Warping Space: High-Speed Rail and Returns to Scale in Local Labor Markets

Daniel F. Heuermann†
Regensburg University

Johannes F. Schmieder‡
Boston University, NBER, and IZA

October 2013

Abstract

Local returns to scale in the labor market have been notoriously difficult to disentangle from increasing returns in the product market and from the spatial sorting of workers and firms as a source for regional variation in productivity. In this paper we use the introduction of high-speed rail (HSR) as a natural experiment in order to isolate the impact of labor market size on urban wages from product market and sorting effects. The key idea underlying our identification approach is that high-speed trains reduce commuting times between regions and thereby effectively increase the size of local labor markets without directly affecting product markets. In addition, the exact timing of the opening of HSR connections can be regarded as exogenous, as an HSR network is very expensive to build, requires a long planning phase and is mainly the result of political decisions. Furthermore, several small towns were connected to the HSR network simply because of their location between major metropolitan hubs and simply got ‘lucky’ compared to neighboring towns. Drawing on a large and novel panel data set on the introduction of ICE-stations and on connection times between regions in Germany, as well as on a full sample of workers’ employment histories, we examine the effect of high-speed trains on commuting behavior, local employment, as well as on individual job finding probabilities and wages. Using a synthetic controls and a difference-in-difference approach we show in a first step that high-speed trains reduce traveling times between regions by seventeen percent on average and, as an effect, significantly raise the number of commuters between labor markets by connecting formerly distinct regions. We find that commuters incur wage gains of about three percent after the opening of an ICE-station, indicating that improved access to larger urban labor markets is associated with productivity gains for workers living in peripheral regions. Given an overall urban wage premium of about eight percent in Germany, these results suggest that between one third and half of overall agglomeration externalities are rooted in increasing returns to scale in local labor markets.

*We would like to thank David Card as well as seminar participants at Boston University and the IAB for many helpful comments on this project. All errors are our own.

† Department of Economics, University of Regensburg, 93040 Regensburg, Germany (email: daniel.heuermann@wiwi.uni-regensburg.de; website: http://www-wiwi.uni-regensburg.de/Persoenen/Daniel-F.-Heuermann.html.en).

‡ Department of Economics, Boston University, 270 Bay State Road, 02115 Boston, USA (email: johannes@bu.edu; website: https://sites.google.com/site/johannesschmieder/).
1 Introduction

It can be regarded as an established fact that workers earn higher wages in cities. This holds true for every country examined so far. For the US, numerous studies find wage premia between five and ten percent in urban SMSAs (e.g., Glaeser/Maré 2001, Yankow 2006), while Lehmer/Möller 2010 provide evidence for the existence of an urban wage premium of about eight percent in Germany, to name only two cases. While the existence of a nominal urban wage premium indicates a higher marginal product of labor in cities, the reasons for such higher productivity are less clear. Three general classes of explanations have been put forth in the literature. First, higher urban wages may result from a sorting of more productive workers and firms into cities. Secondly, the proximity of workers and firms may yield productivity advantages and, hence, increasing returns to scale in product and labor markets. Thirdly, geographic characteristics like natural resources or ports might lead to the founding of cities in economically beneficial locations and thereby lead to higher urban productivity. A number of papers have tried to distinguish between these explanations. The most recent and sophisticated contributions are those by Combes et al (2010) and Combes et al (2012), which both find that agglomeration economies rather than geographic advantage or sorting are the driving forces behind higher urban productivity.

The prominent role of agglomeration economies as an explanation for higher urban wages raises the question whether increasing returns to scale mainly arise in the labor market or in the product market. With respect to product market effects, lower transportation costs, the availability of non-tradable specialized intermediary inputs and a positive selection of firms due to more intense competition might be the driving forces behind a higher urban productivity. In contrast, agglomeration effects might also arise in the labor market because workers benefit from being close to a large number of firms and other workers as this allows for a better matching between workers and firms in thick labor markets, as well as for improved opportunities for mutual learning.

Empirically, it has proven difficult to distinguish between these channels since both are inherently correlated. Ideally, a natural experiment could be found where an exogenous change in the size of one of the markets is not accompanied by a change in the other one. The current literature has employed natural experiments in order to gain insight into the size of agglomeration effects but does usually not differentiate between product and labor market channels (see e.g., Davis/Weinstein 2002 and Greenstone/Hornbeck/Moretti 2010). A notable exception is the paper by Redding/Sturm 2008. Drawing on the case of German division and reunification as a natural experiment, the authors examine the effects of changes in product market access on city growth. In line with predictions from New Economic Geography models they find that the spatial distribution of city sizes can be explained by differences in market access. While being instructive on the role that market access plays for population growth, their results cannot necessarily be attributed solely to product market access as commuting linkages for workers may also have been cut off by the inner-German border.

In this paper we provide a novel approach to disentangling agglomeration economies arising in the product market from those arising in the labor market by using the introduction and expansion of the high-speed rail network in Germany as a natural experiment in order to isolate local increasing returns to labor market size. The key idea behind this approach is that a reduction in commuting cost due to faster train connections effectively increases the size of the labor market without affecting the product market. Since its introduction in 1991, the German high-speed train, the InterCityExpress (ICE), is used only for passenger transport. Hence, the opening of an ICE-station in a city allows
workers to commute over longer distance while it leaves the product market unaffected.\footnote{At a certain point it was considered to use the ICE-technology for freight trains as well, but this idea has been discarded.} As such, the introduction of high-speed rail connections is of prime relevance for workers in smaller cities, as these suddenly gain access to labor markets in larger cities which were not within commuting distance before. With respect medium-sized cities high-speed rail stations were often distributed according to political decisions rather than economic considerations. This led to a situation in which a number of 'lucky' cities were being endowed with a high-speed rail station, while other comparable cities were not. Especially in these cases the exact location and the timing of the opening of an ICE-station can be regarded as exogenous, as both were over long periods of time subject to political haggles between national, regional and local authorities and influenced by petitions and public decisions.

In our identification approach we make use of this quasi-random distribution of shocks to labor market size in order to disentangle the returns to scale arising within labor markets from those arising within product markets. In order to do so we draw on a novel data set which contains information on the establishment of ICE-stations, as well as on connection frequency and commuting times between county pairs. We combine this information with data on employment histories of all workers in Germany between 1975 and 2009. Since these data also contain information on a worker's place of residence and place of work we can observe commuting behavior between counties on the aggregate as well as on the individual level. Combining both data sets allows us to examine the impact that a sudden shock in labor market size has, first, on the commuting behavior and, secondly, on the productivity of workers.

We employ three different identification approaches. First, using county pairs as our unit of analysis we examine whether the introduction of high-speed rail has led to a higher number of commuters between local labor markets. On the individual level we then investigate whether the opening of an ICE-station raises the job change probability for workers and leads to wage increases. Finally, using a difference-in-difference approach with synthetic controls on the county level, we shed light on overall economic benefits arising for regions endowed with a high-speed rail station.

We find that the introduction and expansion of the ICE-network has reduced commuting times on those connections where it has been introduced by seventeen percent on average. These effects have been particularly pronounced for medium-sized and peripheral cities. Estimating the effect of high-speed rail on commuting patterns, we find the numbers of commuters to increase if two regions are connected by direct high-speed trains. Estimating the labor market returns to commuting by exploiting the expansion of the high-speed rail network as an instrument for commuting decisions we provide preliminary evidence that wages of commuters rise by about three percent compared to non-commuters, indicating that about half of the overall size of agglomeration externalities can be attributed to increasing returns in the labor market.

The next section provides an overview of the introduction and the expansion of high-speed rail in Germany and explains our identification approach. Section three contains preliminary empirical results on the effects of the introduction of high-speed rail on the commuting behavior of workers. Section four summarizes the results we have obtained so far and provides an outlook.
2 Data, Descriptives and Identification Approach

2.1 Data

We draw on three different data sets. First, we have extracted data from the electronic train schedules that have been published on CD-ROMs annually from 1993 onwards by the Deutsche Bahn AG, the German railway company. These schedules contain information on train connections between all 5,400 train stations in Germany. From these CDs we then extracted the connection times between all 390 county capitals in Germany between 1993 and 2012. We have taken the first Monday in June each year that is not a holiday and have calculated trip duration, the number of train changes, and the types of trains for the fastest connection between all 76,245 pairs of cities in the time window between 6:00 am and 9:00 am, i.e., at a typical commuting time. The resulting data set contains information on the size of the commutable labor market and can be used to calculate the number of jobs that can be reached within a certain time (e.g., 30 min, 60 min, 90 min) of train commute. In addition, we have extracted the year of establishment of an ICE-station within a county as well as information on the existence of direct ICE-connections between each pair of counties.

These information are merged with administrative data provided by the Institute for Employment Research (IAB), which contains information on all employees in Germany between 1975 and 2009. From this full sample of workers we have generated an annual panel data set on full-time employed workers, containing information on worker and firm characteristics. Unfortunately, information on the county of living is available only from 1999 onwards. As this is a key variable for identifying commuting behavior, most of our analyses refer to the period between 1999 and 2009. For these eleven years, the information on place of residence and place of working allows to gain insight into mobility patterns on the individual as well as on the aggregate level (i.e., between pairs of counties).

In order to control for regional characteristics like county area, population density, unemployment, GDP, land prices, longitude and latitude of the county capital, and natural endowments we make use of information provided by the Federal Statistical Office (Statistisches Bundesamt) in the GENESIS data set.

Throughout this paper we define a region as a city or a county (‘Kreise und Kreisfreie Städte’). These units, which are equal to NUTS-III regions and contain on average about 200,000 inhabitants, are not necessarily self-contained labor markets. Rather, one core city is often surrounded by a number of smaller and more rural counties. In this setting it crucially depends on the existing infrastructural endowments to which extent workers from peripheral counties are able to commute to core cities in order to work there. We use this variation in distances and in connection times in order to shed light on the returns to scale in local labor markets.

2.2 High-Speed Rail in Germany

As in a number of developed countries like Japan, France, Spain and Belgium, transportation policies in Germany have undergone a re-focus from motorized individual transport back to railway transportation in the late 1970s and 1980s, a shift often referred to as the “renaissance of the train”. After two decades of publicly funded research into tracks and propulsion technology, the high-speed train network was introduced in Germany in 1991 and has since then steadily been enlarged. Technically, high-speed trains are running on the same tracks as other passenger trains and, hence, and in most cases stations...
were adjusted to the requirements of the ICE. However, in a number of cases tracks and stations were newly built for high-speed trains. Maps I to IV provide an overview of the introduction and the spatial distribution of the stations that are serviced by ICE-trains. The high-speed rail era started with one major track running from Hamburg via Mannheim and Frankfurt to Munich, servicing a total of eleven stations. In 1997, the existing north-south axis was being complemented by an east-west track, going from Berlin to Cologne and covering a number of large cities in the Rhein-Ruhr area. In addition, a couple of new branches were added to the existing tracks, e.g. to Bremen and to Freiburg, so that in 1998 a total of 34 ICE-stops were in operation. With the opening of the Berlin-Munich line in 2000, a number of cities in Eastern Germany, including Leipzig and Jena, were connected to the high-speed network, leading to a total number of 60 ICE-stations being serviced in 2005. Since then, a number of smaller branches have been added to the network, e.g., to Lübeck and Rostock. Currently, 70 stations are connected by 260 high-speed trains on at least four hour intervals, in most cases every one or two hours. According to information provided by the Deutsche Bahn, nearly 80 million passengers have used the ICE in 2012.

Maps I - IV: Development of the High-Speed Rail Network in Germany

With the introduction of high-speed rail, connection times have decreased substantially between regions in Germany. Figures I and II summarize the development of average connection times and the share of connection that include the use of the ICE between all 76,245 county pairs.
As shown in the first graph, in 1993 the average travel time between a random county pair amounted to five hours and fifteen minutes. Until 2010, this duration had been reduced by about nine percent to
four hours and 47 minutes. These reductions do not arise in a linear way but occur first and foremost at points in time when major ICE-lines were opened, e.g., between 1998 and 2000 when a row of new stations were opened for the EXPO 2000 (World’s Fair) in Hanover and the track Berlin-Munich was inaugurated. The next graph shows that the share of connections which encompass the use of an ICE has increased from 48 to 75 per cent, strongly supporting the notion that the introduction and expansion of the high-speed rail network and an increased regional coverage of long-distance trains has contributed to a better accessibility of regions within Germany.

However, not all regions benefit from high-speed rail alike. Map V contains the average travel time for each county, i.e., the average travel time needed to get from one county to any of the 389 other counties. In 1994, when the main north-south ICE-line was already established, the regions that were best connected were, in fact, those located on the corridor Hamburg - Frankfurt/Nuremberg - Munich. Peripheral regions, especially those in the eastern part of the country, were those that used to be most distant to all other regions in terms of average trip duration. With the further expansion of the high-speed rail network, these regions benefited most from the reductions in travel times. Map VI, which contains the changes in travel times in absolute values, shows that first and foremost regions in the eastern part and in the Ruhr area, which were connected to high-speed rail by the east-west line and the Berlin-Munich line, saw reductions in average travel times, in some cases of more than one hour.

Maps V and VI: Regional Accessibility by Railway Transportation

Among the cities that were characterized by substantial reductions in travel times were a number of 'lucky' cities, which were connected to the high-speed rail network for quasi-random reason. Prime examples of small cities experiencing dramatic reduction in commuting times are, e.g, Montabaur, Limburg, and Ingolstadt, which with the introduction of high-speed rail were suddenly connected to large labor markets in Cologne (Limburg), Frankfurt and Bonn (Montabaur) and Munich (Ingolstadt).
In general, the later phase of the network expansion was characterized by the opening of ICE stations in small and medium-sized German cities which are usually located between the larger metropolises. In contrast to large cities, where economic considerations were the driving force for establishing an ICE-connection, in these cases the reasons for being connected to the network were mainly political ones, as those states (Bundesländer), through which the tracks went, had to contribute to the building costs and therefore demanded the establishment of stops on their territory in return. Table I contains a list of those cities within which an ICE station was opened between 1999 and 2009.

<table>
<thead>
<tr>
<th>Cities where an ICE Station was opened between 1999 and 2009</th>
<th>No of Inhabitants 2012 (in Thousands)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aachen Hbf</td>
<td>240</td>
<td>29</td>
</tr>
<tr>
<td>Bad Hersfeld</td>
<td>28</td>
<td>392</td>
</tr>
<tr>
<td>Bamberg</td>
<td>71</td>
<td>122</td>
</tr>
<tr>
<td>Eisenach</td>
<td>42</td>
<td>245</td>
</tr>
<tr>
<td>Erfurt Hbf</td>
<td>203</td>
<td>37</td>
</tr>
<tr>
<td>Erlangen</td>
<td>105</td>
<td>72</td>
</tr>
<tr>
<td>Gotha</td>
<td>44</td>
<td>215</td>
</tr>
<tr>
<td>Guetersloh Hbf</td>
<td>95</td>
<td>84</td>
</tr>
<tr>
<td>Halle(Saale)Hbf</td>
<td>230</td>
<td>31</td>
</tr>
<tr>
<td>Herford</td>
<td>65</td>
<td>136</td>
</tr>
<tr>
<td>Homburg(Saar)Hbf</td>
<td>41</td>
<td>236</td>
</tr>
<tr>
<td>Jena</td>
<td>107</td>
<td>75</td>
</tr>
<tr>
<td>Kaiserslautern Hbf</td>
<td>97</td>
<td>81</td>
</tr>
<tr>
<td>Koblenz Hbf</td>
<td>108</td>
<td>71</td>
</tr>
<tr>
<td>Koethen</td>
<td>27</td>
<td>436</td>
</tr>
<tr>
<td>Leipzig Hbf</td>
<td>520</td>
<td>12</td>
</tr>
<tr>
<td>Lichtenfels</td>
<td>20</td>
<td>664</td>
</tr>
<tr>
<td>Lutherstadt Wittenberg</td>
<td>47</td>
<td>192</td>
</tr>
<tr>
<td>Luebeck Hbf</td>
<td>211</td>
<td>36</td>
</tr>
<tr>
<td>Lueneburg</td>
<td>70</td>
<td>116</td>
</tr>
<tr>
<td>Minden(Westf)</td>
<td>80</td>
<td>101</td>
</tr>
<tr>
<td>Montabaur</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>Muenster(Westf)Hbf</td>
<td>297</td>
<td>20</td>
</tr>
<tr>
<td>Naumburg(Saale)Hbf</td>
<td>33</td>
<td>340</td>
</tr>
<tr>
<td>Neustadt(Weinstr)Hbf</td>
<td>52</td>
<td>176</td>
</tr>
<tr>
<td>Oberhausen Hbf</td>
<td>210</td>
<td>35</td>
</tr>
<tr>
<td>Osnabrueck Hbf</td>
<td>154</td>
<td>46</td>
</tr>
<tr>
<td>Passau Hbf</td>
<td>49</td>
<td>185</td>
</tr>
<tr>
<td>Regensburg Hbf</td>
<td>138</td>
<td>56</td>
</tr>
<tr>
<td>Rostock Hbf</td>
<td>203</td>
<td>38</td>
</tr>
<tr>
<td>Saalfeld(Saale)Hbf</td>
<td>25</td>
<td>457</td>
</tr>
<tr>
<td>Saarbruecken Hbf</td>
<td>177</td>
<td>43</td>
</tr>
<tr>
<td>Salzwedel</td>
<td>25</td>
<td>517</td>
</tr>
<tr>
<td>Siegburg/Bonn</td>
<td>39</td>
<td>279</td>
</tr>
<tr>
<td>Solingen Hbf</td>
<td>155</td>
<td>50</td>
</tr>
<tr>
<td>Uelzen</td>
<td>33</td>
<td>339</td>
</tr>
<tr>
<td>Weimar</td>
<td>63</td>
<td>130</td>
</tr>
</tbody>
</table>

Rank among all 685 large and medium-sized cities in Germany defined as those with more than 20,000 inhabitants

Data provided by the Federal Statistical Office

The list also provides the rank with respect to the number of inhabitants among all 685 large and medium-sized cities in Germany. It turns out that none of the cities is among the top ten cities and only two are among the top twenty. The median rank is 116 and the average number of inhabitants amounts to 111,000. These are the cases that are of prime interest here, as for these cities and the
the opening of a high-speed railway station constitutes a large, exogenous shock to the labor market.

While ICE-trains are a convenient and fast way of commuting, they are also relatively expensive way and as such probably most relevant for high income individuals. By means of examples, a yearly ticket from Montabaur to Bonn (22 minutes train ride) costs about 3,000 Euro per year. At the same time, this is much more economical than doing the same trip by car, which takes about 50 minutes and costs on average about 8,000 Euro per year. Hence, ICE-trains provide an efficient option mainly for highly productive and mobile workers who commute from small, peripheral cities to large urban labor markets.

2.3 Identification Approach

We make use of a three-tiered identification strategy, i.e., we focus on county pairs, single counties, and individual workers. In all three levels of the analysis we exploit variation in connection times between counties and in the timing of the introduction of ICE-stations in order to identify the returns to scale in local labor markets.

2.3.1 County Pairs

For the 76,245 county pairs in Germany we possess annual information on the existence of a direct ICE-connection, travel duration, train changes, and the number of commuters for the years 1999 until 2009. Using these data we explore whether and to what extent the introduction of a high-speed train on a certain connection has led to a reduction in commuting times and, in turn, to an increase in the number of commuters. This part of the analysis can be thought of as a first stage and, at the same time, as a necessary condition for the further analysis. In this step we ask if workers do take advantage of an enlargement of labor markets by changing jobs without changing their place of residence in an enlarged labor market.

We examine this question by means of case studies, a pooled event study, propensity score matching between treated and untreated counties, and an instrumental variable approach using gravity equations on commuting with ICE-station as an instrument for commuting times. In the case studies we first of all look at the 'obvious' cases, i.e., at those county pairs where reductions in travel times have been substantial and where the two affected counties have de facto become one labor market. In terms of identification, these cases are also the most clear-cut ones in terms of exogenous variation in accessibility, as the affected counties are those that got lucky by receiving an ICE-station for political reasons. Such cases encompass, e.g., Limburg, Montabaur and Ingolstadt.\footnote{The cities Montabaur and Limburg are located in are the counties 'Westerwaldkreis' and 'Limburg-Weilburg', respectively. Throughout the analysis we take the counties rather than the cities within them as our units of observations, but refer to both as 'Montabaur' and 'Limburg' as these are not only the names of the core cities within the counties but also of the two ICE-stations.}

In a pooled event study we then compare the development of commuter numbers on those pairs of counties which have seen substantial reductions in commuting times as a result of the introduction of a direct ICE-connection to cases where commuting times have remained constant. While these first two types of analysis are of a descriptive nature which help us to gain an understanding of the relationship between travel times and commuting, we then approach the question in a causal way by means of propensity score matching, where we compare county pairs which have been treated with a direct...
ICE-connection to counties which have not been treated but are otherwise similar in all relevant characteristics. Finally, we estimate the effect of travel times on the number of commuters in a gravity equation framework where we use the exact timing of the opening of an ICE-station as an instrumental variable.

2.3.2 Individual Workers

On an individual level we corroborate the evidence from the county-pairs by examining whether reduced connection times are associated with a higher job change probability of workers across county borders. Theoretically, with an increased accessibility of potential jobs in adjacent labor markets we expect workers be more likely to change to a better match. To gain insight into matching quality in enlarged labor markets we then estimate whether cross-regional job changes are indeed associated with significant wage gains. In contrast to the county pairs, in this analysis we can avoid selection of the type that people move into a county shortly before an ICE-station is opened, in order to commute to high-paying jobs. We therefore define a cohort of individuals, defined as all individuals who live in a county over the whole period of investigation, i.e., between 1999 and 2009. We then follow these individuals over time in order to examine whether workers in 'treatment' counties exhibit a different job changing behavior and incur higher wage gains when commuting to larger labor markets.

2.3.3 Counties

In this part we examine the effect that the opening of an ICE-station has on commuting, resident population, employment, and wages in a county over time. The objective here is to see whether the opening of an ICE-stations has an overall economic benefit for a county. The key idea is that benefits from such large-scale infrastructure projects can go well beyond an increased matching efficiency for a number of reasons. First, construction works for stations and tracks create demand for labor and thereby raise local wages. In addition, becoming a transportation hub not only leads to an increased demand for goods and services by individuals in transit, but also raise the attractiveness of a location, e.g., as a venue for conferences and meetings. In order to gain insight into such overall effects we resort to case studies with synthetic controls and a pooled event study with instrumental variables.

3 High-Speed Rail and Local Returns to Scale

3.1 County Pairs

Case Studies

In these case studies we first of all look at those connections which come closest to a 'perfect' experimental design. These are connections where ICE-stations in peripheral regions were built mainly for political reasons and where connection times have dropped to an extent that the peripheral region has suddenly gained access to a large urban labor market within a realistic daily commuting time. For these cases we first of all examine the development of commuter numbers.

One of the most definite cases in this respect is the connection between Montabaur and Siegburg (Bonn), as in this case both ICE-stations as well as the tracks were newly built. The track was opened in July 2002 as one central part of the new railway connection between Frankfurt and Cologne, which in turn is part of the trans-European transport network (TEN-T) policies of the European Union. The
The key idea of the project was to better connect the two largest German metropolitan areas, the Rhein-Ruhr area (in the state of North-Rhine Westphalia) with the Rhein-Main area (which lies mostly in the state of Hesse). The track is used only for passenger trains, i.e., travel times for freight trains have remained constant and, hence, product market access between both regions has remained unchanged. As the state Rhineland-Palatinate, through which the tracks also pass, demanded the establishment of a step on its territory in return for its financial contribution, the town of Montabaur with its 13,000 inhabitants suddenly saw itself in the lucky position to also be connected to the high-speed rail network. As a result, the connection time from Montabaur to Siegburg (Bonn) fell by a staggering three hours to just eighteen minutes. As shown in Figure IV, the number of commuters has since then increased steadily from about 300 to 450.

Figure IV: Train Times and Commuters Montabaur - Siegburg

The newly built tracks also connect Siegburg (Bonn) to the city of Mannheim, which is located in the Rhein-Main area. As shown in Figure V, the connection time between these two cities has dropped from three hours to about seventy minutes as a result of the new high-speed track. Accordingly, the number of commuters has gone up from below ten to about forty.
A third and final example is the county pair Nuremberg and Ingolstadt, which were also connected by a newly established high-speed route. In this case, Ingolstadt got lucky by being endowed with an ICE-station when the high-speed track between Nuremberg and Munich was finished in June 2005. As a result, the commuting time between Nuremberg and Ingolstadt plummeted from about ninety minutes to below thirty minutes. Since then, the number of commuters has nearly doubled from about 90 to close to 180.

All three cases have in common a dramatic reduction in traveling times resulting from the establishment of a new high-speed train connection. The increase in the number of commuters, which followed the opening of the connection, is stunning when compared to earlier levels. While in absolute terms the numbers may appear moderate, one needs to consider that commuting by high-speed train is only feasible for workers high-earning (usually highly qualified) workers. As the rise in absolute numbers is compatible with the transport capacity of one or two ICE-trains and as in all three cases commuting time and costs are substantially higher by car than by train, it is very likely that the additional commuters observed in the data are, indeed, using the ICE-train for commuting to work.
Pooled Event Study

In order to more thoroughly test for the effect that a reduction in travel time has on commuting patterns we now compare county pairs which are 'treated' with access to the high-speed rail network to 'control' pairs, i.e., combinations of countries where traveling time has remained constant. The idea here is to see whether rendering two formerly disconnected counties into commuting distance by means of an ICE-connection has led to an increase in the number of commuters between both labor markets. In order to conduct this type of experiment in a plausible way we impose a number of restrictions on the county pairs we examine.

First, we define both the treatment and the control group in a way that the travel time in 1999 between the two counties in a pair lies in the range between 90 and 240 minutes. This restriction makes sure that the two labor markets are not within a daily commuting distance at the beginning of the observation period. At the same time, all those county pairs are excluded which are too far away from each other to attain a commuting relation even by means of a high-speed rail connection. Secondly, the control group is restricted to pairs of counties where travel time is reduced by less than five minutes within the period of observation, thus ensuring that in the control group connection times remain unchanged independent on whether an ICE-connection is introduced or not.

In this overall setting we first define the treatment group such that a direct ICE-connection has been introduced between two counties and where, as a result, travel times have gone down by more than 50 minutes. We compare these county pairs to a group of controls where no ICE-connection has been introduced within the period of observation. Given these definitions we obtain 56 county pairs in the treatment group and 7,796 county pairs in the control group. Figure VII compares both groups with respect to average commuting time and the number of commuters. For the county pairs in the treatment group the average connection time has decreased from 190 minutes in 1999 to slightly below two hours in 2009. The bulk of this reduction occurs between 2002 and 2003, pinpointing the importance that the opening of the track between the Rhein-Ruhr area and the Rhein-Main area has had within the German railway network. Between 2003 and 2009, the average number
of commuters between the 59 treated county pairs rises from about 300 to more than 450. This rise occurs mainly from 2005 onwards, i.e., with a lag of one to three years after the new rail connection has been opened. For the control group the number of commuters remains largely unchanged during the period of observation, indicating that the rise in commuters in the treatment group is not the result of an overall trend, but is associated with the reduction in travel durations resulting from the introduction of a high-speed rail connection.

**Figure VII: Treatment - Control I**

![Graph showing the number of commuters and trip duration by treatment/control.](image)

In a second experiment we compare the treatment group to a control group of county pairs which were also endowed with a direct ICE-connection between 1999 and 2009, but which did not experience time reductions of more than five minutes. As a result, the number of observations in the control group decreases to 95, for which the average connection time ranges at 155 minutes. As shown by Figure VIII, the number of commuters between these counties only slightly increases over time, indicating that there is only a minor tendency for commuter numbers to generally rise on ICE-connections. Hence, the introduction of a high-speed rail connection seems to be only accompanied by rising commuter numbers if it comes along with substantial reductions in travel times.
In a third experiment we examine the extent to which commuter numbers rise with the possibility of daily commuting. We therefore make sure that connection times in the treatment group not only fall dramatically, but that the county pairs are within a daily commuting range thereafter. Hence, we only include county pairs in the treatment group where a direct ICE-connection was introduced between 1999 and 2009, which led to a reduction in travel times by at least 50 minutes, resulting in a travel duration of below 60 minutes in 2009. The control group remains the same as in the second experiment, i.e., we look at county pairs where a direct ICE-connection was introduced between 1999 and 2009 which did, however, not lead to reductions in travel times. This experiment is the most strict one with respect to the treatment group, which now only contains five observations. As before, the fall in travel time mainly occurs in 2003. The number of commuters rise by about 50 percent in the period of observation, namely from about 400 to 600 persons. Despite these larger numbers, the treatment effect is not as clear here as it was before, as commuter numbers were to some extent already on the rise between 1999 and 2002.
**Propensity Score Matching**

Check: is there any literature using PSM on commuting flows?

Insert regression table and interpret; insert table with descriptive statistics for key variables for county-pairs: comm numbers, average commuting distance, distance, comm time, size of origin and destination; for P-Matching report both for treatment and control group

**Gravity Model with Instrumental Variables**

Insert two tables (first and second stage) and interpret

check systematically: are there any papers that have looked at the pull/push factors of commuting in terms of city size (using a gravity model) who has used a gravity model of commuting anyway?

Melo et al (2011): I. there is no specific theory for the parameter of distance in gravity models (distance decay parameters); II. mention: average time for the way to the station (related to county size) is covered by the FE; III. report elasticities and semi-elasticities; IV. key results on distance: raising distance by 10% decreases commuters by 3% (elasticity) and raising distance by 1km decreases commuters by 5 to 6 percent (semi-elasticity): compare these results to our parameter on distance variable; V. coefficients on size of origin and destination are not reported; VI: theory papers on spatial searching behavior and commuting (might be useful for model in 2nd part of the analysis).

McArthur et al (2013): I. parameter in front of distance is called distance delay/deterrence parameter; II. parameter value: .065 (check how to interpret this in structural equation that is used in the paper); III. policy recommendation: pay bridges and tunnels from taxation, not via tolls (can we arrive at similar normative conclusions?); IV: bibliography: three potentially interesting papers

Matha/Wintr (2007): I. effects of size origin/destination: rise in 1% in size leads to 7% (origin) and 5%(destination) more commuting; II a+b: use poisson and negative binomial model.

General form of gravity equation:

\[ C_{ijt} = aR_{it}^{\alpha}W_{jt}^{\beta}f(d_{ijt} \gamma) \]  

Taking logs and adding time and connection FE leads to the following form:

First Stage:

\[ \log(d_{ijt}) = b + o\log R_{it} + \beta \log W_{jt} + \gamma ICE_{ijt} + c_{ij} + t + \epsilon_{ijt} \]
Second Stage:

$$logC_{ijt} = b + \alpha logR_{it} + \beta logW_{jt} + \gamma \log(d_{ijt}) + c_{ij} + t_t + \epsilon_{ijt} \tag{3}$$

3.2 Individual Level

On the individual level we corroborate our findings on rising commuter numbers following the introduction of an ICE-connection. In order to identify the effect of a drop in connection times on commuter numbers we focus on the newly established station in Montabaur and examine whether the access to the high-speed rail network in 2002 is followed by a rise in the number of out-commuters. We construct a sample of all highly-qualified workers who live in the county in 2002 and examine whether these workers are more likely to commute to one of the eleven counties which Montabaur is connected to by a direct ICE. In a first step of this analysis we find that the number of workers commuting out of Montabaur to one of the eleven urban labor markets rises by about fifty percent from 300 to 450 after between 2003, i.e., after the ICE-connection has been introduced.

The average wage of workers commuting out of Montabaur increases sharply in 2003, namely by six percent from 132 to above 140 Euros. A regression analysis with individual fixed effects using all workers living in Montabaur shows that about half of this rise is explained by the introduction of a direct ICE-connection to large labor markets. The wages of workers commuting to these eleven cities show a sudden positive level effect of about three percent, while the other three percent are explained by an overall positive time trend arising for all workers (we cannot provide a regression table here as the results currently are in the process of anonymity screening by the Research Data Centre). These findings provide first evidence on the monetary returns of being connected to a larger labor market. Comparing this wage jump to the results by Lehmer/Möller (2010), who find for a genuine urban wage premium of about eight percent, provides tentative evidence that between one third and half of the urban wage premium is explained by increasing returns in the labor market.

3.3 Counties

In this section we examine whether benefits from increased rail accessibility can also be observed on the level of single counties. We therefore conduct case studies with synthetic controls, a pooled event study and an instrumental variable approach.

As an example for a case study we have again picked Montabaur as the treated observation and created a synthetic control county using a convex combination of other counties in the same state, i.e., in Hesse. We first compare the development of the number of out-commuters to that of the control group. Figure X shows that between 1999 and 2009 the number of workers commuting out of Montabaur has increased from about 24,000 to above 30,000. While the increase in the control group has been substantial as well, both trends start to diverge in 2003, the year after the ICE-station in Montabaur was opened. In 2009 the number of out-commuters in the control group falls short by about 1,500 compared to Montabaur. Certainly, not all 6,000 additional commuters out of Montabaur use ICE-trains on their way to work. However, given that about 150 of the additional commuters are working in Siegburg(Bonn), which is a prime example for a case where commuting by high-speed train is the most efficient solution, it is not unrealistic to assume that a substantial share of the 1,500 workers exceeding the trend in the control group does, in fact, make use of the newly established train
connection. This number also fits very well with the number of daily ICE-passengers which, according to statistics published by the Deutsche Bahn AG, amounts to about 2,400 daily (Frankfurter Allgemeine Zeitung 2012).

**Figure X: Commuters out of Montabaur compared to Synthetic Control**

![Graph showing commuters out of Montabaur compared to Synthetic Control](image)

4 Summary and Outlook

The aim of this paper is to shed light on the extent to which agglomeration externalities are driven by increasing returns in the labor market. Addressing this question we use the German case of the introduction and extension of a high-speed rail network as a natural experiment in order to disentangle increasing returns arising in the labor market from those arising in the product market.

Our results at this stage are preliminary but promising. Focusing mainly on the newly built ICE-connection between the Rhein-Ruhr area and the Rhein-Main area we have provided evidence that access to the high-speed rail network has drastically reduced connection times between counties and has, in some cases dramatically, increased the accessibility of urban labor markets for peripheral counties. For the cases examined to far, the reductions in average connection times are accompanied by sizable increases in the number of commuters. In all three stages of the analysis, i.e., for county-pairs, individuals and single counties, we find that that introduction of an ICE-station is followed by a rise in the number of commuters of about fifty percent. We provide tentative evidence that for these commuters wages rise by about six percent, out of which half is explained by access to larger labor market. Taken together, these findings can be taken as preliminary evidence that between one third and half of the higher urban productivity is rooted in increasing returns to scale in local labor markets.

A lot remains to be done. First, it needs to be clarified whether our results apply to all workers
or are driven by highly qualified workers alone as suggested by theoretical considerations and by the
analysis of wages. Secondly, another key step is to take the analysis to a broader level, i.e., to examine
whether these results also hold for other connections, city pairs and individuals in other regions.
Thirdly, the issue of sorting needs to be addressed more cautiously, i.e., we need to make sure that the
alleged wage effects from commuting are not merely an expression of highly motivated workers taking
the train. This issue can be addressed by examining the job changing behavior of workers who have
been commuters before. This is not a trivial task, as the numbers of observations quickly get very
small, even in a full sample of workers. Finally, we wish to get somewhat closer to getting an idea of
the cost-benefit ratio of such large scale investments into high-speed rail infrastructure. This part is,
however, still quite a way further down the track.
References


