High-Speed Railways and the Geography of Innovation: Evidence from France

Gabrielle Gambuli Fernando Stipanicic*

September 4, 2025

Abstract

This paper provides new causal evidence of the impact of train travel time on patent collaboration between inventors. We construct a novel dataset of train travel times in France between 1980 and 2010 and exploit the roll out of High Speed Railways (HSR) as a quasi-natural experiment. The median decrease in travel time of 12% led to a 2.6% increase in patent collaborations across commuting zones. The effect on the increase in collaborations was stronger for most productive commuting zones and the most productive inventors within them, increasing the collaboration gap across and within regions.
JEL Classification: O31, O33, R12, R41, N72, F60

*Gambuli: Université Gustave Eiffel & INSEE (gabrielle.gambuli@gmail.com)
Stipanicic: University of Oslo & BI Norwegian Business School (fernando.stipanicic@gmail.com)

1 Introduction

Innovation and technological progress are key drivers of economic growth, yet their distribution across space remains highly uneven. Since at least the publication of Vannevar Bush's report, "Science, the Endless Frontier," in 1945, countries have promoted policies to increase innovation. Recent policies and recommendations, such as the CHIPS Act in the United States (U.S. Congress (2022)) and the Draghi Report in the European Union (Draghi (2024)), make a strong emphasis on fostering regional innovation with the goal of reducing regional disparities. A growing body of research highlights the importance of face-to-face interactions in driving innovation thorugh knowledge spillovers (Catalini (2018), Atkin, Chen and Popov (2022)), and that this knowledge spillovers are strongly affected by distance (Jaffe, Trajtenberg and Henderson (1993), Belenzon and Schankerman (2013)). Recent studies show improvements in highway infrastructure (Agrawal, Galasso and Oettl (2017)) and reductions in air travel time (Pauly and Stipanicic (2022)) have facilitated knowledge spillovers, potentially affecting regional innovation. However, less is known about the role of high speed railways (HSR) in shaping the geography of innovation, particularly in Europe.

This paper studies the impact of reductions in travel time due to the expansion of high-speed railways (HSR) in France on patent collaborations, called copatents. The introduction of HSR led to a substantial decrease in travel time, facilitating long-distance face-to-face interactions among inventors. We construct a new dataset of train travel times from 1980 to 2019 and match it geo-referenced patent data from the European Patent Office (1980–2010) aggregated to the commuting zone level (CZ). We estimate a gravity model and find that the 1980-2010 median decrease in travel time of 12% led to 2.6% increase in patent collaborations across CZs. However, we find that the increase in copatents is stronger for CZ pairs in which both CZ are more developed, and between inventors who were previously more productive than average within their CZ. Hence, while the decrease in travel time increased the amount of across-CZ copatents, this may be accompanied by an increase in innovation inequality within and across CZs.

Our identification strategy exploits the staggered rollout of HSR, which introduced variation in travel times across CZ pairs. The identification assumption is that the timing of adoption of HSR is exogenous, which is reasonable given the uncertainty around completion times. We include CZ-pair, origin-time and destination-time fixed effects in the regression, which absorb time invariant characteristics at the CZ-pair level, and time varying shocks at the origin and destination CZs.

The roll out of HSR in France started with the inaguration of the first line in 1981, connecting Paris and Lyon, which are located approximately 400 kilometers apart. This development reduced train travel time from 3 hours and 28 minutes to 1 hour and 53 minutes, making it viable for individuals to complete a return trip within a single day. Over the years multiple other cities became connected to Paris: Le Mans (1989), Tours (1990), Lille (1993), Marseille (2001), Montpellier (2001), Strasbourg (2007-2016), Bordeaux (2017), and Rennes (2017). By 2021, the high-speed rail network covered 2,800 kilometers (Coste (2021)).

We construct a new dataset of train travel time in France. To do so we develop a new method to estimate the unobserved counterfactual travel time before the arrival of HSR. Our method is general and can be applied to reconstruct unobserved travel time in other set ups when the econometrician counts with partial information on the transportation technology. We document in a sample of city-pairs that our method replicates between 62% and 88% of the observed changes in travel time. With our estimated travel times, we document that between 1980 and 2010 the average reduction in train travel time was 13%, while the median and the 75th quantile were 12% and 24%, respectively.

We assemble a dataset of collaborative patents in France granted by the European Patent Office. We take patents with two or more inventors georeferenced by Morrison, Riccaboni and Pammolli (2017) and aggregate it to the CZ-pair year level. Between 1980 and 2010, the share of collaborative patents increased from 43% to 65%, with inter-regional collaborations accounting for 16 percentage points of the growth. By 2010, inter-regional copatents accounted for half of all copatents in France.

We estimate a regression of copatents on travel time. The empirical strategy exploits only changes in travel time and copatents that are differential across CZ pairs, absorbing any aggregate changes in travel time or copatents. Following Silva and Tenreyro (2006) we estimate the regression by Poission Pseudo Maximum Likelihood (PPML) which allows for zeros in the left hand size variable, and gives an unbiased estimate in the case of heteroskedasticity of the underlying multiplicative model. We estimate that the elasticity of copatents to travel time is -0.216, with the effect mostly coming for CZ-pairs at 100km-400km distance. This result is plausible, as inventors may choose to travel by car for short distances, and by airplane for longer distances.[1] We estimate the regression separately for subsamples of CZ pairs where either both, one or none of them has an HSR station by 2021, finding that the effect is largest when both origin and destination have an HSR station.

We split CZs by their initial level of innovativeness based on patents filed in 1980-1985, and find that the elasticity is only negative and significant for collaborations between initially highly innovative locations. In consequence, HSR leads to increased collaborations between CZs that are already the more developed ones.

Additionally, we split copatents based on whether each of the inventors in the inventor-pair is over or under the inventors' region average productivity, and whether she is over or under the collaborators' region average productivity. We proxy productivity by the number of patents filed up to year t-1 weighted by 5 year forward citations. The elasticity is only significantly different from zero when at least one of the inventors is above the average productivity, and it is largest in absolute value when both inventors are over average productivity. Hence, the decrease in travel time increases collaborations between inventors that were already more productive.

Opening up copatents by other characteristics, we estimate an elasticity that is three to four times as large in the case of inventor-pairs that previously collaborated relative to new inventor-pairs. Moreover, while we find a coefficient that is negative for both within and across-firm collaborations, it is imprecisely estimated. The effect remains when weighting copatents by amount of claims and technology classes, suggesting that the increased copatenting does not have reduced scope nor breadth. We find an elasticity not significantly different from zero when weighting copatents by their citations received in a 5-year window, though we are cautious interpreting this result as most patents receive zero citations and hence

[1] In 2010, the average train travel time for distances over 400km in the effective sample was 4 hours and 53 minutes.

the effective sample may be substantially different. Last, we find that the increase in collaborative patents happens across all technology categories and is robust to controlling for technology-time fixed effects.

Literature. This paper contributes to the literature on transportation infrastructure and innovation by providing new evidence on the effect of HSR on patent collaborations. While previous research has shown that highways (Agrawal, Galasso and Oettl (2017)) and air travel (Pauly and Stipanicic (2022); Bahar et al. (2023)) facilitate knowledge spillovers, less is known about the role of HSR.[2] A distinct characteristic of HSR is that it is a means of transportation for people rather than goods, making face to face interactions the relevant mechanism at play. Our results show that reductions in train travel time increase inter-regional patenting, particularly at intermediate distances where neither car nor air travel is the dominant mode of mobility.

Second, we contribute to the literature on spatial inequality and regional innovation (Moretti (2012); Duranton et al. (2009); Carlino and Kerr (2015); Gross and Sampat (2023)). While recent policies emphasize fostering innovation in less developed areas (U.S. Congress (2022); Draghi (2024)), our findings suggest that HSR primarily strengthens collaborations between already developed regions. By showing that increased collaboration is driven by pre-existing and new inventor-pair collaborations, and by more similar inventor-pairs, we also shed light on how reduced travel time shapes the organization of innovation.

Finally, we make a methodological contribution by developing a new approach to estimate historical travel times in the absence of comprehensive records. Our method replicates a large share of the variance in observed changes in travel time and can be applied to other settings where transportation infrastructure has evolved. This provides a useful tool for studying the long-term effects of transportation improvements on economic outcomes.

2 Train travel time data

We construct a new train travel time dataset between CZ of continental France at the yearly frequency between 1980 and 2019. This period includes the roll out of high speed railways, which are around twice as fast as the previous train technology. We document that during this period the median and 3rd quartile reduction in travel time are 19% and 30%, respectively. This reduction in travel time affected only passenger transport, as high speed railways are not used to transport goods. While the plans for construction of high speed railways were publicly advertised, the date of opening of new routes was uncertain, introducing randomness in the timing of travel time reductions. Appendix B provides more details on each of the following subsections.

The method used for constructing this dataset is general and can be applied for constructing other data sets in which previous values of a variable are not observed. This method is especially useful when the analyst knows when the new technology was implemented and either has knowledge of the effi-

[2]Dong, Zheng and Kahn (2020) studies the impact of HSR in China on scientific collaboration and Tsiachtsiras et al. (2022) do so on local innovation and technological specialization. To the best of our knowledge, no study has examined the impact of HSR on innovation in a European setting

ciency of the previous technology (e.g. the speed of trains used previously in those routes, combined with route distance), or the efficiency can be estimated from currently observed values (e.g. the speed can be inferred from older train types that are still used in other routes). Hence, our method is a general contribution about how to construct datasets where past values are not observed, and the current application operates as a proof of concept for such method.

2.1 Roll out of High Speed Railways in France

High speed railways in France have a maximum speed around 320 km/h, compared to around 160 km/h for the alternative Intercités train. Starting with the connection between Lyon and Saint Florentin in 1981, France has progressively expanded its high-speed rail network, reaching 2,800 km by 2021.

Figure 1 shows the roll out of high speed railways by decade. The new high speed railways connected Paris with largest cities in a gradual manner. The first segment to Lyon opened in 1981, then Le Mans (1989), Tours (1990), Lille (1993), Marseille (2001), Montpellier (2001), Strasbourg (2007-2016), Bordeaux (2017), and Rennes (2017). Given the staggered roll out and the network nature of train travel, one city pair could have reductions in travel time multiple times. For example, Paris-Marseille travel time went from 5 hours 47 minutes in 1980, to 4 hours 29 minutes in 1983, to 3 hours 2 minutes in 2001.

The date of opening of high-speed railways was uncertain due to financial constraints, political negotiations, judicial rulings, and environmental concerns.[3] For example, the Paris-Lille line was originally tied to the construction of the Channel Tunnel, but the British government withdrew support in 1975 for financial reasons, leading France to shift its focus to the Paris-Lyon line instead. Political agreements later revived the Paris-Lille project, and despite a relatively quick construction period, public opposition in Lille led to security forces being deployed to control protests. Similarly, the Bordeaux-Toulouse line did not begin construction until three decades after its initial announcement due to legal battles, environmental opposition, and financial disputes, with a court ruling initially halting the project before a later decision allowed it to proceed. Other projects included in national railway plans, such as the Paris-Le Havre line, were ultimately abandoned due to concerns over financial viability, further highlighting the unpredictability of HSR expansions. This uncertainty in completion dates, driven by factors unrelated to regional innovation, introduces arguably exogenous variation in the timing of travel time reductions, which we exploit for identification.

2.2 Data construction: train travel times

We construct a data set on train travel times between all continental French CZ at the yearly frequency between 1980 and 2019. We use 2020 CZ definitions. We use as input the December 2021 travel times from the French railway company SCNF and manually gather the opening dates of each high speed railway segment.[4] Based on the 2021 train network, we estimate the travel speed of each train-railway type.

[3] Details are provided in Appendix A.

[4] High speed trains (TGV – Trains à Grande Vitesse) operate in both high speed railways and normal railways. Other train types do not operate on high speed railways.

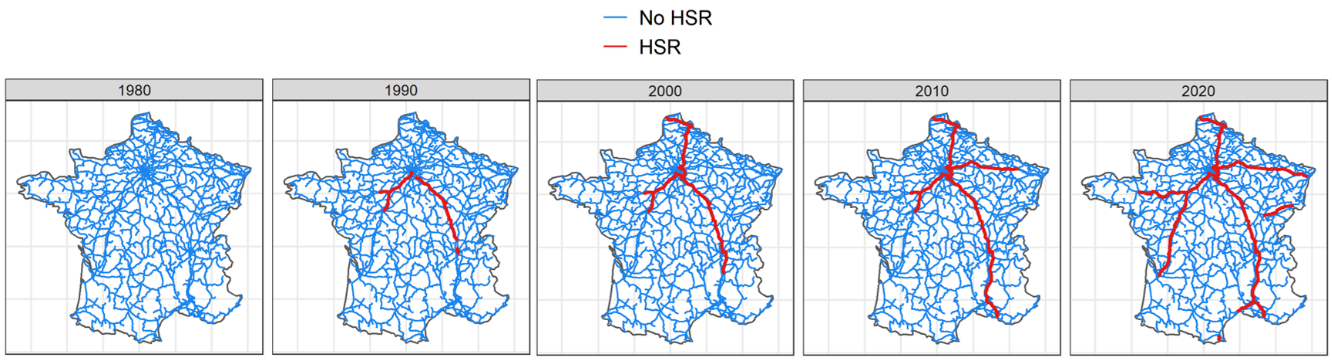

Figure 1: Roll of high speed railways in France

With these estimates we simulate the travel time that would have taken place in each of the segments before the opening of the high speed railway. We validate the travel time against observed travel times for a sample of city-pair years provided by SNCF and document that the constructed dataset accounts for between 62% and 88% of the observed changes in travel time.[5]

We start with all trips by train and railway type during a week of December 2021, adding up to 47,298 trips in 416 routes (origin-destination station pairs). We split each trip, which may include multiple stops, into segments of non-stop station pairs. For each segment and train-railway type, we take the minimum travel time across all trips. We then estimate a linear regression of the minimum travel time on distance for each train-railway type. Appendix Table 3 presents the results for the estimations. For the relevant train types, high speed railways and Intercités, these simple linear models explain around 95% of the variance showing the predicted values would be a good approximation for observed values. The estimated travel speed for high speed trains operating in high speed railways is 229 km/h, while it is 116 km/h for the second fastest train type Intercités.

With the estimated speeds, we impute the counterfactual Intercités travel time for each non-stop station-pair operated with high speed railway in December 2021. We use the counteractual travel time for each non-stop station-pair in all years previous to the year in which the high speed railway opened in that segment, switching then to the observed travel time with high speed railway. For all other station-pairs that are not connected with high speed railways we use the observed travel times from the December 2021 dataset. Hence, across-time variation in travel time comes only from switching from the counterfactual Intercités travel time to the observed travel time with high speed railways, everything else remains constant.

Next, using the travel time between non-stop station pairs for each year, we run the Dijkstra algorithm (Dijkstra (1959)) to obtain the fastest route and travel time between all station pairs in each year.[6]

[5]Appendix Table 4 regresses observed travel time on our predicted travel time for the sample of city-pair years. The regression gives a R2 of 88% when including origin-destination fixed effect. The R2 is 62% when including origin-destination, origin-time and destination-time fixed effects. Hence, our predicted travel times capture a large share of the observed variation of changes in travel time across-destinations with the same origin, which is the variation used for identification in our analysis.

[6]We set a zero-minute penalty for switching trains within a station. We allow for changes of station within a city imputing a travel time using the distance between stations and assuming the change happens at 20 km/h.

Then, we keep only origin-destination stations that belong to the most populated city in each CZ according to 1975 population, leaving us with 273 out of 280 CZ in continental France.[7] Finally, we take the minimum travel time across station-pairs within each CZ-pair for each year.

2.3 Descriptive statistics: Train travel times

For each CZ-pair, we compute the change in travel time relative to 1980. Figure 2 shows the change in travel time within CZ-pair relative to 1980, averaged within 100km distance bins. The change in travel time is non-uniform, having in general larger reductions in travel time for CZ-pairs that are farther apart. The average reduction in travel time is 13% in 2010 and 18% in 2019. The median, 75th percentile and 90th percentile reduction in travel time are, respectively, 12%, 24% and 32% in 2010, and 19%, 30% and 36% in 2019.

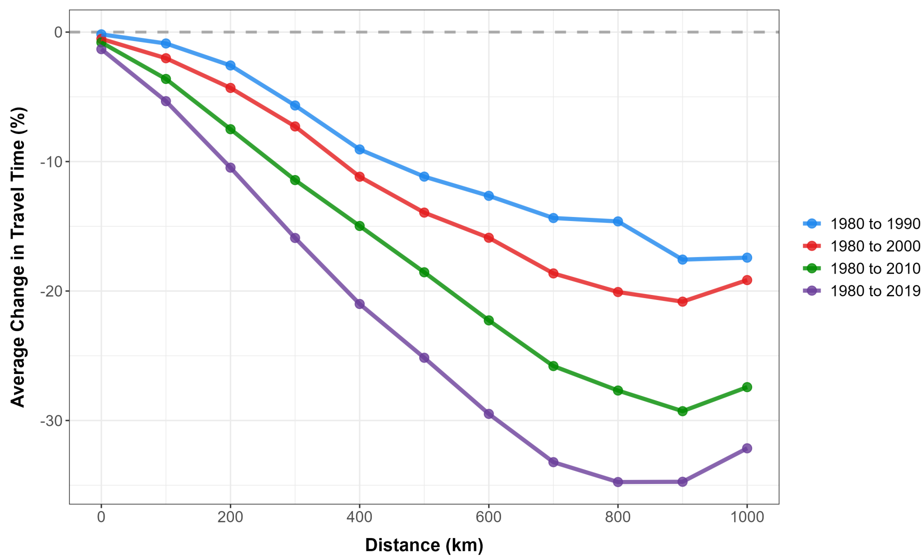

Figure 2: Change in travel time relative to 1980

3 Patent data

We assemble a dataset of patents granted by the European Patent Office (EPO) with inventors residing in France and application year 1980-2010. We obtain geo-referenced patent information from Morrison, Riccaboni and Pammolli (2017), with time invariant identifiers both for inventors and assignees. This dataset covers 1978-2014. We restrict the sample to patents applied between 1980-2010 because, first, we observe a spike in patent applications in 1978, year of creation of the EPO. Second, there is a lag between patent application and patent granting, which introduces measurement error towards the latest years of the sample. We keep only multi-inventor patents that have at least 2 geo-referenced inventors located in

[7]The 7 CZ that did not have a train station in our data accounted for 0.67% of 1975 population.

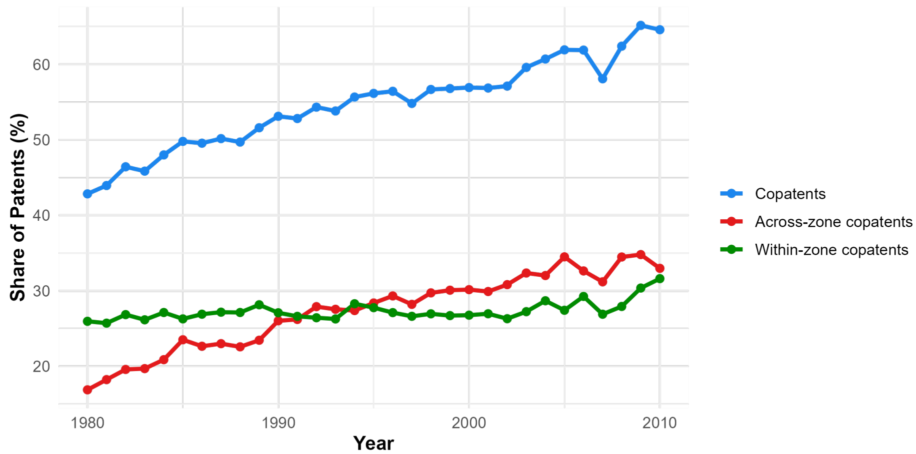

Figure 3: Share of copatents in all copatents

France. Our selected sample includes 88,748 patents done by 106,045 inventors located in 273 CZ. Then, for each CZ-pair year, we count the amount of patents that have inventors in both CZs.

3.1 Descriptive statistics: Patent data

Figure 3 shows the share of copatents in the total number of patents, and splits the share by copatents that have all inventors in the same CZ (within-zone) and copatents that have inventors in multiple CZs (across-zone). The share of copatents increased from 43% in 1980 to 65% in 2010, with the increase in across-zone copatents accounting for 16 out of the 22 percentage points increase. The increase in the share of copatents is accompanied by an increase in the distance between inventors within a team. The average and 90th percentile of distance between collaborators increased respectively by 36% and 70%.

4 Analysis

4.1 Time varying effect of geography

We begin the analysis by providing evidence that the effect of geography on copatents has been decreasing over time. We do so by estimating in gravity equation with, first, the time varying effect of within-CZ collaboration, and second, the time varying effect of distance.

Figure 3 shows that 57% of copatents had all inventors located within the same CZ in 1980, reducing to 35% by 2010. Figure 4 shows that, after controlling for pair, origin-time and destination-time fixed effects, the effect of within-CZ collaboration on copatents has decreased relative to 1980. Figure 5 shows that, relative to 1980, the negative effect of distance on copatents became less negative or "softened" over time. The fixed effects included on both regressions, which are explained in more detail in the Section 4.2, imply that the estimated coefficients are not driven by time-varying factors at origin or destination level, and that it is rather driven by changes at the origin-destination pair level.

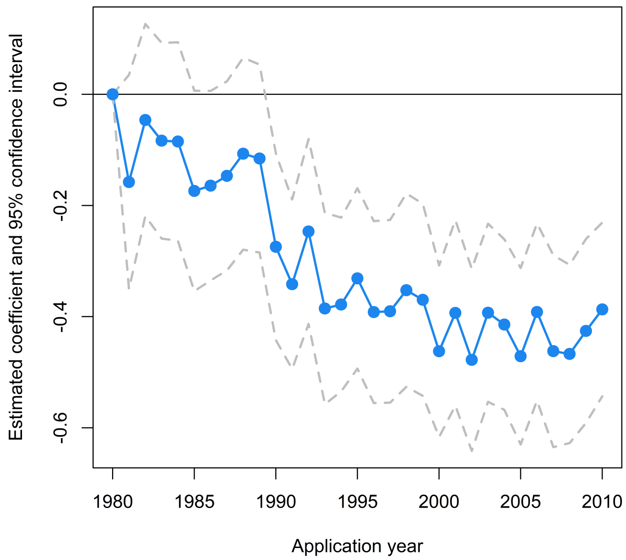

Figure 4: Time varying effect of within-commuting zone on copatents
The plot shows the point estimate and 95% confidence interval of α_t obtained estimating $copatents_{ijt} = exp(\sum_t \alpha_t \mathbb{1}\{i = j\} + \mu_{ij} + \mu_{it} + \mu_{jt}) \times \varepsilon_{ijt}$, normalized to 1980. Confidence intervals are obtained with ij clustered standard errors.

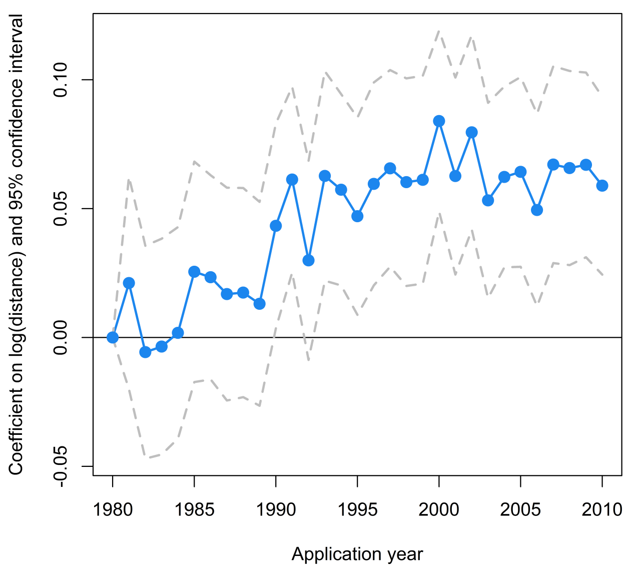

Figure 5: Time varying effect of distance on copatents
The plot shows the point estimate and 95% confidence interval of α_t obtained estimating $copatents_{ijt} = exp(\sum_t \alpha_t \log(\text{distance}_{ij}) + \mu_{ij} + \mu_{it} + \mu_{jt}) \times \varepsilon_{ijt}$, normalized to 1980. Confidence intervals are obtained with ij clustered standard errors.

4.2 Effect of travel time

To study the effect of travel time on copatents we estimate the following gravity equation:

$$\text{copatents}_{ijt} = \exp\left[\beta \, \log(\text{travel time}_{ijt}) + \mu_{ij} + \mu_{it} + \mu_{jt}\right] \times \varepsilon_{ijt} \tag{1}$$

for origin CZ i, destination CZ j and year t. copatents_{ijt} is the number of copatents and $\log(\text{travel time}_{ijt})$ is the log of train travel time.[8,9] β is the elasticity of copatents to travel time. Following Silva and Tenreyro (2006) we estimate equation 1 by Poisson Pseudo Maximum Likelihood (PPML) as this method allows to accomodate zeros in left hand side variable, and is unbiased in the case of heteroskedasticity of the underlying multiplicative gravity model. We cluster standard errors at the CZ-pair level.

The fixed effect μ_{ij} absorbs time invariant factors at the CZ-pair level, such as distance and cultural proximity. In particular, if there are unobserved time-invariant determinants for which CZ-pairs receive a high speed railway that connects them, such as business ties, μ_{ij} would absorb such determinants. The fixed effect μ_{it} absorbs time varying shocks at the origin CZ, like changes in population or local policies which affect the supply or demand of innovation. Similarly, the fixed effect μ_{jt} absorbs time varying shocks at the destination CZ.

In equation 1 the identification of β comes from across-time changes in copatents and travel time within a CZ-pair, relative to other CZ-pairs with the same origin CZ, conditional on time varying shocks to the destination CZ. Hence, the identification is not driven by whether a certain CZ becomes more central in either the copatent or train the network, or whether it has increases or decreases of population or economic activity. Rather, the identification comes from how the train connectivity of a CZ-pair evolves over time relative to other CZ-pairs starting from the same origin CZ.

The identification assumption is that the timing of roll out of high speed railways, and hence reductions in travel time, is exogenous to the copatent activity at the CZ-pair level. This assumption is plausible given the large uncertainty about the opening dates of high speed railways. Appendix C Figure 2 presents an event study version of equation 1 that illustrates little anticipation effect before the reduction in travel time.[10]

Table 1 presents the results of estimating equation 1. Column 1 shows that the elasticity of copatents to travel time is -0.216, significant at the 5% level. In column 2 we open up the elasticity by distance between origin and destination CZs. It is likely that travel time by train is not relevant at short distances, where the relevant measure may be car travel time, or at long distances, where the relevant measure may be airplane travel time. We find that the elasticity is largest in absolute value at distance between 100km and 400km, and it is imprecisely estimated distances under 100km or over 400km. While Figure 2

[8] While copatents are in principle non-directional, we treat them as directional as this allows to estimate the gravity equation in a similar manner as in international trade models. Hence, both copatents_{ijt} and copatents_{jit} appear in the data and $\text{copatents}_{ijt} = \text{copatents}_{jit}$.

[9] We assume passengers take a round trip, hence we make travel time symmetric, i.e. $\text{travel time}_{ijt} = (\widetilde{\text{travel time}}_{ijt} + \widetilde{\text{travel time}}_{jit})/2$, where $\widetilde{\text{travel time}}_{ijt}$ is our constructed travel time.

[10] We adapt the linear version of event study analysis with multiple treatment of different intensities of Schmidheiny and Siegloch (2023) into a non-linear PPML estimation.

shows that the decrease in travel time was larger for longer distances, in 2010 the average travel time for distances over 400km in the effective sample was 4h 53min. Hence, it is likely that total travel time by airplane, accounting for travel time to/from the airport and security checks, may still be lower than by train at distances over 400km.

Table 1: Effect of travel time on number of collaborative patents between Commuting Zones

	Co-patents	
	(1)	(2)
log(Travel time)	-0.216**	
	(0.107)	
log(Travel time) × $\mathbb{1}\{\text{distance} < 100\text{km}\}$		-0.105
		(0.435)
log(Travel time) × $\mathbb{1}\{100\text{km} \leq \text{distance} < 400\text{km}\}$		-0.355***
		(0.123)
log(Travel time) × $\mathbb{1}\{400\text{km} \leq \text{distance}\}$		0.019
		(0.165)
Observations	382,132	382,132
Pseudo R^2	0.80	0.80
μ_{ij} fixed effects	✓	✓
μ_{it} fixed effects	✓	✓
μ_{jt} fixed effects	✓	✓

The table presents the result of estimating by PPML N. co-patents$_{ijt}$ = $\exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

4.3 Effect of travel time: heterogeneity

In this section we open up copatents, inventor and CZ pairs by different characteristics. We find that the reduction in travel time leads to more collaboration between more developed CZ pairs, as proxied by their initial innovativeness, rather than between those and less developed ones, or among less developed ones. We also find that the increase in copatents happens when either both collaborators, or at least one, is highly productive as measured by their previous citation weighted patents. Hence, while reductions in travel time increase copatents, it may also lead to increased collaboration-inequality within and across regions.

Additionally, we uncover that the effect is larger for pre-existing inventor collaborations and inventors with a more similar pre-existing knowledge. When distinguishing between collaborations within and across firms, the effect is negative though imprecisely estimated, with a point estimate about twice as large for across-firm collaborations. The effect also persists once we weight patents by the number of claims or technology classes, which are proxies for patent scope and patent breadth, implying that the increase in copatents is not driven by patents of lower scope or breadth. Finally, the effect is negative for all technology classes, with a larger point estimate for fields Electricity and Physics.

CZ with/without HSR station

We re-estimate equation 1 by whether both origin and destination CZs have an HSR station, only one of them and, none of them. Results are presented in Table 2. We find that the point estimate of elasticity of copatents to travel time is largest in absolute value when both CZ have an HSR station, followed by one CZ with HSR station. The coefficient for CZ-pairs without HSR station is negative but smaller in absolute value and estimate is less precise. This result may highlight the importance of proximity to an HSR station for the reductions in train travel time to have a larger effect. Nonetheless, the loss in precision may be consequence of having smaller variations in travel time: the average decrease in travel time for No-HSR CZ-pairs was 13%, while it was 18% for One-HSR and 27% for Both-HSR CZ-pairs.

Table 2: Sample selection: Commuting Zones with HSR station

	Co-patents			
	(1)	(2)	(3)	(4)
log(Travel time) × Distance < 100km	-0.105	0.167	-0.430	-1.15
	(0.435)	(0.518)	(0.844)	(4.26)
log(Travel time) × Distance 100-400km	-0.355***	-0.518**	-0.407**	-0.234
	(0.123)	(0.212)	(0.194)	(0.365)
log(Travel time) × Distance +400km	0.019	0.005	-0.269	0.033
	(0.165)	(0.372)	(0.223)	(0.310)
Sample selection	All	Both HSR	One HSR	No HSR
Observations	382,132	24,807	138,674	204,318
Pseudo R^2	0.80	0.92	0.85	0.82
μ_{ij} fixed effects	✓	✓	✓	✓
μ_{it} fixed effects	✓	✓	✓	✓
μ_{jt} fixed effects	✓	✓	✓	✓

The table presents the result of estimating by PPML $\text{N. co-patents}_{ijt} = \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

Initial innovativeness

We classify CZ according to the amount of patents filed in 1980-1985 by inventors located in that CZ. We define those in the highest quartile as High initial innovativeness and the rest Low initial innovativeness. Table 3 shows that the elasticity is only statistically significant for High-High pairs. This result shows that reductions in travel time reductions in travel time led to an increase in patent collaboration only in CZ-pairs in which both CZs were already more innovative. Hence, while reductions in travel time may lead to an increase in patent collaborations, it may at the same time lead to increased inequality across regions in terms of collaborative innovation.

Table 3: Initial level of innovativeness

	Co-patents (1)
log(Travel time) × High-High	-0.224**
	(0.114)
log(Travel time) × High-Low	-0.200
	(0.277)
log(Travel time) × Low-Low	2.05
	(1.38)
Observations	382,132
Pseudo R^2	0.80
μ_{ij} fixed effects	✓
μ_{it} fixed effects	✓
μ_{jt} fixed effects	✓

The table presents the result of estimating by PPML $\text{N. co-patents}_{ijt} = \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

Type of collaboration

Table 4 shows the results of estimating equation 1 where copatents are classified into whether the copatent belows to an inventor-pair that did not collaborate in the past (column 2) or it did (column 3), and whether the patent involves only one firm (column 4) or multiple firms (column 5). We find that the point estimate elasticity for new collaborations is comparable to the baseline estimate. In contrast, the estimate for old collaborations is more than three times as large. This result suggests that a reduction in travel time has effects both on an extensive and intensive margin of collaborations: it leads to new inventor-pair collaborations and it increases the collaborative output of inventor-pairs that already collaborated before.

The estimates for intra and inter-firm collaborations, while in the ballpark of the baseline estimate, are not precisely estimated. The significance value for intra-firm estimate to be statistically different from zero is 12.6% and for inter-firm it is 10.5%. Also, although the point estimate for inter-firm collaborations is around twice as large in absolute value as the intra-firm copatents, these coefficients are not significantly different from each other (significance level 52%). This result is related to Giroud (2013) who finds that non-stop flight connections between a subsidiary and its headquarters leads to increased investment in the subsidiary.

Table 4: Nature of collaborations

	Co-patents				
	(1)	(2)	(3)	(4)	(5)
log(Travel time)	-0.216**	-0.212**	-0.787***	-0.183	-0.331
	(0.107)	(0.104)	(0.248)	(0.119)	(0.204)
Sample Co-Patents	All	New collab.	Old collab.	Intra-firm	Inter-firm
Observations	382,132	371,733	81,069	322,592	92,725
Pseudo R^2	0.80	0.77	0.77	0.79	0.57
μ_{ij} fixed effects	✓	✓	✓	✓	✓
μ_{it} fixed effects	✓	✓	✓	✓	✓
μ_{jt} fixed effects	✓	✓	✓	✓	✓

The table presents the result of estimating by PPML $\text{N. co-patents}_{ijt} = \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

Patent characteristics

We investigate whether the decrease in travel time leads to more patents by leading to patents of different quality. We estimate a variation of equation 1 weigthing patents by the amount of citations received in a 5-year window, amount of claims in the patent, and amount of technology classes included in the patent.[11] Table 5 shows the results.

[11] We estimate: $\text{copatents} \times \text{weight}_{ijt} = \exp\left[\beta \log(\text{travel time}_{ijt}) + \mu_{ij} + \mu_{it} + \mu_{jt}\right] \times \varepsilon_{ijt}$ for weight = {N. citations, N. claims, N. Technologies}

When weighting patents by the amount of citations received, we find that the reduction of travel time did not have an effect in the amount of citation-weighted copatents. However, given that most patents have zero citations, the identification of fixed effects drops around two thirds of the CZ-pair-year observations of the baseline estimation in Table 1.[12] As the underlying effective sample is so different, one should be cautious when comparing coefficients across regressions.[13]

Column (2) of Table 5 presents the results with copatents weighted by amount of claims. The amount of claims in a patent is a proxy for its scope, in other words, the extent of coverage over which the patent has intellectual property rights (Lanjouw and Schankerman (2004)). Comparing the coefficient with the baseline estimation, we find a larger coefficient, suggesting that the reduction in travel time led to copatents with larger scope. Column (3) presents the results weighting each copatent by the number of distinct technology classes that it includes.[14] The number of technology classes in a patent is a way to measure the patent breadth, i.e.the number of technological domains that it covers (Lerner (1994)). The reduction in travel time led to a comparable increase in technology-weighted copatents as unweighted copatents, suggesting that the reduction in travel time did not affect the technological breadth of copatents.

Table 5: Weighted by patent characteristics

	Co-patents Citation weighted (1)	Co-patents Claims weighted (2)	Co-patents N. Tech. weighted (3)
log(Travel time)	0.066	-0.347**	-0.289***
	(0.219)	(0.141)	(0.111)
Observations	130,253	382,132	382,132
Pseudo R^2	0.77	0.86	0.82
μ_{ij} fixed effects	✓	✓	✓
μ_{it} fixed effects	✓	✓	✓
μ_{jt} fixed effects	✓	✓	✓

The table presents the result of estimating by PPML $\text{N. co-patents}_{ijt} = \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

[12] In the effective sample of the baseline regression shown in Table 1, the average observation at the CZ-pair-year has 0.57 copatents while it has 0.33 citation weighted copatents. In the effective sample, 85% of observations have zero copatents, while the share increases to 95% when considering citation-weighted copatents.

[13] Estimating equation 1 in the effective sample of CZ-pair-year of column (1) in Table 5 gives an elasticity of copatents to travel time -0.169 with clustered standard error 0.126. Using the same effective sample, the results for claims weighted and technology class weighted copatents are quantitatively similar to the ones in Table 5.

[14] We count the amount of different IPC35 technology classes, having a maximum of 35 technology classes.

Inventor characteristics

We investigate whether the reduction in travel time leads to more copatents among technologically similar or different inventors. We compute the technological similarity of inventors based on their previous patents that do not belong to the inventor-pair and then classify copatents by whether the inventor-pair is over or under the median similarity across inventors.[15] Note that this computation requires that each of the co-inventors has patented before at least once outside that collaboration. Hence, we drop a substantial amount of collaborations. Table 6 shows that the decrease in travel time led to increased co-patents among inventors over the median similarity, but the effect is not statistically different from zero for inventors under the median similarity. We obtain a similar result if we classify inventors over/under average similarity.

Table 6: Collaborations by inventors' technological similarity up to t-1

	Co-patents	
	(1)	(2)
log(Travel time)	-1.09***	-0.263
	(0.272)	(0.258)
Sample selection	Inventors' similarity above median	Inventors' similarity under median
Observations	68,327	80,752
Pseudo R^2	0.75	0.73
μ_{ij} fixed effects	✓	✓
μ_{it} fixed effects	✓	✓
μ_{jt} fixed effects	✓	✓

The table presents the result of estimating by PPML N. co-patents$_{ijt}$ $= \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses. Significance codes: ***=0.01, **=0.05, *=0.10

We compute the productivity of an inventor at year t based on the sum of 5-year forward citations of all patents filed up to $t-1$. We compute each CZ's average productivity and classify inventors under/over the average of her own CZ, and under/over the average of her collaborator's CZ. We find that the elasticity on copatents is largest in absolute value when both inventors in the inventor-pair are over the average productivity, either of theirs and their collaborator's region. The elasticity is not significantly different from zero when none of the inventors is over the region's average. Table 7 and

[15] We compute the cosine similarity of technologies using the IPC35 technologies of previous patents.

Appendix Section C Table 5 present the results. Our results are consistent with Catalini, Fons-Rosen and Gaulé (2020) who find that new low-cost airline routes in the United States disproportionately benefited higher-quality scientists, suggesting that reductions in travel frictions primarily increase collaboration among more productive researchers.

This result highlights that while reductions in travel time increased collaborations, it did so mainly between inventors that were already more productive. As consequence, reductions in travel time may increase inequality not only across regions as suggested by results of Table 3, but also within a region.

Table 7: Collaborations by inventors' productivity relative to the inventor's region's average

	Co-patents			
	(1)	(2)	(3)	(4)
log(Travel time)	-0.216**	-1.22***	-0.605***	-0.113
	(0.107)	(0.398)	(0.187)	(0.108)
Sample selection	All	Top avg. prod. both inv.	Top avg. prod. one inv.	Top avg. prod. none
Observations	382,132	36,823	82,434	332,520
Pseudo R^2	0.80	0.71	0.75	0.76
μ_{ij} fixed effects	✓	✓	✓	✓
μ_{it} fixed effects	✓	✓	✓	✓
μ_{jt} fixed effects	✓	✓	✓	✓

The table presents the result of estimating by PPML $\text{N. co-patents}_{ijt} = \exp\left[\beta \log(\text{Travel time}_{ijt}) + \rho_{ij} + \gamma_{it} + \delta_{it}\right] \times \eta_{ijt}$, for commuting zones i and j, and application year t. Sample includes application year from 1980 to 2010. The regression is estimated as directional (ij different from ji). Standard errors clustered at the non-directional commuting zone pair are presented in parentheses.

Technology categories

We open up copatents at the CZ-pair by technology categories. Table 6 in Appendix section C shows the results for the baseline regressions adding technology-time fixed effects and the elasticity of copatents to travel time by technology category. The table shows that the elasticity is negative and significant for all categories. In particular, the point estimate is largest in absolute value for "Electricity" and "Physics", and smallest for "Chemistry, Metallurgy" and "Fixed Constructions".

5 Conclusion

This paper exploits the roll out of High Speed Railways (HSR) as a quasi-natural experiment to provide new causal evidence of the impact of train travel time on patent collaboration between inventors. To do so, we constructed a new dataset of train travel time in France between 1980 and 2010, documenting a median decrease in travel time between CZ-pairs of 12%. We find that the decrease in travel time led to an increase in patent collaboration across CZs, driven both by pre-existing and new inventor-pairs, and by inventors that are more similar to each other in terms of knowledge. At the same time, we find that the reduction in travel time only leads to an increase in collaborations between CZ pairs that were already highly innovative, and between inventor pairs that were highly innovative within their CZ. As consequence, the decrease in travel time may have increased collaboration gap within and between CZs.

This paper provides evidence on the impact of transportation infrastructure on innovation. Recent policies and recommendations make a strong emphasis on fostering regional innovation. Our results suggest that transportation infrastructure can affect regional innovation by facilitating long-distance face-to-face interactions among inventors. However, the gains from improvements in connectivity may be unevenly distributed, favoring more developed regions.

References

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

Atkin, David, M Keith Chen, and Anton Popov. 2022. "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley." National Bureau of Economic Research.

Bahar, Dany, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W Koo. 2023. "Innovation on wings: Nonstop flights and firm innovation in the global context." *Management Science*, 69(10): 6202–6223.

Belenzon, Sharon, and Mark Schankerman. 2013. "Spreading the word: Geography, policy, and knowledge spillovers." *Review of Economics and Statistics*, 95(3): 884–903.

Carlino, Gerald, and William R Kerr. 2015. "Agglomeration and innovation." *Handbook of regional and urban economics*, 5: 349–404.

Catalini, Christian. 2018. "Microgeography and the Direction of Inventive Activity." *Management Science*, 64(9): 4348–4364.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaulé. 2020. "How do travel costs shape collaboration?" *Management Science*, 66(8): 3340–3360.

Coste, Vincent. 2021. "TGV at 40: The new version of the iconic high-speed train is launching in an era of global rivalry." Euronews. *Available at:* `https://www.euronews.com/next/2021/09/17/the-need-for-speed-the-new-tgv-m-is-being-launched-in-an-era-of-global-high-speed-rail-riv`, With AFP. Accessed 2025-09-02.

Dijkstra, Edsger W. 1959. "A note on two problems in connexion with graphs." *Numerische mathematik*, 1(1): 269–271.

Dong, Xiaofang, Siqi Zheng, and Matthew E Kahn. 2020. "The role of transportation speed in facilitating high skilled teamwork across cities." *Journal of Urban Economics*, 115: 103212.

Draghi, Mario. 2024. "The future of European competitiveness. Part A: A competitiveness strategy for Europe." Accessed: 2025-02-28.

Duranton, Gilles, Philippe Martin, Thierry Mayer, and Florian Mayneris. 2009. "The economics of clusters: evidence from France."

Giroud, Xavier. 2013. "Proximity and investment: Evidence from plant-level data." *The Quarterly Journal of Economics*, 128(2): 861–915.

Gross, Daniel P, and Bhaven N Sampat. 2023. "America, jump-started: World War II R&D and the takeoff of the US innovation system." *American Economic Review*, 113(12): 3323–3356.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics*, 108(3): 577–598.

Lanjouw, Jean O, and Mark Schankerman. 2004. "Patent quality and research productivity: Measuring innovation with multiple indicators." *The economic journal*, 114(495): 441–465.

Lerner, Joshua. 1994. "The importance of patent scope: an empirical analysis." *The RAND Journal of Economics*, 319–333.

Moretti, Enrico. 2012. *The new geography of jobs.* Houghton Mifflin Harcourt.

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli. 2017. "Disambiguation of patent inventors and assignees using high-resolution geolocation data." *Scientific data*, 4(1): 1–21.

Pauly, Stefan, and Fernando Stipanicic. 2022. "The creation and diffusion of knowledge: Evidence from the Jet Age."

Schmidheiny, Kurt, and Sebastian Siegloch. 2023. "On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization." *Journal of Applied Econometrics*, 38(5): 695–713.

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics*, 88(4): 641–658.

Tsiachtsiras, Georgios, Deyun Yin, Ernest Miguelez, and Rosina Moreno. 2022. "Trains of Thought: High-Speed Rail and Innovation in China." *Available at SSRN 4280769*.

U.S. Congress. 2022. "CHIPS and Science Act of 2022." Public Law No: 117-167, Accessed: 2025-02-28.