS3) with a time span ranging from 1995 to 2017.

We exclude those provinces with HSR before 1995 (line Madrid-Sevilla, 4 provinces) and distinguish between four different sectors of activity. Table A.1. shows all the HSR included in the sample and the corresponding opening year: 7 lines and 23 new stations. The sample is composed by 20 provinces with stations and 23 without.

The basic statistics for the number of created new firms are shown in Table 1. We want to point out at two facts:

a. The number of new firms is much larger for service sector, and particularly lower for manufacturing
b. There is a high level of variability among provinces

**Table 1. Descriptive statistics for the dependent variable. Number of new firms**

| | Observations | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Services | 989 | 4489 | 6390 | 296 | 54200 |
| Tourism | 989 | 901 | 1041 | 74 | 8476 |
| KIA | 989 | 1408 | 2312 | 79 | 19239 |
| Manufacturing | 989 | 293 | 481 | 9 | 5839 |

## 4. Econometric strategy

In order to evaluate the impact of HSR we exploit the specific features of the panel data methodology. First, we specify a dynamic equation to allow for temporal adjustments in the data. Second, we control for unobserved individual and temporal effects by including the corresponding dummy variables. Third, we include additional regressors to account for individual specific characteristics that vary over time. Fourth, we include a time specific dummy variable for those provinces that have not benefit from high-speed investment[2]. This way, we are able to account for potential differences in time trends between provinces affected and not affected by HSR. Finally, in order to estimate the impact of HST we create a dummy variable that takes value 1 if the province is connected to the high-speed network and 0 otherwise. We tried up to two lags and leads of this dummy variable to test both for lagged and anticipated changes.

[2] Ideally, we should include a specific time coefficient for each province. However, this would imply estimating 903 additional coefficients (43 provinces * 21 years) rendering the estimation unfeasible.

The initial estimated equation takes the following form:

$$Y_{it} = \gamma Y_{it-1} + X'_{it}\beta + \sum_{n=-m}^{q} \theta \quad F_{it-n} + \alpha_i + \lambda_t + \delta_t D_{it} + \varepsilon_{it} \qquad (1)$$

where: $Y_{it}$ is the number of firms created in province *i* and year *t* and sector

$F_{it}$ is a binary variable that takes value 1 if HSR is available in province *i* and year *t* and 0 otherwise, with m leads and q lags.

$X_{it}$ includes the explanatory variables

$\alpha_i$ are the individual fixed effects

$\lambda_t$ are the temporal fixed effects

$D_{it}$ are the temporal fixed effects referred to those provinces without HSR

This equation was estimated for each of the four considered sectors of activity.

Starting from this general model we carried out a simplification strategy based on the statistical significance of the estimated coefficients. The initial specification includes two lags and two leads for the policy variable to be evaluated. Given that there are 20 provinces with HSR across the sample, that would imply estimating 100 coefficients, plus the rest of the variables in the equation. Data availability prevented such an approximation. As an alternative, we carried out the simplification strategy assuming that the impact of the HSR was the same for all the rail stations. In the final specification we allow for different coefficients.

All equations are estimated using Generalised Least Squares (GLS) and allowing for heteroscedasticity across panels and autocorrelation within the panel.

The initial estimations showed that, for all sectors, the coefficient of the lagged dependent variable took a low value and was not statistically significant except for the industrial sector. Given these results and the problems of correlation between the lagged dependent variable and the error term in a dynamic equation with fixed effects, we excluded the lagged variable and estimated the static equation. The long-term elasticities of the policy variable of interest and the standard errors of the estimated coefficients showed to be very similar in the dynamic and the static equations. Since our interest lies in long term elasticities, we selected the static equation and rely on the lagged values of the policy variable to capture the temporal adjustment. In this regard, we tried from zero to two lags but the results showed that only the first lag for the opening of a HSR station had a significant impact on the creation of new firms. We also tested for anticipated changes including two leads. None of the lead variables was statistically significant.

Hence, the equation finally estimated is:

$$Y_{it} = X_{it}'\beta + \theta_i F_{it-1} + \alpha_i + \lambda_t + \delta_t D_{it} + \varepsilon_{it} \qquad (2)$$

Finally, it has to be said that when estimating the impact of a new infrastructure we have to deal with a potential endogeneity problem provoked by reverse causation. In our case, endogeneity may not be a severe problem since decisions on new railway lines are taken more on political than on economic grounds. Besides, working with panel data makes it possible to address the potential correlation between the regressors and the individual fixed effects. The underlying idea is that any idiosyncratic changes in the number of new firms will be absorbed by the fixed effect.

## 5. Estimation results

Table 2 provides the estimation results of equation (2) assuming that the effect of HSR is the same for all stations. As can be observed, the availability of HSR positively affects the creation of new firms with an average effect between 6% and 7% for the service, tourism and knowledge-intensive activities. As expected, the impact for the manufacturing sector is lower and the coefficient is estimated with a high standard error. The employment cycle, included as a control covariate, is only clearly statistically significant for the tourism activities. The year fixed effects related to those provinces without HSR were not significant for most of the years and sectors. Overall, only the manufacturing activities show a different temporal pattern for the first years in the sample.

Beyond the estimation of an average effect, it is interesting to analyse whether the HSR impact differs among provinces according and whether some specific pattern can be found. Thus, we re-estimated equation 2 allowing for a different coefficient for each province. Table 3 presents the estimated coefficients for the policy variable.

**Table 2. GLS estimates (dependent variable: number of firms)**

| | Services | Tourism | KIA | Manufacturing |
|---|---|---|---|---|
| HSR(-1) | 0.063*** | 0.057*** | 0.072*** | 0.047** |
| | (0.016) | (0.016) | (0.019) | (0.023) |
| Employment cycle | 0.0003* | 0.0005*** | 0.0001 | 0.0001 |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| Constant term | 7.028*** | 5.540*** | 5.990*** | 5.103*** |
| | (0.075) | (0.063) | (0.092) | (0.080) |
| Year-fixed effects no HST | yes | yes | yes | yes |
| Year-fixed effects | yes | yes | yes | yes |
| Province fixed effects | yes | yes | yes | yes |
| Number of observations | 946 | 946 | 946 | 946 |
| chi2 | 80,166.187 | 58,967.318 | 76,106.886 | 40,547.936 |

Note: Standard errors in parenthesis are corrected for heteroskedasticity across panels and for autocorrelation within panels

The estimates clearly show that the impact of a HSR station differs both among provinces and sectors. With regard to services, the impact is positive for almost all provinces (except two), although the standard deviation is high in many of them. In contrast, for the manufacturing sector, the estimated coefficient is only positive in 11 out of 20 provinces and the standard errors are higher. The grouping of tourism activities shows a pattern similar to that of the service sector, whereas in the knowledge-intensive activities the range of variation is higher.

**Table 3. Estimated coefficients for the HSR variable**

| | Services | Tourism | KIA | Manufacturing |
|---|---|---|---|---|
| ***Line Madrid--Barcelona-French border*** | | | | |
| Girona (-1) | 0.158 | 0.210* | 0.205 | 0.217* |
| | (0.120) | (0.116) | (0.142) | (0.117) |
| Barcelona (-1) | 0.162*** | 0.171*** | 0.142** | -0.016 |
| | (0.049) | (0.054) | (0.057) | (0.047) |
| Tarragona (-1) | 0.087 | 0.140** | 0.081 | 0.154*** |
| | (0.061) | (0.056) | (0.061) | (0.057) |
| Lleida (-1) | 0.062 | 0.059* | -0.012 | -0.004 |
| | (0.040) | (0.032) | (0.063) | (0.045) |
| Huesca (-1) | 0.028 | 0.056** | 0.064 | -0.064 |
| | (0.029) | (0.027) | (0.055) | (0.063) |
| Zaragoza (-1) | 0.052 | 0.080* | -0.010 | -0.023 |
| | (0.037) | (0.045) | (0.045) | (0.042) |
| Guadalajara (-1) | 0.403*** | 0.318*** | 0.378*** | 0.395*** |
| | (0.023) | (0.032) | (0.036) | (0.041) |
| ***Line Córdoba-Málaga*** | | | | |
| Málaga (-1) | 0.128*** | 0.087** | 0.167*** | 0.134** |
| | (0.043) | (0.034) | (0.044) | (0.054) |
| ***Line Madrid-Valladolid-León*** | | | | |
| Valladolid (-1) | -0.024 | 0.048 | -0.055 | 0.020 |
| | (0.035) | (0.044) | (0.040) | (0.053) |
| Segovia (-1) | -0.055* | -0.069** | -0.018 | 0.112* |
| | (0.031) | (0.034) | (0.042) | (0.068) |
| Palencia (-1) | 0.073 | 0.139 | 0.048 | -0.209 |
| | (0.091) | (0.125) | (0.124) | (0.150) |
| León (-1) | 0.021 | -0.019 | -0.038 | -0.001 |
| | (0.070) | (0.065) | (0.089) | (0.104) |
| ***Line Madrid-Toledo*** | | | | |
| Toledo (-1) | 0.130*** | 0.082** | 0.152*** | 0.043 |
| | (0.043) | (0.036) | (0.047) | (0.044) |
| ***Line Madrid-Valencia-Murcia*** | | | | |
| Cuenca (-1) | 0.029 | -0.029 | 0.048 | -0.202*** |
| | (0.027) | (0.035) | (0.034) | (0.062) |
| Valencia (-1) | 0.100*** | 0.065** | 0.099*** | 0.018 |
| | (0.032) | (0.031) | (0.034) | (0.046) |
| Albacete (-1) | 0.076** | 0.021 | 0.113*** | 0.064 |
| | (0.035) | (0.039) | (0.036) | (0.068) |
| Alicante (-1) | 0.124** | 0.059 | 0.122*** | 0.130* |
| | (0.057) | (0.069) | (0.044) | (0.070) |
| ***Line Madrid-Ourense*** | | | | |
| Ourense (-1) | 0.030 | 0.049* | 0.010 | 0.131** |
| | (0.036) | (0.029) | (0.037) | (0.055) |
| A Coruña (-1) | 0.028 | 0.022 | 0.012 | 0.115*** |
| | (0.027) | (0.034) | (0.042) | (0.043) |
| ***Line Madrid-Zamora*** | | | | |
| Zamora (-1) | 0.030 | 0.012 | -0.068 | -0.152 |
| | (0.068) | (0.094) | (0.096) | (0.170) |

Note: Standard errors in parenthesis are corrected for heteroskedasticity across panels and for autocorrelation within panels

The low level of accuracy of the estimated coefficients at provincial level introduces uncertainty on the actual impact of the HSR. In order to improve their accuracy we develop a methodological procedure based on a priori information, as explained in the next section.

**6. Distribution functions for the estimated parameters**

Given that we analyse the effect of the infrastructure on the creation of new firms in the province that receives the investment, it can be safely assumed that this effect will be either positive or null. Taking advantage of this a priori information, we propose to estimate the equation under the assumption that $\theta$should be non-negative. A possible way to incorporate this constraint is to estimate a non-linear model (NLS) where the coefficient of the dummy variable that captures the HSR station is the square value of a certain $\gamma$ coefficient. In order to present the methodology, the notation of equation (1) is simplified to:

$$Y_{it} = Z'_{it}\delta + \theta_i F_{it} + \varepsilon_{it} \qquad (3)$$

where $Z_{it}$ includes all the explanatory variables except the dummy for the HSR

Our proposal consists on estimating:

$$Y_{it} = Z'_{it}\delta + \gamma_i^2 F_{it} + \varepsilon_{it} \qquad (4)$$

However, in our case, the estimation of a NLS model with such a high number of coefficients prevented the convergence of the estimator.

As an alternative, it can be proved that estimating the equation by NLS leads to exactly the same results than recovering the non-linear $\gamma$ parameter and its standard error as follows:

$$\hat{\gamma} = \sqrt{\hat{\theta}}$$

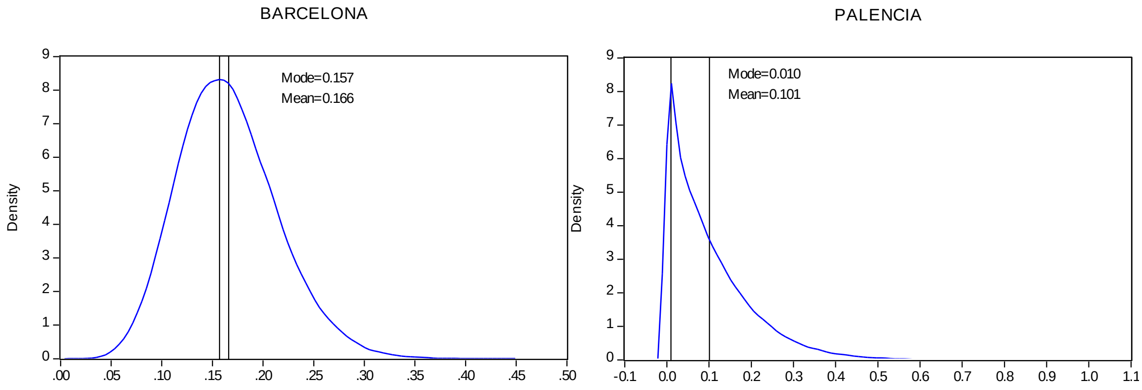
$$\sigma_{\hat{\gamma}} = 0.5 \cdot \frac{\sigma_{\theta}}{\hat{\gamma}}$$

Thus, for those estimated coefficients greater than zero, we computed $\gamma$ and its standard error. With this information, we constructed the one-tail distribution of $\hat{\theta}$ by simulation under the assumption that it only takes positive values. So, in addition to the initial point estimate of $\theta$, the distribution of $\hat{\theta}$ provides more information on the actual impact of being connected to the HSR network.

For each activity sector, we have computed the one-tail distribution of the estimated impact of HSR at provincial level. Given that the distribution is asymmetric, the mean is not the most probable value of the parameter, so we selected the mode as the point estimate of binary policy variable. Besides, we calculated the optimal confidence interval estimate by simulation taking the mode as the central point.

With the objective to illustrate how the degree of asymmetry of the distribution differs among provinces, Figure 2 plots the corresponding distribution of the impact of HSR on the service sector for the provinces of Barcelona and Palencia.

**Figure 2. One-tail distribution of HSR impact for Barcelona and Palencia (service sector)**




In the case of Barcelona the distribution is almost symmetric and the mode and the mean are very similar, whereas the reverse is true for Palencia.

Table 4 displays the mode and the 95% interval upper and lower limits for the one-tail distribution of the service sector. The effect of HSR availability on firm creation shows to be positive for eleven provinces, whereas for the rest it can be considered null. It has to be pointed out that for the province of Guadalajara the estimated effect is much higher than for any other province. This is so for all the four sectors considered. Guadalajara is a province close to Madrid with a relatively low level of economic activity and population. We have checked the data on firm creation with the National statistical office and it seems like there is no mistake. We have also checked for other extraordinary events that might have affected this province, but we are not aware of any. A possible explanation for this result is that the construction of a HSR line has caused a process of decentralization of Madrid, enlarging the metropolitan area. This process could also explain the positive and quite high coefficient for Toledo -province close to Madrid, but with a higher level of economic activity than Guadalajara. The rest of the provinces with a positive impact correspond to those located along the Mediterranean coast.

**Table 4. One-tail distribution of the impact of HSR on the service sector**

| | Lower limit | Upper limit | Mode |
|---|---|---|---|
| Guadalajara | 0.360 | 0.449 | 0.403 |
| Barcelona | 0.080 | 0.272 | 0.157 |
| Toledo | 0.059 | 0.229 | 0.124 |
| Malaga | 0.057 | 0.227 | 0.119 |
| Girona | 0.010 | 0.483 | 0.110 |
| Alicante | 0.037 | 0.260 | 0.110 |
| Valencia | 0.046 | 0.173 | 0.093 |
| Tarragona | 0.009 | 0.248 | 0.069 |
| Albacete | 0.023 | 0.160 | 0.066 |
| Lleida | 0.009 | 0.165 | 0.048 |
| Zaragoza | 0.005 | 0.151 | 0.041 |
| Cuenca | 0.001 | 0.105 | 0.011 |
| Palencia | 0.000 | 0.359 | 0.010 |
| Ourense | 0.000 | 0.141 | 0.004 |
| Huesca | 0.000 | 0.115 | 0.003 |
| A Coruña | 0.006 | 0.012 | 0.003 |
| Leon | 0.000 | 0.395 | 0.000 |

| | | | |
|---|---|---|---|
| Zamora | 0.000 | 0.312 | 0.000 |
| Valladolid | | | 0.000 |
| Segovia | | | 0.000 |

Finally, Table 5 offers the magnitude of the effect of HSR on firm creation for one-tail distribution for the four grouping of sectors. For the tourism related activities the results are similar to those of services although the magnitude of the coefficients varies for some provinces. For knowledge-intensive activities, the same provinces are at the top of the list. In general, all these provinces have also benefitted from the creation of a Science Park.

**Table 5. HSR effect on firm creation (mode of the one-tail distribution)**

| | Service | | Tourism | | KIA | | Manuf. |
|---|---|---|---|---|---|---|---|
| Guadalajara | 0.403 | Guadalajara | 0.316 | Guadalajara | 0.375 | Guadalajara | 0.392 |
| Barcelona | 0.157 | Girona | 0.180 | Malaga | 0.162 | Girona | 0.180 |
| Toledo | 0.124 | Barcelona | 0.164 | Girona | 0.160 | Tarragona | 0.140 |
| Malaga | 0.119 | Tarragona | 0.126 | Toledo | 0.142 | Malaga | 0.120 |
| Girona | 0.110 | Malaga | 0.079 | Barcelona | 0.130 | Ourense | 0.120 |
| Alicante | 0.110 | Toledo | 0.075 | Alicante | 0.116 | Alicante | 0.100 |
| Valencia | 0.093 | Palencia | 0.070 | Albacete | 0.109 | Coruña | 0.100 |
| Tarragona | 0.069 | Zaragoza | 0.068 | Valencia | 0.092 | Segovia | 0.090 |
| Albacete | 0.066 | Valencia | 0.056 | Tarragona | 0.050 | Albacete | 0.010 |
| Lleida | 0.048 | Lleida | 0.051 | Huesca | 0.040 | Toledo | 0.004 |
| Zaragoza | 0.041 | Huesca | 0.048 | Cuenca | 0.035 | Barcelona | 0.000 |
| Cuenca | 0.011 | Ourense | 0.042 | Lleida | 0.000 | Lleida | 0.000 |
| Palencia | 0.010 | Segovia | 0.024 | Zaragoza | 0.000 | Huesca | 0.000 |
| Ourense | 0.004 | Albacete | 0.010 | Valladolid | 0.000 | Zaragoza | 0.000 |
| Huesca | 0.003 | Alicante | 0.010 | Segovia | 0.000 | Valladolid | 0.000 |
| A Coruña | 0.003 | Coruña | 0.005 | Palencia | 0.000 | Palencia | 0.000 |
| Leon | 0.000 | Valladolid | 0.000 | Leon | 0.000 | Leon | 0.000 |
| Zamora | 0.000 | Leon | 0.000 | Ourense | 0.000 | Cuenca | 0.000 |
| Valladolid | 0.000 | Cuenca | 0.000 | Coruña | 0.000 | Valencia | 0.000 |
| Segovia | 0.000 | Zamora | 0.000 | Zamora | 0.000 | Zamora | 0.000 |

Regarding the manufacturing sector, we only find a positive effect in a reduced number of provinces. Moreover, we have to point out that the number of new firms created is much lower in the manufacturing than in the rest of groupings. Consequently, the results can be more erratic.

## 7. Conclusions

- HSR availability seems to have a positive impact on the creation of new firms
- The effects are small but higher for services, knowledge-intensive and tourism-related activities; the effects are almost null for the manufacturing sector
- Geographically, the magnitude of the impact is more relevant in two provinces adjacent to Madrid and in the provinces located along the Mediterranean coast
- In the first case, it may be the result of a process of decentralization of Madrid
- In the second case, the results suggest that HSR may have an impact only on those places able to attract investment
- Contrary to transport authorities' expectations, investing in HSR does not seem to contribute to social cohesion

**Keywords:** *(maximum 6 words) High speed rail, economic impact, location of activities*
**JEL codes: H54, R42, R48**