I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16572

High-Speed Railways and Firms Total Factor Productivity: Evidence from a Quasi-Natural Experiment

Anna Bottasso
Maurizio Conti
Antonella Rita Ferrara
Simone Robbiano

NOVEMBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16572

High-Speed Railways and Firms Total Factor Productivity: Evidence from a Quasi-Natural Experiment

Anna Bottasso
University of Genoa and CIRIEC

Maurizio Conti
University of Genoa and IZA

Antonella Rita Ferrara
University of Calabria

Simone Robbiano
University of Eastern Piedmont

NOVEMBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

High-Speed Railways and Firms Total Factor Productivity: Evidence from a Quasi-Natural Experiment*

The focus of this study is to assess the causal impact of the connection of a local area to a high-speed rail network (HSR) on firms' total factor productivity (TFP). The quasi-random location of the HSR station in the Italian city of Reggio Emilia is exploited in a Difference-in Differences (DiD) research design applied to a large sample of firms, observed over the period 2010-2018. The results suggest that the opening of the HSR station improved treated firms' TFP of about 5%; in particular, such effect is larger for firms closer to the HSR station and slightly increases over the sample period. We also find that the impact of the connection to the HSR station is heterogeneous across industries and depends on firms' size and past productivity. Overall results are robust to a large number of sensitivity checks and falsification tests.

JEL Classification: C50, D24, L92, R30

Keywords: transport infrastructure, Difference-in-Differences, total factor productivity

Corresponding author:
Maurizio Conti
University of Genoa
Department of Economics
5 Via Francesco Vivaldi
Genova
Italy
E-mail: mconti@economia.unige.it

* We wish to thank participants at the 2023 RSAI-BIS, 2023 SBCA-EU, and 2023 SIEP conferences for fruitful discussions. All remaining errors are own own.

1 Introduction

A growing body of empirical evidence in economics, regional science and economic geography has shown that new transportation networks tend to have a broad positive effect on various economic outcomes at local level.[1] Therefore, it is not surprising that transport infrastructure investments score highly among the policy tools considered by policy-makers in place-based policies (Neumark and Simpson, 2015). However, as discussed in Redding and Turner (2015) and more recently in Koster et al. (2022), whether or not the connection to a transportation network increases local economic activity is a more debated issue in the theoretical literature; in other words, it is not clear whether all locations benefit from the construction of a new road/HSR line or if such investments only represent a reorganisation of economic activity across space, thus possibly fostering, rather than alleviating, spatial economic disparities. Koster et al. (2022) have recently proposed a theoretical model which shows that, under certain conditions, the construction of a transport link between any two regions can be detrimental to the intermediate regions bypassed by it (e.g. local areas without HSR stations). Similarly, the model also shows that an intermediate region that is connected to a transport network does not necessarily experience an improvement in its economic conditions compared to an unconnected region.

We argue that the ambiguous predictions of such model are aligned with the empirical literature. Faber (2014) and Baum-Snow et al. (2017) show, for the Chinese case, that the construction of new highways tends to increase economic activity only in large metropolitan areas, with negative effects in more peripheral ones. Similar results have been found for the upgrading of railways in China by Qin (2016) and HSR network in Japan by Koster et al. (2022). In contrast, other studies find no evidence of significant spatial reorganisation of economic activity. For example, Ahlfeldt and Feddersen (2018) find that the opening of HSR stations in three small cities in Germany led to a significant increase in their local economic activity and productivity; similarly, Bernard et al. (2019) find that the opening of new HSR stations in Japan caused an increase in (revenue) TFP in firms that are located near the new station, with no clear evidence of a fall in TFP for more distant firms.

In this study we exploit the quasi-random location of an HSR station in Italy (near the northern city of Reggio Emilia) to causally evaluate the impact of the connection to the HSR network on the evolution of firms' TFP. In particular, we apply a Differences-in-Differences (DiD) research design,

[1] For instance, see Gibbons et al. (2019), Li et al. (2017), Holl (2016) and Ghani et al. (2016) for the case of roads in the UK, China, Spain and India, respectively; Gibbons and Wu (2020) for Chinese airports; Ahlfeldt and Feddersen (2018) and Bernard et al. (2019) for high-speed rail (HSR) connections in Germany and Japan, among others. The paper by Redding and Turner (2015) reviews the early evidence, while the book of Ferrari et al. (2018) surveys also more recent evidence on the effects of different modes of transport and various economic outcomes. See also Blanquart and Koning (2017) for HSR.

where firms near the HSR station (i.e. within a 30 km radious) belong to the treatment group, while those located within a 30-60 km range from the station represent the control group.

Evaluating the causal effect of HSR is interesting for several reasons. First, because important extensions of HSR networks, often mobilising huge financial resources, have been realised and/or are planned in many countries (Koster et al., 2022), while rigorous ex-post evaluations are lacking, especially in the European Union. Second, as noted above, the impact of HSR on peripheral or intermediate regions is still debated in the academic literature. Third, the literature tends to consider the evaluation of HSR extensions to be particularly interesting, as the latter typically move people and not goods; as a result, goods transport costs do not change when a new HSR station is opened, but travel times do. Therefore, the economic impact of HSR, if any, is likely to derive from its effect on labour market expansion. By way of example, Heuermann and Schmieder (2019) report for Germany a reduction of 9.5% of travel time between localities connected with an HSR station and find that a 1% reduction in travel time is associated with an increase in the number of commuters of about 0.25%. Similar results are shown in Baltrunaite and Karmaziene (2020) who show that, in Italy, most city pairs connected by HSR experience drastic reductions in travel time with respect to a connection by car.[2] Moreover, they use labour force data to show that, after the opening of an HSR station, there is a significant increase in the number of people working in the HSR hosting province but living in a different region, particularly in the case of professionals.

Koster et al. (2023) find, for the case of Japan, that trade linkages between firms increase when travel time by train falls, consistently with both the theoretical model and the empirical evidence discussed by Bernard et al. (2019), who report an increase in the number of buyer-supplier relationships after the opening of new HSR stations in Japan, as well as an increase in revenue total factor productivity. This suggests that not only reductions in travel time due to new HSR connections may enlarge firms' relevant labour markets, but they may also expand firms' product markets as well as the market for relevant intermediate inputs, even if such connections do not entail any reduction in shipping costs. This is because establishing trade relations among firms may require face-to-face interactions, which are facilitated by reductions in travel times.

More specifically, the connection to an HSR network can affect the evolution of firms' TFP through different channels. To begin with, HSR nodes might favour face-to-face-interactions between scientists and engineers, which in turn might induce an increase in product and process innovation (e.g. scientists that live in large urban areas where the most important universities are located).[3] In addition, by reducing the isolation of firms, HSR may reduce the spatial barriers to

[2]In the case of Reggio Emilia, the HSR train covers the 60 km distance to Bologna in 30 minutes, compared to 1 hour by car and the 155 km to Milan in 1 hour, compared to 2 hours by car.

[3]See, for relevant empirical evidence for the case of railways, Dong et al. (2020) and Lin (2017); for

the transmission of knowledge from frontier firms (Comin et al., 2012). Furthermore, firms may find higher quality or cheaper suppliers by enlarging the input markets of relatively isolated regions through the reduction of firms' search costs, thus increasing TFP (Bernard et al., 2019). Finally, access to an HSR network should improve the matching process between firms and employees and this effect can be particularly relevant in the case of peripheral areas, which can more easily get access to a larger pool of highly skilled employees, as shown by Heuermann and Schmieder (2019) for Germany. Similarly, Baltrunaite and Karmaziene (2020) report, for the case of Italy, an increase in workers' mobility and assortative matching between firms and directors, with highly productive firms improving the quality of their boards after the opening of a new HSR station.

In this study, we use a large sample of firms observed over the period 2010-2018 to assess the effect of the opening of the HSR station in the city of Reggio Emilia on the TFP evolution of firms located near the new station. This case study is particularly interesting because the area around Reggio Emilia, although part of Italy's manufacturing heartland, has a relatively small population compared with the two cities to which Reggio Emilia is directly linked by HSR, Milan and Bologna.[4] The connection of Reggio Emilia to the Italian HSR network is therefore a good test of one of the predictions of the Koster et al. (2022) model, which suggests that being connected to a network is not enough to improve the performance of an intermediate region relative to an unconnected intermediate one unless the connected region has a sufficiently large market to start with.[5]

Furthermore, the study we propose is a clear example of the incidental treatment approach to identification (Redding and Turner, 2015) in a DiD research design, as Reggio Emilia is almost on the straight line connecting Bologna and Milan (see Figure 1). Moreover, as in the German case analysed by Ahlfeldt and Feddersen (2018) and Heuermann and Schmieder (2019), the decision to stop the HSR in Reggio Emilia was the result of a political negotiation. As the Prime Minister at the time, Romano Prodi, a native of Reggio Emilia, later admitted, the decision was also taken in the face of political hesitation on the part of politicians in Parma.[6] This makes the assumption that the Reggio Emilia HSR station was as good as randomly assigned quite reasonable. Hence, we believe that our case study has a strong internal validity.

---

highways, in turn, see Agrawal et al. (2017) and Bottasso et al. (2022).

[4]The population of the municipality of Reggio Emilia is about 170,000, while the province of Reggio Emilia has about half a million inhabitants. The metropolitan areas of Milan and Bologna have about 3.25 and 1 million inhabitants respectively.

[5]Moreover, the HSR station of Reggio Emilia was the first case, at least in Italy, of a new train station that was built to serve an HSR line.

[6]See the daily "La Gazzetta di Parma", 10 April 2022: https://www.gazzettadiparma.it/parma/2022/04/10/news/prodi-merito-mio-la-mediopadana-a-reggio-emilia-638303/.

Our main results can be summarized as follows. First, we find that the opening of the HSR station in Reggio Emilia improved the TFP of treated firms by about 4% on average, with an effect that has been slightly increasing over time. Moreover, we find that the increase in TFP is about twice as large in the case of firms very close to the station, i.e. within 10 km, and smaller in the case of firms in the 10-30 km range. Our main results are robust to a large battery of sensitivity checks that should make the parallel trend assumption more credible, such as controlling for linear time trends at the local labour markets (LLM) level -which are the Italian equivalent of the UK Travel to Works areas, and that are quite heterogeneous in terms of population density, firm demography, etc.-, or by allowing firms' TFP to follow different trends depending on some predetermined firms' characteristics (such as their distance from other modes of transport) among the others. Given the attrition characterising firm-level balance sheet data, we also test if overall results hold when considering a balanced panel. Moreover, we conduct a placebo analysis by assuming that the HSR station was positioned near the alternative site of Parma and, rather comfortingly, we do not find any positive evolution of TFP for "fake-treated" firms after the "fake-opening" of the HSR station.

Interesting heterogeneous treatment effects emerge from the analysis. In particular, the impact of the opening of the HSR station seems to be concentrated in scale and information intensive and supplier dominated industries [7] with no significant effect for the other industry categories. Although this may seem surprising given that these are not the sectors employing a large proportion of highly skilled workers (who tend to be the more mobile), other mechanisms may be at work, such as the reduction in travel costs to new markets and intermediate suppliers discussed above. Moreover, the impact of the opening of the HSR station is relatively larger for small firms in the 10-50 employees range, while smaller and non-statistically significant effects for both micro and large firms are found. Furthermore, the positive effects on TFP are stronger in the case of firms that, before the opening of the HSR station, used to have low levels of TFP (below the median), thus suggesting that the opening of the new HSR station might have contributed to reduce the dispersion of TFP levels. We argue this result to be consistent with the view that HSR may foster knowledge spillovers from firms closer to the technological frontier.

Finally, our results provide weak evidence of negative spillovers across space, since we find a reduction in the average TFP levels of firms that are located in the 30-40 km range from the HSR station, with respect to firms in the 40-60 km range, a result consistent with a spatial reorganization of economic activity.

Our paper contributes to the literature in several ways. It is one of the very few papers to assess the impact of access to a transport network on productivity using firm-level data, thereby allowing to perform interesting heterogeneity analyses that are challenging when the units of analysis are regions

[7] We use the extension of the traditional Pavitt taxonomy to services that has been recently proposed by Bogliacino and Pianta (2016).

or municipalities, such as differences associated with firm size or the level of TFP.[8] Furthermore, we have granular information on sectors where firms are active, we can explore heterogeneity at the sectoral level, which is possible but not easy without firm-level data unless one has access to very rich industry-level data. Last but not least, with regional level data, productivity is necessarily proxied by labour productivity (e.g. value added per person employed) or, if information on past investments is available, TFP can be considered, but the latter is typically calculated assuming a constant returns to scale technology and perfect competition. However, these assumptions are problematic from a theoretical point of view in a setting where imperfect competition and non-constant returns to scale are key for transport investments to have important wider economic effects. In contrast, we can estimate firm-level TFP without making strong assumptions about technology and market structure using the methods discussed in detail in De Loecker and Syverson (2021).

The remainder of the paper is organized as follows. In Section 2 we describe the construction of the dataset as well the estimation of TFP, while Section 3 explains in more detail our identification strategy. Finally, Section 4 discusses the empirical results while Section 5 concludes.

2 Data

In this section we describe the data used in the analysis, the sampling and geocoding procedure and the approach followed to estimate firms' Total Factor Productivity (TFP).

2.1 Sources and geocoding

We use firm-level balance sheet information extracted from Orbis, a commercial database distributed by Bureau Van Dijk (BvD)/Moody's. Orbis provides economic and financial information (e.g., value added, number of employees, value of fixed intangible assets, expenditure on intermediate goods and services, labor costs) for the non-farm business sector. However, the coverage varies by firm size, sectors of activity, dimensions and over time. Notwithstanding these limitations, Orbis is the largest cross-country firm-level database available for economic and financial research. The data at hand covers approximately 10 years, over which it is possible to construct an unbalanced panel of firms following the procedure highlighted in Kalemli-Ozcan et al. (2023). To better exploit Orbis representativeness, we rely on the findings of Bajgar et al. (2020) that analysed the reliability of the information concerning industry, coverage and over time. Our analysis focuses only on firms operating in Italy, thus avoiding limitations that might arise in cross-country comparisons; moreover, we consider firms operating in the manufacturing and services sectors over the

[8] Among recent relevant exceptions, we can mention the studies by Li et al. (2017) for the case of Chinese highways, Holl (2016) for highways in Spain and Bernard et al. (2019) for HSR in Japan.

period 2006–2019.[9] Bajgar et al. (2020) outlined Italy as a very well-represented country in Orbis characterised by a coverage of about 40% of total value added according to the comparison at the aggregated country-industry cell with respect to OECD-STAN. For our analysis, we select firms operating in a radius of about 60 km from Reggio-Emilia's high-speed railway station and that are at least 30 km away from Bologna - where there is another HSR station - to avoid any confounder effect. Firms are then geolocalised using latitude and longitude coordinates or the address of their headquarters. For the geographical analysis of our data, we rely on the software QGIS.[10] The geocoding procedure envisages several steps. First, we gather information on administrative and Local Labour Market (LLM) boundaries using the shapefiles of Italian municipalities published by the Italian Institute of Statistics (ISTAT).[11] Information on the Italian High-Speed Railway (Alta Velocità Alta Capacità AV-AC) is retrieved from OpenStreetMap, via internal plugins in QGIS; specifically, we build a vector layer containing lines referred to the AV-AC railway, and another vector layer with points for AV-AC stations. This data is then combined with information on other modes of transport, namely the location of motorway toll stations, airports and other non-AV-AC stations, again extracted through OpenStreetMap. This information is then matched with the geolocalised sample of firms. For each unit in our sample, we compute the respective distance from the HSR station of Reggio Emilia and the HSR station of Bologna (to avoid any confounder effect). Furthermore, the distances between the firm and other transport infrastructures(motorway toll stations, airports, and other traditional railway stations) are calculated. In our research design, the identification of treated and control units relies on the geographical distance from the Reggio Emilia station, exploiting its quasi-random location. To this end, we need to draw some buffer zones, defined as vector polygons, describing the areas around the station which is our target. In particular, we define buffer zones in the range 0-10 km, 10-20 km, 20-30 km and 30-60 km from Reggio Emilia HSR station. Firms located less than 60km from Bologna are excluded from the analysis. Figure 1 maps the Italian High-Speed Railway, while Figure 2 focuses on the buffer zones described above.

[9]More precisely, we exclude firms operating in NACE 84–96 which identify public and finance & insurance services. We also exclude sectors with less than 100 observations (2-digit NACE codes=19, 36, 21, 39, 50, 51, 65, 75).

[10]QGIS is a free and open source Geographic Information System, to create, edit, visualise, analyse and publish geospatial information. See https://www.qgis.org/en/site/.

[11]A shapefile is a vector format for geographic information systems (GIS) that can spatially describe points, polylines or polygons, which can be used in various information, environmental and geoscientific fields. Our project is based on the WGS84 - EPSG:32632 data. See https://www.istat.it/it/informazioni-territoriali-e-cartografiche.

Figure 1: Italian High-Speed Railway (AV-AC)

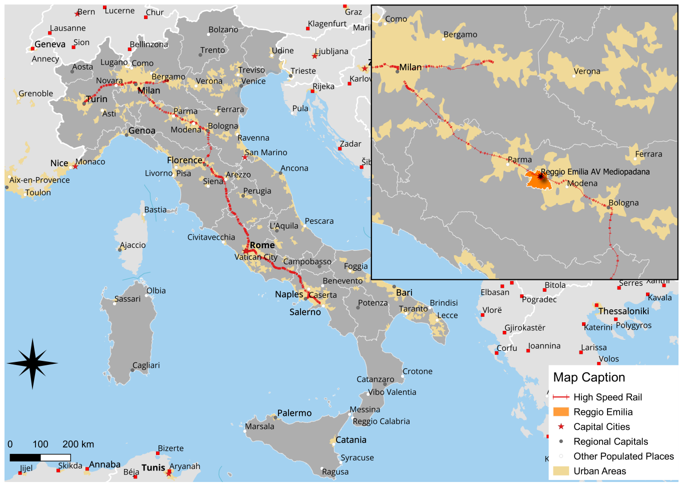
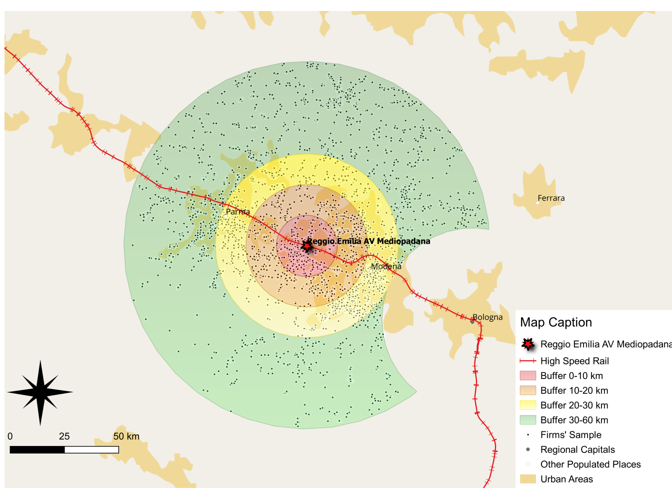


Figure 2: Reggio Emilia HSR station



2.2 TFP estimation

Firms' TFP is derived by estimating a value-added translog production function (with capital and labour as inputs), which is a second-order approximation to any general production function and it is therefore much more flexible than the conventional Cobb-Douglas one (De Loecker and Syverson, 2021). The capital stock is proxied by the value of fixed assets, while the labour input is measured with the number of employees or, as a robustness check, with labour expenditure, to capture possible differences in labour quality across firms. The production function is separately estimated for each NACE 2-digit sector; moreover, all variables are taken in deviation from the yearly average of the sector, which is also a way to deflate monetary variables without using sector-specific deflators, as argued by Bernard et al. (2019). The endogeneity problems that plague the estimation of production functions have been addressed using the control function approach of Levinsohn and Petrin (2003).[12] Finally, given the lack of information on firm-level prices in our dataset (as in most others), we follow De Loecker and Syverson (2021) and De Loecker (2013) and augment the production function with value added at the sectoral level for Italy as a whole. This should capture the evolution of market demand at the sectoral level, thus allowing a separate identification of the demand and production function parameters, which is typically not possible when using a revenue-based production function. Productivity is estimated using the *prodest* routine for Stata (Rovigatti and Mollisi, 2018). Descriptive statistics reported in Table 1 suggests that firms in the treated group had slightly higher TFP with respect to control ones. This difference is statistically significant at 1% and increased slightly after the opening of the HSR station.[13] A thorough description of the research design is presented in the next section.

Table 1: Descriptive statistics on TFP.

	Mean T	**Sd Mean T**	**Mean C**	**Sd Mean C**	**Diff(T-C)**	**P-val Diff**	**N obs T**	**N obs C**
$TFP_{i,t}$ [PRE]	0.4361	0.1198	0.4267	0.1168	0.0093	0.0000	27,483	11,750
$TFP_{i,t}$ [POST]	0.4514	0.1304	0.4399	0.1273	0.0114	0.0000	68,813	29,931

Notes: For ease of interpretation, TFPs are rescaled using a min-max normalisation over the pre and post-periods and range in the [0,1] interval. T and C identify treated and control firms, respectively. Firms located within a 0-30 Km radius from the HSR station are considered as treated. The post period refers to 2013-2018, while the pre covers 2010-2011.

[12] In particular, we exploit intermediate input levels as a proxy for unobservable productivity and we implement the correction proposed by Ackerberg et al. (2015); moreover, following the suggestions in De Loecker and Syverson (2021) and De Loecker et al. (2016), we control for occupation density at the local labour market level, which should proxy for local time-varying shifters to labour demand. Finally, we take into account the attrition in the data associated to firms' exits from the sample.

[13] Detailed descriptive statistics on other relevant variables for treated and controls, before and after treatment, are reported in Table A.1.

3 Identification Strategy

To identify with a DiD research design the effect of the HSR station on (log) firm-level TFP (tfp_{it}) we estimate the following two-way fixed effects model with OLS:[14]

$$tfp_{it} = \sum_{t=1}^{6} \gamma_t Year_t \times Treated_i + \lambda_t + c_i + u_{it}, \qquad (1)$$

where λ_t is a full set of year fixed effects; c_i is a full set of firm fixed effects; u_{it} is a stochastic error term; $Year_t$ is a set of year dummies for the post-opening period, while $Treated_i$ is a dummy equal to 1 for (eventually) treated firms, which in the baseline specification is equal to 1 for firms that are located in the 0-30 km range from the HSR station and 0 otherwise. Such choice is related to the geographical characteristics of the area under scrutiny, where the distance between Reggio Emilia and the adjacent HSR station of Bologna is 60 km, so the 30 km threshold reduces the probability of overlapping effects associated with the presence of the latter.[15] The 30 km threshold for a station is chosen so that total travel time is significantly affected and the newly opened station dominates alternative modes of transport.

In turn, the set of γ_t coefficients measures the average treatment effect on the treated (ATT) of the HSR station for every year since its opening (i.e. 2013=1).[16] As in all DiD research designs, the identification strategy relies on the hypothesis of parallel trends in TFP between treated and control units, so that firms far away from the location of the HSR station can act as the valid control group for firms near the HSR station (the treatment group). We believe this is likely to hold in this case for a variety of reasons. First, because Reggio Emilia is not a central node in the Italian HSR transportation system. Indeed, it is located on a straight line between the already opened HSR stations of Bologna and Milan; therefore, this is an application of the incidental treatment approach to identification that is quite popular in the evaluation literature of transport infrastructures (Koster et al., 2022; Redding and Turner, 2015). Moreover, the choice of Reggio Emilia over the other most

[14]In order to remove possible outliers, we have trimmed extreme values, i.e. those below and above the 1 and 99 percentiles of the TFP distribution, respectively.

[15]Also Bernard et al. (2019) and Inoue et al. (2017) consider treated firms as those located within a distance of 30 km from an HSR station in Japan.

[16]We do not estimate an average effect over the whole sample period. Given the strong unbalancedness of the panel data set (with substantial firm entry and exit as well as improved coverage of the ORBIS dataset over time) we prefer not to impose a common effect over time. Nevertheless, in Tables 2–7, we report an average effect for the whole sample period that, however, rests on the estimation of the more flexible specification of equation (1) above.

likely alternative location (Parma) was made on purely political grounds. In other words, it is very unlikely that the location of the HSR station was motivated by the existence of certain economic characteristics of the area around the Reggio Emilia HSR station that were potentially correlated with the evolution over time of firms' TFP. For these reasons, we are rather confident that the location of the HSR station is almost as good as randomly assigned.

Nevertheless, besides conducting conventional tests for pre-trends in DiD research designs, we also estimate augmented versions of Equation 1, by alternatively including province-by-year fixed effects or Local Labour Market linear time trends that should eliminate possible spatial time-varying heterogeneity, thus possibly making the (conditional) parallel trend assumptions more likely to hold.[17] Along these lines, we also control for important predetermined characteristics of the environment where firms operate that could lead to violation of parallel trends. In particular, we focus on firms' distance from the closest airport, highway exit and traditional railway stations. To this end, we add to Equation 1 the interactions of such distances with all $Year_t$ dummies as well as with the $Year_t$-$Treated_i$ interactions, as suggested by Wooldridge (2021).[18]

Finally, another non-negligible issue in the ex-post evaluation of the economic effect of a newly established transport infrastructure is that the definition of treated and control units, which is based on distance, is necessarily not clear-cut. Moreover, with spatial data, the Stable Unit Treatment Value Assumption (SUTVA) that implicitly or explicitly underpins most research designs is unlikely to be met. Indeed, general equilibrium effects could arise because of the existence of spatial spillovers, whose directions are usually not even clear *a priori*. For this reason, while in our preferred specification, we consider as (eventually) treated only firms that are located between 0 and 30 km from the HSR station, in robustness checks we also consider non-parametric specifications where the $Year_t$ dummies for the post-HSR station opening period are interacted with different distance dummies (0-10 km; 10-20 km; 20-30 km; and also 30-40 km).

4 Results

4.1 Main Results

In this section, we discuss our main results on the effect of the 2013 opening of the HSR station in Reggio Emilia on local firms' TFP. We first estimate Equation 1 on an unbalanced panel of

[17]Conventional tests for pre-trends in DiD research designs do not provide robust evidence for an increase in TFP for treated firms before the opening of the HSR station

[18]We do not include the interaction of distances to other transport infrastructures with the $Treated_i$ dummy because they are time-invariant and therefore perfectly collinear with the firm fixed effects.

28,234 firms observed over the period 2010-2018 and we report the average of the estimated yearly treatment effects for the 2013-2018 time span in column (1) of Table 2. Estimates suggest that the opening of the HSR station in Reggio Emilia generated an average increase in treated firms' TFP of about 4.7%. Moreover, to investigate if the estimated effects on TFP are somewhat affected by firms' entry and exit from the sample, we estimate our baseline model on a balanced panel including 3,961 firms and we find that connecting Reggio Emilia to the Italian HSR network increased firms TFP of about 6.3% (Table 2, column(2)).

As we stressed in the previous Section, the ability of a DiD design to provide causal estimates relies on the validity of its identifying assumptions, hence we first test the absence of any anticipation effect and the validity of the parallel trend assumption by implementing an event study analysis. Results are shown in Fig. 3 and Table A.2 (columns 1 and 2) for both the balanced (black dots) and the unbalanced (blue squares) panel of firms. The dots and the squares are the respective point estimates, the vertical whiskers are 99% confidence intervals. Reassuringly, results confirm the absence of any anticipation effect on the outcome before the opening of the HSR station, as TFP levels are very similar for the treatment and control groups in both balanced and unbalanced samples.

The event study analysis also provides some insights into the dynamics of the effect in the post-treatment period. In particular, we observe that the effect materialises only after two years with a slight increase in magnitude during the last two years of the sample period. Indeed, it seems reasonable to expect that transport infrastructure investments require some time to convey their effects, given that economic agents need some time to adjust their choices to the new setting. The same study conducted on the balanced sample outlines a similar pattern and excludes the presence of any anticipation of the policy, thus confirming the validity of the parallel trend assumption. The event study results are also robust to the inclusion of Local Labor Markets (LLM) linear trends, province-by-year fixed effects as well as interactions between distances from other modes of transport (highways exits, airports and conventional railways stations) and year dummies (Table A.2 columns 3–5).

To further validate our research design, we implement a placebo analysis by estimating our baseline model under the assumption that the HSR line was positioned in Parma, which was planned as an alternative HSR station to Reggio Emilia. In this setting, we should not find any effect of the "fake-opening" of the HSR station on "fake-treated" firms and estimates shown in column (3) of Table 2 suggest that this is indeed the case, thus alleviating possible concerns on the typical nonrandom allocation of transport infrastructure.

As an additional sensitivity analysis, we check how results vary according to the selection of the treatment and control groups and we argue that this analysis can also allow us to deal with possible drawbacks associated with plausible violations of the Stable Unit Treatment Value Assumption, as explained in Section 3. In particular we extend the treated sample in order to include also firms located 30-40 km around the HSR station, leaving only those located further away from the HSR station (40-60 km) in the control group. Evidence from the unbalanced panel, reported in column (1) of Table 3, confirms our previous results for firms closer to the HSR station (within 30 km), while the TFP level of firms in the 30-40km buffer seems to have declined, albeit the latter result is not confirmed in the balanced panel (column 2 of Table 3). The fact that firms in the 30-40 Km buffer zone have been negatively affected by the opening of the HSR station - relatively to more distant firms - may be indicative of some spatial reorganization of economic activity: for instance, firms in the 30-40 km area might have lost some of their high human capital workers, who may have decided to work for firms closer to the HSR station. However, this result needs to be interpreted with some caution, as it does not hold when looking at the balanced sample, which casts some doubt on its robustness.

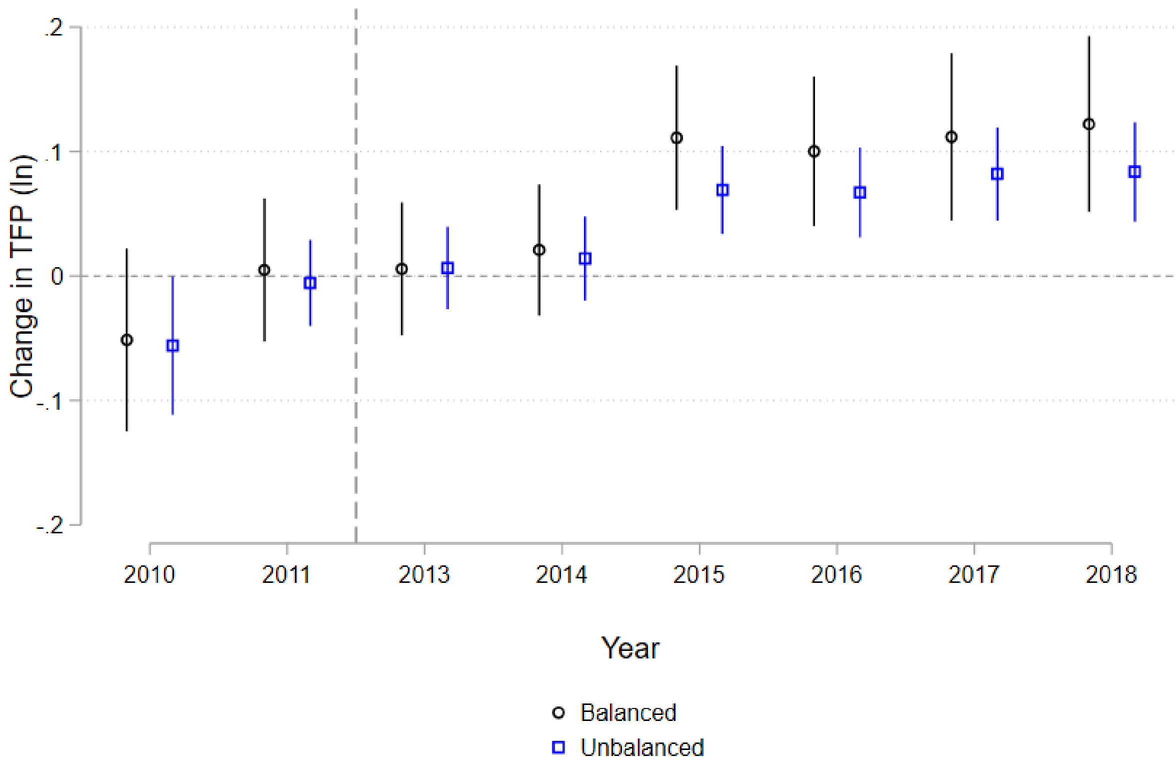
Further insights are provided by estimating Equation 1 by decomposing the treatment sample into smaller distance buffers based on the distance from the HSR station, i.e. distinguishing those located in the ranges 0-10 km, 10-20 km and 20-30 km. Columns (3) and (4) of Table 3 show that the increase in firms' TFP is about twice as large in the case of firms very close to the HSR station (within 10 km) and smaller in the case of firms in the 10-30 km range.

Table 2: Impact of HSR on TFP.

	(1) Baseline	(2) Balanced	(3) Placebo
Distance 0-30 Km	0.04731***	0.06273***	-0.04135
	(0.00970)	(0.01519)	(0.03206)
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
Observations	137,977	35,649	140,791

Notes: All the models include firm and year fixed effects. In Columns (1) and (2) the reported coefficient for *Distance 0-30 Km* is the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Column (3) shows a placebo analysis in which *Distance 0-30 Km*measures the distance from Parma.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3: Event Study of the Impact of HSR on TFP: balanced vs unbalanced



Note: The vertical dashed line identifies the reference year (2012). Black dots and blue squares are the points estimates of the effect of the HSR in each year on the balanced and unbalanced samples, respectively; the vertical whiskers represent the respective 99% confidence intervals. Regressions include fixed effects and controls as from Table 1.

Our main results can be interpreted in the context of recent theoretical models analyzing the spatial distribution of economic activity and using HSR data for model validation.[19] In particular, they could be linked to the study of Bernard et al. (2019) who show that changes in the set of suppliers and sourcing locations generate an improvement in firms' performance. Our findings are also in line with Koster et al. (2023) and Koster et al. (2022), who show that HSR in Japan had local effects on employment, rents and wages, thus facilitating business-to-business services. These studies suggest that transport infrastructure investments can positively affect economic activity by enhancing the scale and efficiency of spatial economic linkages and we argue that they can provide useful insights on the economic mechanisms that might explain our results. Indeed, the Italian HSR network provides only passenger services (even if it was planned for goods transport as well), so that it can ease face-to-face interactions, enhance access to intermediate goods and consumer markets,

[19]These models fit into the more general field of the economic impact of transport infrastructure.

Table 3: Extension by also including 30-40 km

	(1) Unbalanced	(2) Balanced	(3) Multiple Distances	(4) MD-Balanced
Distance 0-30 Km	0.03355***	0.04947***		
	(0.01195)	(0.01902)		
Distance 30-40 Km	-0.03150**	-0.02934		
	(0.01588)	(0.02498)		
Distance 0-10 Km			0.07946***	0.10972***
			(0.01830)	(0.02794)
Distance 10-20 Km			0.03636***	0.03978**
			(0.01151)	(0.01813)
Distance 20-30 Km			0.04340***	0.06257***
			(0.01042)	(0.01645)
Firm FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Observations	137,977	35,649	137,977	35,649

Notes: All the models include firm and year fixed effects. In Columns (1) and (2) the reported coefficients for *Distance 0-30 Km* and *Distance 30-40 Km* are the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located either in the 0-30 Km or in the 30-40 Km radius from Reggio Emilia and at least 60 Km from Bologna. Analogously, Columns (3) and (4) report the average of the estimated yearly treatment effects for the treatment period 2013-2018 and the respective dummies for three different distance buffers: 0-10 Km, 10-20 Km and 20-30 Km. Models are estimated over the unbalanced (Columns 1 and 3) and the balanced (Columns 2 and 4) panels of firms.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

improve location choices and enlarge the labour market pool, thus favouring labour market access and better matching between employers and employees. Furthermore, easier and faster connections can promote knowledge spillover and technology transfer. Indeed, the regional and urban literature considers spatial accessibility as an important characteristic, among other factors, that can support knowledge flows and spillovers, thanks to higher potential opportunities of face-to-face interactions and exposure to foreign or external knowledge (Crescenzi et al., 2011).

We argue that all of the above mechanisms might reflect the existence of agglomeration economies arising from the spatial concentration of economic activity, that, in turn, is favoured by the reduction of transportation costs associated with transport infrastructure investments.[20] Unfortunately, the data at hand prevents us from disentangling all possible channels of transmission that allow the HSR infrastructure to display its economic impacts. Nevertheless, Reggio Emilia's connection to the Italian HSR network makes it a good case study to test one of the Koster et al. (2022) connected model's hypotheses, which predicts that network connectivity alone won't improve an intermediate region's performance over an unconnected one until the connected region has a sizable enough market. Indeed, we find evidence in favour of a positive impact of HSR connection to an intermediate region.

It is worth noting that our empirical results are in line with the economic literature analyzing the extent of the spatial scale where agglomeration economies are observed. Given that such economies are associated with flows of goods, people and information across space, their geographical scope is determined by the rate at which these flows diminish with distance.[21] Our findings suggest that the main effect of the establishment of an HSR station is observable for firms located within 10 km of the station, while such an effect decreases for firms between 20 and 30 km away from the station. Similar results, but in terms of firms' higher innovative capacity, are found by Inoue et al. (2017) on a panel of Japanese-treated firms located within 30 km from an HSR station.[22]

After estimating our baseline specification we extend the analysis by controlling for possible spatial time-varying heterogeneity to prevent that the identifying variation might disproportionately rely on specific trends (e.g. economic or demography trends) in different provinces that may be

[20] Different strands of the economic literature have estimated agglomeration economies by including some measures of locations' attractiveness (market potential or market access) into TFP empirical models or have estimated gravity models where flows between locations depend, among other things, on the distance between them (see Graham et al., 2009, for a short survey).

[21] Among others, Graham et al. (2009) find positive and significant agglomeration effects on firms' total factor productivity in the range of 0 to 25 km and a somewhat smaller less significant effect for the 25 to 50 km buffer.

[22] Miwa et al. (2022) focus instead on Japanese municipality innovative capacity and find significant positive effect of HSR connection only for firms located 10 km away from the HSR station.

systematically correlated with the introduction of the high-speed train. To this aim, we include province-by-year fixed effects and, in a similar vein, we include local Labour Market linear time trends into Equation 1.[23] We argue that it is important to control for time-varying heterogeneity common to firms belonging to the same Local Labour Market that might have specific common trends of productivity, also in light of the predictions of theoretical models on the impact of HSR, based on the enlargement of the labour market pool, on possible better matching between workers and firms and on higher knowledge spillovers and technology transfer. Results reported in columns (1) and (5) in Table 4 confirm our previous findings, both in terms of sign and magnitude of our DiD coefficient of interest.

Furthermore, we account for possible predetermined environmental factors that may cause a violation of the parallel trend assumption. In particular, for each firm, we consider its distance to the nearest airport, tollbooth exit and traditional railroad station, as measures of the accessibility to other transport modes. As recommended by Wooldridge (2021), we include in Equation 1 the interactions of such distances with all $Year_t$ dummies as well as with the $Year_t$-$Treated_i$ interactions. Indeed, firms' access to other transport infrastructure might determine specific TFP patterns associated with the opening of the HSR station. Interestingly, the TFP of firms located within 30 km from the HSR station is found to be 3% higher with respect to control firms located in the 30-60 km buffer (Table 4, column 2), i.e. the average treatment effect on the treated is still positive; however, treated firms' distance from other transport infrastructures differentially affect firms' TFP patterns according to the transport mode. On the one hand, treated firms located relatively far from an airport benefited more from the opening of the HSR station, probably because the reduction of travel time to an international airport might have reduced, among other factors, the costs to explore new foreign markets.[24] On the other hand, treated firms' distance to conventional railway stations or highway tollbooths does not provide an additional boost to their TFP.[25] In addition to this, when we jointly control for Local Labour Market specific trends and firms' distance to other transport infrastructure, the estimated DID coefficient is found to be almost

[23]According to the Italian Statistical Office, Local Labour Market represent a territorial grid whose boundaries, regardless of the administrative articulation of the territory, are defined using the flows of daily home-to-work (commuting) movements recorded during the General Population and Housing Censuses and can be quite heterogeneous in terms of population density, firm demographics, etc. Since each local system is the place where the population lives and works and exercises most of its social and economic relations, home-to-work commuting is used as a proxy for the relations existing in the territory.

[24]HSR might have reduced the distance from nearby airports, or might have favoured long distance travels *per se*; by way of example, Alderighi and Gaggero (2017) suggest that air transport service positively affects the export of Italian manufacturers.

[25]Point estimates are available upon request.

identical to the estimates in our basic specification (Table 4, column 3). Finally, while it would be important to control for municipal-specific time-varying fixed effects, such an approach would be computationally unfeasible, given the high number of municipalities observed over the sample period. Hence, following Bernard et al. (2019), we include a proxy for municipality-year average performance, namely the log of municipality-by-year sales;[26] comfortingly, estimates reported in Column (4) of Table 4 confirm our overall results.

Table 4: Impact of HSR on TFP. Controls.

	(1) LLM Trends	(2) Other Transport	(3) LLM Trends + Other Transport	(4) Bernard AV Sales	(5) Prov-by-Year FE
Distance 0-30 Km	0.03963***	0.02948***	0.04717***	0.04717***	0.05301***
	(0.01041)	(0.01110)	(0.01328)	(0.00640)	(0.01211)
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Observations	137,977	137,977	137,977	137,977	137,977

Notes: All the models include firm and year fixed effects. Across columns, the reported coefficient for *Distance 0-30 Km* is the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Columns 1 - 5 differ from one another for the set of controls considered. Column (1) includes local labour market linear trends; Column (2) considers the distance of each firm to the nearest airport, highway exit and traditional railroad station, as measures of the accessibility to other transport infrastructure, these covariates are demeaned and interacted with year dummies and with the treatment following Wooldridge (2021). Column (3) combines the previous set of controls. Column (4) includes a proxy for municipality-year average performance, whereas Column (5) adds province-by-year fixed effects.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.2 Heterogeneity Analysis

To further improve the understanding of our results we explore different possible layers of heterogeneity. First, we estimate our baseline model after splitting the sample according to the revised Pavitt taxonomy, as proposed by Bogliacino and Pianta (2016), that "can capture major structural differences in the relationship between innovation and performance". Indeed, as mentioned in the previous Section, our findings of improved total factor productivity for firms close to the Reggio Emila HSR station might be related to an increase in treated firms' innovation performance associated with improved knowledge flows and spillovers among human capital and researchers in

[26]We build municipality sales using our firm-level sample; moreover, to alleviate endogeneity concerns, we exclude, for the generic firm i, its sales when computing the municipality-level sales

different regions generated by the enhanced regional accessibility.[27]

In particular, the innovation literature has shown that knowledge spillovers are localized in the range of about 80 km (e.g. Murata et al., 2014), thus suggesting the importance of geographical proximity for firms, especially for knowledge-intensive ones. Moreover, some authors (e.g. Nakajima et al., 2010) argued that knowledge movements take place with that of its inventors, thus suggesting a new channel of long-distance knowledge transfer if inventors or researchers can travel more easily. Hence, transport infrastructure investments that increase potential opportunities for face-to-face interactions and exposure to foreign or external knowledge in any region, might improve firms' innovative performance and therefore total factor productivity.

The Pavitt taxonomy is based on the differences in the way technology is developed and used across industries that, in turn, might affect the impact of HSR infrastructure on firms' innovative performance. The taxonomy originally included Science-Based Industries (SB), Specialized Suppliers Industries (SS), Scale Intensive Industries (SI) and Supplier Dominated Industries (SD) and was subsequently modified to classify service industries.[28] In particular, Scale Intensive industries have been extended to include Information Intensive industries (now Scale and Information Intensive (SII)), whose key technological focus is on the design, usage and enhancement of big information-processing ICT systems. Estimates reported in columns from (3) to (6) of Table 5 suggest that the impact of the opening of the Reggio Emilia HSR station is concentrated in SII and SD industries, with no significant effect for the other sectors. On the one hand, these results may appear counter-intuitive, given that these industries are not those employing a large fraction of high human capital workers, which are typically more mobile. However, firms belonging to SB industries are significantly under-represented in our sample (just about 5% of total observations), so that the respective coefficient is likely to be poorly estimated. On the other hand, SD and SII industries that are more focused on process innovation might have benefited from the reductions in travel costs that have favoured access to a wider pool of suppliers from other industries and/or made it possible to reach new markets. A possible increase in process innovation induced by HSR infrastructure might have induced a growth in firms' total factor productivity through a reduction in production cost and optimization of the time to market. Moreover, it is important to remind

[27] A positive impact of HSR on local innovative capacity has been found, among other authors, by Hanley et al. (2022) and Dong et al. (2020) for the case of China and by Miwa et al. (2022) for Japan.

[28] Science-based industries include sectors where innovation is based on advances in science and research and development leading to significant product innovation and patent activity; Specialized Suppliers industries include the sectors whose products are new processes for other industries. An important innovative input is tacit knowledge and design skills embodied in the labour force; Scale Intensive industries include industries where scale economies are relevant. Important process innovation coexists with new product development; Supplier Dominated industries include traditional sectors where small firms are prevalent and technological change is introduced through the inputs and machinery provided by suppliers from other industries.

that SII industries include non-metallic mineral products and motor vehicles, while SD industries include the food and drink sectors that are all well represented in the neighbourhood of Reggio Emilia, where some important industrial districts are located (e.g. the motor valley, the ceramic and tile districts and some food districts).

Another factor that might drive the impact of HSR infrastructure on firms' performance is the nature of the firm activity, namely if the firm is operating in the manufacturing or service sectors. As predicted by some theoretical models (e.g. Koster et al. (2022)) business to business services might benefit relatively more from the HSR service that moves people instead of goods. Estimates reported in Columns (1) and (2) of Table 5 show that there is no difference in the impact of HSR on manufacturing/services firms' TFP.

Table 5: Heterogeneity by sector

	(1) Manufacturing	(2) Services	(3) Science Based	(4) Specialised Suppliers	(5) S.I. Intensive	(6) Suppliers Dominated
Distance 0-30 Km	0.03236**	0.03960**	-0.04044	0.01844	0.05612**	0.07282***
	(0.01290)	(0.01668)	(0.05408)	(0.02176)	(0.02381)	(0.01288)
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Observations	61,461	66,205	5,864	30,251	12,906	71,230

Notes: All the models include firm and year fixed effects. Across columns, the reported coefficient for *Distance 0-30 Km* is the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Columns 1 - 6 show the estimated effect on the subsample of firms grouped by their main sector of activity. In particular, Columns (3) - (6) refer to the revised Pavit taxonomy proposed by Bogliacino and Pianta (2016).
Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The size of the firm is an additional characteristic that might provide interesting insights in terms of the heterogeneity of the HSR impact. Firms' size has been often considered the origin of the low TFP performance of Italian firms over the past decades, as noted in Calvino et al. (2022). To analyse this aspect, we consider firms' size observed before the opening of the HSR station (2010) and we consider the balanced panel in order to get rid of firms' size dynamics that cannot be properly accounted for, such as entry or exit from the panel that may be due either to genuine exit from the market or not. Estimates reported in Table 6 suggest that firms that have benefited more from HSR investments are those in the 10–50 employees size class. The DiD coefficient of interest implies that such firms have experienced a boost in TFP of about 10%, while for firms with more than 50 employees the coefficient, although positive, is not precisely estimated, probably because of the relatively small sample size of such group of firms. Interestingly, we do not find any significant effect for micro firms with less than 10 employees: a possible reason for this result is that micro firms may be just too small to benefit significantly from the wider labour and intermediate goods markets that the opening of a new HSR station may have favoured.

Table 6: Heterogeneity by size

	(1) 0-10 Employees	(2) 10-50 Employees	(3) >50 Employees
Distance 0-30 Km	0.03123	0.09661***	0.04233
	(0.02132)	(0.02469)	(0.04839)
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
Observations	17,883	13,392	3,852

Notes: All the models include firm and year fixed effects. Across columns, the reported coefficient for *Distance 0-30 Km* is the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Columns 1 - 3 show the estimated effect on the subsample of firms grouped by their size class.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Finally, we find that the positive effects on firms' TFP are larger in the case of firms that, before the opening of the HSR station, had relatively lower levels of TFP (below the sample median). In particular, the greater the distance from the TFP frontier, the greater the impact of HSR on firms' TFP. Also in this case we estimate our baseline model over the balanced sample, after computing the distribution of firms' TFP in 2010. We believe that this result is particularly interesting since it suggests that the opening of the new HSR station might have contributed to lower TFP dispersion across firms, thanks to the reduction of spatial barriers to the transmission of knowledge from more innovative firms (Comin et al., 2012) and/or from private or public research institutions. In particular, access to the HSR network might have supported the improvement of those factors that favour firms' absorptive capacity (e.g. market knowledge, managerial resources, innovation cooperation, workers' training), thus facilitating knowledge spillovers and technology transfer. Indeed, technology adoption needs knowledge acquisition, which frequently arises from interactions with other agents, whose frequency and success are likely to be shaped by geography and transportation costs.

5 Conclusions

Using a Difference-in-Differences (DiD) research design, we leverage on the quasi-random location of the Italian HSR station of Reggio Emilia to analyze the impact of the connection to the HSR network on the evolution of local firms' total factor productivity. We consider as treated firms

Table 7: Heterogeneity by TFP quartiles

	(1) p(25)	(2) p(50)	(3) p(75)	(4) p(100)
Distance 0-30 Km	0.11042***	0.05786**	0.02586	0.05719
	(0.02689)	(0.02382)	(0.02724)	(0.04226)
Firm FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Observations	8,919	8,910	8,910	8,910

Notes: All the models include firm and year fixed effects. Across columns, the reported coefficient for *Distance 0-30 Km* is the average of the estimated yearly treatment effects for the treatment period 2013-2018. Yearly treatment effects are given by the interaction between year dummies for the post-treatment period and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Columns 1 - 3 show the estimated effect on the subsample of firms grouped by their pre-treatment TFP quartile.
Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

those located within 30 km from the HSR station, while those between 30 and 60 km of distance from the station represent the control group. Indeed, our setting is a clear illustration of the incidental treatment strategy to identification (Redding and Turner, 2015) in a DiD study design, since Reggio Emilia is almost on the straight line linking Bologna and Milan, two other important HSR stations. Additionally, much like in the German case examined by Ahlfeldt and Feddersen (2018) and Heuermann and Schmieder (2019), the decision to have the HSR stop at Reggio Emilia was the outcome of political negotiations, so it is plausible to assume that the location of the HSR near Reggio Emilia was as good as randomly assigned.

According to our study, treated firms' TFP grew by an average of 4.7% as a result of the HSR station in Reggio Emilia, with an effect that is gradually getting stronger over the sample period, as suggested by the event study estimates. Overall results are confirmed after controlling for possible spatial time-varying heterogeneity, for firms' accessibility to other transport networks or municipal-specific time-varying characteristics. Interestingly, we find a higher impact of HSR connection for firms located relatively far from an airport and for those that are very close to the HSR station (within 10 km). Furthermore, the heterogeneity analysis provides valuable insights since it suggests that firms whose TFP levels were below the sample median before the opening of the HSR station have been those to be more positively affected by the connection to the HSR network.

Our findings are in line with the predictions of those theoretical models analysing the impact of transport infrastructure investments on the spatial distribution of economic activity, with a particular focus on HSR networks (Bernard et al., 2019; Koster et al., 2023). Even if we do not directly observe the employment pattern of our investigated area, we do find a positive impact on treated firms' TFP that is broadly consistent also with the model by Koster et al. (2023). Finally, our results are also broadly in line with Crescenzi et al. (2011) who suggest that improved accessibility associated to transport infrastructure investments can improve firms' innovative performance, through higher knowledge spillovers and information flows, which in turn can foster firms' TFP.

We argue that overall findings can be interpreted as the result of *Marshallian* economies, associated to better access to inputs, people and ideas, generating firms' productivity gains. Indeed, some authors have shown that the benefits of agglomeration are greater for firms having a higher level of internal resources, as proxied by age, size, and ownership status, among others (e.g. Rigby and Brown (2015)), so that we can interpret results stemming from the heterogeneity analysis along these lines. In particular, the finding of higher productivity gains for firms belonging to the lower part of the TFP distribution adds further insights into this issue and implies that HSR investments can reduce TFP dispersion across firms by favouring the catch-up of laggard ones.

It is worth noting that, because the opening of the HSR station in Reggio Emilia was the result of a political bargain, the random assignment assumption seems quite reasonable, which in turn ensures a strong internal validity to our study, even if this is not necessarily true as far as the external validity is concerned. Indeed, the specific characteristics of the region around Reggio Emilia as well as the fact that we only focus on one single station of the Italian HSR network can raise concerns about the external validity of our analysis. In this respect, it seems interesting to develop a follow-up study that considers the entire Italian HSR network; such analysis might also shed further light on the possible displacement effects generated by transport infrastructure investments. Indeed, such investments aim to strengthen EU connectivity and ease peripheral isolation, but they may also increase the dominance of large urban areas. Moreover, we believe that it would be interesting to analyze in detail whether the effects on firms' TFP are associated with the interactions between changes in workers' flows across space and innovative activity.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. Econometrica, 83(6):2411–2451.

Agrawal, A., Galasso, A., and Oettl, A. (2017). Roads and innovation. Review of Economics and Statistics, 99(3):417–434.

Ahlfeldt, G. M. and Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. Journal of Economic Geography, 18(2):355–390.

Alderighi, M. and Gaggero, A. A. (2017). Fly and trade: Evidence from the italian manufacturing industry. Economics of transportation, 9:51–60.

Bajgar, M., Berlingieri, G., Calligaris, S., Criscuolo, C., and Timmis, J. (2020). Coverage and representativeness of Orbis data. OECD Science, Technology and Industry Working Papers 2020/06, OECD Publishing.

Baltrunaite, A. and Karmaziene, E. (2020). Trainspotting: board appointments in private firms. Bank of Italy Temi di Discussione (Working Paper), 1278.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics, 99(3):435–448.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019). Production networks, geography, and firm performance. Journal of Political Economy, 127(2):639–688.

Blanquart, C. and Koning, M. (2017). The local economic impacts of high-speed railways: theories and facts. European Transport Research Review, 9(2):1–14.

Bogliacino, F. and Pianta, M. (2016). The pavitt taxonomy, revisited: patterns of innovation in manufacturing and services. Economia Politica, 33:153–180.

Bottasso, A., Conti, M., Robbiano, S., and Santagata, M. (2022). Roads to innovation: Evidence from italy. Journal of Regional Science, 62(4):981–1005.

Calvino, F., DeSantis, S., Desnoyers-James, I., Formai, S., Goretti, I., Lombardi, S., Manaresi, F., and Perani, G. (2022). Closing the italian digital gap: the role of skills, intangibles and policies. OECD Science, Technology and Industry Policy Papers, (126).

Comin, D. A., Dmitriev, M., and Rossi-Hansberg, E. (2012). The spatial diffusion of technology. National Bureau of Economic Research - WP 18534.

Crescenzi, R., Rodríguez-Pose, A., Crescenzi, R., and Rodríguez-Pose, A. (2011). Geographical accessibility and human capital accumulation. Innovation and Regional Growth in the European Union, pages 31–49.

De Loecker, J. (2013). Detecting learning by exporting. American Economic Journal: Microeconomics, 5(3):1–21.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. Econometrica, 84(2):445–510.

De Loecker, J. and Syverson, C. (2021). Chapter 3 - An industrial organization perspective on productivity. In Ho, K., Hortaçsu, A., and Lizzeri, A., editors, Handbook of Industrial Organization, volume 4 of Handbook of Industrial Organization, pages 141–223. Elsevier.

Dong, X., Zheng, S., and Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. Journal of Urban Economics, 115:103212.

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.

Ferrari, C., Bottasso, A., Conti, M., and Tei, A. (2018). Economic role of transport infrastructure: theory and models. Elsevier.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to success: the impact of the golden quadrilateral project for the location and performance of Indian manufacturing. The Economic Journal, 126(591):317–357.

Gibbons, S., Lyytikäinen, T., Overman, H. G., and Sanchis-Guarner, R. (2019). New road infrastructure: the effects on firms. Journal of Urban Economics, 110:35–50.

Gibbons, S. and Wu, W. (2020). Airports, access and local economic performance: evidence from China. Journal of Economic Geography, 20(4):903–937.

Graham, D. J., Gibbons, S., and Martin, R. (2009). Transport investment and the distance decay of agglomeration benefits. Report to the Department of Transport.

Hanley, D., Li, J., and Wu, M. (2022). High-speed railways and collaborative innovation. Regional Science and Urban Economics, 93:103717.

Heuermann, D. F. and Schmieder, J. F. (2019). The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in Germany. Journal of Economic Geography, 19(2):335–372.

Holl, A. (2016). Highways and productivity in manufacturing firms. Journal of Urban Economics, 93:131–151.

Inoue, H., Nakajima, K., and Saito, Y. U. (2017). The impact of the opening of high-speed rail on innovation. RIETI.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., and Yesiltas, S. (2023). How to construct nationally representative firm level data from the orbis global database: New facts on smes and aggregate implications for industry concentration. American Economic Journal: Macroeconomics.

Koster, H. R., Hayakawa, K., and Tabuchi, T. (2023). High'speed rail and the spatial distribution of economic activity: Evidence from japanrs shinkansen. WP.

Koster, H. R., Tabuchi, T., and Thisse, J.-F. (2022). To be connected or not to be connected? The role of long-haul economies. Journal of Economic Geography, 22(4):711–753.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. The review of economic studies, 70(2):317–341.

Li, Z., Wu, M., and Chen, B. R. (2017). Is road infrastructure investment in China excessive? Evidence from productivity of firms. Regional Science and Urban Economics, 65:116–126.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from china's high speed railway system. Journal of Urban Economics, 98:98–123.

Miwa, N., Bhatt, A., Morikawa, S., and Kato, H. (2022). High-speed rail and the knowledge economy: Evidence from japan. Transportation Research Part A: Policy and Practice, 159:398–416.

Murata, Y., Nakajima, R., Okamoto, R., and Tamura, R. (2014). Localized knowledge spillovers and patent citations: A distance-based approach. Review of Economics and Statistics, 96(5):967–985.

Nakajima, R., Tamura, R., and Hanaki, N. (2010). The effect of collaboration network on inventors' job match, productivity and tenure. Labour economics, 17(4):723–734.

Neumark, D. and Simpson, H. (2015). Chapter 18 - Place-based policies. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, Handbook of Regional and Urban Economics, volume 5 of Handbook of Regional and Urban Economics, pages 1197–1287. Elsevier.

Qin, Y. (2016). 'No county left behind?' The distributional impact of high-speed rail upgrades in China. Journal of Economic Geography, 17(3):489–520.

Redding, S. J. and Turner, M. A. (2015). Chapter 20 - Transportation costs and the spatial organization of economic activity. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, Handbook of Regional and Urban Economics, volume 5 of Handbook of Regional and Urban Economics, pages 1339–1398. Elsevier.

Rigby, D. L. and Brown, W. M. (2015). Who benefits from agglomeration? Regional Studies, 49(1):28–43.

Rovigatti, G. and Mollisi, V. (2018). Theory and practice of total-factor productivity estimation: The control function approach using stata. The Stata Journal, 18(3):618–662.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. Available at SSRN 3906345.

Appendix A

Table A.1: Descriptive statistics.

	Mean T	Sd Mean T	Mean C	Sd Mean C	Diff(T-C)	P-val Diff	N obs T	N obs C
Firm characteristics								
Number of employees [PRE]	30.596	193.120	33.290	103.616	-2.690	0.420	8,791	3,760
Number of employees [POST]	28.892	175.399	30.893	98.857	-2.000	0.359	17,377	7,328
Sector of activity								
Manufacturing [PRE]	0.453	0.498	0.460	0.498	-0.010	0.197	27,483	11,750
Manufacturing [POST]	0.434	0.496	0.459	0.498	-0.020	0.000	68,813	29,931
Services [PRE]	0.491	0.500	0.425	0.494	0.070	0.000	27,483	11,750
Services [POST]	0.507	0.500	0.430	0.495	0.080	0.000	68,813	29,931
Science based [PRE]	0.046	0.209	0.032	0.177	0.010	0.000	27,483	11,750
Science based [POST]	0.047	0.213	0.032	0.176	0.020	0.000	68,813	29,931
Specialised suppliers [PRE]	0.238	0.426	0.194	0.396	0.040	0.000	27,483	11,750
Specialised suppliers [POST]	0.228	0.419	0.192	0.394	0.040	0.000	68,813	29,931
S.I. intensive [PRE]	0.103	0.304	0.087	0.282	0.020	0.000	27,483	11,750
S.I. intensive [POST]	0.098	0.297	0.078	0.268	0.020	0.000	68,813	29,931
Suppliers dominated [PRE]	0.495	0.500	0.539	0.498	-0.040	0.000	27,483	11,750
Suppliers dominated [POST]	0.509	0.500	0.542	0.498	-0.030	0.000	68,813	29,931
Time invariant controls								
Distance tollbooth	6.955	4.606	12.995	8.845	-6.040	0.000	27,483	11,750
Distance railway station	2.565	2.185	5.366	5.362	-2.800	0.000	27,483	11,750
Distance airport	28.435	12.977	28.367	12.598	0.070	0.629	27,483	11,750

Notes: T and C identify treated and control firms, respectively. Firms located within a 0-30 Km radius from the HSR station are considered as treated. The post period refers to 2013-2018, while the pre covers 2010-2011. For consistency with the analysis, statistics on the number of employees are computed over the balanced sample. Statistics on time-invariant controls refer to the pre-treatment period.

Table A.2: Event Study

	(1) Unbalanced	(2) Balanced	(3) LLM Trends	(4) Province-by-Year FE	(5) Other Transport
Distance 0-30 Km x 2010	-0.05591**	-0.05131*	-0.05288**	-0.04453	-0.04676*
	(0.02195)	(0.02854)	(0.02273)	(0.02787)	(0.02788)
Distance 0-30 Km x 2011	-0.00561	0.00489	-0.00504	-0.00389	-0.00401
	(0.01362)	(0.02234)	(0.01392)	(0.01720)	(0.01720)
Distance 0-30 Km x 2013	0.00644	0.00571	0.00571	0.01034	-0.03424*
	(0.01304)	(0.02076)	(0.01332)	(0.01623)	(0.01757)
Distance 0-30 Km x 2014	0.01419	0.02094	0.01242	0.02428	-0.02023
	(0.01331)	(0.02041)	(0.01449)	(0.01636)	(0.01772)
Distance 0-30 Km x 2015	0.06915***	0.11108***	0.06639***	0.07980***	0.03517*
	(0.01390)	(0.02252)	(0.01634)	(0.01744)	(0.01869)
Distance 0-30 Km x 2016	0.06716***	0.10019***	0.06342***	0.07776***	0.03312*
	(0.01419)	(0.02333)	(0.01831)	(0.01784)	(0.01913)
Distance 0-30 Km x 2017	0.08208***	0.11178***	0.07773***	0.09059***	0.04603**
	(0.01478)	(0.02602)	(0.02065)	(0.01838)	(0.01963)
Distance 0-30 Km x 2018	0.08372***	0.12203***	0.07877***	0.10267***	0.05810***
	(0.01572)	(0.02737)	(0.02354)	(0.01965)	(0.02081)
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Observations	137,977	35,649	137,977	137,977	137,977

Notes: All the models include firm and year fixed effects. Across columns, the reported coefficients fare the early treatment effects given by the interaction between year dummies and a dummy variable which is equal to 1 if a firm is located in the 0-30 Km radius from Reggio Emilia and at least 60 Km from Bologna. Columns 1 and (2) show the results for the unbalanced and balanced panel, respectively. Columns (3) – (5) include additional covariates as from 4, Columns (1), (2) and (5), respectively.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1