The Innovation Dividend of High Speed Rail: Evidence from Italy*

Anna Bottasso[1], Federico Mij[1], and Simone Robbiano[2]

[1]Department of Economics, University of Genoa
[2]Department of Economics, University of Genoa and Global Labor Organization

October 2025

Abstract

In this work we analyze the impact of high-speed rail (HSR) connection on regional innovation performance in Italy. Using a balanced panel of 89 NUTS-3 regions observed over the period 1980–2019, we measure innovation with forward citation–weighted patent fractional counts from the EPO PATSTAT database. We conduct a causal analysis that exploits the staggered opening of HSR stations in a extended two-way fixed effects (ETWFE) difference-in-differences (DiD) design. We find that HSR access increases regional innovation by economically large and statistically significant amounts (overall ATT on the order of 0.24–0.33 log points), with dynamics that remain positive several years after openings. A more detailed analysis finds gains along both quality (citations) and extensive margins (patent share and inventors). Moreover, estimates of a dyadic gravity model show that HSR boosts inter-regional inventor collaboration by about 30%, suggesting reduced effective distance as a key mechanism for improved innovative performance, while a social network analysis (SNA) reveals an increase in the degree of centrality in the regional innovation network associated to better HSR access. Sectoral results point to especially strong responses in chemistry, electrical and mechanical engineering.

Keywords: Transport infrastructures and planning; High speed rail; Government investments; Difference in differences; Innovation and invention; Regional economic activity.

JEL Codes: R11; R42; L92; O31; O33; C23.

1 Introduction

The contribution of transport infrastructure investments to growth has been widely examined in economics and regional science. Recent surveys [e.g. Ferrari et al., 2018] and meta-analyses [e.g. Melo et al., 2013] report significant effects across multiple outcomes. By lowering transport costs, such investments can raise the productivity of other inputs, reduce production costs, expand trade and competition by enlarging relevant markets, and enable the exploitation of scale economies. Greater accessibility also increases places' market potential, shaping the spatial allocation of human capital and economic activity via agglomeration economies. A further channel operates through innovation: cheaper and faster

*Declarations of interest: none. All of the sources of funding for the work described in this publication are acknowledged below: PRIN (Research Projects of National Relevance) - Italian Ministry of Education, University and Research (MIUR) - Project "INFINITE - INfrastructure: Firms and Individual ouTcomEs" ID D53D23006270006. The authors would like to thank the volunteers at the Ligurian Railway Museum (`https://www.museoferroviarioligure.it/index.html`) for kindly providing access to the original historical volumes relating to the planning and construction of the Italian high-speed railway network.

connections facilitate interactions among inventors, favoring the creation and diffusion of knowledge and helping localized ideas to spread across space.

Among other transport modes, high-speed railway (HSR) has emerged as a cornerstone of spatial and industrial policy across Europe and China, with far-reaching consequences for regional economic development, based on expected productivity gains, deeper market integration and reductions in generalized transport costs.[1] By lowering travel time between cities, HSR reduces the effective distance that separates firms, laboratories and inventors thus increasing the frequency and quality of face-to-face meetings, which are critical for exchanging tacit knowledge; moreover, improved accessibility also enlarges local matching markets for specialized skills, labs, and equipment, raising the probability of productive firm–researcher pairings and shortening project cycle times. It is therefore unsurprising that such transport investments sit at the heart of place-based policy [Neumark and Simpson, 2015]; however, economic theory remains divided on whether they reliably spur local growth or instead reallocate activity across space, potentially exacerbating disparities [Redding and Turner, 2015]. Moreover, HSR schemes involve multi-billion-euro, largely irreversible public outlays with long-lived distributional effects, so that rigorous causal impact evaluations become essential for policymakers to assess value for money, verify realized effects and calibrate complementary measures.

The literature has documented both direct and indirect economic effects of major transport projects, yet identifying the latter remains challenging because of simultaneity, causal direction, and mixed findings across scales. Reductions in transport costs and travel times are tightly connected to agglomeration mechanisms [Koster et al., 2021]. As far as HSR is concerned, scholars point to multiple channels—from short-run construction demand to longer-run shifts in productivity, consumption geography, and the relocation of firms and households—but evidence on magnitudes and persistence is mixed [Blanquart and Koning, 2017, Bottasso et al., 2025]. Moreover, better transport infrastructures can relax frictions in knowledge creation and diffusion [Agrawal et al., 2017] and may amplify knowledge interactions by enhancing inter-regional accessibility [Crescenzi and Rodríguez-Pose, 2011]. However, these gains might not be evenly distributed across space. Network-wide reallocation, general-equilibrium price adjustments, and changes in firm-to-firm matching can shift benefits across places, thus yielding heterogeneous local impacts, even when aggregate benefits rise [Bernard et al., 2019]. Theory indicates that lower transport costs may draw especially skilled workers toward larger cities, potentially reducing employment in smaller regions [Koster et al., 2021], while related models emphasize sizable average gains alongside marked local heterogeneity [Hayakawa et al., 2021].

Within the literature that has examined the relationship between transport infrastructure and innovation, HSR has emerged as a salient catalyst of knowledge exchange, with a growing body of quasi-experimental evidence—often from China, linking HSR access to higher inventive activity, improved patent quality, and denser cross-city collaboration networks [e.g. Dong et al., 2020, Hanley et al., 2022, Sun et al., 2021]. These studies typically identify reductions in effective distance that lower coordina-

[1] Modern HSR networks began with Japan's Shinkansen in 1964, which set the template for subsequent systems worldwide. In Europe, early adoption followed in Italy with the Florence–Rome *Direttissima* (first sections opened in 1977), then France's TGV (1981), Germany's ICE (1991) and Spain's AVE (1992). The European Union's dedicated HSR network has expanded rapidly over the past decade to 8,556 km in 2023, with Spain leading by network length and France second. China's build–out since 2008 is unprecedented: by end–2024 it operated about 48,000 km of HSR, with plans for roughly 60,000 km by 2030. Recent global developments include the first African and Middle Eastern HSR lines—Morocco's Al Boraq (2018) and Saudi Arabia's Haramain (2018)—and Southeast Asia's first true HSR, Jakarta–Bandung (commercial service from October 2023). New projects are advancing or under construction in multiple regions, e.g. Brightline West (Las Vegas–Southern California; U.S. groundbreaking April 2024), Portugal's Lisbon–Porto corridor (Phase 1 financing approved), Egypt's 2,000 km network with Siemens Mobility, Thailand's Bangkok–Nong Khai line targeting a China linkage around 2030, and Vietnam's approved 1,541 km North–South HSR.

tion costs and expand market and labor-market catchment areas, thereby facilitating matching, learning, and knowledge recombination; moreover, they underscore impact heterogeneity by city size, pre-existing specialization, and network position, with some places specializing while others diversifying in response to altered market access [e.g. Chen and Guo, 2023]. Recent contributions refine the mechanisms through which HSR reshapes the geography of ideas showing, for instance, how new links "bridge" innovation distances via institutional intermediation such as patent agencies [Wu and Yu, 2025], and how accessibility shocks propagate through wider spatial equilibria with potential siphoning towards well-connected hubs [Gao and Zheng, 2020], thus raising distributional and cohesion concerns. Beyond aggregate patenting, recent works also examine environmental and organizational dimensions, associating HSR with urban green innovation [Wen et al., 2025] and firm-level capacity utilization [Lu and Li, 2022]. In contrast, studies from Japan reveal effects that tend to be context-specific and localized [Inoue et al., 2017, Miwa et al., 2022]. What is more, [Cao et al., 2024] suggest that "distance is not dying uniformly", meaning that HSR may cut travel times, but its innovation pay-offs vary by activity and place. Indeed, codified knowledge moves easily, while tacit, trust-based know-how often still needs co-location, so gains are uneven. Moreover, network position and local complements (skills, R&D assets, last-mile links) determine who can convert faster trips into collaboration and ideas, so that HSR can catalyze innovation only where complementary capabilities and institutions are in place.

Finally, credible identification of the research design remains a central challenge in the literature, given non-random network localization and spatial spillovers, motivating the methodological shift towards counterfactual impact evaluation approaches, like difference-in-differences, instrumental variables and related designs, alongside careful attention to network endogeneity and general-equilibrium feedbacks [e.g. Lin et al., 2021, Bhatt and Kato, 2021].

Despite the extensive research on the aggregate impacts of HSR, often focused on Chinese and Japanese contexts, the channels through which it relates to innovative activity are largely unexplored in the European HSR context, apart from the work of Gambuli and Stipanicic Márquez [2025] as far as we know, showing how a reduction in travel time of 12% led to a 2.6% increase in patent collaborations across French commuting zones.[2] In this study we start filling this gap by exploring whether—and through which channels—HSR access might raise Italian regional innovative performance.[3] In order to address this issue, we build a 1980–2019 panel for 89 provinces, regionalising EPO PATSTAT patents by inventor address and constructing forward citation-weighted patent fractional counts (PFC) as our main outcome. We complement these data with both unweighted PFC and inventors counts, WIPO technology classes classification, institutional quality, multimodal accessibility, as well as a rich set of economic covariates, and we geocode the national HSR network and station openings from official sources.[4]

[2]It is worth noting that, for Europe, an eminent work, even if not directly related to innovation, is that of Ahlfeldt and Feddersen [2018], which 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. For Italy, related works include Baltrunaite and Karmaziene [2020], that show how the introduction of HSR service resulted in a positive shock to the non-local managers supply, through their increased mobility, thus improving the quality of high-efficiency firms' boards; Bottasso et al. [2023], focusing on the opening of Reggio Emilia Mediopadana HSR station, suggest instead that the opening of the latter improved treated firms' total factor productivity of about 5%. It is worth noting that Italy is also historically salient, since it was the first European country to operate a high-speed line. The "Direttissima" opened indeed in 1977 as the first high-speed rail route in Italy and Europe, connecting Rome with Florence. The top speed on the line was 250 km/h (160 mph), giving an end-to-end journey time of about 90 minutes with an average speed of 200 km/h (120 mph).

[3]From here on, both the terms "region" and "NUTS-3 region", as well as "province" will be used interchangeably to indicate the Italian NUTS-3 statistical territorial unit.

[4]World Intellectual Property Organization (WIPO); see `https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=117672`.

The estimation strategy relies on an extended two-way fixed-effects (ETWFE) staggered DiD model [Wooldridge, 2021, 2025] that allows the HSR connection's influence on regional innovation to vary with predetermined regional characteristics and temporally, net of province and year fixed effects. Our identification strategy considers as treated those regions hosting an HSR station, while other regions represent the control sample and its validity is thoroughly analyzed by means of event study estimates and by the implementation of the Rambachan and Roth [2023]'s Honest DID approach. Moreover, in order to address potential simultaneity in treatment selection and timing, we instrument realized access to the HSR network with the planned one, exploiting delays between blueprint and opening. We also probe the robustness of our results by estimating our baseline model with alternative control groups (propensity score matching, PS-DiD) and by applying a synthetic DiD (SC-DiD) estimator; moreover, we check the robustness of our main findings by enlarging the treatment group to include also regions which border those that host the HSR stations, and by comparing main estimates to classic two-way fixed effects (TWFE) models.

Estimated average treatment effects on the treated (ATT) indicate that HSR connection raises forward citation-weighted PFC by about 0.24 to 0.33 log points, and classic TWFE yields nearly the same magnitudes. When analyzing our results by cohort, we observe that sizable gains are detected for the 2008 one (about 33–51%), which includes the opening of the HSR stations in Milan, Bologna, Naples and Salerno that marked a significant step toward the completion of the national HSR network; similarly, for the 2013 cohort (Reggio Emilia), estimates show that forward citation-weighted PFC increased by about 75–96%, while lower but still significant effects emerge for both 2007 and 2016 cohorts, namely Padua, Venice and Brescia, respectively (about 29-36%); finally, the impact of the completion of the Turin-Milan corridor in 2009 is poorly estimated.[5]

A more in depth analysis of the impact of HSR on regional innovation performance suggests movements of both the intensive and extensive margins: the unweighted patent share rises by about 17%, and the inventor share by about 21–23%, alongside the citation-weighted gains. Moreover, by considering an heterogeneity analysis by technological sectors, effects concentrate in chemistry (overall ATT equal to 0.48), electrical engineering (about 0.44) and mechanical engineering (about 0.44), with weaker or null responses elsewhere—consistent with domains where iterative testing, supply-chain coordination and tacit, face-to-face problem-solving benefit the most from transport time reduction.

In order to shed some light on the mechanisms behind the estimated improvement in the regional innovation performance, we estimate a dyadic gravity model of inventor collaborations between NUTS-3 region pairs and we find that inter-regional co-inventorship rises by around 30% following HSR station openings, consistently with lower frictions to face-to-face interactions, richer tacit knowledge exchange and a wider spatial reach of routine research mobility. As a complement of these flow-based results, a social-network analysis (SNA) of a co-patenting graph shows that HSR connected regions become more central within the national collaboration network: except for the 2007 cohort, regions' centrality rises across different metrics (overall 2–9%), with regions hosting an HSR station lying about 2.7% more often on shortest paths (betweenness), being roughly 9% closer to all other nodes (closeness), and recording an increase of around 2% in high-quality connections (eigenvector). Finally, we do not find evidence in favor of displacement effects at work in our sample, thus indicating that HSR strengthens innovation in connected regions rather than siphoning activity from nearby areas. Taken together, this evidence indicates that HSR access has strengthened pass-through roles, tightened collaboration ties,

[5]IV estimates that instruments realized HSR access with the planned network delivers comparable directions of the effects, but higher in magnitude, while passing weak-IV diagnostics (effective first-stage F = 49.6) and remaining robust under Conley et al. [2012]'s plausibly-exogenous bounds.

and deepened regional embeddedness in influential knowledge hubs, thereby amplifying knowledge flows, diffusion and innovation performance. Moreover, our findings yield clear policy implications and indicate that HSR functions as “knowledge infrastructure”: by expanding effective collaboration markets, it improves both the quality and the quantity, as well as the organization of innovative activity. These benefits are not uniform across places and sectors as they are strongest in intermediate and smaller regions, supporting the view that transport infrastructure can be an effective policy lever for narrowing regional economic disparities.

This work makes several contributions to the literature. It is the first paper that analyzes the impact of the opening of HSR stations on regional innovation performance in Italy and is particularly salient given that Italy was the first European country to invest in high-speed rail and has, over time, continued to build a network that links the north and south of the country. Moreover, our empirical findings help rationalize evidence from [Bottasso et al., 2023]'s firm-level study, which, for Reggio Emilia, documents an increase in total factor productivity (TFP) among firms located near the HSR station. This suggests that at least part of the observed TFP gains may be attributable to improvements in local innovative performance. It is worth noting that this study is one of the few papers that uses a particularly precise measure of innovation based on the specific contribution of local inventors which accounts for the quality of the invention, as proxied by the number of forward citations received. Finally, by providing well-identified causal estimates of the innovative pay-offs to high-speed rail access, this study offers a robust empirical yardstick that policymakers can set against capital and operating costs within formal cost–benefit appraisal. The estimated ATTs and their dynamics inform realistic benefit time paths, while the sectoral pattern of gains indicates where value is most likely to accrue. Together with extensive robustness analysis, the evidence provided by our analysis narrows uncertainty around the benefits of HSR investments and enables more transparent value-for-money assessments, supporting evidence-based decisions on whether, where, and how to prioritize comparable public investments. In particular, our findings of no displacement effects contribute to the debate of the role of transport infrastructure investments in mitigating regional economic disparities.

The reminder of the paper is as follows; Section 2 describes the data and our identification approach, while Section 3 provides results of our analyses. Finally, Section 4 concludes.

2 Data and Identification Strategy

2.1 Data and Descriptive Statistics

Our study scrutinizes the impact of HSR access on regional innovation, as proxied by per capita patent fractional counts, for 89 Italian NUTS-3 regions as defined in the national 1974 administrative setting.[6] Specifically, we primarily build upon per capita fractional patent counts weighted for forward citations as a measure of regional innovative activity.[7] The economic literature recognises patents as key instru-

[6]Given that the number of Italian NUTS-3 regions has changed frequently over time, we will consider the local administrative structure of 1974, which had 95 NUTS-3 regions. Moreover, we exclude from the latter the Sardinian provinces of Cagliari, Nuoro, Oristano and Sassari because they have no rail continuity with the mainland and thus effectively zero exposure to Italy's HSR network; their island-specific transport regime (air and Ro-Ro ferries) makes common support and parallel trends less credible. By contrast, we retain Sicilian regions because there is rail interoperability across the Strait of Messina: passenger and freight trains are ferried between Messina and Villa San Giovanni, so upgrades on the mainland HSR spine plausibly lower generalised travel costs for Sicilian firms via a ferry–rail chain. Furthermore, we exclude the provinces of Rome and Florence, given that the pre-existing Florence–Rome "Direttissima" HSR line–whose first section was inaugurated on 24 February 1977–makes them ever-treated within our sample window, leaving no clean pre-treatment period.

[7]Patent applications are allocated using the inventor criterion, i.e., they are attributed to the inventor's place of residence. Where a patent lists multiple inventors, the application is apportioned equally across inventors and, correspondingly, across

ments for appropriating the returns to inventive activity; moreover, technologies with greater welfare gains and growth potential are more likely to be patented [Pakes and Griliches, 1980]. Within this tradition, forward citations have been identified as an indirect measure of an invention's value, since the number of citations a patent receives proxies its relevance for subsequent technological developments [Trajtenberg, 1990, Hall et al., 2005]. That said, patents capture inventions rather than the entirety of innovative activity [Smith, 2006], and not all inventions are patented; however, the consensus in the innovation literature is that patents remain an effective indicator of local technological capacity.

We rely on annual patent data from the European Patent Office (EPO)'s PATSTAT® database, which provides bibliographic and legal-status information for multiple countries at the NUTS-3 level.[8] The dataset covers applications filed directly under the European Patent Convention and Euro-PCT applications designating the EPO and it includes detailed fields such as forward citations, applicants and inventors (and their attributes), the International Patent Classification (IPC) codes, and the Statistical Classification of Economic Activities in the European Community (NACE-2 sector classification).[9] We recover data for the 1980–2019 time-span and "regionalise" patents by the inventors' addresses (NUTS-3 codes) and we also calculate the number of inventors patenting in each region in a given year. What is more, following the WIPO technology classification derived from IPC codes, we also group patents into five technological classes, namely electrical engineering, instruments, chemistry, mechanical engineering, and a residual category.[10]

Turning to HSR access information, we leverage on a geocoding procedure envisaging the use of the software QGIS.[11] We firstly obtain the 1974 administrative boundaries from the Italian National Institute of Statistics (ISTAT) shapefiles.[12] Information on Italy's High-Speed/High-Capacity rail network (Alta Velocità–Alta Capacità, AV–AC) is sourced from OpenStreetMap via QGIS plugins.[13] We then construct a vector line layer for HSR AV–AC routes and a vector point layer for AV–AC stations. Finally, information on the staggered opening of HSR stations have been recovered from official sources of the Italian network operator, namely Rete Ferroviaria Italiana (RFI).[14] Figure 1 depicts the Italian HSR network.

Another primary data source is the ARDECO dataset (Joint Research Centre, European Commission), which provides harmonized provincial accounts for consistent long-run, policy-oriented economic analysis.[15] The dataset covers a number of time series, allowing to control for predetermined regional characteristics and to track sectoral and demographic shifts associated with high-speed rail access. In

their NUTS-3 regions of residence.

[8]See https://www.epo.org/en/searching-for-patents/business/patstat.

[9]WIPO IPC-based technology field classification. Source: WIPO IPC Technology Concordance Table (https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=117672). For details on NACE-2 sector classification see https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-07-015.

[10]The patent panel dataset ends in 2019 given that COVID-19 pandemic strongly negatively affected innovation spending in the short run [Trunschke et al., 2024].

[11]QGIS is an open-source geographic information system (GIS) software that supports viewing, editing, printing, and analyzing geospatial data. See https://qgis.org/.

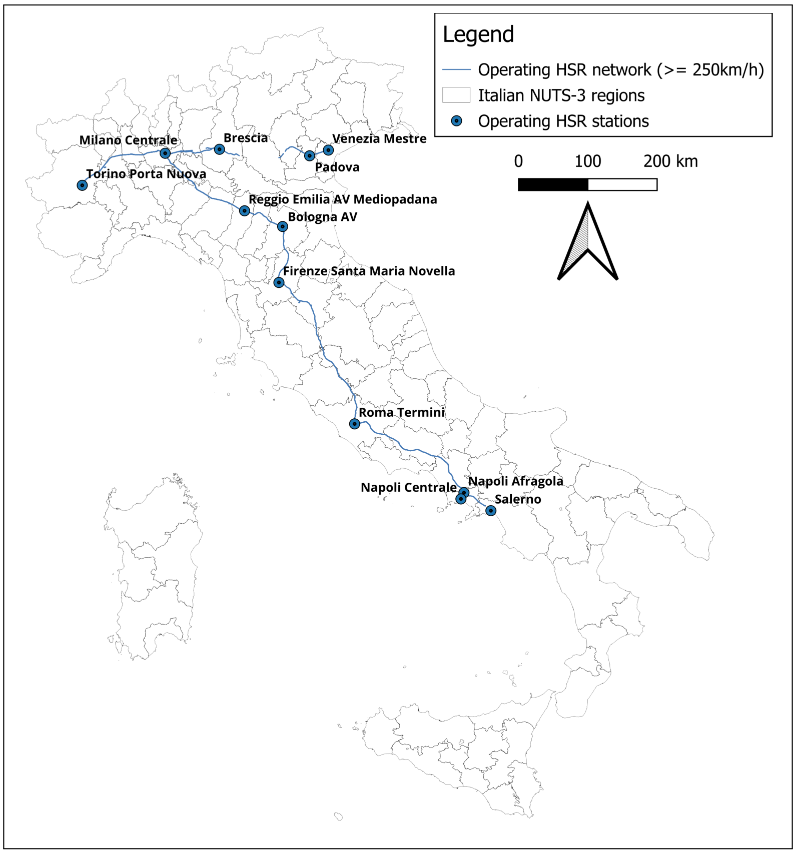
[12]A shapefile is a vector GIS format that represents spatial features as points, polylines, or polygons and is widely used across environmental, geoscientific, and other spatial applications. Our project uses the WGS 84 / UTM zone 32N coordinate reference system (EPSG:32632). Source: https://www.istat.it/it/informazioni-territoriali-e-cartografiche.

[13]High-speed rail generally refers to any line with a maximum speed exceeding 250 km/h (155 mph).

[14]See https://www.rfi.it/.

[15]ARDECO—the Annual Regional Database—is compiled by the European Commission's Directorate-General for Regional and Urban Policy and regularly maintained by the Joint Research Centre. It provides extensive time-series indicators for EU regions and selected EFTA and candidate countries across multiple statistical levels (NUTS-1, NUTS-2, NUTS-3, and metropolitan areas). The data are organized into thematic chapters covering population, GDP, employment, labor costs, labor productivity, capital formation, capital stock, and households. See https://territorial.ec.europa.eu/ardeco/explorer?lng=en.

Figure 1: HSR network in Italy.



Notes: Authors' elaboration via QGIS software.

particular, information used in this paper refer to both regional size and production structure (per capita GDP, industry share on GVA), as well as to labor costs (unit labor costs and compensations per hour worked).

We also recover the Institutional Quality Index (IQI) developed by Nifo and Vecchione [2015], that measures governance quality across Italian regions and is based on five pillars: voice and accountability (civic participation and cultural engagement), government effectiveness (quality of services and infrastructure), regulatory quality (business environment and market openness), rule of law (crime, judicial efficiency, tax evasion), and control of corruption (misuse of public power, municipal dissolutions, Golden–Picci index). Each component is built from multiple indicators, normalized, and aggregated into a composite score, with higher values reflecting better institutional quality.[16]

Finally, we consider an indicator of potential multimodal accessibility to capture absolute changes in accessibility across regions over time. This ESPON's measure is a gravity-type index summarizing how easily a region can interact with population and economic mass via the European transport system. It integrates road, rail, and air networks (with intermodal transfers) into a single normalized NUTS-3 metric that reflects both the size of reachable destinations and the generalized travel times to reach them. Higher values denote faster access to larger markets; lower values indicate relative isolation.[17]

The final result of this data collection is a balanced panel covering the aforementioned 89 Italian NUTS-3 regions over the 40-year period 1980-2019 (3,560 observations); Table 1 reports descriptive statistics of the main variables related to regional innovation, namely the forward citations-weighted

[16] See `https://sites.google.com/site/institutionalqualityindex/home`.

[17] See `https://database.espon.eu/indicator/1541/#metadata-download` for methodological details.

patent fractional count described above and the number of inventors patenting in each region, for both the full sample and separately for the regions with/without an HSR station. On average, the NUTS-3 regions connected to the HSR network exhibits higher values for both variables with respect to unconnected ones. It is worth noting that these differences do not represent an identification issue as long as the parallel trends assumption between treated and control groups holds.

Table 1: Descriptive statistics of the main variables

Variable	Mean	SD	Min	p25	Median	p75	Max	Obs
				Full Sample				
Forward citations-weighted PFC	444.037	719.833	0.830	43.344	186.873	579.562	14,101.392	3,560
Number of Inventors	88.300	107.160	0.830	14.854	45.289	120.048	872.754	3,560
				Regions without HSR access				
Forward citations-weighted PFC	438.310	722.065	0.830	41.948	180.910	571.597	14,101.392	3,464
Number of Inventors	84.649	103.554	0.830	14.362	43.788	114.020	872.754	3,464
				Regions with HSR access				
Forward citations-weighted PFC	650.665	601.615	21.224	162.523	449.556	998.974	2,993.178	96
PFC Number of Inventors	220.012	145.891	24.337	80.158	225.959	336.027	572.811	96

Notes: PFC stands for patent fractional count (per million inhabitants). The number of inventors should be understood as per million inhabitants.

2.2 Identification strategy

Following Wooldridge [2021, 2025], we analyze the impact of the opening of HSR stations on regional innovation using a staggered Difference-in-Differences (DiD) research design by estimating the following Extended Two-Way Fixed Effect (ETWFE) Equation:

$$y^w_{i,t} = \alpha + \sum_{k\in K} \left(\beta_{k,t} \cdot T_i \cdot Post_{k,t}\right) + \sum_{k\in K} \left(T_i \cdot Post_{k,t} \cdot \dot{\mathbf{X}}'_{i,t_0} \cdot \gamma_{k,t}\right) + \left(\eta_t \cdot \mathbf{X}'_{i,t_0} \cdot \lambda_{k,t}\right) + \mu_i + \eta_t + \varepsilon_{i,t}, \tag{1}$$

where $y^w_{i,t}$ is the natural logarithm of the per capita forward citation-weighted patent fractional count measured in NUTS-3 region i in year t. T_i is the treatment defined as equal to 1 if the NUTS-3 region has, at least, one HSR station and 0 otherwise. $K = \{2007, 2008, 2009, 2013, 2016\}$ is the set of years, from now on defined as cohorts, in which the treatment occurs and $Post_{k,t}$ is a post treatment dummy variable for each cohort such that $Post_{k,t} = 1$ if $t \geq k$ and 0 otherwise. The set of $\beta_{k,t}$ coefficients measures the ATT to be connected to the HSR network for each pair of cohort and year (i.e. $\beta_{k=2013,t=2014}$) after the opening of the HSR station. In line with Wooldridge [2021, 2025], we include in Equation (1) a set of predetermined control variables, i.e. $\mathbf{X}_{i,t_0}$, as measured at the beginning of the sample period, where $\dot{\mathbf{X}}_{i,t_0}$ is the within-cohort mean transformation.[18] The inclusion of $\mathbf{X}_{i,t_0}$ is aimed to capture different regional predetermined characteristics and to increase the comparability of the treatment and control groups. In the same vein, our specification includes also a full set of NUTS-3 region

[18]The incremental set of controls includes the number of inventors patenting in 1982, that represents a proxy for the regional human capital and knowledge endowment of each province and that is a key control variable explaining the variability of our innovation index, the per capita GDP and industry share on GVA in 1982, unit labor costs and compensations per hour worked in 1982, the IQI index in 2004 and the potential multimodal accessibility index in 2001. Control variables are log transformed except for the IQI, multimodal accessibility index and the share of industry over GVA. If the control variables were not centered around the cohorts mean, the $\beta_{k,t}$ coefficients would not directly measure the ATTs, but would also capture the heterogeneity linked to deviation from the cohort mean [Wooldridge, 2021, 2025].

(μ_i) and year fixed effects (η_t). From the one hand, year fixed effects capture unobserved determinants of the outcome that are common to all NUTS-3 regions in a given year—such as EU/national demand conditions, aggregate business cycles, inflation or economy-wide technological change, and nationwide regulatory shifts. On the other hand, NUTS-3 region fixed effects absorb time-invariant regional heterogeneity (e.g., long-run industrial structure, geography and transport endowments, baseline market access, and persistent institutional quality) that may be correlated with the regressors and could otherwise drive regional outcomes. What is more, $\left(\eta_t \cdot \mathbf{X}'_{i,t_0}\right)$ allows for heterogeneous time trends by $\mathbf{X}_{i,t_0}$, while $\left(T_i \cdot Post_{k,t} \cdot \dot{\mathbf{X}}'_{i,t_0}\right)$ allows the treatment effect to vary with $\mathbf{X}_{i,t_0}$. Finally, $\varepsilon_{i,t}$ represents an error term.

The specification in Equation (1) accounts for treatment effect heterogeneity over time and for the staggered nature of the treatment design. However, as a robustness test, we also estimate a simpler DiD two-way fixed effects (TWFE) model, which provides a unique ATT coefficient for all periods and cohorts, in order to analyze whether the asynchronous opening of the HSR stations in different regions represents a major issue when estimating ATTs. Estimates of this specification are reported in all tables, together with a statistical t-test aimed to prove if the difference between these two models are equal to zero.[19]

As in all DiD research designs, the identification strategy relies on the crucial assumption of parallel trends in per capita forward citation-weighted patent fractional counts between the treated and the control units in the periods before the treatments occur. Considering the specification of Equation (1), we rely on the parallel trends assumption conditional on both the set of predetermined controls and time/regional fixed effects. We reasonably believe that, to isolate the causal impact of the opening of HSR stations on regional innovation, it is important to control for the economic, institutional and territorial accessibility characteristics of each NUTS-3 region. Economic control variables allow us to account for the fact that some NUTS-3 regions are more likely to generate innovation given the composition of their economic activity. Moreover, the institutional quality might affect the impact of HSR investments, given that better institutional quality can foster innovation by strengthening appropriability and contract enforcement–which lowers the risk of investing in R&D and technology transfer–and by providing predictable, transparent regulation that reduces entry and compliance costs for new and expanding firms, as well as by curbing corruption so that both public funds and private capital are allocated more efficiently to high-productivity projects. Well-functioning institutions also improve the diffusion of knowledge through more reliable public procurement, clearer IP regimes, and greater trust among market participants, all of which support the formation of innovative firms and the scaling of new technologies. Finally, we add a regional multimodal accessibility index that measures the degree of connection with other regions.

In order to verify the conditional parallel trends assumption, we estimate an event-study DiD. Let $K = \{2007, 2008, 2009, 2013, 2016\}$ denote the cohorts, and define for each $k \in K$ the indicators $\text{Lead}_{k,t}^{(-\ell)} = \mathbf{1}\{t = k - \ell\}$ for $\ell = 1, 2, 3, 4, 5, \quad \text{Lead}_{k,t}^{(-6^-)} = \mathbf{1}\{t \leq k - 6\}$. The estimating Equation is:

[19] All models are estimated with the *jwdid* Stata® command, that allows us to compare ETWFE and TWFE models. See Rios-Avila et al. [2024] for details.

$$
\begin{aligned}
y_{i,t}^{w} = \alpha &+ \sum_{k\in K}\left(\sum_{\ell=1}^{5}\beta_{k,-\ell}\,T_i\cdot \mathrm{Lead}_{k,t}^{(-\ell)} + \beta_{k,-6^-}\,T_i\cdot \mathrm{Lead}_{k,t}^{(-6^-)}\right)\\
&+ \sum_{k\in K}\left(\sum_{\ell=1}^{5} T_i\cdot \mathrm{Lead}_{k,t}^{(-\ell)}\cdot \dot{\mathbf{X}}'_{i,t_0}\cdot\gamma_{k,-\ell} + T_i\cdot \mathrm{Lead}_{k,t}^{(-6^-)}\cdot \dot{\mathbf{X}}'_{i,t_0}\cdot\gamma_{k,-6^-}\right)\\
&+ \left(\eta_t\cdot \mathbf{X}'_{i,t_0}\cdot\lambda\right) + \mu_i + \eta_t + \varepsilon_{i,t}
\end{aligned}
\tag{2}
$$

where the coefficients $\beta_{k,-\ell}$ measure differences between treated and control units ℓ periods before treatment for cohort k; if the parallel trends assumption holds, they should be individually and jointly equal to zero.[20]

Some authors note that pre-trend tests can have limited statistical power. We therefore complement them with the Honest DiD procedure of Rambachan and Roth [2023], which quantifies how large post-treatment departures from parallel trends would be needed to drive the estimated effect to zero. Following their approach, we allow post-treatment deviations up to twice the magnitude of the largest pre-treatment discrepancy.[21]

Moreover, in the context of spatial data, potential spillover effects on untreated units may challenge the validity of the Stable Unit Treatment Value Assumption (SUTVA). Specifically, HSR openings might positively affect innovation in treated NUTS-3 regions by siphoning activity from untreated, especially adjacent, NUTS-3 regions, or, alternatively, generate positive spillovers that benefit neighbors. In both cases the SUTVA would be violated since the control group would be affected by the treatment. In order to analyze this issue, we re-estimate Equation 1 after including in the treated group also the NUTS-3 regions contiguous to those hosting an HSR station; moreover, we implement a "donut-hole" design whereby treatment is defined solely for neighboring NUTS-3 regions, while the NUTS-3 regions actually hosting an HSR station are excluded from the sample.

Another important issue in our study is related to the relatively low number of treated units. Indeed, in a DiD research design with only a few treated units, standard cluster-robust inference might break down because variance is estimated from very few treated clusters, which biases standard errors downward and inflates rejection rates. Statistical power might be low, so moderate effects are hard to detect and pre-trend tests lack precision. Estimates might crucially depend on a single treated unit or outlier that might drive results, thus making them sensitive to small specification changes. What is more, staggered adoption might provide little identifying variation if only few cohorts exist, so timing-based comparisons could be noisy; finally, parallel trends checks might be weak because of short or volatile pre-periods for the treated units prevent credible validation. From the one hand, to mitigate such problems one possible solution has been proposed by Arkhangelsky et al. [2019], who suggest to implement a Synthetic Control DiD (SC-DiD) approach that estimates the ATT by contrasting treated units with their synthetic counterparts before and after the treatment. The synthetic control sample is built by giving more weight to controls with pre-treatment path closest to the treated units and time windows most comparable to the treated period. From the other hand, in the same vein we also assess the sensitivity of the results when a different control group is considered. In order to increase the comparability between treated and control groups, we match regions on the basis of a battery of observable characteristics from our set of controls by estimating the propensity score (nearest neighbor[s] within

[20] Equation (2) is estimated on the pre-treatment subsample for ever-treated units ($t < k$ for unit i in cohort k), using as controls the set of never-treated regions (for which $T_i = 0$). We deliberately exclude not-yet-treated observations from the control group in this diagnostic to keep the comparison set stable and uncontaminated by impending treatment. By contrast, when estimating our baseline specification in Equation (1) we include in the control sample also not-yet-treated units.

[21] This analysis leverages the *honestdid* Stata® command. See Rambachan and Roth [2023], Bravo et al. [2024] for details.

caliper = 0.05); then, the sample has been restricted to observations with a propensity score in the interval [0.1,0.9]. Finally, we estimate Equation 1 leveraging on this propensity-score-matched sample, using the propensity score to compute weights in the subsequent DiD estimation.

Despite the identification of the causal effect of HSR on regional innovation relies on the validity of our staggered DiD research design, it is worth addressing another identification challenge associated with the possible simultaneity between regional innovation capacity and infrastructure expansion. As noted by Cao et al. [2024, page 10], "the central government might plan to build HSRs between cities with stronger economic and innovation ties... Another possibility is that the government might plan to link highly populated cities with other cities to decentralize population". If this were the case, the ATTs estimates would be upward or downward biased, respectively. In addition, unobserved factors could simultaneously influence both infrastructure development and innovation outcomes, potentially biasing ordinary least squares estimates of the causal impact. To address this potential issue, we apply one of the approaches commonly used in the applied literature on the economic effects of infrastructure, namely the planned HSR IV one.[22] Our identification strategy exploits the discrepancy between the planned HSR network and the actual timing of station openings. Although the national plan determined which regions would eventually receive a station, the precise timing of realization was shaped by administrative, financial, and engineering factors largely unrelated to local innovation trajectories. We use the planned network as an instrument for actual HSR access, so that causal identification is driven by variation in the timing of openings relative to the plan, rather than by endogenous regional characteristics.[23]

In this regard, the aforementioned instrument is relevant because corridor blueprints and station plans strongly predict subsequent openings given multi-year funding, engineering constraints and regulatory sequencing; the relevance of the instrument is confirmed by a strong first stage, and by the reported the Kleibergen–Paap rk Wald F statistic. The instrument's orthogonality (i.e. instrument exogeneity) rests on plans being set many years before actual HSR access and driven primarily by national corridor design and engineering feasibility rather than short-run, region-specific shocks to innovation; conditional on rich pre-determined covariates plus NUTS-3 region and year fixed effects, the correlation between planned HSR and regional innovation performance pass through realised HSR access. For weak-IV diagnostics, we complement the Kleibergen–Paap rk F with the effective F of Olea and Pflueger [2013] using the Stata® *weakivtest* routine, which delivers weak-instrument-robust inference under heteroskedasticity, autocorrelation and clustering; where relevant, we reference Stock and Yogo [2002]'s critical values. In addition, we perform inference relying on the *plausexog* Stata® command to relax the exclusion restriction, as suggested by Conley et al. [2012].[24]

Finally, it is worth noting that Section 3.3 provides and comments a full set of placebo and falsification tests, whose aim is to test the validity of our research design.

[22] The other approaches are the inconsequential units and the historical railway IVs approaches.

[23] The instrument is based on several official documents of past years: the FS Technical Proposal of 1986, Italian Law 910/1986, the municipal management resolution n. 2334/2002 of Reggio Emilia, the national program for the redevelopment of main Italian stations managed by "Grandi Stazioni", and the deliberation of the "Comitato interministeriale per la programmazione economica e lo sviluppo sostenibile" (CIPE) n. 94/2006.

[24] The command has been developed by Clarke [2020].

3 Empirical results

3.1 Main results

In this Section, we discuss the main results on the effect of the staggered access to the HSR network on regional innovation. Estimates of our baseline specification in Equation 1, with a full sequence of incremental predetermined controls, are reported in Table 2. In the first five rows we present the ATTs at cohort level, while in the sixth row we show the overall ATT aggregated across cohorts and years.[25] Across specifications, the estimated ATTs are positive and statistically significant at 1% level. In particular, estimates of the bare-bones model suggest that NUTS-3 regions connected to the HSR experienced an average increase in forward citation-weighted PFC of about 27%; the inclusion of controls for institutional quality, regional accessibility labor market and economic size only induces a marginal increase in the magnitude of the estimates, ranging from 35.39% to 39.24%, as highlighted in columns from (2) to (5) of Table 2.

It is worth analyzing how each cohort contributes to our findings, thereby clarifying the spatial heterogeneity of the ATT. The staggered opening of HSR stations has generated a positive and statistically significant effect on regional innovation for most cohorts. In particular, the 2007 cohort is related to the opening of the HSR stations in Padua and Venice, which however are still disconnected from the rest of the HSR network (see Figure 1); this might explain the relatively lower magnitude of the estimated ATT with respect to the other cohorts (from about 12% to 29%). Conversely, the estimated ATTs for the 2008 cohort (from about 33% to 51%) exceeds the overall ATT and reflects the opening of the HSR stations in Milan, Bologna, Naples, and Salerno, which marked a significant step toward the completion of the national HSR network. In contrast, the ATT for the 2009 cohort (completion of the Turin–Milan connection) is generally positive but not statistically significant, a result plausibly attributable to the earlier partial opening of the line in 2006, which may have attenuated any additional impact in 2009. It is worth noting that the opening of Reggio Emilia AV Mediopadana station in 2013 generates a noticeable positive impact on regional innovative performance, whose amount (from about 75% to 96%) is significantly above the overall average and robust at 1% level of statistical significance across all estimated specifications. Finally, the 2016 cohort represents the opening of the relatively short segment between Milan and Brescia, as part of the wider project to connect Milan and Venice. In this case the estimated ATT is significantly positive (about 36%) and broadly in line with the overall effect. These findings show that the benefits from HSR access are not uniform across places; in particular they are stronger for intermediate and smaller regions like Reggio Emilia, thus supporting the view that transport infrastructure can be an effective policy lever for narrowing regional economic disparities. This result is in line with Miwa et al. [2022] showing that HSR raises regional innovation more in smaller jurisdictions: when they split Japanese municipalities by size, the estimated treatment effects are substantially larger for towns and villages (under 50,000 residents) than for cities, which the authors attribute to stronger marginal gains from improved inter-regional accessibility and knowledge absorption in rural/peripheral places. A complementary pattern appears in Hanley et al. [2022] who find that HSR-induced collaborative innovation (co-patenting and citation quality) is stronger in less-developed regions, again pointing to disproportionate benefits outside core hubs. By contrast, Wen et al. [2025] on green innovation reports larger effects in bigger cities (consistent with agglomeration).

[25] More formally, the aggregated ATT by cohort is given by $ATT_k = \sum_{t \geq k} w_{k,t} \beta_{k,t}$, with $w_{k,t} = \frac{N_{k,t}}{\sum_{s \geq k} N_{k,s}}$ and N is the number of treated regions. The overall ATT aggregated by cohort and year is given by $ATT_{\text{overall}} = \sum_{k \in K} \sum_{t \geq k} \omega_{k,t} \beta_{k,t}$, with $\omega_{k,t} = \frac{N_{k,t}}{\sum_{g \in K} \sum_{s \geq g} N_{g,s}}$.

Table 2: DiD estimates of the impact of HSR access on regional innovation.

Dependent variable (log):	Forward-Citation-Weighted PFC					Unweighted PFC	Inventors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSR – cohort 2007	0.128**	0.113*	0.125	0.172*	0.257**	0.230***	0.136
	[0.059]	[0.063]	[0.092]	[0.094]	[0.119]	[0.087]	[0.100]
HSR – cohort 2008	0.287**	0.398**	0.410***	0.410***	0.339*	0.168	0.271
	[0.113]	[0.171]	[0.156]	[0.158]	[0.193]	[0.133]	[0.171]
HSR – cohort 2009	-0.012	0.102	0.101	0.181	0.258	-0.062	0.164
	[0.159]	[0.200]	[0.216]	[0.250]	[0.241]	[0.147]	[0.183]
HSR – cohort 2013	0.675***	0.673***	0.658***	0.631***	0.560***	0.190**	0.129
	[0.067]	[0.085]	[0.083]	[0.110]	[0.116]	[0.087]	[0.095]
HSR – cohort 2016	0.368***	0.321***	0.317***	0.308***	0.231*	0.169*	0.233**
	[0.080]	[0.089]	[0.097]	[0.103]	[0.136]	[0.088]	[0.097]
HSR – Overall ATT	0.241***	0.303***	0.311***	0.331***	0.319***	0.160*	0.210**
	[0.089]	[0.113]	[0.109]	[0.112]	[0.118]	[0.083]	[0.107]
HSR – ATT_{TWFE}	0.245**	0.311**	0.315***	0.332***	0.321**	0.162**	0.210**
	[0.100]	[0.122]	[0.119]	[0.120]	[0.125]	[0.081]	[0.105]
NUTS 3 region FE	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓
Inventors	✓	✓	✓	✓	✓	✓	✓
Institutional quality	✗	✓	✓	✓	✓	✓	✓
Multimodal accessibility	✗	✗	✓	✓	✓	✓	✓
Labor market controls	✗	✗	✗	✓	✓	✓	✓
Economic controls	✗	✗	✗	✗	✓	✓	✓
Observations	3,560	3,560	3,560	3,560	3,560	3,560	3,560
H_0: ATT=ATT_{TWFE}	0.981	0.964	0.980	0.997	0.990	0.990	0.998

Notes: This Table presents DiD estimates of Equation 1 on the effects of HSR access on regional innovation. The latter is measured by per capita forward citations-weighted patents fractional counts (columns from 1 to 5), or per capita unweighted patent fractional counts (column 6), or by the per capita number of inventors patenting in the region (column 7). All specifications include both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). The aggregated ATTs by cohorts are shown by *HSR – cohort k*, while the overall ATT aggregated by cohort and year by *HSR – Overall ATT*. *HSR* – ATT_{TWFE} reports the results from a two-way fixed effect (TWFE) specification that does not account neither for treatment effect heterogeneity over time, nor for the staggered nature of the treatment. The *t*-test reported in the last row is aimed to verify whether the average treatment effects estimated under the two specifications do not significantly differ from each other, namely H_0: ATT=ATT_{TWFE}. All specifications are estimated by OLS. Robust standard errors, clustered at NUTS-3 region level, are shown in parentheses: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

It is worth noting that our treatment effect cannot be interpreted as a rise in each province's absolute patent count. Our innovation indicator is indeed a composite measure that captures each NUTS-3 region's annual share of national patenting—proxied by the region's share of resident inventors—and weights each patent by the forward citations it accrues over the sample to reflect quality [Hall et al., 2005]. Movements in this index may therefore stem from (i) more patents, (ii) more inventors, and/or (iii) higher citations per patent. To separate quantity from quality, we re-estimate the more extended version of Equation 1 using as the dependent variable both the per capita unweighted patent fractional count (i.e., the region's share of Italy's annual patents) and the per capita number of inventors. The ATTs by cohorts reported in column (6) of Table 2 are generally statistically significant at the usual levels of significance, except for the 2008 cohort, and indicate a positive effect on the patent share that, aggregating by cohorts and years, has increased of about 17%. Our results of an improvement in both quantity and quality of the patenting activity is in line with Bhatt and Kato [2021] who find a positive and significant impact of the HSR opening on innovation, measured as the number of patent applications, for an international sample of countries observed over the period 1980-2018; also some studies for the Japanese case exhibit similar results: Inoue et al. [2017] study patent publications/quality and co-inventing patterns and Miwa et al. [2022] measure innovation via patents/citations. Also Hanley et al. [2022] find similar results focusing on co-patenting and citation quality for China.

Moreover, if one considers estimates on the number of inventors in column (7) of Table 2, it can be observed that the HSR access has favored also the regional share of inventors of about 23%.[26] Taken together, these results imply that the observed improvement in regional innovative performance is not driven predominantly by a specific channel, but instead by an increase in both patent quality—as captured by forward citations—and by an expansion in the region's share of patenting, as well as in the number of inventors residing in the region. Indeed, HSR access may compress effective distance to national R&D hubs, possibly cutting search and coordination costs for tacit, face-to-face knowledge exchange. Firms may reorganize R&D across sites and form larger, more diverse inventor teams, hopefully raising the novelty and influence of inventions (higher forward citations). Simultaneously, improved market reach and visibility may increase the returns to locating and remaining in the region, attracting and retaining inventors and R&D-active firms and expanding the pipeline of patentable projects. These complementary agglomeration and collaboration effects therefore generate both quality upgrading and an extensive-margin rise in patenting, rather than a single dominant channel.

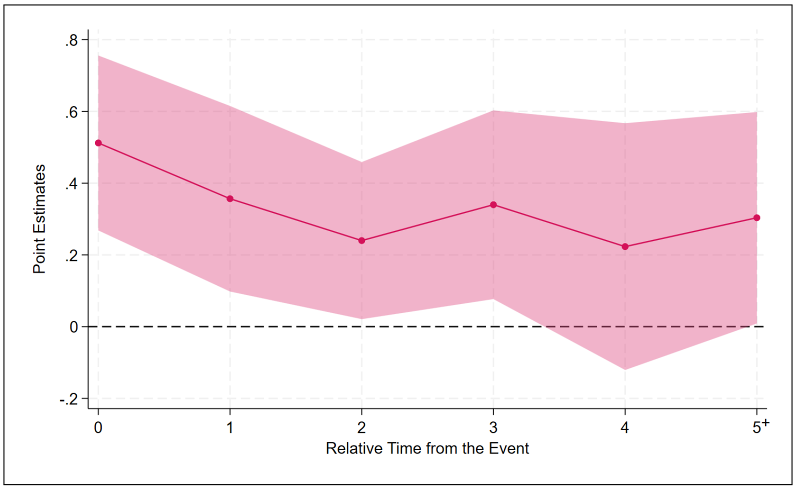
It is worth noting that, given the staggered nature of our treatment, estimates of Equation 1 in Table 2 rely on a ETWFE model [Wooldridge, 2021, 2025], isolating cohort-by-relative-time effects and alleviating concerns about heterogeneous dynamics, differently from classic TWFE that can use already-treated units as (improper) controls and that place non-convex—sometimes negative—weights on cohort-time effects, biasing the average effect. Testing differences between the two estimation methods thus serves as a diagnostic for "forbidden-comparison" contamination; to this end, in the seventh row of Table 2, we additionally report the results from a TWFE specification that does not account neither for treatment effect heterogeneity over time, nor for the staggered nature of the treatment. Moreover, we perform a t-test in order to analyze whether the asynchronous opening of the HSR stations in different regions does significantly affect the estimates. The t-test reported in the last row of Table 2 (p-value) indicates that the average treatment effects estimated under the two specifications, namely the Overall ATT and the ATT_{TWFE}, do not significantly differ from each other.

An issue that is worth analyzing is the possibility that the effects of the treatment speed up, stabilize, or mean revert over time. In order to explore this issue, we also adapt Equation (1) to estimate post-treatment dynamics after binning from period five onward. Figure 2 shows the impact of HSR on regional innovation over time, where the plotted ATTs have been aggregated for each post-treatment year. Results indicate that the estimated ATTs are positive and decreasing over time; moreover, estimates provide evidence in favor of a medium-long term statistically significant positive effect, notwithstanding the poorly identified coefficient in the fourth post-treatment year, as shown by the estimated ATT for the fifth year, where we bin all the remaining post-treatment years.

As emphasized in Section 2.2, the validity of our identification strategy relies on the parallel trends assumption between the treated and control units, which must hold in order to interpret the estimates as causal effects of HSR openings on regional innovation. To this end, we estimate Equation 2, whose results are reported in Figure 3; the latter suggests the absence of any anticipating effect, given that all leads coefficients are individually and jointly not statistically different from zero, thus making the assumption of parallel trends plausible. In particular, to ensure model identification and reduce

[26] To the best of our knowledge there are no papers in the previous literature that has investigated the effect of HSR access on the number of inventors.

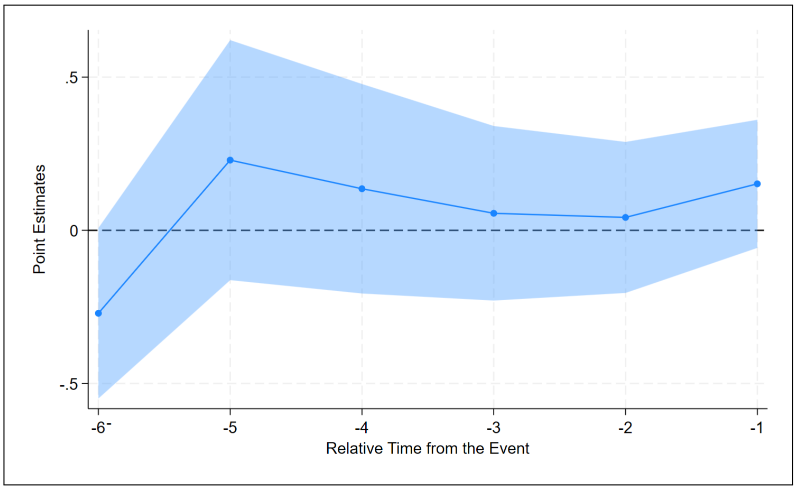
Figure 2: Dynamic ATTs of HSR stations opening over time.



Notes: OLS estimates of Equation (1) on the effects of HSR access on regional innovation aimed at estimating post-treatment dynamics after binning from period five onward. Regional innovation is measured by per capita forward citations-weighted patents fractional counts. The specification includes both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). The red dots show estimated coefficients of the ℓ lags of the treatment interaction, spanning from the first-treatment year of cohort k to 5 years after the treatment and onward, reflecting dynamic ATTs. The red area represents the respective 95% confidence intervals. Robust standard errors are clustered at NUTS-3 region level.

noise in distant leads, we bin all years earlier than six periods before treatment into a single indicator; consequently, the coefficient for the period "-6^{-}" represents the mean difference between treated and control units in all pre-treatment periods before "–6", relative to the omitted reference period.

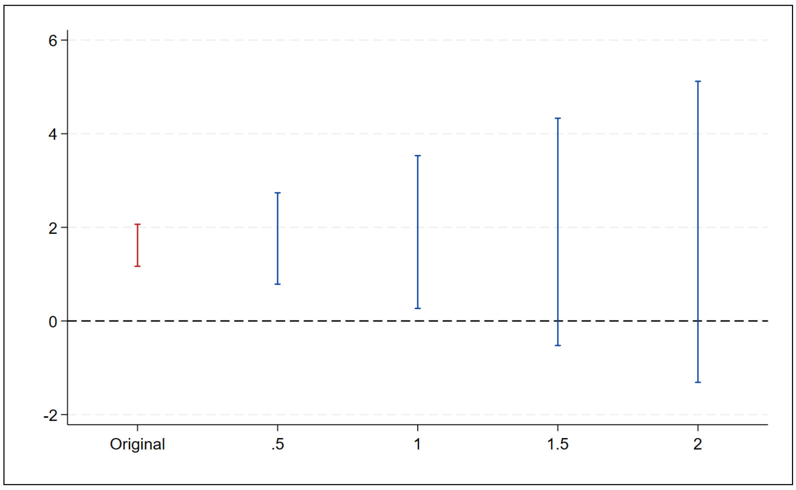
Figure 3: Dynamic ATTs of HSR stations opening over time (pre-treatment period).



Notes: OLS estimates of Equation 2 on the effects of HSR access on regional innovation. The latter is measured by per capita forward citations-weighted patents fractional counts. The specification includes both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). The blue dots show estimated pre-trend coefficients of the ℓ leads of the treatment interaction, spanning from the first period before the treatment of cohort k up to 6 years before the treatment and backward. The blue area represents the respective 95% confidence intervals. Robust standard errors are clustered at NUTS-3 region level.

In order to further analyze the parallel trends assumption, we also apply Rambachan and Roth [2023]'s Honest DiD procedure, which measures the magnitude of post-treatment deviations from parallel trends required to reduce the estimated effect to zero. Following their method, we allow post-treatment deviations up to twice the size of the biggest pre-treatment violation. Figure 4 provides results of this analysis and highlights robust confidence intervals for the treatment effect of the HSR access on regional innovation in the post-treatment period, following an event-study approach, and relying on relative magnitude bounds ($\Delta^{RM}\bar{M}$). It is worth noting that the analysis provides evidence that, by restricting post-treatment violations of parallel trends up to the maximum pre-treatment deviation (i.e. setting $\bar{M} = 1$), a robust confidence interval can be derived. Even though these intervals are wider than conventional OLS confidence intervals—which are valid only under strict parallel trends—they still exclude a zero innovative effect of HSR access in the post-treatment period. Overall, the positive treatment effects are robust to medium violations of the parallel trend assumption.

Figure 4: Robust confidence intervals for HSR's effect on regional innovation.



Notes: Robust confidence intervals for HSR's effect on regional innovation. The latter is measured by per capita forward citations-weighted patents fractional counts. The x-axis reports different values for the relative magnitudes bounds, i.e. $\Delta^{RM}\bar{M}$. Vertical whiskers represent the respective 90% confidence intervals. $\bar{M} = 1$ is equivalent to limit post-treatment violations of parallel trends to be no greater than the maximum pre-treatment deviation.

Summarizing, our main findings show that HSR connection improves the regional innovative performances. In particular, connection to the HSR network can expand market reach and intensify competition, thus raising incentives to innovate; moreover, it can enhance regional innovation performance by reducing effective travel times, thereby lowering coordination costs for research and development. Improved accessibility can enlarge matching markets for skills, laboratories, universities, investors, and specialized suppliers, strengthening collaborative opportunities and knowledge spillovers; furthermore, greater inter-regional connectivity can foster co-patenting and the absorption of external knowledge.

To further explore the effects of HSR access on regional innovation, we perform an heterogeneity analysis in order to address the issue of possible unobserved heterogeneity associated to different patents technology sectors. We adopt the World Intellectual Property Organization (WIPO) classification that identifies five industries, namely Electrical engineering, Instruments, Chemistry, Mechanical engineering, and Other sectors. Table 3 shows the presence of heterogeneity in the impact of HSR on regional innovation across industries and cohorts; in particular, for the cohorts 2013 and 2016 most of industries are positively affected though to different extents, while for the other cohorts only some industries benefited from the opening of the HSR stations.

Table 3: DiD estimates of the heterogeneous impact of HSR access on regional innovation by technological sector.

Dependent variable (log): Forward-Citation-Weighted PFC					
	(1)	(2)	(3)	(4)	(5)
Tech. Sector:	Electrical engineering	Instruments	Chemistry	Mechanical engineering	Other sectors
HSR – cohort 2007	0.621***	-0.008	-0.039	0.131	0.301*
	[0.187]	[0.159]	[0.160]	[0.158]	[0.172]
HSR – cohort 2008	0.480*	0.247	0.661***	0.626***	-0.176
	[0.255]	[0.251]	[0.219]	[0.186]	[0.185]
HSR – cohort 2009	-0.010	0.084	0.500***	0.492	-0.002
	[0.551]	[0.603]	[0.193]	[0.304]	[0.266]
HSR – cohort 2013	-0.019	0.872***	1.409***	0.410***	0.596***
	[0.167]	[0.222]	[0.133]	[0.134]	[0.106]
HSR – cohort 2016	0.716***	0.744***	0.091	0.012	0.516***
	[0.163]	[0.187]	[0.188]	[0.181]	[0.149]
HSR – Overall ATT	0.436**	0.225	0.484***	0.435***	0.058
	[0.174]	[0.193]	[0.124]	[0.121]	[0.143]
HSR – ATT_{TWFE}	0.439**	0.221	0.483***	0.439***	0.067
	[0.171]	[0.189]	[0.144]	[0.129]	[0.140]
NUTS 3 region FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Inventors	✓	✓	✓	✓	✓
Institutional quality	✓	✓	✓	✓	✓
Multimodal accessibility	✓	✓	✓	✓	✓
Labor market controls	✓	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓	✓
Observations	3,560	3,560	3,560	3,560	3,560
H_0: ATT=ATT_{TWFE}	0.990	0.987	0.996	0.980	0.963

Notes: This Table presents DiD estimates of Equation 1 on the heterogeneous effects of HSR access on regional innovation by technological sectors. Regional innovation is proxied by forward citations-weighted patents fractional counts by technological sector. Following the WIPO classification, we consider five technological sectors, namely electrical engineering, instruments, chemistry, mechanical engineering, and other sectors. All specifications include both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). The aggregated ATTs by cohorts is shown by *HSR – cohort k*, while the overall ATT aggregated by cohort and year by *HSR – Overall ATT*. *HSR* – ATT_{TWFE} reports the results from a two-way fixed effect (TWFE) specification that does not account neither for treatment effect heterogeneity over time, nor for the staggered nature of the treatment. The t-test reported in the last row is aimed to verify whether the average treatment effects estimated under the two specifications do not significantly differ from each other, namely H_0: ATT=ATT_{TWFE}. All specifications are estimated by OLS. Robust standard errors, clustered at NUTS-3 region level, are shown in parentheses: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

The estimated overall ATT and the ATT_{TWFE} are positive and statistically significant for both the electrical and the mechanical engineering, as well as chemistry industries. A similar result has been reported by Bottasso et al. [2022] who found a positive and significant impact of highways on the forward citation-weighted patent fractional count for the case of Italian NUTS-3 regions. In particular, in the context of HSR access, in chemistry (including pharmaceuticals), where research is science-based, cumulative and highly appropriable, better connectivity intensifies collaboration and accelerates the translation of marginal advances into patentable outputs [Levin et al., 1987]. In electrical engineering (e.g., power electronics and semiconductors), where innovation is equipment-intensive and relies on rapid prototyping, iterative testing and certification, improved accessibility enables frequent lab-to-fab interactions, shortens design–test–redesign cycles, and eases coordination with component suppliers and systems integrators [Thomke, 1998, Macher and Mowery, 2004]. Conversely, in mechanical engineering, characterized by tacit, incremental and production-embedded innovation, HSR enables same-day, face-to-face coordination across supplier–client networks, speeding problem-solving and turning shop-floor learning into codified IP [Cohen et al., 2000]. Weaker responses in other fields (e.g., ICT's standard-driven, platform dynamics) are consistent with regimes where travel-time compression shifts collaboration margins less.

3.2 Patents collaborations and network positioning

In this Section we discuss the analysis dedicated to investigate possible channels through which the opening of HSR stations can positively affect the quality of the regional innovation performance.

A possible transmission channel might be related to the intensity of collaboration across inventors located in different NUTS-3 regions. Indeed, HSR reduces the generalized cost of face-to-face knowledge exchange, so bilateral innovation flows should rise most between places that (i) have larger "innovation mass" (R&D stock, inventor base, innovative firms) and (ii) experience the largest fall in effective distance (rail travel time/frequency). Specifically, we analyze whether enhanced cross-regional high-speed travel opportunities facilitate inventor interactions and improve idea matching, thereby yielding an innovation process that is both higher in quality and greater in scale.

Strictly following the approach in Picci [2010] and Bottasso et al. [2022], it is worth defining the subsequent measure, which is based on our standard patent fractional count weighted for forward citations:

$$\psi_{a,i,t} = PFC_{i,a,t} \cdot Cit_a \tag{3}$$

in which $PFC_{i,a,t}$ is the fraction of patent a attributable to NUTS-3 region i in year t according to the inventors address, and Cit_a is the sum of the relative forward citations the patent received.[27]

Given Equation 3, for each patent application a and region pair $i-j$, the inventors' collaboration intensity—i.e. the co-authorship intensity— can be defined as:

$$Collaboration_{a,i,j,t} = \psi_{a,i,t} \cdot \psi_{a,j,t} \tag{4}$$

This metric identifies whether a patent application is intra-regional—its inventors reside in the same region—or inter-regional—its inventors come from different regions—and, in the latter case, quantifies the degree of collaboration; specifically, it takes on a zero value if the patent is not co-authored by inventors from different regions, while it takes on a positive value (and increasing according to the

[27] Note that $\sum_{a=1}^{A} \psi_{a,i,t} = y^{w}_{i,t}$, our main dependent variable.

intensity of the collaboration) otherwise. Lastly, assume the aggregate measure of collaboration among inventors belonging to different regions as the sum of Equation 4 for all patent applications:

$$Collaboration_{i,j,t} = \sum_{a=1}^{A} Collaboration_{a,i,j,t} \tag{5}$$

Following Dong et al. [2020], we then estimate a gravity model to analyze the impact of the staggered HSR opening on the degree of collaboration among inventors located in different regions [Anderson and Van Wincoop, 2003]:

$$\begin{aligned} \log\left(Collaboration_{i,j,t}\right) = \alpha + \sum_{k \in K} \left(\beta_k \cdot T_{i,j} \cdot Post_{k,t}\right) + \sum_{k \in K} \left(T_{i,j} \cdot Post_{k,t} \cdot \dot{\mathbf{X}}'_{i,j,t_0} \cdot \gamma_k\right) \\ + \left(\mathbf{X}'_{i,j,t_0} \cdot \tau_t\right) + \sigma\left(\log Distance_{i,j}\right) + \mu_i + \mu_j + \tau_t + \epsilon_{i,j,t} \end{aligned} \tag{6}$$

where $Collaboration_{i,j,t}$ (log) is the dependent variable, namely the intensity of innovative collaborations between region pairs $i-j$ in year t, $T_{i,j}$ is a dummy equal to 1 if a region pair is connected by the HSR network, and $Post_{k,t}$ is the usual post-treatment dummy variable for each cohort such that $Post_{t,k} = 1$ if $t \geq k$ and 0 otherwise, with $K = \{2007, 2008, 2009, 2013, 2016\}$. This gravity specification also includes the log distance between each NUTS-3 region pair ($Distance_{i,j}$) and all controls in both $\mathbf{X}_{i,j,t_0}$ and $\dot{\mathbf{X}}_{i,j,t_0}$ are constructed as cross-products of the corresponding variables for regions i and j. Both origin and destination fixed effects, namely μ_i and μ_j, as well as time fixed effects (τ_t) are also included in the model.

Estimates of Equation 6 are reported in column (1) of Table 4 and suggest that the opening of HSR stations has favored an increase of the collaboration activities across inventors located in different regions of about 30%. Such result can be interpreted by arguing that HSR access reduces effective distances, thus lowering frictions to face-to-face interactions, increasing the frequency and richness of communication, and facilitating the exchange of complex, tacit information that is hard to codify [Audretsch and Feldman, 2004]. Moreover, by cutting transportation costs, HSR expands the spatial reach of daily and periodic mobility, raising inter-regional contact rates. In turn, this intensifies knowledge flows, diffusion, and spillovers, and increases the probability that researchers and firms engage in joint projects.

Our results are consistent with the ones in Wang et al. [2022] and Hanley et al. [2022], who find that HSR connection between different samples of Chinese cities, observed between 2000 and 2016, induces a 2.7% increase in the number of cross-city co-developed patents (neither weighted for forward citations, nor for city contribution). Another interesting result, more directly comparable to our analysis, is found by Cao et al. [2024], who show that connecting two minor cities via HSR increases by about 21% the total number of citations received by a cited patent from a citing city.

Building on this result, we analyze how the presence of an HSR station relates to a region's embeddedness within the broader innovation network. By embeddedness, we mean the extent to which a region is well connected to, and positioned within, the overall structure of inter-regional research ties. To study this issue, we adopt a social network analysis (SNA) perspective, since research collaboration can be represented as a network composed of nodes and links [Galaso and Kovářík, 2021]. Specifically, we examine whether being connected to the HSR network facilitates deeper regional embeddedness, as proxied by different node-centrality measures [Wanzenboeck et al., 2014, Wanzenböck et al., 2015, Mitze and Strotebeck, 2018, Wanzenböck and Piribauer, 2018]. Indeed, a region's centrality in collaboration networks has been shown to boost innovation by enhancing knowledge flows and strengthening

Table 4: DiD estimates of the impact of HSR access on possible transmission channels. Patents collaboration and network positioning.

Dependent variables (log):	Collaborations	Katz-Bonacich	Betweenness	Closeness	Eigenvector
	(1)	(2)	(3)	(4)	(5)
HSR – cohort 2007	0.271***	0.008	-0.005	-0.037	-0.046***
	[0.028]	[0.017]	[0.008]	[0.038]	[0.011]
HSR – cohort 2008	0.915***	-0.061**	0.036***	0.121**	0.058***
	[0.028]	[0.029]	[0.009]	[0.056]	[0.013]
HSR – cohort 2009	-0.102*	0.104***	0.064***	0.163**	0.005
	[0.060]	[0.039]	[0.012]	[0.074]	[0.019]
HSR – cohort 2013	0.813***	0.403***	0.049***	0.151***	0.041***
	[0.023]	[0.022]	[0.005]	[0.035]	[0.007]
HSR – cohort 2016	0.216***	-0.053*	-0.001	0.114***	0.003
	[0.027]	[0.030]	[0.007]	[0.044]	[0.015]
HSR – Overall ATT	0.265***	0.011	0.027***	0.085***	0.020*
	[0.038]	[0.021]	[0.008]	[0.030]	[0.011]
HSR – ATT_{TWFE}	0.275***	0.009	0.028***	0.085***	0.020
	[0.072]	[0.022]	[0.009]	[0.031]	[0.015]
NUTS 3 region FE	✗	✓	✓	✓	✓
NUTS 3 region i FE	✓	✗	✗	✗	✗
NUTS 3 region j FE	✓	✗	✗	✗	✗
Year FE	✓	✓	✓	✓	✓
Inventors	✓	✓	✓	✓	✓
Institutional quality	✓	✓	✓	✓	✓
Multimodal accessibility	✓	✓	✓	✓	✓
Labor market controls	✓	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓	✓
Observations	327,520	3,520	3,520	3,520	3,520
H_0: ATT=ATT_{TWFE}	0.903	0.955	0.974	0.990	0.979

Notes: This Table presents DiD estimates of Equation 1 on the effects of HSR access on possible transmission channels. Column (1) shows results related to the aggregate weighted forward citation measure of collaboration among inventors of regions i and j. Columns from (2) to (5) show results related to different centrality measures to proxy network embeddedness of region i into the interregional innovation network. The specification in column (1) includes both NUTS-3 region i and NUTS-3 region j fixed effects, as well as time fixed effects. Specifications in columns from (2) to (5) include both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). Controls for the specification in column (1) are constructed as cross-products of the corresponding variables for regions i and j. The aggregated ATTs by cohorts is shown by *HSR – cohort* k, while the overall ATT aggregated by cohort and year by *HSR – Overall ATT*. *HSR –* ATT_{TWFE} reports the results from a two-way fixed effect (TWFE) specification that does not account neither for treatment effect heterogeneity over time, nor for the staggered nature of the treatment. The t-test reported in the last row is aimed to verify whether the average treatment effects estimated under the two specifications do not significantly differ from each other, namely H_0: ATT=ATT_{TWFE}. All specifications are estimated by OLS. Robust standard errors, clustered at NUTS-3 region level (NUTS-3 region-pairs in column 1), are shown in parentheses: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

absorptive capacity for spillovers [Owen-Smith and Powell, 2004, Whittington et al., 2009].[28]

In particular, we build a co-patenting network in which nodes are the NUTS-3 regions and links

[28] [Galaso and Kovářík, 2021] show that the national networks of collaboration play a crucial positive role in patenting performances.

represent observed collaborations between inventors located in different regions, as defined in Equation 5. This network allows us to compute multiple centrality measures for each region and to assess whether regions with an HSR station systematically occupy more central positions in the collaboration network. In our analysis, we rely on measures like the Katz-Bonacich, betweenness, closeness and eigenvector centrality. The first index is based on the number of direct and indirect connections of a node, while Freeman [1978]'s betweenness centrality relies on how often a node lies on the shortest paths between other nodes, so that nodes with high betweenness act as bridges or bottlenecks.[29] Closeness centrality is based instead on how close a node is, on average, to all other nodes in the network (via shortest path distances). Finally, eigenvector centrality is based not just on the number of connections, but also on the quality of those connections, in terms of being connected to other well-connected nodes [Bonacich, 1987, Wanzenboeck et al., 2014].

The aforementioned centrality metrics represent the dependent variables for the estimation of a model in which we scrutinize the role of HSR access on node properties and network structure. Specifically, we estimate the following Equation:

$$\log Centrality_{i,t} = \alpha + \sum_{k \in K} \left(\beta_{k,t} \cdot T_i \cdot Post_{k,t}\right) + \sum_{k \in K} \left(T_i \cdot Post_{k,t} \cdot \dot{\mathbf{X}}'_{i,t_0} \cdot \gamma_{k,t}\right) + \left(\eta_t \cdot \mathbf{X}'_{i,t_0} \cdot \lambda_{k,t}\right) + \mu_i + \eta_t + \varepsilon_{i,t} \quad (7)$$

where $Centrality_{i,t}$ is the dependent variable, namely the centrality measures discussed above.

In columns from (2) to (5) of Table 4 we report results from estimating the impact of the HSR station opening on these different centrality measures of a region within the inter-regional collaboration network. The results show that for all cohorts except the 2007 one (Venice, Padua) the degree of centrality increases after the opening of HSR stations and this result is robust across different measures of centrality, with the exception of the Katz-Bonacich index. Estimates suggest an overall effect in the range 2-9%, but the impact varies according to the different centrality indexes considered. In particular, regions hosting an HSR station lie about 2.7% more often on the shortest paths between node pairs (betwenneess), they are about 9% closer to the other nodes of the network (closeness) and show an increase of the number of high quality connections by about 2% (eigenvector). These results imply that hosting an HSR station makes the region a natural corridor for collaborations and boosts connectivity, knowledge flows, and innovation performance, through an increase of pass-through centrality; moreover, in regions with an HSR station, effective distances are reduced so that collaboration ties are strengthened and of higher quality, since the number of connection to high-profile collaborative hubs increases, thereby expanding access to influential knowledge and partners.

Our results are in line with prior evidence showing that higher centrality can enhance regional innovation by strengthening knowledge flows and regions absorptive capacity for spillovers. Dosso and Lebert [2020] analyze a sample of 495 worldwide equivalent NUTS-2 regions for the period 2010-2012 and find evidence in favor of a positive correlation between different centrality measures and the innovation performance in terms of patents fractional count weighted for backward citations. For instance, Galaso and Kovářík [2021] compare the effects of regional vs. country-level Spanish networks on future patenting, showing that collaboration networks impact future patenting; moreover, they show that the embeddedness at different geographical scales exerts differing influence on innovation. Last but not least, Bottasso et al. [2022] highlight the importance of roads infrastructure in fostering regional

[29]As suggested by Bottasso et al. [2022], "a higher value of Betweenness Centrality characterizes those regions that are more likely to influence the transfer of knowledge within the whole network and the creation of knowledge spillovers".

embeddedness (as proxied by centrality measures) into an inter-regional collaboration network.

3.3 Robustness checks

In this Section we discuss the results of a wide set of robustness checks aimed at testing the validity of our findings illustrated in the previous subsections.

First, we analyze the validity of our research design in terms of identification of treated and control regions. In particular, we study the possible existence of displacement/spillover effects that could undermine the SUTVA, since control regions should not be affected by the treatment in order to represent a valid counterfactual sample. Furthermore, if spillover (displacement) effects occur, it becomes crucial to understand whether neighboring regions are positively (negatively) affected by the treatment. Indeed, the opening of HSR stations may raise innovation in treated NUTS-3 regions at the expense of untreated, especially adjacent, NUTS-3 ones or, alternatively, might generate positive spillovers that benefit also neighboring areas. In order to shed some light on this issue, we extend the sample of treated regions by including also NUTS-3 regions adjacent to those hosting HSR stations and we re-estimate Equation 1. Estimates of such specification are reported in column (1) of Table 5 and suggest that spillover effects are not present in our sample, thus confirming the validity of the SUTVA in our research design. Furthermore, we implement a "donuts-hole" design in which only neighboring contiguous regions are considered as the treated sample, while regions actually hosting an HSR station are excluded from both the treated and the control samples. The results of this analysis is shown in column (2) of Table 5, confirming the latter conclusion of no-treatment effect on neighboring regions. Such results implies that the improvement of the regional innovation performance did not come at the expenses of neighboring regions. Firm- and city-level studies in China typically find net gains without zero-sum crowding out: manufacturing firms and peripheral areas experience higher innovative output after connection, consistent with additive spillovers rather than siphoning from neighbours [Gao and Zheng, 2020, Dong et al., 2020]. Moeover, even where HSR nudges places towards specialisation, notably in core cities, this is portrayed as intra-urban portfolio reshaping rather than evidence of neighbouring decline [Chen and Guo, 2023].

Another important issue in our study is related to the relatively low number of treated units, which may undermine standard cluster-robust inference (downward-biased SEs and over-rejection), may lower power, potentially raising leverage of single treated units, thus possibly limiting identifying variation under staggered adoption, and weakens parallel-trends checks. To mitigate these issues, we apply the Arkhangelsky et al. [2019]'s Synthetic Control DiD (SC-DiD) approach, which estimates the ATT by contrasting treated units with weighted synthetic counterparts, assigning higher weights to controls with similar pre-treatment paths and to time windows most comparable to the treated period. The results of SC-DiD estimates, reported in column (3) of Table 5, suggest that our measure of regional innovation increases by about 31.5%, quite in line with our findings for the baseline specification shown in Table 2. The same conclusions arise when estimating a model that preserves the staggered nature of the treatment and uses a control sample built with Propensity Score Matching techniques, as shown in column (4) of Table 2.[30]

In order to further validate our research design, we perform three falsification tests. On the one hand, we firstly estimate the more extended specification of Equation 1, including predetermined control variables, after introducing both *fake* treated areas and *fake* treatment timings. These latter are drawn

[30]Specifically, the treated and control groups are matched by using all control variables measured in the pre-treatment period and with a caliper equal to 0.05, as discussed in Section 2.2.

Table 5: Robustness tests and placebo inference.

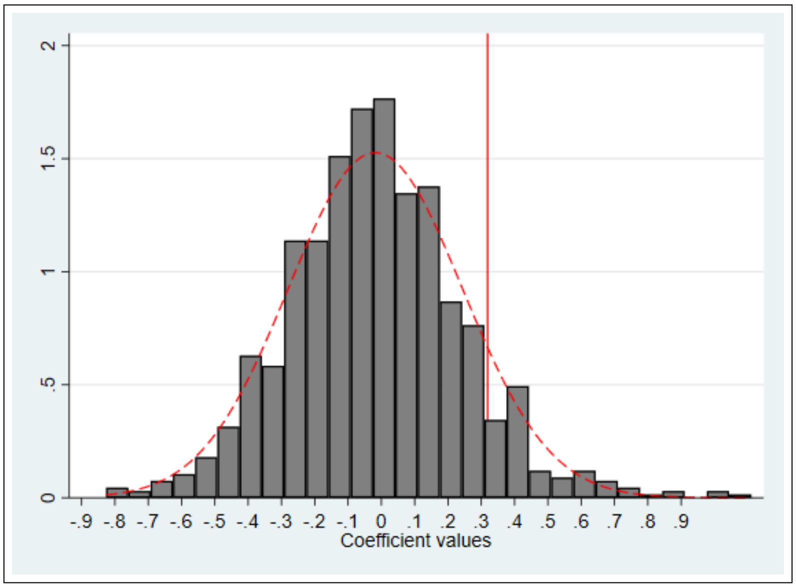
Dependent variables (log):	Forward Citation Weighted Patent Fractional Count							*Fake* Forward Citation Weighted Patent Fractional Count
	(1) Neighbours	(2) Donuts Hole	(3) SC-DiD	(4) PS-DiD	(5) IV	(6) 2SRI	(7) Placebo I	(8) Placebo II
HSR – Overall ATT	0.087	0.070	0.274*	0.238***	0.724***	0.671***		-0.135
	[0.105]	[0.117]	[0.153]	[0.080]	[0.250]	[0.185]		[0.119]
Fake HSR – Overall ATT							0.002	
							[0.003]	
NUTS 3 region FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
#Inventors	✓	✓	✓	✓	✓	✓	✓	✓
Institutional quality	✓	✓	✓	✓	✓	✓	✓	✓
Multimodal accessibility	✓	✓	✓	✓	✓	✓	✓	✓
Labor market controls	✓	✓	✓	✓	✓	✓	✓	✓
Economic controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3,560	3,200	3,560	2,466	3,560	3,560	3,520	3,560
Kleibergen–Paap F Statistic					49.55			

Notes: This Table presents robustness analyses on the effects of HSR access on regional innovation. Column (1) provides estimates of Equation 1 after extending the sample of treated regions to include also NUTS 3 regions contiguous to those actually hosting HSR stations; column (2) shows result of a "donuts-hole" exercise in which we re-estimate Equation 1 after considering only neighboring contiguous regions as the treated sample, while regions actually hosting an HSR station are excluded from both the treated and the control samples; column (3) provides Synthetic DiD estimates [Arkhangelsky et al., 2019]; in column (4), to improve comparability between treated and control regions, we estimate a propensity score using the observable characteristics from baseline controls (nearest neighbor[s] within caliper = 0.05), and we keep observations within the common support [0.1, 0.9], then running the DiD analysis of Equation 1 on the matched sample, applying propensity-score weights in the DiD estimation; in column (5) we instrument the actual HSR opening with the planned HSR network, so that our identification strategy leverages the gap between the planned HSR network and the realized opening dates of stations; in column (6) we rely on the aforementioned instrument to apply a two-stage residual inclusion (2SRI) approach [Wooldridge, 2015]; column (7) provides placebo estimates by replacing in Equation 1 the treatment interaction with a *fake* one, where it is drawn from Bernoulli distributions with parameters t (probability of success), derived from the original sample distributions; column (8) shows Equation 1's estimation results of another placebo test, in which the original dependent variable is replaced with a *fake* one that is drawn from random forward citations-weighted PFC distributions resembling sample ones (same mean and variance). All specifications include both NUTS-3 region and time fixed effects. Controls include predetermined regional characteristics at the beginning of the sample period, i.e. Inventors (log number of inventors patenting in 1982), Institutional quality (institutional quality index in 2004, as in Nifo and Vecchione [2015]), Multimodal accessibility (potential multimodal regional accessibility index in 2001), Labor market controls (log unit labor cost in 1982 and log compensation per hour worked in 1982), Economic controls (log per capita GDP in 1982 and industry share on GDP in 1982). The overall ATT aggregated by cohort and year is shown. Specifications in columns (1), (2), (4), (7) and (8) are estimated by OLS; the specification in column (3) is estimated via SC-DID, which learns unit and time weights from the pre-treatment period; the specification in column (5) is estimated by 2SLS; the specification in column (6) is estimated by 2SRI. Robust standard errors for specifications in columns (1), (2), (4), (5), (7) and (8) are clustered at NUTS-3 region level; standard errors for the specification in column (3) are obtained via unit-level block bootstrap with re-estimation of the weights (robust to heteroskedasticity and serial correlation); standard errors for the specification in column (6) are clustered at NUTS-3 region level and bootstrapped (1,000 replications) to account for first-stage estimation. Standard errors are shown in parentheses: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

from Bernoulli distributions, with parameter t (namely the probability of success) derived from both T_i and $Post_{k,t}$ original sampling distribution. In this setting, we should not observe any significant effect on regional innovative levels; comfortingly, results reported in column (7) of Table 5 confirm this prediction. On the other hand, we also implement a random-allocation (placebo) test by constructing a *fake* treatment indicator that equals 1 when a uniform distribution $[0,1]$ random draw exceeds the unit's sampling treatment probability. Using this placebo indicator, we re-estimate Equation 1 and repeat the procedure 1,000 times to obtain a distribution of placebo average treatment effects. We then compare this distribution with the estimated ATT reported in column (5) of Table 2 (ATT = 0.319). The logic is that a truly significant treatment effect should differ from effects generated under random (placebo) assignments. Rather comfortingly, in Figure 5, where the dark bars show the distribution of placebo coefficients from the *fake* iterations while the vertical red line marks the ATT based on actual treatments, the true effect falls well outside the core of the placebo distribution, on the right tail; moreover, the placebo estimates are approximately normally distributed and centered at zero, indicating no effect under *fake* assignments. As expected, some placebo draws may, by chance, approach the "true" estimate, occasionally placing it within the placebo range. Overall, the random-allocation test corroborates the robustness of our main results. Last but not least, we also estimate our baseline specification in Equation 1 after randomly assigning a value for the patent fractional count for each region; in particular, *fake* dependent variable values are drown from NUTS-3 regions specific random innovation distributions resembling sampling ones (same mean and variance). Again, estimated ATTs, shown in columns (8) of Table 5, are no longer significant, thus further confirming the robustness of our main results.

Finally, despite our analysis support the validity of the chosen staggered DiD research design, we

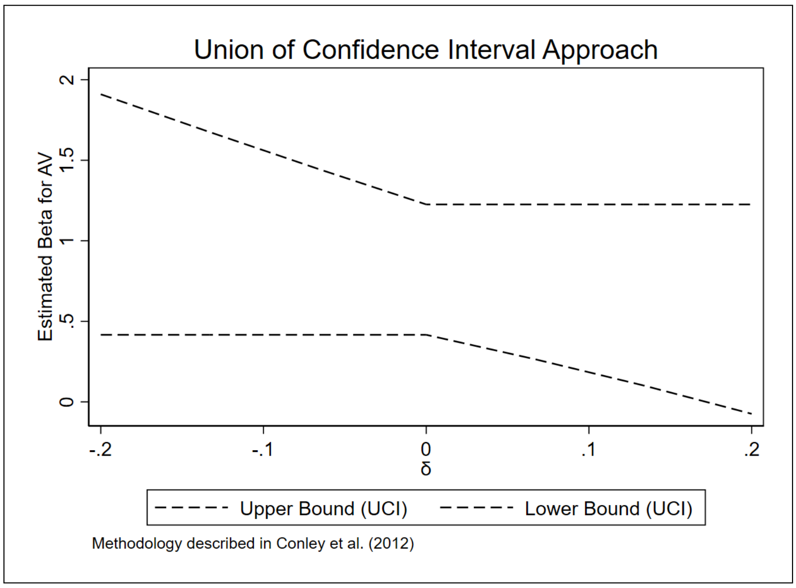
Figure 5: Random treatment allocation. Placebo plot test



Notes: Random treatment allocation. In this test we randomly assign treatments by generating simulated values for the treatment interaction. In particular, we build a *fake* treatment dummy (for 5 *fake* cohorts) equal to 1 if random numbers drawn from a uniform distribution [0, 1] are greater than the sample treatment probability. We then estimate Equation 1 after including the new *fake* treatment indicator. We iterate such procedure 1,000 times in order to obtain a distribution of placebo coefficients to compare with the actual estimated value shown in column (5) of Table 2. Dark bars represent the distribution of estimated placebo coefficients. The vertical solid red line represents the overall estimated average treatment effect (ATT = 0.319). The dashed red line fits a Normal distribution.

explore possible endogeneity of the treatment arising from reverse causality or from omitted variable issues. In particular, we instrument the actual HSR opening with the planned HSR network. Our identification strategy leverages the gap between the planned HSR network and the realized opening dates of stations. While the national plan fixed which regions would ultimately host a station, the actual roll-out schedule was steered by administrative, financial, and engineering contingencies that are largely orthogonal to local innovation dynamics. We instrument actual HSR access with the planned network so that causality is identified from deviations in opening timing relative to the plan, rather than from endogenous regional characteristics. Column (5) of Table 5 presents the limited information maximum likelihood (LIML) estimates, while column (6) shows the ATT estimated through a two-stage residual inclusion (2SRI) regression [Wooldridge, 2015], yielding a coefficient of quite similar magnitude. The value of the Kleibergen-Paap rk Wald F statistic, reported at the bottom of column (5) in Table 5, confirms that the chosen instrument is relevant, since it is strongly correlated with the endogenous variable (after including a set of controls), thus supporting the validity of the IV approach. Moreover, the Olea and Pflueger [2013]'s robust weak-instrument test yields an effective first-stage statistic of 49.55. At $\alpha = 10\%$, this exceeds all reported critical values for TSLS/LIML—33.105 (guaranteeing $\leq 5\%$ worst-case relative bias), 19.748 ($\leq 10\%$), 12.374 ($\leq 20\%$), and 9.650 ($\leq 30\%$)—so the instrument is rather strong and weak-IV concerns are minimal. Last but not least, to further scrutinize the exclusion restriction, we follow Conley et al. [2012], which relax the strict exclusion assumption by allowing the instrument to have a small, direct effect on the outcome. Their union of confidence intervals (UCI) approach allows us to estimate the upper and the lower bounds of the effect of HSR access on regional innovation, remaining valid even if the instrument is only approximately exogenous. Results in Figure 6, where the horizontal axis represents the tolerance of the deviation from "strict exogeneity," and the vertical axis represents the estimated interval of the HSR access coefficient, suggest that none of those confidence intervals includes zero, so that we can argue that our findings are not sensitive to possible violations of the exclusion restriction, and thus that the chosen instrument can be considered as "plausibly exogenous" [Conley et al., 2012].

Figure 6: Bounds of the 90% confidence intervals of β, associated with different values of γ.



Overall results are consistent with our main analysis and are in line with Hanley et al. [2022] who find that the IV estimates are higher with respect to OLS results. According to Cao et al. [2024], a likely source of the downward OLS bias is the government's focus on connecting core to peripheral areas instead of reinforcing existing economic–innovation networks.[31]

4 Conclusions

This paper examines whether connecting Italian provinces to the high–speed rail (HSR) network raises regional innovation performance, on the premise that faster intercity travel reduce effective distance and lowers frictions to face-to-face knowledge exchange. We build a balanced panel for 89 NUTS-3 regions observed over the period 1980–2019 and we measure innovative performance with forward citation-weighted patent fractional counts allocated by inventor residence and complemented by inventor information and predetermined regional covariates (institutional quality, multimodal accessibility, labor market and economic indicators).

Identification relies on an extended two-way fixed-effects (ETWFE) difference-in-differences where treated regions are those hosting an HSR station; a full set of robustness test validates our research design and support the parallel trend assumption. The results point to economically meaningful and statistically significant gains in regional citation-weighted patenting by about 0.24 to 0.33 log points following HSR access, with post-treatment dynamics remaining positive in the medium run. Moreover, estimates of a dyadic gravity model based on inter-regional collaboration indicates an increase of about 30% in co-inventorship after station openings. Further analysis based on a social network approach shows higher betweenness, closeness and eigenvector centrality (on the order of 2–9%) for hosting regions, thus suggesting deeper embedding in the national collaboration network. The heterogeneity analysis based on technological sectors highlights that effects are concentrated in chemistry, electrical

[31] Also Chen and Guo [2023] instrument station assignment at city level with the 2004 national HSR blueprint in China.

and mechanical engineering, aligning with domains where iterative testing, supply-chain coordination and tacit problem-solving benefit most from time compression. Finally, our estimates do not provide evidence in favor of spillover/displacement effects, thus suggesting that innovation gains associated to HSR investments did not generate crowding-out effects in neighboring regions. Overall results are confirmed by an IV analysis where we instrument realized access with the planned HSR network.

Our findings generate different policy implications and suggest that HSR acts as knowledge infrastructure: by enlarging effective collaboration markets it raises both the quality and the quantity, as well as the organization of innovative activity; however such benefits are distributed across places and sectors with different intensity. In particular, our results suggest that the higher impacts of HSR investments on regional innovation are observed for intermediate and smaller regions, thus supporting the role of transport infrastructure as policy instruments able to reduce regional economic divide. Moreover, the evidence of heterogeneous sectoral responses to HSR investments suggests that policies should be sector-matched rather than generic. In chemistry and life sciences, certification facilities and shared instrumentation near stations accelerate lab-to-market cycles; in electrical and mechanical engineering, testbeds, metrology centres and supplier–client prototyping hubs exploit same-day coordination.

References

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

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

J. E. Anderson and E. Van Wincoop. Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93(1):170–192, 2003.

D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference in differences. Technical report, National Bureau of Economic Research, 2019.

D. B. Audretsch and M. P. Feldman. Chapter 61 - Knowledge spillovers and the geography of innovation. In J. V. Henderson and J.-F. Thisse, editors, *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2713–2739. Elsevier, 2004. doi: https://doi.org/10.1016/S1574-0080(04)80018-X. URL `https://www.sciencedirect.com/science/article/pii/S157400800480018X`.

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

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

A. Bhatt and H. Kato. High-speed rails and knowledge productivity: a global perspective. *Transport Policy*, 101:174–186, 2021.

C. Blanquart and M. Koning. The local economic impacts of high-speed railways: theories and facts. *European Transport Research Review*, 9(2):12, 2017.

P. Bonacich. Power and centrality: a family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

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

A. Bottasso, M. Conti, A. Ferrara, and S. Robbiano. High-speed railways and firms total factor productivity: evidence from a quasi-natural experiment. Technical report, IZA Discussion Papers Series, 2023.

A. Bottasso, M. Conti, A. Fiduccia, S. Robbiano, and M. Tartaglia. Does the opening of a high-speed-rail station impact house prices? Evidence from a quasi-natural experiment: Reggio Emilia Mediopadana. In F. Pagliara, editor, *Socioeconomic Impacts of High-Speed Rail Systems*, pages 319–341, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-82528-6.

M. C. Bravo, J. Roth, and A. Rambachan. Honestdid: Stata module implementing the honestdid r package. Technical report, Boston College Department of Economics, 2024.

S. Cao, G. Ma, and H. Mao. Is distance really dying? Transportation and knowledge spillovers. *Journal of Regional Science*, 64(2):355–405, 2024.

J. Chen and G. Guo. Diversification or specialization: measuring the impact of high-speed rail connection on technological diversity in China. *Journal of Innovation & Knowledge*, 8(1):100306, 2023.

D. Clarke. Plausexog: Stata module to implement Conley et al's plausibly exogenous bounds. Technical report, Boston College Department of Economics, 2020.

W. M. Cohen, R. Nelson, and J. P. Walsh. Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent (or not), 2000.

T. G. Conley, C. B. Hansen, and P. E. Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012.

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

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

M. Dosso and D. Lebert. The centrality of regions in corporate knowledge flows and the implications for smart specialisation strategies. *Regional Studies*, 54(10):1366–1376, 2020. doi: 10.1080/00343404.2019.1619923. URL https://doi.org/10.1080/00343404.2019.1619923.

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

L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

P. Galaso and J. Kovářík. Collaboration networks, geography and innovation: local and national embeddedness. *Papers in Regional Science*, 100(2):349–378, 2021. ISSN 1056-8190. doi: https://doi.org/10.1111/pirs.12578. URL https://www.sciencedirect.com/science/article/pii/S1056819023003160.

G. Gambuli and F. Stipanicic Márquez. High-speed railways and the geography of innovation: evidence from France. *Available at SSRN 5495056*, 2025.

Y. Gao and J. Zheng. The impact of high-speed rail on innovation: an empirical test of the companion innovation hypothesis of transportation improvement with China's manufacturing firms. *World Development*, 127:104838, 2020.

B. H. Hall, A. Jaffe, and M. Trajtenberg. Market value and patent citations. *RAND Journal of Economics*, 36(1):16–39, 2005.

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

K. Hayakawa, H. Koster, T. Tabuchi, and J.-F. Thisse. High-speed rail and the spatial distribution of economic activity: evidence from Japan's Shinkansen. *CEPR discussion paper no. DP15771*, 2021.

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

H. R. A. Koster, T. Tabuchi, and J.-F. Thisse. To be connected or not to be connected? The role of long-haul economies. *Journal of Economic Geography*, 22(4):711–753, 12 2021. ISSN 1468-2702. doi: 10.1093/jeg/lbab042. URL `https://doi.org/10.1093/jeg/lbab042`.

R. C. Levin, A. K. Klevorick, R. R. Nelson, S. G. Winter, R. Gilbert, and Z. Griliches. Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity*, 1987(3): 783–831, 1987.

Y. Lin, Y. Qin, and Z. Xie. Does foreign technology transfer spur domestic innovation? Evidence from the high-speed rail sector in China. *Journal of Comparative Economics*, 49(1):212–229, 2021.

J. Lu and H. Li. Can high-speed rail improve enterprise capacity utilization? A perspective of supply side and demand side. *Transport Policy*, 115:152–163, 2022.

J. T. Macher and D. C. Mowery. Vertical specialization and industry structure in high technology industries. In *Business Strategy Over the Industry Lifecycle*, pages 317–355. Emerald Group Publishing Limited, 2004.

P. C. Melo, D. J. Graham, and R. Brage-Ardao. The productivity of transport infrastructure investment: a meta-analysis of empirical evidence. *Regional Science and Urban Economics*, 43(5):695–706, 2013.

T. Mitze and F. Strotebeck. Centrality and get-richer mechanisms in interregional knowledge networks. *Regional Studies*, 52(11):1477–1489, 2018.

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

D. Neumark and H. Simpson. Chapter 18 - Place-Based Policies. In G. Duranton, J. V. Henderson, and W. C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1197–1287. Elsevier, 2015. doi: https://doi.org/10.1016/B978-0-444-59531-7.00018-1. URL `https://www.sciencedirect.com/science/article/pii/B9780444595317000181`.

A. Nifo and G. Vecchione. Measuring institutional quality in Italy. *Rivista Economica del Mezzogiorno*, 1–2(1–2):157–182, 2015. doi: 10.1432/80447:y:2015:i:1-2:p:157-182.

J. L. M. Olea and C. Pflueger. A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369, 2013.

J. Owen-Smith and W. W. Powell. Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. *Organization Science*, 15(1):5–21, 2004.

A. Pakes and Z. Griliches. Patents and R&D at the firm level: a first report. *Economics Letters*, 5(4): 377–381, 1980.

L. Picci. The internationalization of inventive activity: a gravity model using patent data. *Research Policy*, 39(8):1070–1081, 2010.

A. Rambachan and J. Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.

S. J. Redding and M. A. Turner. Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, 5:1339–1398, 2015.

F. Rios-Avila, A. Nagengast, and Y. Yotov. Jwdid: Stata module to estimate difference-in-difference models using Mundlak approach, 2024.

K. Smith. Chapter 6 - Measuring Innovation. In *The Oxford Handbook of Innovation*. Oxford University Press, 01 2006. ISBN 9780199286805. doi: 10.1093/oxfordhb/9780199286805.003.0006. URL `https://doi.org/10.1093/oxfordhb/9780199286805.003.0006`.

J. H. Stock and M. Yogo. Testing for weak instruments in linear IV regression, 2002.

D. Sun, S. Zeng, H. Ma, and J. J. Shi. How do high-speed railways spur innovation? *IEEE Transactions on Engineering Management*, 70(11):3944–3957, 2021.

S. H. Thomke. Managing experimentation in the design of new products. *Management science*, 44(6): 743–762, 1998.

M. Trajtenberg. A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, pages 172–187, 1990.

M. Trunschke, B. Peters, D. Czarnitzki, and C. Rammer. Pandemic effects: do innovation activities of firms suffer from long COVID? *Research Policy*, 53(7):105024, 2024. ISSN 0048-7333. doi: https://doi.org/10.1016/j.respol.2024.105024. URL `https://www.sciencedirect.com/science/article/pii/S0048733324000738`.

Y. Wang, G. Cao, Y. Yan, and J. Wang. Does high-speed rail stimulate cross-city technological innovation collaboration? Evidence from China. *Transport Policy*, 116:119–131, 2022.

I. Wanzenböck and P. Piribauer. R&D networks and regional knowledge production in Europe: evidence from a space-time model. *Papers in Regional Science*, 97:S1–S24, 2018.

I. Wanzenböck, T. Scherngell, and R. Lata. Embeddedness of European regions in European Union-funded research and development (R&D) networks: a spatial econometric perspective. *Regional Studies*, 49(10):1685–1705, 2015.

I. Wanzenboeck, T. Scherngell, and T. Brenner. Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. *The Annals of Regional Science*, 53(2):337–368, 2014.

Y. Wen, Y. Liu, L. Sheng, and Z. Yu. The impacts of high-speed rail networks on urban green innovation. *Transport Policy*, 163:168–184, 2025.

K. B. Whittington, J. Owen-Smith, and W. W. Powell. Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative Science Quarterly*, 54(1):90–122, 2009.

J. M. Wooldridge. Control function methods in applied econometrics. *The Journal of Human Resources*, 50(2):420–445, 2015. ISSN 0022166X. URL `http://www.jstor.org/stable/24735991`.

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

J. M. Wooldridge. Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. *Empirical Economics*, pages 1–43, 2025.

X. Wu and J. Yu. Bridging innovation distances: high-speed rail and patent agency utilization in China. *Journal of Regional Science*, 65(2):446–470, 2025.