Transportation Infrastructure and Knowledge Diffusion

Adrian Kulesza
Yale University
April 2021

Abstract
This paper investigates the role of infrastructure in enabling the diffusion of knowledge by helping overcome geographical constraints. We consider the development of Germany’s high-speed rail network and its impact on the pattern of patent production and the aggregate economy. Using data on regional patent authorship and economic outcomes, we first establish that collaboration on patents is constrained by travel times and second show that high-speed rail connections increased regional employment and GDP. To understand the impact of the infrastructure network on the German economy we develop a model of inter-regional transfers of knowledge for the creation of patents and embed this setup in a spatial economy with labor mobility. Estimating the model with patent data, we find that the first wave expansion of Germany’s high-speed rail system increased welfare by 3.5 percent and the second wave expansion increased welfare by an additional 1.0 percent.

I am grateful for the guidance and discussion provided by Professor Costas Arkolakis who contributed greatly to my passion for spatial economics. All errors are my own.
1 Introduction

Geography is a well documented barrier to the creation and diffusion of knowledge. Head et al. (2019) summarizes some of the major findings regarding the relationship between geography and knowledge noting that patent citations are geographically concentrated, research spillovers are declining with distance, and spatial separation between firms and inventors reduces the likelihood of new technological adoption.1 In particular, a key component of knowledge production which is often subject to geographical constraints is the spatial interactions of inventors and researchers.2 In this context, commuting infrastructure can help alleviate some of the barriers posed by geography and enable the diffusion of knowledge. For example, Perlman (2015) find an increase in patent production as a response to local agglomeration and patenting cost reductions caused by rail network connections in the US, and Bernard et al. (2020) find increased collaboration as well as improved productivity and team quality for inventors who experienced a reduction in travel time as a result of new bridge connections in Japan. Ultimately, this topic is largely unexplored as measures of knowledge diffusion aren’t readily available and any consideration of changes in the underlying determinants is likely to face some endogeneity concerns.

This paper contributes to the literature of geography and knowledge diffusion by using data on patent authorship and economic outcomes to model the impact of changes in Germany’s high-speed rail (HSR) network on the pattern of patent production and the macro economy. We first use spatial data on the location of patent filing and the location of inventors as well as data on regional employment and production to characterize the aggregate relationship between travel times, infrastructure developments, and changes in

---

1See Jaffe et al. (1993), Keller (2002), and Comin et al. (2012) respectively.
2Akcigit et al. (2018) highlight how important interactions are by demonstrating that the majority of patents are produced collaboratively and that not only more interactions, but also interactions with better inventors are correlated with innovation quality.
patenting and economic variables. We establish that travel times between inventors and the location of patenting filing is an important component of the knowledge production network. We also demonstrate that connection to the HSR network meaningfully impacts the GDP and employment of regions. Motivated by these findings, we then construct a model where knowledge is sourced and transferred for the creation of patents. This setup is nested in a spatial economy with labor mobility where patents as well as labor are used in the production of freely traded final goods. Using the observed pattern of patent production alongside regional wages and labor, we are able to recover key parameters in the model including the determinants of fixed sourcing costs, location-specific productivities and amenities, and the quality of ideas in each region. This allows us to finally consider the implications of counterfactual HSR networks on welfare, patenting, and economic outcomes.

We begin by describing the policy setting for Germany’s HSR network expansion. The HSR system was built over two phases: the first targeting the majority of major cities and the second connecting mostly smaller and peripheral cities. We focus on the second wave expansion which precipitated an average 13% drop in travel times between locations which featured a direct connection in the network. In order to understand how this policy change affects the pattern of patent production and economic outcomes, we start by describing the data we have which characterizes these variables. Most notably, we use data which contains the location of patent filing and the locations of listed inventors for all patent applications to the European Patent Office to investigate how knowledge diffuses across space. There are two main facts which characterize the data: the majority of patents are the result of collaborative behavior (Figure 1) and inventors tend to be located near the location of patent application (Figure 2). These facts suggest that the number of inventors and their spatial distribution is an important aspect of knowledge production. They also invite the question of how changing the geographical barriers which inventors
face might affect where patents get produced and who works on those patents.

Figure 1: Number of Inventors Listed per Patent

![Histogram of Number of Inventors Listed per Patent](image1.png)

Figure 2: Distance of Inventors from Application Location

![Histogram of Distance of Inventors from Application Location](image2.png)

With the policy setting and data established, we turn to quantifying aggregate trends as well as changes in knowledge production and the economy. We start by using a reduced-form gravity equation to demonstrate how the flow of authors (equivalently referred to as inventors) between their location of residence and the location of patent filing is decreasing in travel times and increasing in demand for authors as measured by the number of patents produced. We next use a difference-in-difference approach to show how connection to the high-speed rail network increased regional GDP as well as altered the distribution of authors and labor. We measure connectivity using a straightforward geometric definition as well as a model-implied measure known in the spatial literature as commuter market access. Given that the high-speed rail network generally targets economically and politically important regions, we use an instrumental variable strategy based on a hypothesized least-cost corridor between locations to get at the causal effects of network connections.

We next turn to our model which begins with establishing a multi-region spatial econ-
omy populated by consumers who derive utility from constant elasticity of substitution (CES) preferences for a variety of goods as well as location-specific non-pecuniary amenities. Next, we describe the setup for patent production where inventors, the patent producing entities, combine a measure of tasks which differ in their quality and are subject to production and trade costs. These tasks can be sourced from other regions after inventors first pay a fixed search cost to observe their task quality. Each inventor simultaneously operates a firm which uses their flow of patent production as well as local labor to produce a single differentiated variety of the freely traded final goods in a monopolistically competitive market. The inventor’s problem can then be written as choosing the locations to source from so that profits net of the fixed search costs are maximized. A free entry condition for inventors as well as a labor supply equation based on wages and amenities will pin down the equilibrium wage and labor for each location.

We map our model to the data by letting each patent represent an inventor and by having each of the cited authors with their respective locations represent the imported tasks and sourced locations. From the patent-level import shares we can use a fixed-effect regression to recover the region-specific task quality and project trade costs onto geographic variables including HSR connections. Then, after some relatively smaller intermediate steps, we can use the simulated method of moments to recover the fixed costs of sourcing, which we also allow to depend on HSR connections. We do so by simulating a large sample of inventors in each region and having them solve the inventor’s problem given the recovered task qualities. Aggregating their behavior, we target moments in the patenting data including the fraction of patents which import inventors and the share of imported inventors coming from each region. The model is calibrated to the German economy prior to the second wave HSR expansion, so we simulate the construction of the new HSR connections and find consistency with the results from our empirical estimates. We additionally consider the effect of expanding the network to cover all locations and
the effect of removing the entire network, ultimately finding that the network is a meaningful component of aggregate production and welfare.

This paper is primarily related to the literature on knowledge diffusion and geography (e.g. Jaffe et al. (1993), Keller (2002), and Comin et al. (2012)). In particular, it complements the subset which uses empirical methods to investigate the effects of substantive infrastructure projects on intra-country trends in patent production (Perlman (2015), Dong et al. (2018), and Bernard et al. (2020)). We build on this literature by developing a model of inter-regional transfers of knowledge for the production of patents which can be structurally estimated and used to run counterfactuals. The collaboration component of our model also provides a novel application of the the intermediate input sourcing framework of Antràs et al. (2017) which is likewise applied by Bernard et al. (2019) to describe intra-country buyer-supplier relationships. The model also ties behavior in patent production to a macro economy with spatial elements, relating it to a large body of literature on spatial models (e.g. Allen and Arkolakis (2014), Ahlfeldt et al. (2015), and Tsivanidis (2019)). The analysis of high-speed rail in particular also contributes to the ongoing study of its general efficacy and adjacent benefits (Chen and Haynes (2017), Heuermann and Schmieder (2018), Ahlfeldt and Feddersen (2018), Dong et al. (2018), and Bernard et al. (2019)).

Our paper is organized as follows. In Section 2 we overview the history of Germany’s HSR development and describe the geographical area of study as well as the data used in our analysis. In Section 3 we quantify aggregate patterns by first considering a reduced-form gravity model for inventor flows and by second performing a difference-in-difference analysis on regions which received connections to the HSR network. In Section 4 we develop our model of inventor sourcing for patents which underpins production in an aggregate economy with labor mobility. In Section 5 we structurally estimate the model and consider the effects of counterfactual high-speed rail networks. Section 6 con-
2 Background and Data

The policy setting of our empirical and structural analysis is the second wave expansion of Germany’s HSR network completed between 1999 and 2010. We first provide context for the development of the national rail system up to this point and convey some of HSR’s contributions as a commuting technology. We then turn to describing the data used in our analysis and provide some baseline regional statistics to set up our empirical analysis.

2.1 Intercity Express

In 1991, the first regularly scheduled high-speed trains, known as Intercity Express (ICE) trains, began to run through Germany on the newly finished Hanover-Würzburg and Mannheim-Stuttgart lines. In the decades that followed, numerous additional lines were either constructed or upgraded to expand ICE operations to every major city in Germany. Heuermann and Schmieder (2018) categorizes the history of ICE development in two phases. An initial expansion, conducted between 1991 and 1998, saw connections established between major urban centers in Germany (Figure 3). And a second phase, conducted between 1999 and 2010, saw the expansion of the network to relatively smaller and peripheral regions as well as additional connections established along existing routes (Figure 4). With each of these innovations to the ICE network, the rolling stock of ICE trains underwent upgrades with trains capable of reaching speeds of 330 km/h in use by the late 90s. Meanwhile, local trains had maximum speeds of 160 km/h, but typically averaged much slower speeds.

High-speed rail is able to deliver reduced travel times not only by operating at high
speeds, but also by operating at those high speeds for long durations at time with infrequent stops. Therefore, despite the fact that a high-speed line might run through a region, that region need not have direct access to the ICE network. However, travelers in unconnected regions can still transfer to an ICE train after traveling to another connected station, thereby still experiencing a reduction in travel time. During the second wave, travel times fell broadly with trips featuring a direct ICE connection seeing a 13 percent average reduction in travel time while other trips underwent a more moderate 3.5 percent reduction (Heuermann and Schmieder (2018)). Travel in the ICE network is also incredibly popular with a reported ridership of nearly 80 million in 2010. For comparison, Amtrak, the state-sponsored US rail network had ridership of 30 million in the same year.

---

3See Deutsche Bahn (2020) for ICE ridership statistics.
4See Bureau of Transportation Statistics (2014) for Amtrak ridership statistics.
2.2 Data

The spatial data used in this paper is aggregated to the NUTS 3 level. The Nomenclature of Territorial Units for Statistics or NUTS is a geographic standard for referencing regions within countries maintained by the European Union. Under the NUTS 3 classification, Germany is split into 429 regions known as districts. These 429 regions can further be aggregated to the NUTS 1 level, or equivalently to the states of Germany. German states have significant legislative power so keeping track of the aggregation will be relevant for later estimation. Over the course of the second wave ICE expansion (1999-2010), several changes were made to the NUTS 3 structure of Germany including the complete restructuring and merging of a state. In most of these cases old districts are matched and aggregated to new districts by referencing maps from before and after, but in some cases the regions either lack the data or correspondence to be appropriately matched so they are instead dropped from the following exercises.

The geocoded locations of the German railway lines are derived from OpenStreetMap and imposed on shapefiles of Germany’s NUTS 3 regions available from Eurostat. OpenStreetMap doesn’t discriminate between high-speed, freight, and other types of rail lines so high-speed lines are manually selected from the rail network by referencing Wikipedia maps of the various lines. The location of all major cities which ICE lines connect were provided by the Google Maps API.

Demographic data and aggregate statistics for all NUTS 3 regions are taken from Eurostat. This includes data from 2000 through 2012 on total and sectoral employment as well as GDP. Data on commuters between LAU regions comes from the German Federal Employment Agency (BA), but data is only available in a suitable format from 2002 onwards. Correspondence between the LAU and NUTS regional system is taken from Eurostat. Specific data on patent applications to the European Patent Office comes from
the OECD REGPAT Database which crucially contains the NUTS 3 location for the patent applications and their listed inventors. From this data, we consider only authors and patents located in Germany, focusing our attention on intra-country knowledge diffusion. Indicators for patent quality come from the OECD Patent Quality Indicators Database.

As an initial check, Table 1 compares some characteristics of the regions which received an ICE connection in the first wave, regions which received an ICE connection in the second wave, regions which lie along the path of an ICE line constructed in the second wave, and the remaining regions. Specifically, we compare the number of patent applications filed in a region, the number of inventors living in a region, the total employment, the fraction of employment in financial, insurance, professional, scientific, technical, administrative, and support service activities (FRPS), and GDP. The first wave targeted the majority of major cities so we see large differences between the first column and the other three. The second wave finished establishing direct connections between the remaining cities of meaningful size. Otherwise the intersected and other remaining regions are not significantly different from one another or even from Wave 2 regions, especially if we consider the patenting related variables.

<table>
<thead>
<tr>
<th></th>
<th>ICE Wave 1</th>
<th>ICE Wave 2</th>
<th>ICE Wave 2 Intersection</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent Applications</td>
<td>217.83</td>
<td>34.82</td>
<td>29.78</td>
<td>31.18</td>
</tr>
<tr>
<td>Inventor References</td>
<td>318.85</td>
<td>127.02</td>
<td>111.13</td>
<td>110.56</td>
</tr>
<tr>
<td>Employment (thousand)</td>
<td>219.35</td>
<td>86.24</td>
<td>54.38</td>
<td>67.81</td>
</tr>
<tr>
<td>FRPS Employment Share</td>
<td>0.16</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>GDP (million Euro)</td>
<td>15724.88</td>
<td>4762.73</td>
<td>3264.15</td>
<td>3866.55</td>
</tr>
</tbody>
</table>

Finally, as HSR is a commuting technology, it is important to quantify the time-saving
benefits of a high-speed rail connection. Unfortunately, travel times between NUTS 3 regions are not documented anywhere. So, we turn to the estimates of Heuermann and Schmieder (2018) who use timetable data for over a million journeys taken on the ICE network between 1999 and 2010 to project travel times onto commuter and distance variables. Essentially, for each trip beginning in location $i$ and ending in location $j$, the authors estimate:

\[
\log \text{Duration}_{ij} = \beta_0 \text{Direct Ice}_{ij} + \beta_1 \log \text{Residents}_i + \beta_2 \log \text{Workers}_j + \beta_3 \log \text{Distance}_{ij} + \epsilon_{ij}.
\]

We then use their estimates of $\beta_0, \beta_1, \beta_2,$ and $\beta_3$ along with the commuter and employment data described above and distances calculated between the centroids of all NUTS 3 regions to construct a matrix of direct travel times. Since travelers can make intermediate stops in the rail network by, for example, taking a local train to an ICE connected region and then using that ICE line to take a longer trip, the direct times may not reflect the actual times faced by travelers. Therefore, we use Dijkstra’s algorithm to find the least cost path through the travel time matrix between any two locations. When we consider changes in the ICE network later, we will recalculate this matrix as an ICE connection will not only enable lower direct travel times between the newly endowed regions and other ICE connected regions, but also it will allow other regions to make intermediate stops and likewise reduce their travel times as well.

3 Aggregate Estimation

In this section we attempt to quantify some determinants of knowledge dispersion by using a gravity regression on the flows of inventors from their location of residence to the location of patent filing. Then, we introduce some definitions of connectivity as well as
an instrument for the placement of ICE routes to estimate the aggregate effects of high-speed rail connections on the German economy. This section’s findings will motivate the construction of our model in Section 4.

3.1 Gravity Regression

When considering flows of either goods or people between any two locations, gravity equations have long provided a standard empirical framework for evaluating the determinants of these flows. Theoretical derivations for these equations in a commuting context can be formally done as in Persyn and Torfs (2015), but they largely follow a similar structure. For our purpose, we define the following gravity equation for the flow of inventors $X_{ij}$ between their location of residence $i$ and the location where the patent they are referenced as authors on is filed $j$:

$$\log X_{ij} = \beta_0 + \beta_1 \log \text{Time}_{ij} + \beta_2 \log \text{Applications}_j + \beta_3 \log \text{Inventors}_i + \alpha_i + \alpha_j + \epsilon_{ij}. \quad (2)$$

Note that when we use the term “flow” to describe inventors we do not literally mean the movement of an inventor between locations, but rather we mean the contribution of knowledge from an inventor living in one location to a patent produced in another location. Here $\text{Time}_{ij}$ is the travel time from location $i$ to $j$, $\text{Inventors}_i$ is the number of inventors living in location $i$, $\text{Applications}_j$ is the number of patent applications filed in location $j$, $\alpha_i$ and $\alpha_j$ are origin and destination fixed effects, and $\epsilon_{ij}$ is a location-pair specific error term. Intuitively, we can think of $\text{Applications}_j$ as the demand for the knowledge of inventors and $\text{Inventors}_i$ as the supply. In practice, using log transformations of all covariates in our gravity equation poses issues as a large share of flows are zero and as Santos Silva and Tenreyro (2006) point out, in the presence of heteroskedasticity these models lead to biased estimates when estimated by OLS. Therefore, as the authors sug-
gest, we use the Poisson pseudo-maximum-likelihood (PPML) estimator to estimate our gravity equation (2).

Table 2 features the results of the PPML estimation. Most importantly, travel times are significant in describing the flows of inventors between regions. If we reduce the travel time from location $i$ to location $j$ by 1%, the number of inventors in $i$ being listed as authors on patents filed in $j$ rises by approximately 1.38%. Moreover, on the demand side, the larger a destination is in terms of patent filing, the more inventors it will attract as authors. However, the size of an origin in terms of the supply of referenced inventors does not predict flows to other destinations. With the gravity relationship established between travel times and inventor flows, we now turn to considering the possible impact of ICE connections on aggregate outcomes.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>log(Inventors$_{ij}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Time$_{ij}$)</td>
<td>$-1.380^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
</tr>
<tr>
<td>log(Applications$_j$)</td>
<td>$1.042^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
</tr>
<tr>
<td>log(Inventors$_i$)</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
</tr>
</tbody>
</table>

Origin FE  Yes
Destination FE Yes
Observations 156025

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust Standard Errors in parentheses.

3.2 Measuring Aggregate Outcomes

This section uses some established regression frameworks to evaluate the aggregate effects of the expansion of the ICE network on economic and patenting outcomes. The first
specification and estimation procedure closely follows Faber (2014) where we define the following estimating equation:

\[
\log y_{i,s}^{2013} - \log y_{i,s}^{2002} = \alpha_s + \beta ICE_i + \gamma X_{i,s} + \epsilon_{i,s}
\]  

(3)

where \( i \) indexes the NUTS 3 district in state \( s \). \( y_{i,s} \) is any district-level relevant outcome, \( \alpha_s \) is a state fixed effect, and \( X_{i,s} \) is a series of controls. Most importantly, \( ICE_i \) is an indicator variable which takes a value of 1 if district \( i \) gained connection to the ICE network during the second wave expansion between 1999 and 2010. Both in this estimation and those going forward, we choose to exclude the major cities which are the main nodes of the HSR network. The relevant comparison here is between peripheral regions as the major cities which received direct ICE connections differ fundamentally along various economic and patenting related dimensions (Table 1). Moreover, when we consider the potential causal effects of ICE connections, we can not establish an argument that network connections are exogenous for major cities as their connection is precisely the goal of the infrastructure project.

We begin with a loose definition of connectivity. To have direct access to the high-speed rail network a region requires a station which services ICE trains, and not every region that the high-speed rail line crosses over has such a station. As we argued before, this does not mean that lacking an ICE servicing station precludes a region from benefiting from the high-speed rail network. Travelers from peripheral regions can make intermediate stops at ICE servicing stations and then finish their trip utilizing the benefits of the high-speed rail lines. And, in many cases, the trains which service the periphery run in parallel with the ICE lines. Therefore, we first define connectivity for a region as having an intersection with an ICE line completed between 1999 and 2010 (Figure 5).

Table 3 features the results from estimating (3) with OLS. Generally, ICE connectivity
is not a significant predictor of the change in number of patent applications or inventors referenced in a region. The only significant predictor for either of these variables is the log distance to the nearest major city for the number of inventors cited. The sign on the coefficient suggests that regions further away from major cities saw more growth in the number of inventors they housed. On the other hand, access to the ICE network was correlated with positive economics outcomes, raising employment by 2.8% and GDP by 3.4%. These results complement the findings of Heuermann and Schmieder (2018) who suggest that workers prefer to live in larger cities and use reductions in travel time to find employment in nearby regions.

Next, we consider a continuous measure of network connectivity. For this, we draw on the notion of market access in the spatial economics literature. In particular we follow the work of Tsivanidis (2019) who define the following two terms known as Residential
Commuter Market Access (RCMA) and Firm Commuter Market Access (FCMA):

\[ \Phi_{Ri} = \sum_j \tau_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}} \]

\[ \Phi_{Fj} = \sum_i \tau_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}} \]

where \( \tau_{ij} \) denotes travel times, \( L_F \) employment, and \( L_R \) residents. To give some intuition as to how to think about these variables, the RCMA is high for a location \( i \) which is close to locations \( j \) (\( \tau_{ij}^{-\theta} \) is low) that are large employers (\( L_{Fj} \) is high) and have trouble employing from other locations (\( \Phi_{Fj} \) is low). Likewise FCMA is high for a location \( j \) which is close to locations \( i \) (\( \tau_{ij}^{-\theta} \) is low) where lots of employees reside (\( L_{Ri} \) is high) and those employees have trouble traveling to other locations (\( \Phi_{Ri} \) is low). The parameter \( \theta \) captures the sensitivity of workers to travel times. In an intra-city context, Tsivandis estimates \( \theta = 2.724 \) for high-skilled workers, and being that in Japan at least HSR is mostly used by technical workers and business travelers we use this estimate assuming a similar pattern for Germany.\(^5\) Using our procedure for calculating travel times described in Section 2.2, we can find the change in travel times induced by the ICE expansion and then calculate the change in \( \Phi_{Fi} \) for each location. Figure 7 illustrates the log change in FCMA for each region; note the qualitative correlation with the second wave ICE lines in blue.

From here on out, when we refer to Market Access, we will be referring to FCMA since we are centrally concerned with the changes in patent production which require access to inventors from other regions. Now we can consider the same regression as before (3), but instead use the change in Market Access that results from the second wave expansion of the ICE network as the regressor:

\[ \log y_{i,s}^{2013} - \log y_{i,s}^{2002} = \alpha_s + \beta \Delta \log \Phi_{Fi,s} + \gamma X_{i,s} + \epsilon_{i,s}. \]  \(^{(4)}\)

\(^5\)See Hayakawa et al. (2021b) for Japanese HSR ridership demographic statistics.
Table 3 features the results from estimating (4) with OLS. As with the previous regression, Market Access is not significant in predicting either patenting related variable. If we consider a weaker level of significance, it is however correlated with economic outcomes as a 1% increase in market access correlates with a rise in employment and GDP by approximately 0.065% and 0.089% respectively.\footnote{These coefficients come short of the 5% significance level, but it is worth still discussing the results as they will become significant when we consider IV estimation later.} To contextualize these results, consider that the largest increase in Market Access is 92% with the mean increase for regions which intersect a new ICE line being 22%. Roughly speaking, this would correspond to a
maximum GDP increase of 8% and a mean increase of 2%.

Table 3: OLS Results

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>∆ log(Applications)</th>
<th>∆ log(Inventors)</th>
<th>∆ log(Employment)</th>
<th>∆ log(GDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICE</td>
<td>−0.089 (0.126)</td>
<td>−0.139* (0.071)</td>
<td>0.028** (0.014)</td>
<td>0.034* (0.018)</td>
</tr>
<tr>
<td>∆ log(MA)</td>
<td>0.106 (0.386)</td>
<td>−0.118 (0.202)</td>
<td>0.065* (0.037)</td>
<td>0.089* (0.045)</td>
</tr>
<tr>
<td>log(Distance City)</td>
<td>−0.003 (0.073)</td>
<td>0.106** (0.067)</td>
<td>0.120*** (0.045)</td>
<td>−0.003 (0.042)</td>
</tr>
<tr>
<td>log(Employment) 2002</td>
<td>0.082 (0.048)</td>
<td>0.020 (0.007)</td>
<td>0.036*** (0.007)</td>
<td>−0.002 (0.009)</td>
</tr>
<tr>
<td>FRPS Share 2002</td>
<td>−1.989 (0.919)</td>
<td>−0.587 (0.630)</td>
<td>−0.438 (0.556)</td>
<td>−0.077 (0.101)</td>
</tr>
</tbody>
</table>

State FE: Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes
Observations: 369, 369, 369, 369, 369, 369, 369

Notes: "* p < 0.1; ** p < 0.05; *** p < 0.01. Cluster Robust Standard Errors at NUTS 1 (state) level. We control for the distance to the nearest targeted city. This hopes to address the concern that regions closer to major cities benefit from economics or knowledge spillovers. We control for the population size of regions, hoping to capture the agglomeration benefits of a larger population. We add a control for the share of employment in FRPS which is the labor category that includes scientific and research employment. Since German regions are organized into 16 states which each have legislative power, we add state fixed effects.

In order to make a causal interpretation of the regressions presented, we would have to be comfortable with the assumption that ICE connections were exogenously determined. Intuitively, this is a questionable assumption to make as policy makers could have targeted the path of an ICE line to intersect districts based on unobservable characteristics that might easily be correlated with the outcomes we are considering. Therefore, we follow Faber (2014) and use the least-cost path implied by the Euclidean straight-line connection between the targeted major cities as an instrument for the actual placement of the HSR lines. If the objective of policy makers was to construct the least amount of rail lines that connect the nodes of the network, then this would be the resulting path (Figure 6).

Table 4 presents the first stage results of regressing the actual ICE line intersection indicator on the Euclidean line intersection indicator. In part just by mechanical design of this instrument and the scale of distance we are considering, the instrument is a very strong predictor of actual line placement. We will also consider instrumenting for the change in Market Access with the Euclidean line intersection. One issue to note with
using the binary variable of Euclidean intersection to instrument for Market Access is that the instrumented regressor, which is typically continuous, will only be able to take two values.

### Table 4: First Stage Results

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>ICE</th>
<th>Δ log(MA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean Path</td>
<td>0.776***</td>
<td>0.112***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>log(Distance City)</td>
<td>−0.053</td>
<td>−0.011</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>log(Employment) 2002</td>
<td>0.063</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>FRPS Share 2002</td>
<td>−0.857</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>(0.440)</td>
<td>(0.263)</td>
</tr>
<tr>
<td>State FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>369</td>
<td>369</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.661</td>
<td>0.212</td>
</tr>
</tbody>
</table>

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster Robust Standard Errors at NUTS 1 (state) level. Major Cities excluded.

The results from estimating (3) and (4) using two-stage least squares with the Euclidean intersection as the instrumental variable are presented in Table 5. First considering the economic outcomes, both regressors are significant predictors of employment and GDP with ICE connectivity suggesting a 5.2% increase in employment and a 5.9% increase in GDP. For Market Access, the coefficients are almost an order of magnitude larger. However, in terms of predicted regional outcomes, since the IV is binary the predicted changes in Market Access takes two values of 2.1% for regions not intersected by the instrument and 13.3% for those which are. Therefore, the coefficients suggest a 4.7% and 5.5% rise in employment and GDP respectively for treated regions, which isn’t inconsistent with our other estimates.\(^7\)

Turning to patenting related outcomes, we now observe a significant and negative co-

\(^7\) Ahlfeldt and Feddersen (2018) evaluate the impact of the ICE connection between Cologne and Frankfurt (which was an entirely new line constructed between 1995 and 2002) on three intermediate regions and find that after a three year period and six year period GDP rose by 6.6% and 8.4% respectively relative to a synthetic control group of three other regions. These values are broadly consistent with those produced by our analysis.
efficient for both measures of connectivity on the change in inventors. Given the gravity regression from earlier, we would expect that an increase in connectivity and the corresponding reduction in travel times would enable a location to better outsource inventors. However, in a similar vein to our earlier reference to Heuermann and Schmieder (2018), the ICE network may have incentivized a reorganization in the distribution of labor with inventors moving out of peripheral regions and into major cities.\(^8\) Likewise, to the extent that patent production requires inventors and that the ICE network enables access to a broader inventor pool, the non-significant coefficients on connectivity measures for the patent application regressions doesn’t square well. While excluded from the table here, we also consider the impact of these connectivity measures on a variety of other patent-related outcomes including the share of inventors from other regions, the mean distance of inventors, the average quality of patents, and others. Even before correcting for the number of hypotheses we are testing, these all give a non-significant coefficient to our connectivity measures.

Table 5: IV Results

<table>
<thead>
<tr>
<th></th>
<th>(\Delta \log(\text{Applications}))</th>
<th>(\Delta \log(\text{Inventors}))</th>
<th>(\Delta \log(\text{Employment}))</th>
<th>(\Delta \log(\text{GDP}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICE Predicted</td>
<td>(-0.087) (0.149)</td>
<td>(-0.182^{**}) (0.080)</td>
<td>(0.052^{*}) (0.023)</td>
<td>(0.059^{**}) (0.029)</td>
</tr>
<tr>
<td>(\Delta \log(\text{MA})) Predicted</td>
<td>(-0.605) (1.039)</td>
<td>(-1.269^{**}) (0.556)</td>
<td>(0.360^{**}) (0.158)</td>
<td>(0.414^{**}) (0.201)</td>
</tr>
<tr>
<td>(\log(\text{Distance City}))</td>
<td>(0.002) (0.068)</td>
<td>(0.110^{**}) (0.043)</td>
<td>(-0.003) (0.006)</td>
<td>(0.017^{**}) (0.008)</td>
</tr>
<tr>
<td>(\log(\text{Employment})) 2002</td>
<td>(0.076) (0.068)</td>
<td>(0.016) (0.043)</td>
<td>(0.038^{**}) (0.006)</td>
<td>(0.043^{**}) (0.008)</td>
</tr>
<tr>
<td>FRPS Share 2002</td>
<td>(-1.913) (0.928)</td>
<td>(-0.472) (0.627)</td>
<td>(-0.099) (0.103)</td>
<td>(-0.469^{**}) (0.138)</td>
</tr>
<tr>
<td>State FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>369</td>
<td>369</td>
<td>369</td>
<td>369</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.032</td>
<td>0.032</td>
<td>0.095</td>
<td>0.255</td>
</tr>
</tbody>
</table>

Notes: \(^*p < 0.1; \^{**}p < 0.05; \^{***}p < 0.01.\) Cluster Robust Standard Errors at NUTS 1 (state) level. Major Cities excluded. Because we use the same IV for both regressors with no other change in model specification, the non-instrumented regressors have the exact same coefficients and standard errors across the two cases.

\(^8\)Note that the employment variable keeps track of where workers are employed, but the inventor variable keeps track of where inventors live. Therefore, in the face of these results, it is consistent to hold the position that the ICE network creates outflows of inhabitants and inflows of workers.
Given how significant travel time is in determining the flow of inventors between locations (Table 2), it’s rather surprising to see that the ICE connections had no significant effect on the productive capacity of a region in creating patent application. And, as we mentioned in the previous paragraph, it is similarly surprising to see no measurable effect on region-to-region patent collaboration or quality. However, it may be that in the aggregate, other variables, which we do not control for, are much more significant determinants of patent production and the margin for which spatial connectivity can influence the outcome is small. For example, R&D funding, changes in human capital accumulation, and trends in industries which are heavy in innovation are likely to play a larger role in patent production than the changes in travel times induced by the ICE network. This will especially be the case for regions where the number of patents produced is small, which is the case for most regions which received ICE connections in the second wave (Table 1). That is, if a region is only annually producing roughly 30 patents (as is the average for Wave 2 region) and the margin for which travel times can increase production is small, then it would be statistically difficult to observe a result distinguishable from zero. Regions connected in Wave 1 have a much higher patent production capacity, but the data required to do a similar estimation procedure is not available for the years required.

Bernard et al. (2020) also study the effects of infrastructure developments on patenting-related outcomes and are able to demonstrate significant, positive results. They use a generally comparable estimation procedure, however their case study is for connecting an island in Japan, which was previously only accessible by ferry or plane, with a population of 4 million people to main-land Japan. Their case-study represents a much more drastic change in spatial connectivity and concerns a region of relatively larger magni-
tude. With this in mind, it may be that the development and regions we are considering don’t lend themselves to a statistically-significant observable result if one exists. Despite these findings, we are sure that travel time is a significant component of the flow of inventors (Table 2) and that ICE connections not only reduce travel times (Equation 1), but also positively affect GDP and employment (Table 5). With these results in hand, we now turn to our model which intends to describe the process for collaboration on patents and use this to underpin production in an economy with labor mobility.

4 Model

We now develop a model of patent production which crucially requires that inventors utilize a continuum of tasks which can be sourced from other regions. The sourcing framework for our model is adapted from Antràs et al. (2017) who present a model for input sourcing in an international trade context. Given a mechanism for how inventors produce patents, we introduce a final goods producing sector which takes as input the flow of patents from inventors and labor. Finally, we nest these components in a spatial model allowing for labor mobility as in Allen and Arkolakis (2014). Despite the fact that our model relies on Antràs et al. (2017) to provide the core component of inter-region sourcing, we make a variety of alternate assumptions that allow us to use the model for a novel purpose. Most importantly, we allow wages and labor to be endogenously determined in equilibrium which allows for rich linkages in a spatial model.

Bernard et al. (2019) presents a simplified version of this model to describe input sourcing for firms in a closed domestic economy.
4.1 Consumer Preference

Consider an economy with $N$ regions, each populated by consumers with identical constant elasticity of substitution (CES) preferences for a variety of goods:

$$C_i = \left( \int_{\omega \in S} c_i(\omega) \frac{\sigma - 1}{\sigma} d\omega \right)^{\frac{1}{\sigma - 1}}$$

where $S$ denotes the set of varieties available for consumption. We assume the welfare of a consumer is then given by the product of consumption and a non-pecuniary amenity which diminishes with the population of a region:

$$W_i = C_i \bar{u}_i L_i^{-\beta}.$$ (5)

Aggregate consumer income in a location is given by $L_i w_i$ where $L_i$ denotes the inhabitants in a region $i$ and $w_i$ their wage. Optimization leads to the following expression for the total demand of each variety in location $i$:

$$c_i(\omega) = \frac{p_i(w)^{-\sigma}}{P_i^{1-\sigma} - w_i L_i}$$

where $P_i$ is the CES price index of varieties and $p_i(\omega)$ is specifically the price of variety $\omega$ in region $i$.

We make several simplifying assumptions going forward: each location consumes the same set of varieties, labor can freely move across regions, and there is free trade of the full set of varieties. These assumptions make the model tractable as they ultimately provide us with price and welfare equalization which we denote by $P$ and $\bar{W}$ respectively. Additionally, they are appropriate for our policy setting since we are considering a single country where, relative to an inter-country context, consumers are more homogenous,
there is ease of mobility between locations for employment, and the trade costs of goods are negligible. With these assumptions, we can finally write the aggregate demand in the economy for a variety \( \omega \) as

\[
X(\omega) = \frac{p(w)^{1-\sigma}}{p^{1-\sigma}} \sum_{i=1}^{N} w_i L_i = \frac{p(w)^{1-\sigma}}{p^{1-\sigma}} \bar{E}
\]

where \( \sum_{i=1}^{N} w_i L_i \) is the total spending in the economy which we denote as \( \bar{E} \).

### 4.2 Patent Creation

In each region \( i \) there are inventors who produce patents by combining a measure of tasks which differ in their quality. In particular, each inventor requires a specific constant elasticity of substitution unit continuum of tasks. Inventors themselves vary in terms of efficiency \( z \), which dictates how effectively they are able to turn those tasks into patents. Along these lines, we use \( z \) to index inventors differentiated by their productivity.

A key feature of tasks is that they can be sourced from other regions. Regions are differentiated by their quality of tasks where \( q_j(z, \iota) \) denotes the quality of inventor \( z \)'s task \( \iota \) sourced from region \( j \). These tasks are produced by perfectly competitive suppliers using labor, which is paid a region-specific wage \( w_j \). In order for an inventor in \( i \) to observe the quality of tasks in a region \( j \), they must pay a fixed cost \( f_{ij} \) in terms of labor. \( J_i(z) \) is then the set of regions for which inventor \( z \) in region \( i \) has paid the fixed cost to search. Finally, when an inventor chooses to source a task, some of the task’s quality is lost as an iceberg trade cost \( \tau_{ij} \geq 1 \). Since tasks will ultimately be represented by authors in the patent data, we can think of trade costs as the loss in knowledge that occurs when someone communicates an idea to another person with only a fraction of the author’s

\[11\]We use the model to evaluate the effect of changes in the ICE network which is notably a commuting-based infrastructure technology. In Germany, the HSR network is not used to transport goods so we do not require a role for trade costs in the model as it pertains to final goods.
original idea being preserved.

Now consider how quality enters the CES production function for patents:

\[ y_i(z) = z \left[ \int_0^1 (\bar{q}_i(z, \iota) x(\iota | q))^{\frac{\rho - 1}{\rho}} d\iota \right]^{\frac{\rho}{\rho - 1}} \]  \hspace{1cm} (7)

where \( y_i(z) \) can be thought of as the flow of patent production and \( \rho \) is the elasticity of substitution between tasks. Here \( \bar{q}_i(z, \iota) \) denotes the quality of task \( \iota \) that inventor \( z \) in region \( i \) ends up using in production after solving the following maximization problem:

\[ \gamma_i(z, \iota; J_i(z)) = \max_{j \in J_i(z)} \left\{ \frac{q_j(z, \iota)}{\tau_{ij} w_i} \right\} \]

where they choose the task of highest quality net of its production and importing cost. The quantity being maximized here corresponds to a minimization of the unit cost function of the inventor:

\[ c_i(z) = \frac{1}{z} \left[ \int_0^1 \gamma_i(z, \iota; J_i(z))^{1 - \rho} d\iota \right]^{\frac{1}{1 - \rho}}. \]

The quality of task \( \iota \) in location \( j \) is a realization of a random variable \( Q \) from a Frechet distribution with scale parameter \( A_j \): \( F(Q) = e^{-A_j(Q)^{-\theta}} \). This formulation follows from Eaton and Kortum (2002), which itself develops on earlier work (Eaton and Kortum (2001)) where the Frechet distribution is used to describe the distribution of good qualities and \( A_j \geq 0 \) represents the stock of knowledge accumulated over time. Accordingly, \( A_j \) informs the average quality of tasks in a region \( j \), and \( \theta \) the inverse of variability in quality.

Continuing to follow Eaton and Kortum (2002), given the set of regions which inventor \( z \) in region \( i \) has surveyed, \( J_i(z) \), their purchasing shares from each region \( j \in J_i(z) \) takes
the form
\[ \lambda_{ij}(z) = \frac{A_j(\tau_{ij}w_j)^{-\theta}}{\Phi_i(z)}. \] (8)

The term \( \Phi_i(z) \) captures the access to quality tasks:
\[ \Phi_i(z) = \sum_{j' \in J_i(z)} A_{j'}(\tau_{ij'}w_{i})^{-\theta}. \]

It also allows us to write the price index an inventor faces for the tasks it uses in production as
\[ c_i(z) = \frac{1}{z} \left( \lambda \Phi_i(z) \right)^{-1/\theta}, \]
where \( \lambda^{1-\rho} = \Gamma \left( \frac{1-\rho}{\theta} + 1 \right)^{\theta} \) and \( \Gamma \) is the gamma function.

When we take the model to the data, we map each patent to an inventor. Patents are filed in a location \( i \) and contain authors who come from various regions \( j \); these authors represent the tasks which are sourced. For each patent, the fraction of authors from each location is what gives \( \lambda_{ij}(z) \), the import shares, and the set of locations is what gives \( J_i(z) \), the sourcing strategy. In Section 5.1 we show how to use a regression with fixed effects to recover estimates for \( A_j(\tau_{ij}w_j)^{-\theta} \), the sourcing potential of a region \( j \) from the perspective of region \( i \).

### 4.3 Firms

Each inventor simultaneously operates a firm which produces a single differentiated variety \( \omega \) in a monopolistically competitive market using labor as well as their patents with Cobb Douglas technology:
\[ x_i(z) = T_i L_i(z)^{1-\alpha} y_i(z)^{\alpha}. \]
where \( T_i \) is a location specific technology term, \( L_i(z) \) is the labor employed in the firm operated by inventor \( z \) in location \( i \), and \( y_i(z) \) is the patent flow of said inventor. Inventors own all equity in their firms, and thus receive all profits. Facing a demand for variety \( \omega \), (6), the firm earns a profit of

\[
\pi_i(z) = c_i(z)^{\alpha (1-\sigma)} B_i
\]

where \( c_i(z) \) is the price index for the inventor’s patent production and \( B_i \) is given by

\[
B_i = \left( \bar{EP}^{\sigma-1} \right)^{1-\sigma} \left( \frac{\sigma}{\sigma-1} \right)^{1-\sigma} \left( \left( \frac{1}{\alpha} \right)^{\alpha} \left( \frac{w_i}{1-\alpha} \right)^{1-\alpha} \frac{1}{T_i} \right)^{1-\sigma}.
\]

Taking wages as given, the only margin the inventor has for affecting their firm’s profits is minimizing the cost of their patent production. In this regard, the only choice made by the inventor is which locations they will observe quality from, i.e. \( J_i(z) \). Therefore, we can write the problem of the inventor as

\[
\max_{I_{ij} \in \{0,1\}_{j=1}^N} \pi(z, I_{i1}, ..., I_{iN}) = \left( \frac{1}{\lambda} (\bar{\Phi}_i(z))^{-1/\theta} \right)^{a(1-\sigma)} B_i - w_i \sum_{j=1}^N I_{ij} f_{ij}.
\] (9)

That is, the inventor wants to maximize their access to quality tasks net of the cost of gathering information on the quality distribution in other regions. This expression is functionally equivalent to the optimal sourcing strategy firms face in Antràs et al. (2017), allowing us to use similar procedures to implement a structural estimation. What remains is describing the equilibrium of the economy.

We assume that inventors must pay a location-specific fixed cost in terms of labor, \( f_{ei} \).
in order to determine their productivity. Therefore, we can write a free entry-condition:

\[
\int_{z_i}^{\infty} \left[ \left( \frac{1}{z} \left( \lambda \Phi_i(z) \right)^{-1/\theta} \right)^{a(1-\sigma)} B_i - w_i \sum_{j=1}^{N} f_{ij} \right] dG_i(z) = w_i f_{ei} \tag{10}
\]

where \( z_i \) is the lower-bound for productivity of operational inventors in region \( i \) and \( G_i(z) \) is the cumulative distribution of inventor productivities.\(^{12}\) Inventors who draw productivities less that \( z_i \) will not operate. We assume that the mass of inventors in each location will remain fixed and wages will adjust as firms alter labor demand to make sure \( (10) \) is satisfied.

### 4.4 Aggregation

Finally, solving \( (5) \) for labor, using the fact that welfare equalizes under our assumption of labor mobility, and normalizing by the aggregate labor supply gives the equilibrium distribution of labor given wages:

\[
\frac{L_i}{L} = \frac{\left[ \left( \left( \frac{w_i}{P} \right) \bar{u}_i \right) / \bar{W} \right]^{1/\beta}}{\sum_{i'} \left[ \left( \left( \frac{w_{i'}}{P} \right) \bar{u}_{i'} \right) / \bar{W} \right]^{1/\beta}} = \frac{\left[ (w_i \bar{u}_i) \right]^{1/\beta}}{\sum_{i'} \left[ (w_{i'} \bar{u}_{i'}) \right]^{1/\beta}}. \tag{11}
\]

After one price normalization in \( P \), this leaves us with two sets of \( N \) equations (\( (10) \) & \( (11) \)) in two sets of \( N \) unknowns (\( \{ w_i \} \) & \( \{ L_i \} \)), which uniquely pins down the equilibrium wage and labor. The equalized welfare for the economy can then be written as the solution to \( \bar{W} = w_i \bar{u}_i L_i^{-\beta} \) for any location \( i \). For the purpose of later discussion, and to make the presence of endogenous variable more clear, it is useful to rewrite \( (10) \) in order to

---

\(^{12}\)Proposition 4 of Antràs et al. (2017) claims that this free entry condition delivers a unique \( B_i \) for each region \( i \) when wages are taken as given. This will allow us to use the observed distribution of wages to recover \( B_i \).
explicitly show how wage and labor enters the equation:

\[
\int_{z_i}^{\infty} \left[ \left( \frac{1}{z} \left( \sum_{j' \in J(z)} A_{j'} \left( \tau_{ij} w_{j'} \right)^{-\theta} \right) \right)^{-1/\theta} \right]^{a(1-\sigma)} \left( w_i^{1-\alpha}(1-\sigma) \right) \left( \sum_i w_i L_i \right) \tilde{B}_i - w_i \sum_{j=1}^N f_{ij} \right] dG_i(z) = w_if_{ei}
\]

(12)

where, using our price normalization, \( \tilde{B}_i \) now only depends on exogenous parameters

\[
\tilde{B}_i = \lambda a(\sigma-1)/\theta \left( \frac{\sigma}{\sigma-1} \right)^{1-\sigma} \left( \frac{1}{\alpha} \right) \left( \frac{1}{1-\alpha} \right) \left( \frac{1}{T_i} \right) \]

(13)

There are several key differences between the model we present here and that of Antràs et al. (2017) outside the contextual differences of knowledge production instead of trade. First, we assume that final goods can be freely traded so that the CES price index for consumption varieties is equal across regions. Second, we do not assume wages are exogenously determined by an un-modeled sector and instead allow wages to be endogenously determined. Estimation of this model is only made possible by jointly making these assumptions. If not, then we wouldn’t have the required equations to pin down an additional \( N \) unknowns in the endogenous price indices. Third, we assume labor is mobile, rather than being fixed for each region, and thus distributes so that welfare equalizes across locations. The assumption of labor mobility is motivated by the fact that this model is applied to describe patenting and production activity within a country where people are more free to move between locations relative to movement between countries, which would be the case if mobility was allowed in Antràs et al. (2017). Additionally, as we see in Table 5, the distribution of labor responds to changes in infrastructure so accounting for this behavior is ideal if the key changes we will make in counterfactuals is altering the determinants of fixed sourcing costs and trade costs through changes in the ICE network.

\[13\] The simplified model of Bernard et al. (2019) also makes this assumption, but they additionally assume wages equalize across locations.
5 Model Estimation

The process for estimating our model is based off of Antràs et al. (2017), but due to the assumptions made and data availability we require additional steps to recover parameters and run counterfactuals. One limitation of the author’s estimation strategy is that it must be performed with the importing location fixed. However, not only do we have data on the input usage of every single inventor in each region in our data, but also our model requires we estimate sourcing outcomes for every region as in equilibrium these outcomes are jointly determined. Estimation will broadly follow a three part procedure. First, we use a regression on patent-level import shares with fixed effects to partially recover the sourcing potential of each region and the determinants of trade costs. We can use our estimates to fully reconstruct sourcing potentials for each region as measured from each importing location. Second, we estimate the Frechet dispersion parameter for task quality from our partially recovered sourcing potentials. Third, we use the simulated method of moments to implicitly estimate the productivity in final goods production of each region as well as the fixed costs of sourcing. With these estimates in hand, we can consider counterfactual setups for the ICE network and their aggregate implications for economic and patenting outcomes.

The key dataset used in our estimation is the OECD REGPAT Database for patent applications to the European Patent Office. It crucially contains the location of application for each patent and the location of each listed author. To map the data to the model, we let every patent application represent an “inventor” and the authors of that patent are the “tasks” used as inputs in its creation. We initially calibrate the model considering patent applications filed between two years prior and two years after 2002. For our initial estimation, we only include regions where at least 100 applications are submitted or where at least 100 authors are listed as from. This leaves us with 331 of the 396 available
regions. As we will discuss later, the sample of regions considered will have to be ratified to accommodate some of the methods used.

5.1 Parameter Estimation

We first need to recover the sourcing potential for each region pair, $A_j (\tau_{ij}w_i)^{-\theta}$, as these values are crucial in later solving the inventor’s problem when implementing the simulated method of moments. Since the quality of tasks in a region, $A_j$, is unobserved we follow Eaton and Kortum (2002) to estimate the qualities from import shares, but just as Antràs et al. (2017) use firm-level shares we use inventor-level shares.\(^{14}\) Let $n$ denote the inventor, and we begin by normalizing the import shares, as given in (8), of inventors in region $i$ from region $j$ by the import share from the domestic region:

$$\frac{\lambda_{ij}^n(z)}{\lambda_{ii}^n(z)} = \frac{A_j (\tau_{ij}w_i)^{-\theta}}{A_i (\tau_{ii}w_i)^{-\theta}} = \frac{A_j (\tau_{ij}w_i)^{-\theta}}{A_i (\tau_{ii}w_i)^{-\theta}}$$

In practice, the import shares are the fraction of authors in location $j$ cited on patent application $n$ filed in location $i$. One important assumption we make here is that a minimum share of tasks come from the domestic region.\(^{15}\) In the patent data there is nothing that explicitly motivates this assumption, but it is necessary in order to make sure the estimated sourcing potentials are sensible. Otherwise, we run into cases where there are patents which either never or rarely source from the domestic region and thus imply either undefined or excessively large sourcing potentials. When we then turn to the simulated method of moments, it is hard to rationalize these large sourcing potentials with

\(^{14}\)Antràs et al. (2017) estimate their model holding the importing location $i$ fixed. They therefore normalize the domestic sourcing potential to 1. Since we need to recover sourcing potentials for each destination-origin pair we have to adopt a slightly different approach, allowing for two fixed effects when running our regression on input shares.

\(^{15}\)We end up imposing that at least 75% of tasks come from the domestic region. Lower values for this minimum share produced similar results for the trade cost coefficients and for $\theta$, but increasingly make the simulated method of moments more difficult to implement.
the moments. Intuitively, we can think of this assumed minimum share as capturing the non-inventor tasks required in producing a patent including the necessary bureaucratic, legal, and support tasks which might be provided at the location of application.\textsuperscript{16}

We continue by log-linearizing the expression for relative shares:

$$
\log \left( \frac{\lambda_{ij}^n(z)}{\lambda_{ii}^n(z)} \right) = -\theta \log \tau_{ij} + \log A_j (w_j)^{-\theta} - \log A_i (\tau_{ii} w_i)^{-\theta}
$$

We normalize the cost of trade to the domestic location, $\tau_{ii}$, to be 1 and then define the following variable:

$$
S_j = \log A_j (w_j)^{-\theta}
$$

We also posit that the trade cost term $\tau_{ij}$ can be written as

$$
\log \tau_{ij} = \log \beta_0 + \beta_d \log \text{distance}_{ij} + \text{ICE}_{ij} \log \beta_I + \text{continguous}_{ij} \log \beta_c
$$

where $\text{distance}_{ij}$ is the distance in kilometers between the centroids of regions $i$ and $j$, $\text{ICE}_{ij}$ is an indicator for the presence of a direct ICE connection between $i$ and $j$, and $\text{continguous}_{ij}$ is an indicator for whether $i$ and $j$ share a border. We can then write the log-linearized expression for normalized import shares as a regression equation:

$$
\log \left( \frac{\lambda_{ij}^n(z)}{\lambda_{ii}^n(z)} \right) = S_j - S_i - \theta \left( \log \beta_0 + \beta_d \log \text{distance}_{ij} + \text{ICE}_{ij} \log \beta_I + \text{continguous}_{ij} \log \beta_c \right) + \epsilon_{ij}^n
$$

We can estimate (14) using OLS, imposing equality in the importing and exporting fixed effects for each region, to recover a portion of every region’s sourcing potential $\hat{S}_j$ and

\textsuperscript{16}A majority of patents (71\%) received by the EPO are from large enterprises and the remainder is split between small and medium size business as well as individuals and universities (EPO (2018)). Large enterprises undoubtedly have vast support networks for R&D which go beyond the listed authors on a patent. We can think of this imposed domestic import share as reflecting this fact. However, a more complete implementation of this procedure should attempt to more precisely pin down the non-author contributions to patent production.
the coefficients for the determinants of trade costs \( \{\hat{\beta}_0, \hat{\beta}_d, \hat{\beta}_I, \hat{\beta}_c\} \). From here we can reconstruct an estimate for the sourcing potential for each region pair \( i,j \)

\[
\log A_j (\tau_{ij} w_i)^{-\theta} = \hat{S}_j + \log \hat{\beta}_0 + \hat{\beta}_d \log \text{distance}_{ij} + \text{ICE}_{ij} \log \hat{\beta}_I + \text{contiguous}_{ij} \log \hat{\beta}_c \quad (15)
\]

Column 1 in Table 6 presents the coefficients on the trade cost terms obtained from estimating (15) and Figure 8 maps the estimated sourcing potential components \( \hat{S}_j \). As expected, relative imports are decreasing with distance and rising with the presence of an ICE connection. However, counterintuitively, relative imports are decreasing for regions which share a border. In the data we see that a small number of regions account for a large share of the applications filed, so it may be that these regions already contain a large fraction of the labor force of nearby inventors and, after controlling for distance, collaboration mostly happens between these distant hubs rather than between contiguous regions.

We additionally require estimates for several parameters: \( \alpha, \beta, \sigma, \) and \( \theta \). The labor share of income for Germany was 70% in 2007 so given our production function we take \( 1 - \alpha \) to be 0.70.\(^{17}\) For \( \beta \) we consider the isomorphism of the spatial model of Allen and Arkolakis (2014), who model amenities as we do, to the Redding (2016) framework where \( \beta = \frac{1-\delta}{\delta} \) and \( 1 - \delta \) equals the expenditure share on housing. In 2014, 35% of German private consumption expenditures were on housing, implying a value of \( \beta = .53. \(^{18}\) For \( \sigma \) we follow Antràs et al. (2017) who use the fact that with CES preferences and monopolistic competition, the markup over marginal costs is given by \( \sigma / (\sigma - 1) \). Over the period 1996 to 2014 the average price-cost margin for Germany was 1.39, implying a value 3.56 for \( \sigma. \(^{19}\) We additionally follow the authors in obtaining our estimate for \( \theta \). Recall from our

\(^{17}\)See OECD (2015) for G20 labor shares.
\(^{18}\)See Statistisches Bundesamt (2020) for German consumption expenditures.
\(^{19}\)See Deutsche Bundesbank (2017) for mark-ups in European countries.
Table 6: Sourcing Potential Estimation

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable:</td>
<td>[ \log \left( \frac{\lambda_{ij}}{\lambda_{ii}} \right) ]</td>
<td>[ \log \left( \frac{\lambda_{ij}}{\lambda_{ii}} \right) ]</td>
</tr>
<tr>
<td>log(distance)</td>
<td>-0.174***</td>
<td>-0.158***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>ICE_{ij}</td>
<td>0.028***</td>
<td>0.189***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>contiguous_{ij}</td>
<td>-0.233***</td>
<td>-0.314***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Location FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>194235</td>
<td>79933</td>
</tr>
<tr>
<td>R²</td>
<td>0.744</td>
<td>0.768</td>
</tr>
</tbody>
</table>

Notes: * p < 0.1; ** p < 0.05; *** p < 0.01. Robust Standard Errors in parentheses.

regressions for sourcing potential we obtained estimates for \( S_j \) which can be written as:

\[
S_j = \log A_j - \theta \log w_j
\] (16)

Here we again posit some determinants of task quality including the number of patent applications, number of inventors, total employment, and employment in the sector including scientific and technical services (FRPS).\(^{20}\) Proxying for wage with the GDP per capita of each region, we can then recover \( \theta \) by estimating:

\[
\hat{S}_j = \beta_0 + \beta_A \log \text{Applications}_j + \beta_{INV} \log \text{Inventors}_j + \beta_E \log \text{Employment}_j + \beta_{FRPS} \log \text{FRPS Employment}_j - \theta \log w_j + \epsilon_j
\] (17)

\(^{20}\)There are some other potential determinants of task quality (for instance R&D expenditure and average level of education) that seem obvious, but due to data limitations these aren’t available at the NUTS 3 level for the years being considered.
Column 1 in Table 7 presents the coefficients we obtain from estimating (17) with OLS. We get an estimate of $\theta = 0.193$ which is low relative to the estimate Antràs et al. (2017) get of $\theta = 0.537$, implying a large dispersion in tasks qualities across location. The authors go on to instrument for wage with population, arguing that wages may reflect differences in productivity across countries, and get an even larger estimate $\theta = 1.789$. Since population is endogenous in our model, we can not follow in utilizing it as an instrument. Being that we consider an intra-country context of patent production, the scope for unobserved confounders to act on regional wages as well as the sourcing potential of authors in a region is arguably limited so we forgo instrumenting for wage and continue with our OLS estimates. If we are to believe the OLS estimated dispersion parameter, then the quality of tasks which authors across regions contribute to their patents is subject to much more heterogeneity than the productivity across countries.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{S}_j$</td>
<td>$-0.193^{**}$</td>
<td>$-0.371^{**}$</td>
</tr>
<tr>
<td></td>
<td>$(0.075)$</td>
<td>$(0.154)$</td>
</tr>
<tr>
<td>$\log(\text{Wage})$</td>
<td>$0.026^*$</td>
<td>$0.054$</td>
</tr>
<tr>
<td></td>
<td>$(0.015)$</td>
<td>$(0.034)$</td>
</tr>
<tr>
<td>$\log(\text{Applications})$</td>
<td>$-0.001$</td>
<td>$0.036$</td>
</tr>
<tr>
<td></td>
<td>$(0.015)$</td>
<td>$(0.042)$</td>
</tr>
<tr>
<td>$\log(\text{Inventors})$</td>
<td>$0.066$</td>
<td>$0.113$</td>
</tr>
<tr>
<td></td>
<td>$(0.045)$</td>
<td>$(0.128)$</td>
</tr>
<tr>
<td>$\log(\text{Employment})$</td>
<td>$-0.012$</td>
<td>$0.121$</td>
</tr>
<tr>
<td></td>
<td>$(0.032)$</td>
<td>$(0.098)$</td>
</tr>
<tr>
<td>Observations</td>
<td>331</td>
<td>49</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.101</td>
<td>0.209</td>
</tr>
</tbody>
</table>

Notes: $^* p < 0.1; ^{**} p < 0.05; ^{***} p < 0.01$. Robust Standard Errors in parentheses.

The final stage in the estimation procedure requires that we use the simulated method of moments to recover a large set of parameters that underlie fixed costs and productivities. This process is computationally intensive, so we have to reduce the number of regions included in the analysis. So far we’ve considered a fairly large sample with only
the regions lacking a significant amount of observations excluded. We reduce the regions included in the sample to those with either at least 2000 inventors or at least 1000 applications filed. This leaves us with 49 of the 396 possible regions. Column 2 in both regression tables features the coefficients we get from using this new sample. The major changes in our results are that the presence of ICE connections now places a larger role in determining trade costs and the task quality dispersion is higher at $\theta = 0.371$.

What remains to be recovered is the region-specific productivities and amenities as well as the fixed costs of observing a region’s task quality and the fixed cost of entry. Amenities, $\mu_i$, are easily recovered given the observed wage (again proxied for by the GDP per capita) and employment for each region from the equation we have for the equilibrium distribution of labor (11). See Table 8 for some statistics of the distribution of amenities. Productivities as well as fixed costs are recovered using the method of simulated moments as in Antràs et al. (2017). But since our model additionally requires that we estimate parameters for each importing location, we allow the determinants of fixed costs to vary by importing location. Moreover, fixed costs are allowed to vary by inventor where each is drawn from a log-normal distribution with dispersion parameter $\beta_{disp,i}^f$ and scale parameter similarly modeled to trade costs: $\log \beta_0^f + \beta_{d,i}^f \log$ distance$_{ij} + \text{ICE}_{ij} \log \beta_{ij}^f + \text{contiguous}_{ij} \log \beta_{c,i}^f$. Therefore, the remaining $N \times 6$ estimates required are $\left\{ \tilde{\beta}_i, \beta_0^f, \beta_{d,i}^f, \beta_{ij}^f, \beta_{c,i}^f, \beta_{disp,i}^f \right\}_{i=1}^N$ which collectively we denote by $\delta$.\footnote{We also follow the authors in setting the lower bound on productivity for active inventors to 1: $\bar{z}_i = 1 \forall i$.} Note that $\tilde{B}_i$, as defined in equation (13), is a function of exogenous parameters with the only region-specific term being the productivity in final goods production $T_i$.\footnote{In order to recover $T_i$ from $\tilde{B}_i$ we have to posit some value for $\rho$, the elasticity of substitution between tasks in the inventor’s patent production function (7). Since all we need to run counterfactuals is $\tilde{B}_i$ we forgo this step and just report estimates for $\tilde{B}_i$.}

In order to estimate $\delta$ we simulate a large sample of inventors in each region and solve the inventor’s problem in hopes of targeting a set of moments. To start, we now assume
that each inventor’s productivity in creating the knowledge flow required for final good production is a draw from a Pareto distribution with shape parameter $\kappa = 4.25$.\footnote{This value for $\kappa$ follows from Melitz and Redding (2015) as an estimate for a firm’s core productivity. Admittedly, a more complete analysis would include an estimate for $\kappa$ given our alternate context of patent production. But, since inventors underpin firms we continue with this assumption.} For each inventor we draw one of these productivities as well as a vector of fixed costs for observing the task quality of other locations which comes for the log-normal distribution described above and determined by the $\beta_i$ components in $\delta$. Given all these parameters, we can now solve the inventor’s profit maximization problem where it decides the set of regions to source from. To do so, we must evaluate $2^N$ possible sourcing options for each inventor which poses a computation issue for large $N$, but as in Antràs et al. (2017) we can use Jia’s algorithm to ease the computational burden. Jia’s algorithm will effectively create bounds on the regions that we need to consider by evaluating the marginal benefit of including each region by itself in the set sourcing locations. Crucially, the algorithm requires that the inventor’s profit function (9) is super-modular in $\Phi_i(z)$, which will only be the case if $\alpha (\sigma - 1) > \theta$.\footnote{For a longer discussion on Jia’s algorithm and its appropriateness in this kind of sourcing model please see Antràs et al. (2017). In particular, see Proposition 2 for the relevant restriction of parameters.} The current calibration of our parameters does in-fact satisfy this restriction.

Given the optimal sourcing strategy for each simulated inventor, we can calculate two sets of moments. The first set of moments is a column vector of length $N$ where each entry is the share of inventors in region $i$ (row) that engage in importing. The second set of moments is a matrix of size $N \times N$ where each entry is the share of inventors in region $i$ (column) which import from region $j$ (row). In using the simulated method moments, we are effectively looking for the parameters $\delta$ which minimize the squared difference between the model generated moments and the moment as taken from the data. Given the $\delta$ which satisfies this criteria, we can finally determine the fixed cost of entry for each location, $f_{ei}$, by setting it equal to the model-generated average profits for each region.
divided by it’s wage.

Table 8 features statistics on the estimated components of fixed costs as well as $\tilde{B}_i$. Note that with the way fixed costs are constructed, a value for any parameter $[\beta_{f_1,i}, \beta_{f_2,i}]$ that is less than 1 implies a reduction in fixed costs. So, fixed costs are increasing in distance as well as decreasing for contiguous locations and for locations which share an ICE connection. Recall that earlier Table 6 demonstrated that relative import shares were decreasing in contiguity for inventors who had contiguous regions in their sourcing strategy. This result for $\beta_{c,i}$ instead says that the fixed cost of including a contiguous region in a sourcing strategy is low. Taken together, this means that lots of patents have authors from nearby regions, but those authors constitute a small share of the imported authors on a patent. Notice also that fixed costs are about 5 times lower for regions which share an ICE connection. Alongside the coefficient we observe on ICE connections in Table 6, these estimates suggest that changes in the ICE network have the opportunity to drastically change how inventors source their patent production.

Table 8: Region Specific Parameters

<table>
<thead>
<tr>
<th></th>
<th>$\mu_i$</th>
<th>$\tilde{B}_i$</th>
<th>$\beta_{0,i}^f$</th>
<th>$\beta_{d,i}^f$</th>
<th>$\beta_{1,i}^f$</th>
<th>$\beta_{c,i}^f$</th>
<th>$\beta_{disp,i}^f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>0.990</td>
<td>0.554</td>
<td>0.945</td>
<td>2.040</td>
<td>0.212</td>
<td>0.109</td>
<td>3.012</td>
</tr>
<tr>
<td>10th Percentile</td>
<td>0.641</td>
<td>0.454</td>
<td>0.804</td>
<td>1.769</td>
<td>0.097</td>
<td>0.017</td>
<td>2.838</td>
</tr>
<tr>
<td>90th Percentile</td>
<td>1.353</td>
<td>0.745</td>
<td>1.087</td>
<td>2.255</td>
<td>0.329</td>
<td>0.162</td>
<td>3.213</td>
</tr>
</tbody>
</table>

Note that for regions which don’t have a contiguous region in the sample or for regions which don’t have ICE connections, the estimates for $[\beta_{1,i}^f, \beta_{c,i}^f]$ are meaningless since there is no variation in the corresponding variables.\textsuperscript{25} Table 8 excludes these cases. Moreover,

\textsuperscript{25}Keeping these parameters in the simulated method of moments is harmless since they will always be
when we later turn to running counterfactuals with alterations to the ICE network, we use the 75th percentile ICE fixed cost parameter for the regions which gained connectivity as their previous parameter wasn’t informative. We assume that the original network was chosen somewhat optimally among possible locations so that the most economically impactful network connections were built. The estimates for $\beta_{1,i}$ would therefore reflect this fact, so we choose a value higher than the median which represents a relatively less effective connection.\footnote{The 75th percentile value is 0.245, only slightly different than the median which would have been the most straightforward estimate to use. The counterfactual results do not meaningfully change between these two choices.} Table 9 summarizes our discussion of the procedure for the calibrating the parameters for the structural model.

### Table 9: Structural Parameter Summary

<table>
<thead>
<tr>
<th>Structural Parameter</th>
<th>Value</th>
<th>Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$ Frechet Dispersion</td>
<td>0.37</td>
<td>Equation (15)</td>
</tr>
<tr>
<td>$\sigma$ Elasticity of Substitution</td>
<td>3.56</td>
<td>Price Markup</td>
</tr>
<tr>
<td>$\beta$ Amenity Dispersion</td>
<td>0.53</td>
<td>Housing Expenditure</td>
</tr>
<tr>
<td>$1 - \alpha$ Production Parameter</td>
<td>0.70</td>
<td>Labor Share</td>
</tr>
<tr>
<td>$A_i (w_i)^{-\theta}$ Sourcing Potential</td>
<td>See Figure 8</td>
<td>Equation (12)</td>
</tr>
<tr>
<td>$\tau_{ij}$ Trade Costs</td>
<td>See Table 6</td>
<td>Equation (12)</td>
</tr>
<tr>
<td>$\mu_i$ Amenity</td>
<td>See Table 8</td>
<td>Equation (9)</td>
</tr>
<tr>
<td>$B_i$ Implicit Productivity</td>
<td>See Table 8</td>
<td>Simulated Method of Moments</td>
</tr>
<tr>
<td>$f_{ij}$ Fixed Costs</td>
<td>See Table 8</td>
<td>Simulated Method of Moments</td>
</tr>
</tbody>
</table>

When it comes to fitting the model to the data, there are some notable obstacles. First, there is the problem of dimensionality. For every region included we gain 6 additional parameters to adjust and an additional $2 \times N$ moments to target. If we included the full raised to the power of 0 and thus produce no change when multiplied with the other fixed cost components.
set of NUTS 3 regions this would leave us with roughly 65 moments for every degree of freedom. Therefore, as we explained earlier, we reduced the set of regions included to only the 49 most active regions. Second, the pattern of patent production, as it appears in the data, does not lend itself to be modeled as cleanly as trade, the original topic of focus for this sort of model. The nature of the patent data, with a few discrete number of inventors being sourced on each patent, leads to sometimes peculiar results in terms of sourcing potentials which are hard to reconcile with the moments. This is especially the case for our second set of moments which is the share of patents in \( i \) sourcing an author from \( j \). For many location pairs, this share is near-zero and there simply is not enough finesse in the model to accurately target these moments. Third, in the model we end up estimating, we are optimizing a non-linear and non-continuous function with respect to 294 parameters. The non-continuity limits the algorithms available for solving this problem with the only reliable and practically implementable procedure being a Nelder-Mead simplex search. Nelder-Mead is however not guaranteed to find the global minimum and is a relatively slow method.\(^{27}\) Despite these obstacles, the model is able to reasonably target the moments. The correlation coefficient for the first set of \( N \) moments is 0.99 (Figure 9) and for the second set of \( N^2 \) moments is 0.97 (Figure 10).

### 5.2 Counterfactuals

Given the estimates from the previous section, we can use equations (11) and (12) to determine counterfactual outcomes in response to underlying changes in the determinants.

\(^{27}\)The results presented in this section come from running Nelder-Mead simplex search for one week while parallelized across 12 processor cores. At this point it had still not converged to a minimum, but the iteration steps were small enough that it warranted we stop. This computational bottleneck additionally means that we were not able to try alternate starting points so there is definitely room to improve on this model’s ability to target the moments. Moreover, one possibility is to follow Bruins et al. (2018) and smooth the objective surface thereby allowing the calculation of first and second derivatives, but this is not a trivial procedure.
of import flows and sourcing strategies. Specifically, we can consider alternate arrangements of ICE high-speed rail connections and calculate changes in wages, employment, GDP, the share of inventors who engage in importing authors (tasks), and welfare. In Section 3.2 we documented the empirical relationship between ICE connections established in the second wave expansion and aggregate outcomes. Since our model is calibrated to the ICE network that existed before the second wave, we first consider the changes that result from building the second wave and compare the model implied outcomes to the observed outcomes. We consider mean changes for three sets of regions: those that received an ICE connection in the first wave (prior to 1999), those that received a connection in the second wave (1999-2010), and those that were not connected as of 2010. Respectively, these groups represent 72%, 5%, and 22% of total GDP.

Scenario A in Table 10 presents outcomes for the actual ICE network that was built in the second wave expansion. Regions which received a new ICE station saw a roughly 6% rise in GDP and a 3% rise in employment on average. This is broadly consistent with the results in Table 5 which suggested a roughly 6% change in GDP and 5% change in
employment for regions which experienced ICE connectivity. ICE stations significantly reduce the fixed cost of sourcing from other connected regions, so we also see a significant increase in the share of inventors which engage in importing authors from other regions. Recall that the only margin the inventor has for improving firm profits is maximizing task quality subject to labor and trade costs by choosing locations to source from (see Equation (9)). Therefore, the increase in sourcing activity is precisely the mechanism through which economic gains arise. These gains, targeted at the Wave 2 regions, cause reallocations in labor from the unconnected regions as they become comparatively worse at patent production. Wave 1 regions also see economic gains since they already had ICE connections and now simply have more regions to travel to on the network. Overall, there is a 1.0% rise in welfare from the change in infrastructure.

We next consider the effects of removing the entire ICE network (Table 10 Scenario B). Removing the entire network promotes significant labor outflows from the connected regions and improves the GDP of unconnected regions. There is however an aggregate reduction in wages which combined with the negative externalities on amenities caused by the new inflows of employment reduces the equilibrium welfare and output. Recall from Table 8 the estimates for $\beta^{f}_{L,i}$, which are only identified for Wave 1 regions since only these locations have variation in $\text{ICE}_{ij}$. For many of these regions the ICE connections are incredibly important as they reduce the fixed cost of sourcing by upwards of a factor of 10. When these connections are removed, patent production costs will rise creating these large deteriorations in the economy. We can think of the decline in welfare caused by removing all connections as conversely the increase in welfare from the construction

---

28There is a distinction between the treatment in the model and the treatment in the empirical exercise. For endogeneity concerns we removed targeted cities in the empirical exercise and defined connectivity as intersection with the Euclidean line which instrumented for the ICE route. In the model we include targeted cities and introduce connectivity by endowing regions with an ICE station. So, there is overlap between the connectivity definitions, but they are not exact. Either way, the empirical and model estimates are attempting to get at the causal effect of an ICE connection so consistency in the results is reassuring.
of the first wave. Put in exact terms, the first wave of ICE connections raised welfare by 3.5%. For comparison, Hayakawa et al. (2021a), using a quantitative spatial model which includes spatial linkages between firms and labor mobility, find that removing Japan’s HSR network would reduce aggregate welfare by 6.5%. Using ridership and population statistics we see that Japan’s HSR network averaged 2.87 rides per person in 2015, and Germany’s HSR network averaged 0.81 rides per person in 2005. Given that Japan’s network is roughly 3.5 times larger in terms usage and the welfare gains are nearly 2 times larger, assuming some diminishing benefits to a larger network, these figures square well with our welfare estimates.

We finally consider expanding the entire ICE network to cover every location (Table 10 Scenario C). Developing the network to this scale will undoubtedly run into some diminishing returns as the existing routes represent the most important connections between the largest cities and the estimates for \( \beta_I \) and \( \beta_I^f \) reflect this fact. Full connection also implies the creation of nonsensical routes between regions that are so close that a high-speed trains wouldn’t even have time to reach top speeds before it needed to decelerate. To impose that the fixed effects would fall for these regions just as much as they would for regions hundreds of kilometers apart is not reflecting reality. Therefore, we can think of Scenario C as an upper bound for the potential gains from a full network where we see significant improvements in the GDP, collaboration, and welfare for all regions. The real benefits however, especially after accounting for the cost of constructing such a network which would expand the total length of HSR lines four fold, may be substantially lower.

---

29The authors calibrate their model to 2014 and ours is calibrated to 2002. These are the closest available ridership figures as they appear on Wikipedia.
Table 10: Counterfactual ICE Networks

<table>
<thead>
<tr>
<th>ICE Connection Status</th>
<th>Change in Wage</th>
<th>Change in Employment</th>
<th>Change in GDP</th>
<th>Change in Importing Share</th>
<th>Change in Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario A: ICE Wave 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1</td>
<td>1.119%</td>
<td>0.206%</td>
<td>1.330%</td>
<td>3.526%</td>
<td>1.009%</td>
</tr>
<tr>
<td>Wave 2</td>
<td>2.652%</td>
<td>3.096%</td>
<td>5.840%</td>
<td>31.346%</td>
<td>1.009%</td>
</tr>
<tr>
<td>Other</td>
<td>0.299%</td>
<td>-1.322%</td>
<td>-1.027%</td>
<td>-0.040%</td>
<td>-0.040%</td>
</tr>
<tr>
<td><strong>Scenario B: No ICE Connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1</td>
<td>-5.054%</td>
<td>-2.937%</td>
<td>-7.747%</td>
<td>-26.949%</td>
<td>-3.518%</td>
</tr>
<tr>
<td>Wave 2</td>
<td>-1.193%</td>
<td>4.595%</td>
<td>3.346%</td>
<td>0.228%</td>
<td>0.102%</td>
</tr>
<tr>
<td>Other</td>
<td>-1.151%</td>
<td>4.680%</td>
<td>3.475%</td>
<td>0.102%</td>
<td>0.102%</td>
</tr>
<tr>
<td><strong>Scenario C: Full ICE Connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1</td>
<td>7.582%</td>
<td>-1.028%</td>
<td>6.606%</td>
<td>16.194%</td>
<td>8.204%</td>
</tr>
<tr>
<td>Wave 2</td>
<td>7.196%</td>
<td>-1.741%</td>
<td>5.355%</td>
<td>46.213%</td>
<td>8.204%</td>
</tr>
<tr>
<td>Other</td>
<td>11.726%</td>
<td>6.319%</td>
<td>19.008%</td>
<td>32.682%</td>
<td>8.204%</td>
</tr>
</tbody>
</table>

Notes: ICE Connection Status defines regions by their connection to the ICE network as it appears in reality. Wave 1 regions are those that were connected prior to 1999 and Wave 2 regions are those connected between 1999 and 2010. Other regions are those that never received an ICE connection as of 2010. The sample of 49 regions which we consider includes 21 Wave 1 regions, 4 Wave 2 regions, and 24 other regions. In terms of GDP, Wave 1 regions account for 72%, Wave 2 regions account 5%, and other region account for 22%. Where recall that the 49 regions are those that produced at least 1000 patents or housed at least 2000 authors in the period 2001-2005. The figures reported are for the mean change in outcome for each respective group.

6 Conclusion

The majority of patent production is done by teams and collaboration among inventors is subject to geographical constraints. This paper considers how infrastructure can enable the diffusion of knowledge, as measured by inter-regional patent collaboration. We do so using a comprehensive dataset of patent applications and the locations of listed authors. Simultaneously, we consider the development of Germany’s high-speed rail network and evaluate how the pattern of patent production and aggregate economic outcomes react to developments in the network. Guided by what we observe in the aggregate, we develop a model to describe how patent producers source authors to work with and use this sourcing framework to underpin a spatial economy with mobile labor. We find the existence of large production, collaboration, and welfare gains from the presence of the high-speed rail network and find the potential for further gains from expanding the number of connections in the network. There are numerous ways to improve on the work done here
from allowing for more spatial linkages between labor, inventors, and firms to considering more team-level responses to the development of the ICE network to investigating more precisely how firms employ patents in production.
References


