WORKING PAPERS

N° 1385

June 2022

"Agglomeration Transport and Productivity: Evidence from Toulouse Metropolitan Area."

Marc Ivaldi, Emile Quinet, Celia Ruiz Meija

Toulouse School of Economics

Agglomeration, Transport and Productivity: Evidence from Toulouse Metropolitan Area.

Marc Ivaldi[1], Emile Quinet[2], Celia Ruiz Mejia[3]

June 2022

Abstract

The objective of this paper is to estimate the extent of agglomeration externalities taking into account the direct and indirect impacts of transport exposure on productivity. To do so, we take advantage of a rich data infrastructure that combines very fine georeferenced infra-municipality data on more than one million employees with detailed data on the public-transport and road networks of a typical European metropolitan area, namely the Toulouse Metropolitan Area (TMA). We recover the productivity effects of agglomeration and transport measures by the implementation and estimation of a wage determination model in two stages. The first stage assesses the importance of industrial concentration and employees' characteristics against true productivity differences across zones on the average of local industrial wages. The second stage explains local productivity differences on our local factors of interest: agglomeration and transport. Finally, and to have a full representation of transport impacts, we investigate the size of the indirect effect of transport exposure on productivity by its impact on the distribution of metropolitan employment. We exploit the panel nature of our data and apply instrumentation techniques to cope with the endogeneity of agglomeration and transport measures. Our results suggest that both agglomeration and transport exposure measures have a substantial and significant effect on local productivity. Indeed, when density of employment doubles, productivity increases by 1.6%. Further, the effects of transport exposure measures differ for the two modes considered, private vehicle and public transport. In both cases, a higher exposure to transport supply implies higher levels of employment an productivity.
Keywords: agglomeration economies; accessibility; transport exposure; public transport network; road network; productivity; transport infrastructure; density; cities; commuting costs; urban economics; transport economics.

1. Introduction

Urban economic activity tends to be highly concentrated, despite such concentration may involve high costs for inhabitants. It is well established in the economics literature that these agglomeration or concentration forms are driven by spatial externalities on productivity and are named agglomeration economies. In the same time, urban transport has also an important role on determining the productivity structure of a city. Firstly, transport may extend the scope of agglomeration externalities by reducing the interaction cost between economic agents placed in different locations. Secondly, transport may affect firms' and employees' location decisions by making some places more attractive than others. Also, it may disproportionally attract the high skilled labor force and high value-added firms, who have a comparative advantage to locate in better connected areas. This leads to local productivity gains in two ways, i.e. increased *quantity* and increased *quality* of the workforce.Thus, the objective of this paper is to estimate the extent of agglomeration externalities at the urban scale taking into account these impacts of transport on productivity.

[1] *Toulouse School of Economics, Toulouse, France*
[2] *Paris School of Economics, Paris, France*
[3] *Toulouse School of Economics, Toulouse, France*

To do so, and to our knowledge for the first time in this literature, we take advantage of a rich data infrastructure combining very fine georeferenced infra-municipality data on more than one million employees with detailed data on the public-transport and road networks of a typical European metropolitan area, namely the Toulouse Metropolitan Area (TMA). Thanks to the availability of these data, we develop an econometric method inspired by Combes *et al.* (2008) to decipher the complex relations between agglomeration effects, supply of transport and productivity. This aims at providing a more precise tool to an urban planner who wants to maximize city's wealth by identifying the disaggregated productivity impact of transport infrastructure.

This investigation brings together two important branches of the economics literature: the agglomeration economies literature and the literature investigating the effect of transport on several employment-related outcomes. Regarding the former, there is a long list of articles establishing the positive relationship between agglomeration measures and productivity using across cities variation (some examples are Ciccone, 2002; Glaeser and Mare, 2001; Combes *et al.*, 2010; De La Roca and Puga, 2012; D'Costa and Overman, 2014; Groot *et al.*, 2014; for a recent review, see Combes and Gobillon, 2015). Yet, agglomeration literature exploiting within-city variation is very scarce (Arzaghi and Henderson, 2008), and only few studies include transport-induced agglomeration measures (e.g. Graham, 2007; Holl, 2012; Gibbons *et al.*, 2019). Regarding the second branch, there are several articles investigating the relationship between transport infrastructure and the spatial distribution of economic activity (see Redding and Turner, 2015 for a survey). These papers have mainly focused on the effect of railways and roads on outcomes such as employment and population (e.g. Chandra and Thompson, 2000; Holl, 2004; Baum-Snow, 2007; Michaels, 2008; Atack *et al.*, 2010 ; Duranton and Turner, 2012; Gibbons *et al.*, 2012 ; Ghani *et al.*, 2016, Mayer and Trevien, 2017). To the best of our knowledge, only Gibbons *et al.*, (2019) study the effect of roads on labor productivity. We contribute to both literatures by simultaneously investigating the productivity effects of agglomeration and transport measures using variation across very small-scale geographical areas, and identifying the effect of both, public transport and private vehicle transport measures on the productivity structure of Toulouse metropolitan area.

Our econometric approach consists on the implementation a two-stage wage determination model. The first stage allows us to assess the importance of average employees' characteristics and industrial factors against those highlighting true productivity differences across zones. Formally, in this first stage we regress the local industrial average of the logarithm of individual wages on the local industrial average of time-varying employees' characteristics, zone-year fixed effects, industry fixed effects, and a set of variables related to the local characteristics of the specific industry, which account for *localization economies*. In the second stage, we use as dependent variable the vector of zone-year fixed effects estimated in the first stage. This vector is considered a local wage index net of employees' characteristics and industry effects. Formally, we regress the estimated zone-year fixed effects, on a set of time dummies, several agglomeration measures, which account for *urbanization economies*, and finally, several transport exposure measures. Furthermore, and to have a full representation of transport impacts, we investigate the effect of the later on local employment density, since transport exposure variation may affect firms' and employees' location decisions by making some places more attractive than others.

We measure transport exposure by two continuous indexes of accessibility calculated at a small geographical scale. The first one measures the *accessibility to employment* from a given origin to all potential destinations along the road and public transport network, and the second one measures the *reachability level* of a particular destination. Both indexes project transport improvements through changes in commuting times, and they represent the local and global exposure to transport of each location. In particular, *accessibility to employment* is measured by the weighted sum of inverse optimal travel times, where the weights are measures of employment density at destination. This index has been called in previous literature *index of accessibility* (e.g., El-Geneidy and Levinson, 2006; Vickerman *et al*., 1999), *population or market potential* (e.g., Harris, 1954), *effective density* (Graham, 2007), or *market access* (e.g., Donaldson and Hornbeck, 2016). The basic idea is that transport exposure, by reducing the interaction cost between economic agents placed in different locations, extend the geographic scope of agglomeration externalities on productivity. Likewise, *reachability levels* are computed by the average of optimal travel times from all potential origins along the road and public transport network. The index is computed so that reachability improvements are achieved when the index decrease, reflecting a decline in the average of optimal travel times to get to a particular destination.[4] Reachability levels have two potential effects on productivity. On the one hand, they may lead to productivity gains by locally attracting more and higher skilled employment activity. On the other hand, reachability improvements may affect productivity negatively when the last is measured by average wages, since companies may compensate employees for the higher commutes.

The main challenge to deal with is the endogeneity of agglomeration and transport exposure measures. Firstly, there is an omitted variable bias when firms' quality and/or employees' skills are sorted by unobserved local characteristics correlated with observed local factors. For instance, differences in local amenities may influence the location of both firms and workers: places with higher amenities may disproportionally attract the high skilled workforce and the higher-value added firms relying on skilled workers. When those amenities also determine agglomeration measures and the supply of transport, the estimates of both measures end up being biased upwards. Secondly, the reverse-causality between productivity and agglomeration, and between productivity and transport exposure, threatens identification. On one hand, employees are attracted by the higher wages of denser areas. On the other hand, the allocation decision of transport infrastructure is not at random. Indeed, the planner may intentionally decide to connect dynamic (deprived) areas, biasing the estimates upwards (downwards).

To solve for the omitted variable bias, we exploit the high geographical disaggregation of our data and we include a rich vector of local fixed effects into our estimation. Specifically, we combine fixed effects on EPCI areas - Établissement Public de Coopération Intercommunale - with Toulouse sector fixed-effects. The first corresponds to a French administrative classification that bundles a relatively small number of municipalities who exercise jointly most of their competencies. The second corresponds to the main six sectors comprising the city of Toulouse (Toulouse Centre, Rive Gauche, Toulouse Nord, Toulouse Est, Toulouse Sud-Est, Toulouse Ouest).

To solve for the reverse-causality bias, we perform an instrumental variables approach. We use several instruments proposed by previous literature, such as historical levels of agglomeration and transport exposure measures. Further, we add others, such as the physical-distance-based counterparts of the accessibility and reachability indexes.[5] We also use several indicators of the

[4] Even if the structure of the index may seem counterintuitive, it is very convenient when interpreting its coefficients, since they reflect the change on productivity when the time to get to a particular destination changes, i.e. the effect on productivity when a commuter is able to arrive to a particular destination in a shorter period of time.

[5] These two instruments are inspired by the peripherality instrument used by Combes *et al.* (2008), where they compute the average distance of an employment zone with respect to all the others.

distance to historical infrastructure, i.e. the distance to the historical public-transport plan during the years 1863 and 1957, the distance to two ancient roads, the roman and Cassini roads, and the distance to the 1870 railway network. We further deal with the non-randomness of infrastructure location by controlling for infrastructure-type measures in our regressions. By doing so, we are able to evaluate the accessibility and reachability effects orthogonal to the endogenous presence of infrastructure. In other words, we exploit the continuous spatial variation of our transport exposure measures, that is partly unrelated to the local presence of infrastructure, to identify transport effects on productivity. This allows us to identify transport exposure effects independently from the advantages or disadvantages of chosen locations.[6]

Our key findings suggest that both agglomeration and transport exposure measures have an important role in determining the internal productivity structure of a metropolitan area. Firstly, and according to our preferred estimates obtained after correcting for endogeneity, if local density is doubled, productivity increases by 1.6%. Secondly, the effects of transport exposure measures differ for the two modes considered, private vehicle and public transport. Regarding accessibility to employment by the road network, an average increase of 10% is associated with an increase of 0.49% in local productivity. If accessibility to employment is calculated through the public transport network, an average increase of 10% is associated with an average increase of 0.53% in local productivity levels. Regarding reachability measures, only reachability by the road network has a direct impact on productivity. Specifically, when the average of optimal commuting times to arrive to a particular destination by the road network decreases by 10%, local productivity increases by 1.6%. With respect to the indirect impact of reachability levels on productivity, through the attraction of employment concentrations, we find that when optimal travel times by public transport (private vehicle) decrease by 1%, productivity increases by 0.38% (0.6%). All these results together suggest that agglomeration externalities are locally happening at the metropolitan level, and that private vehicle and public transport exposure measures affects productivity levels directly and indirectly, by locally increasing the *quality* and the *quantity* of the workforce, respectively.

From our findings we also document the importance of controlling by unobserved local effects and applying instrumentation techniques to cope with the omitted variable and the reverse-causality biases present in OLS estimators. Indeed, when the econometric specification solve for those two sources of biases, agglomeration effects significantly decline, and both accessibility and reachability indexes increase by more than half. The decline in the size of agglomeration effects reveals that the skills-sorting issue remains an important matter at the metropolitan level. Also, the underestimation of transport effects are in line with a planner that tends to assign transport infrastructure to more secluded areas to stimulate its growth.

Finally, we contribute to the literature in several ways. First, we evaluate agglomeration externalities together with transport exposure effects on productivity at the urban level, using very fine georeferenced data. Second, we estimate the full-set of productivity effects of transport exposure: *indirect*, through the change in the interaction cost among economic agents and their location decisions, and *direct*, through the increased quality of the labor force by the boost in local amenities. In order to account for these effects, we wide our definition of transport exposure; we introduce a new variable accounting for the local reachability level *by* the transport network *to the location considered*, apart from the more traditional measures investigated in the literature – i.e. the access *to* the transport network and the access *by* the transport network *to other locations*.

[6] Specifically, we control by the local distance to the closest node of the public transport and road network from each zone centroid.

Finally, we combine public transport and road accessibility-type indexes, namely, accessibility by the road and public transport networks to economic concentrations and reachability by the road and public transport networks of a particular location. [7]

The paper is organized as follows. In the next section, we introduce the data sources and the metropolitan area of the study. Section 4 displays the theory supporting our econometric methodology, explained in Section 5. The results are presented in Section 6. Finally, Section 7 concludes and introduces some basis for future research.

2. Study Area and Data

2.1. Toulouse Metropolitan Area.

This paper estimates agglomeration and transport effects within the metropolitan area of Toulouse. The core urban unit of Toulouse Metropolitan Area, the city of Toulouse, is the fourth biggest in France and it has experienced an important population growth since the 1970s. Today, its transport networks are well-developed and dense, both the public transport and the road networks.

Regarding its public transport network, several modes operate in the city: two metro lines of a total of 28.2 kilometres, two tramway lines of 16.7 kilometres and 137 bus lines amounting 3809.8 kilometres in routes. [8, 9] Table 1 presents some statistics on the economic activity and public transport network evolution in Toulouse over the period 2004 – 2015.[10] Compared to 2004, the number of operating establishments increased by 115% in 2015. Also, the workforce increased by 81,093 employees and the transport network experienced an important expansion; Starting with a length of 1,443.46 kilometers, it reaches the 3,742.84 kilometers in 2015, with 689 new metro, tramway and bus stations. Further, and focusing on our two years of analysis, 2013 and 2015, the network has also experienced many changes: 67 new public transport stations and 100 more kilometers of public transport network, the total number of establishments increased by 52,155 and the labor force by 4,000 employees.

---

[7] Chatman and Nolan (2014) and Melo and Graham (2018) also include road together with transit measures. Although, their focus is on the aggregated metropolitan effects, ours is to investigate the economic dynamics happening within the city.

[8] It also counts with three lines of urban train and 20 lines of inter-urban train that passes through the metropolitan area and connects other parts of the region, i.e. Haute-Garonne.

[9] Developing such analysis within a metropolitan area could raise one concern: there may be relevant economic concentrations outside the metropolitan frontiers interacting with inner ones, which leads to a measurement error of some agglomeration measures, as for instance, employment accessibility of those zones lying in the limits of the metropolitan area. Thus, it is important to recall the French definition of a metropolitan area ("*aire urbaine"*) and its implications regarding this concern. A metropolitan area is defined by a core centre called urban unit ("*unité urbaine*") plus its commuter municipalities. First, the urban unit consists of a city ("*commune"*), or a group of cities, that have over 2000 inhabitants and contain a centre of dense construction, i.e. the buildings are spaced no more than 200 metres apart. In our case, this core centre is the city of Toulouse. Second, the commuter municipalities are all these municipalities whose at least 40% of the population is attracted by the economic activity of the central urban unit. Therefore, the metropolitan area of Toulouse is described by its central urban core, the city of Toulouse, and its economic influence on the surrounding commuter municipalities. This implies that the main local economic activity happens inside the limits Toulouse Metropolitan Area, where the majority of inhabitants works in either their own municipality, or in any other municipality of the metropolitan area.

[10] The data used to construct this table has been provided by Tisseo Collectivités (TC), the company providing and managing public transport services in Toulouse Metropolitan Area.

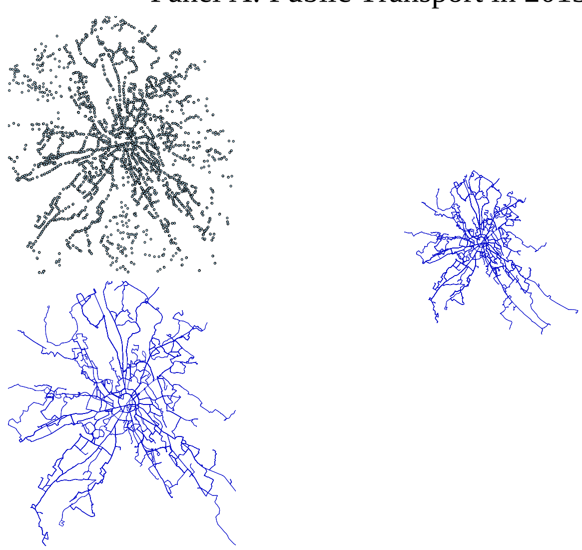
Table 1. Evolution of internal structure of Toulouse Metropolitan Area over the period 2004 – 2015.

Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Establishments	72994	76073	78602	83550	86413	89253	97150	105189	104797	119004	173573	156952
Employees	451055	451429	478778	512783	519088	452371	459064	474030	387430	528142	529036	532148
Public Transport Stations	3120	32766	3470	3563	3740	3740	3509	2984	3619	3742	3799	3809
Length of the Public Transport Network (km)	1443.46	1491.77	1867.52	1827.97	1942.55	1982.55	1769.76	1566.05	3257.85	3643.46	3884.62	3742.84

Regarding the structure of the public transport system, a new tramway line was opened, connecting the city center with the airport. These changes, together with many other small changes on already operating public transport services, creates enough variation in our transport exposure measures over the sample period to be able to identify productivity effects. Figure 1 displays the structure of the network in 2013 (Panel A) and in 2015 (Panel B). The network looks slightly denser in 2015, and in both years the network does not directly supply the whole space of the metropolitan area. It is located in the center and extends its arms towards the periphery.[11]

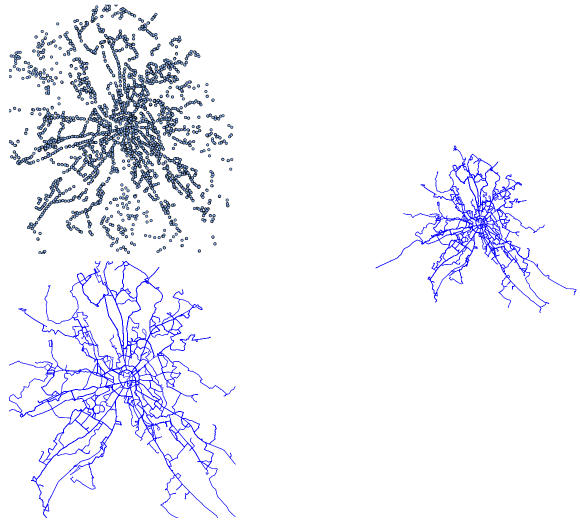
Figure 1: Public Transport Structure in 2013 and 2015, lines and stations.

Panel A: Public Transport in 2013.



Panel B: Public Transport in 2015.

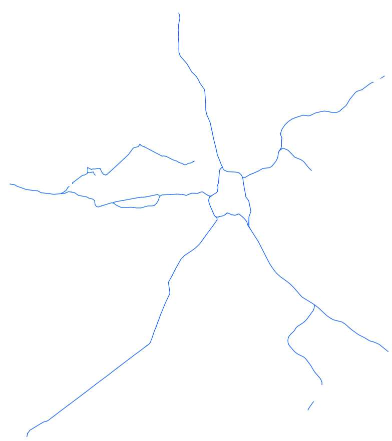
[11] This have important implications on the construction of public transport commuting times, which are combined with other means of transport for the zones not directly accessible by the network.



Source: Own elaboration from data provided by the *Agence d´Urbanisme et Amenagement de Toulouse (AUAT).*

Regarding the highway network, it is approximately 377 kilometers long and it is composed by highway and national roads. Figure 2 displays the representation of this network in the Metropolitan Area of Toulouse for the year 2013.

Figure 2: Highway Network in 2013, highways and national roads.



Source: Own elaboration from data ROUTE500, from the Institut National de l'Information Géographique et Forestière. [12]

[12] https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html

2.2. Data Sources.

We construct a unique database from three institutional sources: first, the Agency of Urbanism and Planning of Toulouse (Agence d'Urbanisme et d'Aménagement Toulouse, AUAT), second, Tisseo, the company in charge of the public transport service in the metropolitan area of Toulouse, and finally, the French Institute for Economic and Statistical Studies (Institut national de la statistique et des études économiques, INSEE).

In the first place, from the AUAT we obtain (ii) the exact location of each establishment located in the metropolitan area of Toulouse for the years 2013 and 2015, together with its legal number, i.e. the SIRET number, and (iii) the matrix of optimal trajectory travel times by different modes between each pair of zones for 2013 and 2015. The transport modes considered are private vehicle, public transport and bike.

Secondly, Tisseo facilitates the geographical placement of the Toulouse public transport network from 2004 to 2016, including the exact coordinates of every metro, bus and tram stations, together with the lines operating in each of these stations.

And finally, INSEE provides information collected from the Déclaration Annuelle des Données Sociales (DADS) survey, which gathers data from all French employers and self-employed. We have access to the establishments´ and employees´ files from 2004 to 2015. These contain, on one hand, a cross-section of public and private employees in manufacturing and services industries working in France for each year and, on the other hand, a cross-section of the establishments operating in France per year. In the employees file, personal data, establishment data, number of days worked, and several measures of earnings are incorporated. In the establishment one, information on the number of employees working, the start date of operations, together with the cessation date if it would be the case, is included.

We geographically merge the data coming from all different data sources. From the first and the second source, we use the available geographical coordinates of each public transport line and station to identify the disaggregated local supply of public transport services in each of the 867 zones. Similarly, we use the geo-coordinates of each establishment for identifying the study zone they operate in. Once public transport services and establishments are localized among the study space, we geographically identify employees´ workplace by exploiting information about the legal number of the establishment they work for, i.e. the SIRET number. This information, together with individual earning measures, is collected from INSEE data source. Finally, we end up with an original data set of very disaggregated and georeferenced data on more than one million employees working in Toulouse Metropolitan Area in 2013 and 2015.

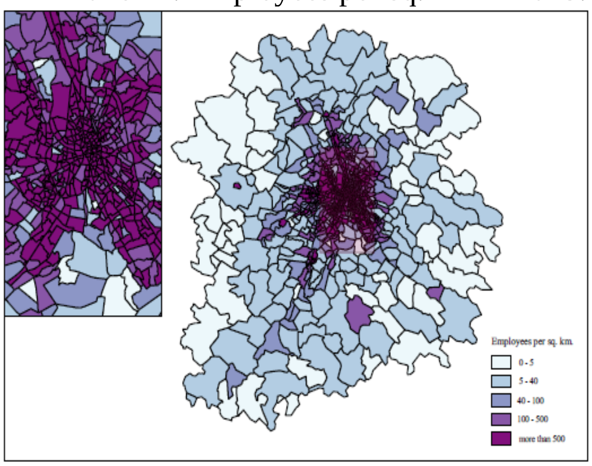
2.3. Geographical Distribution of the Study Space.

In this subsection, we provide maps with the geographical distribution of the main variables of the study. From figures 3 and 4, the geographical decomposition in 867 zones of the metropolitan area of Toulouse can be identified. These zones are outlined based on the local population level, established at a rate of approximately 2000 inhabitants.

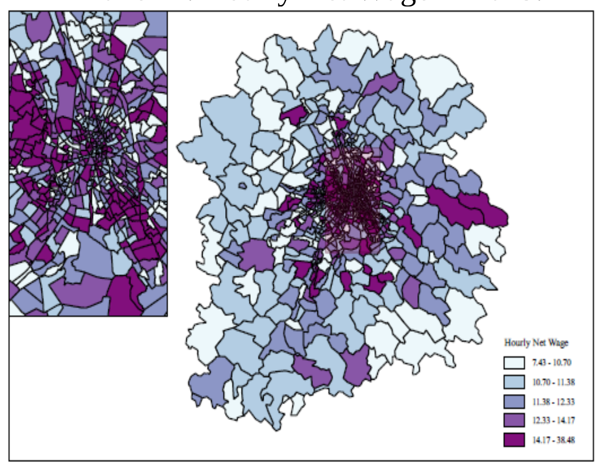
Further, Figure 3 displays the density and productivity distributions over the 867 zones of the study space for 2013 in Panels A and B, and for 2015 in Panels C and D. Density levels are mapped in Panel A and Panel C. They are measured by the number of employees per squared kilometer. In Panels B and D, productivity levels are mapped, and they are expressed by the average of individual hourly net wages per zone. From the joint inspection of the graphs, we conclude that the geographical distribution of productivity and density are very similar, hosting the highest levels around the center, and decreasing the farther you go from it. Therefore, zones with a higher density of employment also experience higher levels of productivity rates in both years. Furthermore, from the joint examination of Figures 1, 2, 3 and 4, those zones with a higher local supply of public transportation and road infrastructure, have higher levels of employment density. Moreover, average net hourly wages seem to follow the same pattern, although it is less clear than the case of employment density. Also, the effect of public transport expands in the direction of its constructed infrastructure, making higher the density and productivity levels on nearby zones.

Figure 3: Density and productivity distributions for 2013 and 2015.

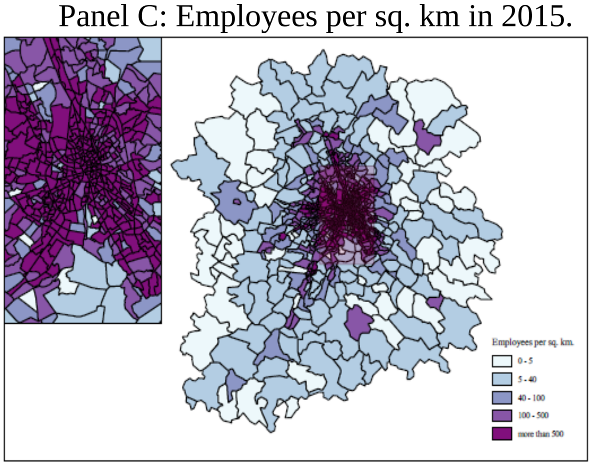
Panel A: Employees per sq. km in 2013.



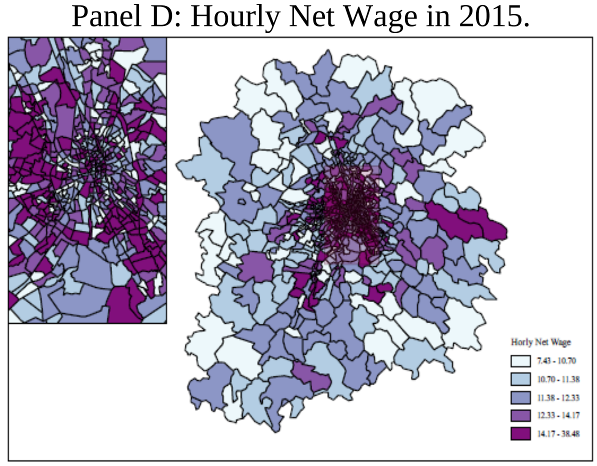
Panel B: Hourly Net Wage in 2013.



Panel C: Employees per sq. km in 2015.



Panel D: Hourly Net Wage in 2015.

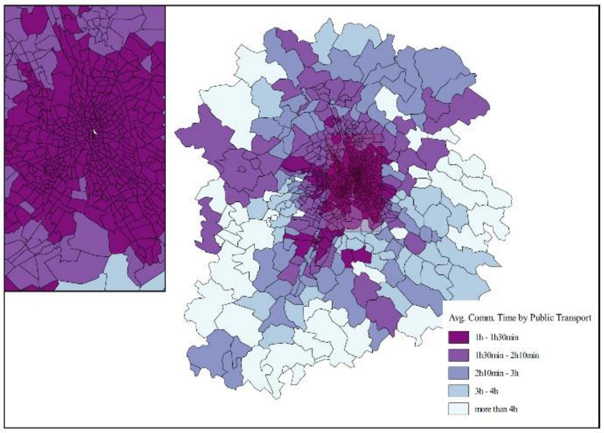


Source: Own elaboration from data of the survey *Declaration Annuelle de Donnees Sociales (DADS)*, provided by the Institut National de la Statistique et des Études Économiques *(INSEE)* .

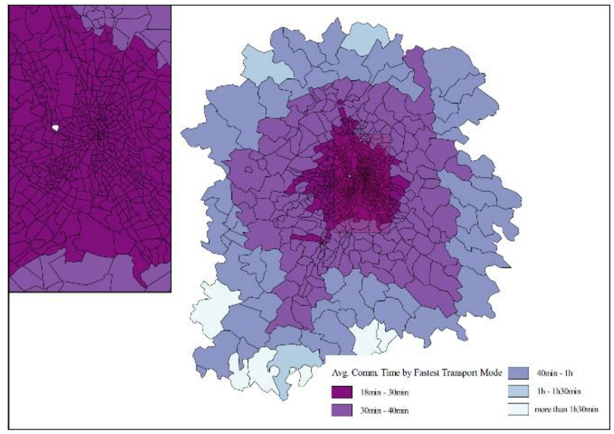
Finally, Figures 4 represents the distribution of local average optimal travel times by the two modes available, representing our reachability indexes.[13] From the inspection of the two figures, we confirm the negative correlation between average commuting times and employment density and productivity levels. Therefore, the more it takes to arrive to a zone, the less reachable it is, the lower is the level of productivity and employment density.

Figure 4: Average Travel Times Distribution for 2013 and 2015.

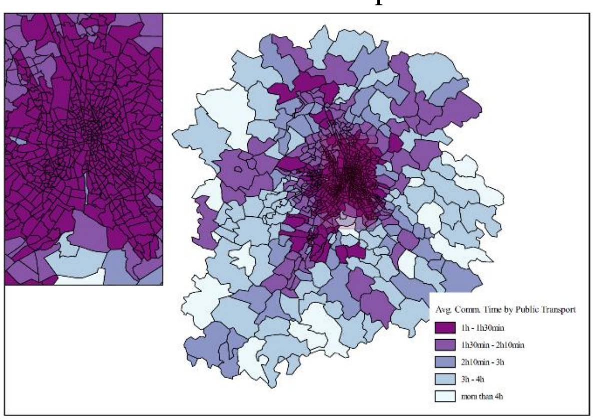
Panel A: Public Transport 2013



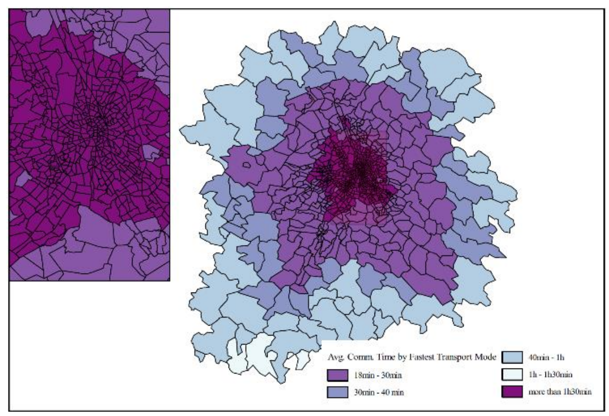
Panel B: Private Vehicle Mode 2013



Panel A: Public Transport 2015



Panel B: Private Vehicle Mode 2015



Source: Own elaboration from data on optimal travel times provided by the *Agence d´Urbanisme et Amenagement de Toulouse (AUAT).*

[13] The individual origin-destination commuting time by the public transport network has been computed by the sum of (i) the optimal access time to the transport network from the origin's centroid, (ii) the average in-vehicle time of the optimal connection, and (iii) the optimal egress time from the transport network to destination's centroid. The access and egress times are computed by the most convenient mode. If the zone lies into the space directly supplied by the transport network, access and egress times are computed by walking, given that this is the mode minimizing the time to access and egress the network. For zones sufficiently far away from the network, the times are computed by car. This way, any pair of zones is connected by the public transport network.

2.4. The model.

The objective of the model is to obtain an expression of individual wages with respect to local factors that can be estimated. With this purpose, the maximization problem of a competitive firm is carried out.[14] We take the profit of a competitive firm operating in zone *z* and industry *k* in year *t*:

$$\Pi_{z,k,t} = p_{z,k,t} q_{z,k,t} - \sum_{i \in (z,k,t)} w_{i,t} l_{i,t} - r_{z,k,t} k_{z,k,t} \tag{1}$$

where $p_{z,k,t}$ is the price of its output $q_{z,k,t}$, $w_{i,t}$ is the daily wage received by employee *i*, $l_{i,t}$ is the number of working days and, finally, $k_{z,k,t}$ represents the quantity of other production factors used, being $r_{z,k,t}$ its price. The production function is assumed to follow a Cobb-Douglas function in labor and in all the other production factors:

$$q_{z,k,t} = A_{z,k,t} \left(\sum_{i \in (z,k,t)} s_{i,t} l_{i,t} \right)^{b} \left(k_{z,k,t} \right)^{1-b} \tag{2}$$

where the coefficient *b* is such that $0 < b \leq 1$, $s_{i,t}$ denotes the skills of employee *i* in year *t* and $A_{z,k,t}$ is the total factor productivity. At the competitive equilibrium, employee *i*, located in zone *z* and operating in industry *k* at year *t* receives a wage equal to her marginal product:

$$w_{i,t} = b p_{z,k,t} A_{z,k,t} \times \left(\frac{k_{z,k,t}}{\sum_{i \in (z,k,t)} s_{i,t} l_{i,t}} \right)^{1-b} s_{i,t} \tag{3}$$

Using the first-order condition for profit maximization with respect to the other factors and inserting it in Equation (3) yields:

$$\begin{aligned} w_{i,t} &= b(1-b)^{\frac{1-b}{b}} \times \left(p_{z,k,t} \frac{A_{z,k,t}}{(r_{z,k,t})^{1-b}} \right)^{\frac{1}{b}} s_{i,t} \\ &= B_{z,k,t} s_{i,t} \end{aligned} \tag{4}$$

From the expression above, wages are defined by two types of components. First, a location-industry-specific component, $B_{z,k,t}$, and second, an individual-based component, s_i. Therefore, wage differences across employees can reflect differences in their characteristics and/or individual skills, or alternatively, they can reflect productivity differences across locations and industries. Further, this last source of wage variability can operate throughout three different channels: total

---

[14] The assumption of a competitive wage-setting mechanism may seem too restrictive. Still, in any imperfect competition framework, where the wage is a mark-up on marginal productivity, similar results would be obtained; given the use of a log specification, this mark-up would enter the constant or the industry fixed effects if such mark-ups vary between industries but not between areas. In France, there is some empirical support for the competitive/fixed-mark-up assumption (see Abowd et al., 1999).

factor productivity, $A_{z,k,t}$, the price of the output, $p_{z,k,t}$ and, finally, the price of non-labor inputs, $r_{z,k,t}$. The first objective of the following section is to build up an estimation equation for the expression of wages defined by Equation (4), and to estimate the effects of local factors on productivity. Hereinafter, we refer to this analysis as *productivity analysis*, and estimates, inter alia, the direct effect on productivity of transport exposure measures. The second objective of the following section lies in the development of a *density analysis*, where we recover the indirect effects of transport exposure measures on productivity.

3. Econometric Methodology

3.1. Productivity Analysis.

The econometric methodology performed for the productivity analysis is based on a two-stage approach, inspired by Combes *et al.* (2008). Essentially, we perform a two-stage estimation approach because of two reasons. First, to properly take into account correlations between zone and industry variables and error terms at the zone level. And second, to account for zone-specific error terms when computing the standard errors for our coefficients of interest. In a single-stage estimation, the computation of the variance of local shocks is not possible, i.e. the error term of the second stage. Moreover, this variance has to be ignored when computing the covariance matrix of estimators. As shown by Moulton (1990), this creates large biases in the standard errors for the estimated coefficients of aggregate explanatory variables.

3.1.1. Micro-econometric Specification for the Wage Equation: First Stage Specification.

We take Equation (4) into the data. First, we apply logarithms to both sides of the equation:

$$\log(w_{i,t}) = \log(B_{z,k,t}) + \log(s_{i,t}) \tag{5}$$

Therefore, two specifications are needed: one for the logarithm of the individual-specific term, $\log(s_{i,t})$, and another for the one of the local-industry-productivity term, $\log(B_{z,k,t})$. Assume first that the logarithm of the individual-specific term of worker *i* at time *t* is explained by:

$$\log(s_{i,t}) = X_{i,t}a + g_i + e_{i,t} \tag{6}$$

where $X_{i,t}$ is a vector of employee characteristics, g_i is an *i.i.d.* measurement error across employees, and $e_{i,t}$ is the time-varying unobserved component of individual skills.

Second, assume that the logarithm of the term reflecting productivity differences across locations and industries, $\log(B_{z,k,t})$, follows the subsequent expression:

$$\log(B_{z,k,t}) = d_{z,t} + d_{k,t} + X_{z,k,t}b_k \tag{7}$$

where $\delta_{k,t}$ is the time-varying industry-specific effect and it is defined by a vector of dummies taking value one if employee *i* operates in industry *k* at time *t*, and zero otherwise; $\delta_{z,t}$ is the time-varying location-specific effect and it corresponds to a vector of dummies taking value one if employee *i* is located in zone *z* at time *t*, and zero otherwise, and finally, $X_{z,k,t}$ is the group of local characteristics of industry *k* at time *t*, measuring local interactions between employees at the industry level, and β_k its associated vector of coefficients.

Summing the logarithm of the individual-specific term and the logarithm of the local-industry-productivity term – i.e. Equation (6) and Equation (7), the econometric specification for the wage equation is:

$$\begin{aligned} \log(w_{i,t}) &= \log(s_{i,t}) + \log(B_{z,k,t}) \\ &= \gamma_i + \delta_{z,t} + \delta_{k,t} + X_{z,k,t}\beta_k + X_{i,t}\alpha + \varepsilon_{i,t} \end{aligned} \tag{8}$$

The direct estimation of Eq. (8) is not possible due to data limitations. Particularly, because we do not follow employees over time. Therefore, the systematic individual effect, γ_i, is not identifiable. For this reason, we aggregate individual wages by industry and zone to define our final first stage specification:

$$\begin{aligned} w_{z,k,t} &= \frac{1}{N_{z,k,t}} \sum_{i\in(z,k,t)} \log(w_{i,t}) \\ &= \delta_{z,t} + \delta_{k,t} + X_{z,k,t}\beta_k + \left(\frac{1}{N_{z,k,t}} \sum_{i\in(z,k,t)} X_{i,t} \right)\alpha + \nu_{z,k,t} \\ &= \delta_{z,t} + \delta_{k,t} + X_{z,k,t}\beta_k + I_{z,k,t}\alpha + \nu_{z,k,t} \end{aligned} \tag{9}$$

Where $I_{z,k,t}$ is the vector of the average of individual time-varying characteristics, $X_{i,t}$, of the $N_{z,k,t}$ employees working in industry *k* and located in zone *z* at time *t*, and $\nu_{z,k,t} = \frac{1}{N_{z,k,t}} \sum_{i\in(z,k,t)} (\gamma_i + \varepsilon_{i,t})$ is the new error term.

As discuss in Combes *et al.* (2008), some variables in equation (9) exhibit interpretation issues. Specifically, $X_{z,k,t}\beta_k$ and $I_{z,k,t}\alpha$. These vectors of variables include an industry component non-dependent on location. For instance, the reader may think about an industry employing more female employees. The average proportion of women in this industry should be systematically higher in almost all locations. Therefore, and assuming wages are lower for women on average, this industry will pay lower wages, all else equal. This systematic industry component should be indeed captured by the "industry effect", $\delta_{k,t}$. In order to properly account for these non-location-dependent industry characteristics, we center the variables included in vectors $X_{z,k,t}$ and $I_{z,k,t}$ around their industrial metropolitan means. This way, the systematic industry components in $X_{z,k,t}\beta_k$ and $I_{z,k,t}\alpha$ are added to the industry fixed effect to build the "total industry effect".

We also assume that the time trend is the same for all industries. Therefore, the total industry-year effect can be decomposed into the sum of an industry fixed effect and a year fixed effect. Further, the last one is normalized to zero for all years, since time evolution is absorbed by the zone-year fixed effect, $\delta_{z,t}$[15]. The final specification is:

$$w_{z,k,t} = \delta_{z,t} + \delta_k + \tilde{X}_{z,k,t}\beta_k + \tilde{I}_{z,k,t}\alpha + \varsigma_{z,k,t} \qquad (10)$$

where $\tilde{X}_{z,k,t}$ is the centered vector of local characteristics of industry *k* at time *t* and $\tilde{I}_{z,k,t}$ is the centred vector of local average time-varying individual characteristics of employees working in industry *k* at time *t.*

Recapitulating, the first stage of our econometric methodology estimates local industrial productivity, measured by the *average of the logarithm of net hourly wages* of employees working in industry *k* and zone *z* at time *t*, as a function of (i) a vector of average observed employees characteristics – i.e. the *average age*, the *average of the squared ages*, the *proportion of woman* and the *proportion of full-timers,* (ii) an *industry* fixed effect, (iii) a *zone-year* fixed effect and (iv) a long vector of local industrial characteristics: the logarithm of the industrial share of local employment, henceforth the *industrial specialization index,* the *log of the number of establishments* locally operating in the industry, and finally, the *log of the proportion of professionals* working locally in the industry, used as a proxy of the level of education locally in the industry. The effects of this last vector of variables on productivity aims at identifying the productivity gains emerging from employees' interactions within the same industry (*localization economies*).

The derived specification allows us to identify separately the effects on productivity of "employees" (observed employees' characteristics), versus "industries" (industry-specific effects and local industrial interactions), versus "places" (location-specific effects). The unbiased and consistent estimators of the location-specific effects on productivity provided by this first stage, $\hat{\delta}_{z,t}$, corresponds to the dependent variable of the second stage. This vector is considered as a local

---

[15] Formally, as specified by Combes *et al.* (2008), the effects of local industry characteristics, $X_{z,k,t}\beta_k$, can be decomposed into a non-location-dependent effect, $X_{\circ,k,t}\beta_k$, and a location-dependent effect net of metropolitan effects, $\tilde{X}_{z,k,t}\beta_k \equiv X_{z,k,t}\beta_k - X_{\bullet,k,t}\beta_k = (X_{z,k,t} - X_{\bullet,k,t})\beta_k$. Respectively, $X_{z,k,t}$ is the location dependent vector of local industry characteristics and $X_{\circ,k,t}$ is the average of those industry characteristics weighted by the local number of industrial employees, i.e. $X_{\circ,k,t} = \frac{1}{N_{k,t}} \sum_{z \in (k,t)} N_{z,k,t} X_{z,k,t}$, where $N_{z,k,t}$ is the number of employees working in industry *k* at time *t* in zone *z*, and $N_{k,t}$ is the number of employees working in industry *k* at time *t* in the full territory. Likewise, the effects of local average time-varying employees characteristics, $I_{z,k,t}\alpha$, may be decomposed into a non-location-dependent effect, $I_{\circ,k,t}\beta_k$, and a location-dependent effect net of metropolitan effect $\tilde{I}_{z,k,t}\alpha \equiv I_{z,k,t}\alpha - I_{\bullet,k,t}\alpha = (I_{z,k,t} - I_{\bullet,k,t})\alpha$, where $I_{\circ,k,t} = \frac{1}{N_{k,t}} \sum_{z \in (k,t)} N_{z,k,t} I_{z,k,t}$. The total industry effect is thus $\delta_k + X_{\circ,k,t}\beta_k + I_{\circ,k,t}\alpha = \delta_k + \tau_t = \delta_k$, once the year fixed effect has been normalized to zero, $\tau_t = 0$, since τ_t and $\delta_{z,t}$ cannot be identified separately in the same estimation equation.

wage index once controlled for employees' characteristics and industry effects – or *localization effects*.

3.1.2. Econometric Specification of the Second Stage.

The objective of the second stage is to assess the relative importance of agglomeration measures, which shape *urbanization economies*, and transport endowments or amenities, on our measure for local productivity – i.e. the estimation of the location-specific term of the wage equation (Eq. 10). The econometric specification assumed is:

$$\hat{d}_{z,t} = k_t + k_c + E_{z,t}\rho_1 + X_{z,t}\rho_2 + Controls_{z,t} + u_{z,t} \tag{11}$$

Where k_t are time fixed-effects, k_c are sector fixed-effects for the Toulouse municipality and EPCI fixed-effects for the rest of municipalities, $E_{z,t}$ represents the vector of observed local transport endowments or amenities – i.e. transport exposure measures, and ρ_1 its vector of coefficients, $X_{z,t}$ is the vector of location-specific variables representing agglomeration characteristics, and ρ_2 its vector of coefficients, $Controls_{z,t}$ represent a vector of control variables relevant for productivity and correlated with our regressors of interest, and finally, $v_{z,t}$ is the error term reflecting local shocks assumed to be *i.i.d.* across zones and periods.

Regarding our agglomeration measures, $X_{z,t}$, the main one is the logarithm of the density of local employment, $density_{z,t} = \log\left[\frac{emp_{z,t}}{area_z}\right]$, where $emp_{z,t}$ represents the number of employees working in zone *z* at time *t*, and $area_z$ is the area of the zone in squared kilometers. Additionally, some of the specifications include an alternative agglomeration control traditionally used by the literature, i.e. industrial diversity. Jacobs (1969) made popular the intuition that industrial diversity could be favorable for productivity as there could be cross-fertilization of ideas and transmission of innovations between industries. This has been for instance formalized by Duranton and Puga (2001) and many measures of diversity have been proposed. The most used is the inverse of the Herfindahl index constructed from the shares of industrial employment within the local workforce, $\text{diversity}_{z,t} = \frac{1}{HHI_{z,t}} = \frac{1}{\sum_k \left(\frac{emp_{z,k,t}}{emp_{z,t}}\right)^2}$, where $emp_{z,k,t}$ is the number of industry *k* employees working in zone *z* at time *t*. Since specialization is also introduced in the specification, interpretation is easier if the own industry is removed from the computation of *diversity*. In that case, whereas specialization relates to the role of the industrial local share, diversity relates to the role of the distribution of employment over all other industries, and the two indices clearly capture two different types of mechanisms. [16]

[16] Since the Herfindahl index has the inconvenient property of being largely influenced by the number of units from which the index is computed, the number of industries in our case, we introduce the number of local active industries in our specification, as proposed by Combes *et al.* (2004).

Regarding transport exposure measures, which account for local transport endowments or amenities, two continuous indexes of accessibility are included: a reachability index, measured by the inverse of the average of optimal travel times, and a transport-based accessibility to employment measure. Both are computed for the two transport modes available – public transport and private vehicle. The former follows the following formula: $\text{Reachability Index}_{z,m,t} = \frac{\sum_{j \neq z} d_{jz(m)}}{J}$ where J is the total number of zones except zone z, and $d_{jz,t(m)}$ is the commuting time from zone j to zone z by mode m at time t. The second index captures the local accessibility level to economic concentrations. This measure is the logarithm of the weighted average of optimal travel times, where the weights are levels of neighboring employment density, $\text{Accessibility to Employment}_{z,m,t} = \sum_{j \neq z} \left(\frac{dens_j}{d_{zj,t(m)}} \right)$, where $dens_j$ is the density of employment in zone j.

Certainly, there are more ways to measure transport exposure. Most articles have opted for the use of infrastructure-type measures. For instance, articles analysing road-based transport improvements have used the local length of roads (Melo *et al*., 2010), whether an area is crossed by a highway (e.g. Chandra and Thompson, 2000; Michaels, 2008; Faber, 2014; Duranton and Turner, 2011), the distance to the closest highway (Baum-Snow, 2007; Ghani *et al*., 2016; Holl, 2016), the local number of radial roads from a city centre (Baum- Snow, 2007, 2010; Baum-Snow *et al.*, 2016a), or the local public expenditure on roads (Fernald, 1999). Regarding articles studying public-transport improvements, they use distance to infrastructure (Gibbons and Machin, 2005; Billings, 2011; Baum-Snow and Kahn, 2000; Glaeser *et al.*, 2008) or the local number of public-transport stations and lines (Gonzalez-Navarro and Turner, 2016).

However, and in the context of our study, both networks are already well-developed and dense, leading to a very small variation on infrastructure-type measures from one year to the other. Indeed, in 2013, the public transport network was 3643.46 kilometres long, and it did not experience a large physical expansion by the end of 2015: it increased by 0.027% to 3742.84 kilometres, with the construction of a second tram line. Regarding the road network, it is composed by 377 kilometres of highway and national roads and it experience no changes from the initial to the final date of this study. Nonetheless, we find much more variation on our accessibility-type indexes which measure transport exposure. Particularly, accessibility to employment by the public transport and road network change on average 11.93% and 14.25%, and reachability levels do so by 5.91% and 26.58%, respectively. This higher level of accessibility and reachability variation is explained by the multiple sources of travel time adjustments, where the construction of transport infrastructure plays a limited role. [17]

The second reason why we use accessibility-type measures is the serious endogeneity concerns raised by infrastructure-type measures and caused by the non-randomness of their location. Therefore, instead of using infrastructure-type measures as study indexes, we control for them in our analysis. By doing so, we are able to compute the accessibility and reachability effects orthogonal to the endogenous presence of infrastructure. In essence, we exploit the continuous variation over the space of our transport exposure measures that is partly unrelated to the local

[17] Even if the road network does not physically change from one year of the study to the other, its optimal travel times distribution changes as a consequence of several factors. For instance, the road congestion level. Congestion of roads may be affected by different determinants as changes in the public transport usage given a change in its supply, changes in individual preferences for private vehicles, changes in individual concerns about the environment, etc. Therefore, it is not surprising that travel times changes when the network has remained the same.

presence of infrastructure. This allows to identify transport exposure effects separately from the advantages or disadvantages of chosen locations. We control for the distance to the closest station and the distance to the closest highway from each zone centroid.

Together with these infrastructure-type measures, included in the vector $Controls_{z,t}$, we include an additional group of controls that are relevant for productivity and are correlated with agglomeration and transport exposure measures. These variables account for the urban structure of the metropolitan area. They control for any correlation between agglomeration and/or transport exposure levels, and urbanity. These two additional variables are a dummy variable indicating if the IRIS code belongs to an *urban* or a *rural* municipality and a continuous variable measuring the driving time from each zone centroid to the Central Business District (CBD) of Toulouse.

3.1.3. Estimation Issues.

There are several issues with respect to the estimation of equations (10) and (11). Manifestly, the true value of the dependent variable of the second stage, $d_{z,t}$, is unknown. Therefore, we use the estimators $\hat{d}_{z,t}$ obtained in the first stage. Yet, if there is any correlation between our local characteristics of interest, i.e. agglomeration and transport exposure measures, and the average of local skills, included in the error term of the first stage, these estimators would be biased and inconsistent. Certainly, this issue becomes more relevant the higher the spatial aggregation of the data. For instance, when using cities as spatial units, entire cities are constrained to one unique unit of observation. This poses relevant concerns for unobserved variable bias, since there are variables persistently and simultaneously determining the quality of the urban workforce together with transport and agglomeration effects.

Still, this skill-sorting scenario, even if less troublesome, it can still happen at the infra-city level. Especially within big cities, that host a wide variety of neighborhoods characterized by an heterogeneous composition of local amenities. Therefore, in the estimation of the second stage we include sector effects for the six sectors comprising the municipality of Toulouse (Toulouse Centre, Rive Gauche, Toulouse Nord, Toulouse Est, Toulouse Sud-Est, Toulouse Ouest) and EPCI effects for the rest of smaller municipalities. Through this strategy, we are able to independently identify agglomeration and transport exposure effects on local productivity apart from potential employees´ sorting at the sector level within Toulouse, and at EPCI levels for the rest of the metropolitan area.

We are also cognizant of the reverse-causality between local productivity rates and observed transport exposure and agglomeration measures. Employees, and more likely those with higher skills, are attracted by the higher productivity of denser areas. Further, the allocation decision of transport infrastructure is not random, the planner may intend to connect dynamic (deprived) areas, biasing the estimates upwards (downwards). The direction of the bias is unclear and therefore endogeneity becomes a serious concern.

To deal with this issue, we perform an instrumental variables approach. We use traditional instruments from the literature, i.e. past levels of agglomeration and transport exposure measures together with physical-distance-based counterparts of our accessibility and reachability indexes. Regarding the first group of instruments, we use 2004 levels of employment density, 2004 predicted levels of accessibility to employment and local reachability, and the distance from every zone centroid to several historical networks: the distance to the closest roman and Cassini roads, and the distance to the closest line of the 1870 railway network. Regarding the second group of instruments, we compute our accessibility to employment and reachability indexes using simple straight line distances in kilometers.[18] The rationale behind is the following: geographical proximity alone should not have an independent impact on productivity as physical proximity does not affect productivity independently, it is instead the interaction between economic agents what generates productivity gains. Therefore, this interaction would not be possible without a network connecting the nodes. Therefore, we use as instruments synthetic measures of accessibility to employment and reachability that substitute travel times by simple physical distances between each pair of nodes.

3.2. Density Analysis.

In order to have a full representation of transport effects, we perform a density analysis by which we estimate the effects of transport exposure on local density of employment.

Transport exposure levels may make some places more attractive than others, affecting employment location decisions. Employees may prefer to work in better connected areas, where commuting is relatively shorter. Therefore, it is interesting to investigate the effect of our *reachability index* on employment location.

Further, by estimating the effect of transport exposure levels on employment density, we are able to obtain the indirect effect of transport exposure on productivity through employment location. By interacting the *reachability index* coefficient from the density analysis with the *density* coefficient from the productivity analysis, we recover the indirect effect of transport on productivity through agglomeration and employment location.

Certainly, the estimation of transport exposure effects on employment density arises endogeneity issues. The reverse-causality between transport exposure and employment density, caused by the non-randomness of transport location, may bias the estimates both, upwards and/or downwards. If the planner decides to invest on transport within deprived and less dense areas to boost their growth, the estimate of transport exposure on density is bias downwards. If the planner decides to better connect dense and dynamic areas, the estimate is bias upwards. To solve for this bias, we follow the same method as in the productivity analysis. Firstly, we include variables measuring the local presence of physical transport infrastructure. This way, we control for the endogeneity of local infrastructure placement. Secondly, we perform an instrumental variables approach where the instruments are a sub-group of the ones proposed in the productivity analysis: levels of past reachability, the physical-distance counterpart of this index, and several indicators of the distance to historical infrastructure.

4. Results

4.1. Productivity Analysis Results.

The results correspond to the estimation of the two-stage econometric specification presented in Section 5, i.e. the estimation of equations (10) and (11).

Table 2. First Stage: Estimation Results.

Stage I Results - Dependent Variable: Average of Log Wages of Industry K at Zone Z

[18] These instruments are inspired by the *peripherality index* introduced by Combes *et al.* (2008) to instrument current levels of employment accessibility.

	Coefficient	St. Deviation
Specialization Agriculture	0.386**	(0.186)
Specialization Manufacture	0.261***	(0.035)
Specialization Construction	0.160***	(0.030)
Specialization Commerce and Transport	0.097***	(0.020)
Specialization Communication	0.010	(0.020)
Specialization Education and Health	0.048***	(0.016)
Num. Establishments Agriculture	-0.025	(0.030)
Num. Establishments Manufacture	0.002**	(0.001)
Num. Establishments Construction	0.003***	(0.003)
Num. Establishments Commerce and Transport	0.000	(0.0002)
Num. Establishments Communication	-0.0002	(0.0002)
Num. Establishments Education and Health	-0.0008	(0.0008)
Prop. Professionals Agriculture	0.472***	(0.078)
Prop. Professionals Manufacture	0.639***	(0.035)
Prop. Professionals Construction	0.805***	(0.038)
Prop. Professionals Commerce and Transport	0.785***	(0.042)
Prop. Professionals Communication	0.546***	(0.023)
Prop. Professionals Education and Health	0.344***	(0.023)
Mean Age	0.040***	(0.002)
Mean Square Age	-0.0003***	(0.00003)
Prop. Woman	-0.076***	(0.010)
Prop. Full-timer	0.165***	(0.009)
Industry-Time Effects	Yes	
Zone-Time Effects	Yes	
Observations	7,329	
Adjusted R^2	0.562	
F Statistic	166.353	

Firstly, Table 2 presents the estimation results for Equation (10), i.e. the first stage. The dependent variable is the local industrial average of individual log wages. Regarding the regressors, all variables referring to individual characteristics are statistically significant and has the expected signs. The coefficients for the mean age and the mean of the squared ages in the industry are 0.040 and -0.0003, respectively. Both significant at 99%. Furthermore, the proportion of woman in the industry decreases average industrial log wages by 0.076, and the proportion of full-timers increases it by 0.165, respectively. On the other hand, even though they are not displayed here, none of the industry indicators [19] are significant, contrary to the high significance of local industry characteristics. This points out to the hypothesis that variation on individual productivity is not explained by the actual industry the employee is working for, but by the local characteristics of this industry. This hypothesis supports the existence of *localization economies*. The elasticity of industry specialization on productivity is constantly positive and significant

[19] The industry classification is the following: AZ refers to agriculture, silviculture and fishing, BE to manufacturing and extractive industries, FZ to construction, GI to commerce, transport and restauration, JU to information and communication, financing activities, and scientific, medical and social aid activities, and finally, OQ refers to the public administration and teaching activities.

except for communication, with an average of 0.1603. The log number of establishments in the same industry is positive for Manufacturing, Construction, Commerce and Transport. Whereas, it is negative for Agriculture, Communication and Education and Health. The average is 0.008. And finally, the proportion of professionals locally in the industry is very positive and significant for all industries, with an average effect of 0.5985. All these coefficients are of the expected sign and consistent with previous literature in localization economies.

Focusing now on the second stage, the results are presented in Table 3. The dependent variable is the estimated zone-year fixed effect from the first stage. The instrumental variables (IV) estimations in columns (4) and (5) apply the nine instruments proposed in this paper, i.e. past density levels, the synthetic index for reachability and accessibility to employment, both past predicted levels of transport exposure measures, the distance to the closest railway line in 1870, the distance to the closest Cassini road, the distance to the closest Roman road, and the distance to the closest public transport station from 1863 to 1957. Table 3 also displays the over-identification tests computed by the Sargan method. Their values of 0.517 and 0.493 allows us to reject the null hypothesis of having at least one endogenous instrument. Therefore, we reject the alternative hypothesis of all instruments being not valid. The size of the first stage F-statistics displayed in Table 3 and in all specifications of Table 4, corresponding to the first stages of all endogenous regressors on the instruments, allows us to conclude that the instruments are relevant for all endogenous regressors.

From Table 3, we identify several interesting findings. Firstly, employment density has a positive and significant effect on local productivity consistent among all specifications. Before instrumenting (columns 1 to 3), the estimate is between 0.024 and 0.025, depending on the controls and fixed-effects included in the specification. After instrumenting (columns 3 and 4), it decreases to 0.023 when local fixed-effects are not included, and to 0.016 when fixed-effects on Toulouse sectors and EPCI areas are included. The decline in the size of the coefficient reveals the importance of controlling by the skills-sorting and the reverse causality issues. Therefore, according to our preferred specification in column (5), if local density is doubled, productivity increases by 1.6%. This estimate is in the lower bound of the estimates found by previous literature, where transport exposure measures were not included.

Secondly, transport exposure measures have different effects on productivity depending on the mode considered, i.e. public transport or private vehicle. On one hand, both accessibility to employment measures have a positive and significant effect on local productivity. Specifically, when accessibility to employment by the road network increases by 10%, local productivity does by 0.49%. Further, when accessibility to employment by the public transport network increases by 10%, productivity does by 0.56%. Both estimates are lower (and in the case of public transport measures non-significant) before the introduction of Toulouse sectors and EPCI effects. This suggests a better identification of accessibility effects when exploiting the infra-metropolitan (within) variation of accessibility rates. In other words, while we find lower or non-significant accessibility effects when exploiting variation in the average accessibility from one IRIS code to the next, we find positive and significant accessibility effects when exploiting the variation in accessibility within each area (Toulouse sectors and EPCI areas) over time. This is because by introducing local fixed-effects into the estimation, we remove the pernicious effect of omitted variable by comparing IRIS codes with similar unobserved characteristics.

Table 3. Productivity Results: Public Transport and Private Vehicle Measures.

Stage II Results: Public Transport and Private Vehicle Measures.

	Dependent Variable: First Stage Zone-Year Effects				
	OLS	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)
log(Density)	0.025***	0.025***	0.024***	0.023***	0.016***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)
log(Acc. to Emp. by Private Vehicle)		0.011***	0.010**	0.023*	0.049***
		(0.003)	(0.004)	(0.013)	(0.015)
log(Acc. to Emp. by Public Transport)		0.001	-0.001	0.014	0.056**
		(0.003)	(0.003)	(0.017)	(0.021)
log(Reachability by Private Vehicle)		-0.004	-0.038	0.111	-0.165**
		(0.022)	(0.029)	(0.193)	(0.077)
log(Reachability by Public Transport)		-0.001	-0.011	0.120	0.162
		(0.009)	(0.011)	(0.125)	(0.153)
Urban Controls	No	Yes	Yes	Yes	Yes
Physical Infrastructure Controls	No	Yes	Yes	Yes	Yes
Agglomeration Controls	No	Yes	Yes	Yes	Yes
Time Effects	No	Yes	Yes	Yes	Yes
EPCI Effects	No	No	Yes	No	Yes
Sector Effects	No	No	Yes	No	Yes
Instruments:					
Past Levels of Employment Density				x	x
Past Levels of Reachability				x	x
Past Levels of Acc. to Emp.				x	x
Physical Reachability (Peripherality)				x	x
Distance to Roman Roads				x	x
Distance to PT Network from 1863 to 1957				x	x
Distance to Railway lines in 1870				x	x
Distance to Casini Roads				x	x
Observations	1,723	1,663	1,663	1,629	1,629
Adjusted R^2	0.174	0.166	0.173	0.052	0.017
Over-identification p-value				0.517	0.493
First Stage Statistics for:					
log(Density)				85.752	78.059
log(Acc. to Emp. by Private Vehicle)				160.508	59.720
log(Acc. to Emp. by Public Transport)				41.955	18.418
log(Reachability by Private Vehicle)				38.661	43.511
log(Reachability by Public Transport)				15.942	25.318

On the other hand, only local reachability levels by the road network have a significant impact on productivity.[20] Specifically, if the average time it takes to arrive to a particular destination by the road network decreases by 10%, local productivity increases by 1.65%. Again, road-reachability effects are not significant before instrumenting for their endogeneity and before controlling by Toulouse sectors and EPCI effects. This result is in line with the accessibility to employment results, pointing out at a better identification coming from the comparison of zones with similar unobserved characteristics. Further, it suggests that reachability improvements are associated with a larger increase in productivity in a randomly chosen IRIS code than in a selected zone under the prevailing political process. Therefore, this result is consistent with a planner that tends to assign transport infrastructure to more secluded areas to stimulate its growth.

Table 4. First Stage IV Results for All Endogenous Regressors in Productivity Analysis.

First Stage Results for Productivity Analysis

	Dependent variable:				
	log(Density)	*log(Acc.to Employment by Private Vehicle)*	*log(Acc. to Employment by Public Transport)*	*log(Reachability by Private Vehicle)*	*log(Reachability by Public Transport)*
log(Past Density)	0.418***	0.001	0.079***	-0.005***	-0.008**
	(0.016)	(0.007)	(0.030)	(0.001)	(0.003)
log(Past Acc. To Employment)	0.013	-0.105***	0.117	0.015***	-0.037**
	(0.070)	(0.030)	(0.131)	(0.005)	(0.015)
log(Past Reachability)	0.556	-0.540***	-1.377*	-0.088***	0.313***
	(0.390)	(0.165)	(0.723)	(0.030)	(0.085)
log(Physical Reachability)	-6.816***	-7.688***	4.604	1.157***	-0.424
	(1.565)	(0.662)	(2.903)	(0.119)	(0.334)
log(Synthetic Acc. to Emp. Index)	0.115	-0.378***	0.828**	-0.069***	-0.022
	(0.220)	(0.093)	(0.408)	(0.017)	(0.046)
Distance to 1870 Railway	-0.036	-0.084***	0.028	-0.015***	0.008
	(0.026)	(0.011)	(0.048)	(0.002)	(0.006)
Distance to Cassini Roads	-0.032	0.030***	-0.006	0.002	0.019***
	(0.022)	(0.009)	(0.041)	(0.002)	(0.005)
Distance to Roman Roads	-0.092***	0.025**	-0.060	0.010***	0.006
	(0.025)	(0.010)	(0.046)	(0.002)	(0.005)
Distance to PT Network from 1863 to 1957	0.067*	0.037**	0.014	-0.004	-0.009
	(0.040)	(0.017)	(0.075)	(0.003)	(0.008)
Time Effects	Yes	Yes	Yes	Yes	Yes
EPCI Effects	Yes	Yes	Yes	Yes	Yes
Sector Effects	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.859	0.977	0.438	0.922	0.835

[20] Although, from the density analysis results presented in Subsection 6.2, Table 4, we learn the significant impact of reachability by the public transport network on employment density. Therefore, it is important to precise that our results does not suggest a direct effect of public transport reachability levels on productivity, but they indeed suggest an indirect effect on productivity, explained by the significant impact on the metropolitan distribution of employment.

F-Statistic for Instruments	78.059	59.720	18.418	43.511	25.318
F Statistic	183.971	1,302.613	24.542	362.697	153.955

4.2. Density Analysis Results.

The results from the density analysis are presented in Table 5. The first columns display the OLS results, and column (4) displays the results after applying instrumentation techniques. The instruments used are a subset of the group proposed for the productivity analysis. The subset of candidates are: the past predicted reachability levels by the public transport network computed for 2004, the synthetic reachability index computed with simple straight line distances and the distance to several historical networks, i.e. the public transport network in Toulouse during the period 1863-1957, the 1870 railway network, the Cassini roads and the Roman roads. Table 5 also displays the over-identification p-value computed by the Sargan method. Its value of 0.467 allows us to reject the null hypothesis of having at least one endogenous instrument, concluding that all instruments are valid. Also, Table 3 and in all specifications of Table 4 display the first stage F-statistics corresponding to the first stages of all endogenous regressors on the instruments. The size of these estimates allows us to conclude that the instruments are relevant for all endogenous regressors.

Table 5. Density Results: Public Transport and Private Vehicle Measures.

Density Analysis Results: Private Vehicle and Public Transport Measures.

	Dependent Variable: First Stage Zone-Year Effects			
	OLS	OLS	OLS	IV
	(1)	(2)	(3)	(4)
log(Reachability by Private Vehicle)	-4.777***	-1.003***	-2.213***	-2.382*
	(0.260)	(0.357)	(0.430)	(1.238)
log(Reachability by Public Transport)	-1.751***	-0.099	-0.292*	-3.763***
	(0.122)	(0.139)	(0.165)	(0.641)
Urban Controls	No	Yes	Yes	Yes
Physical Infrastructure Controls	No	Yes	Yes	Yes
Agglomeration Controls	No	Yes	Yes	Yes
Emp. Accessibility Controls	No	Yes	Yes	Yes
Time Effects	No	Yes	Yes	Yes
EPCI Effects	No	No	Yes	Yes
Sector Effects	No	No	Yes	Yes
Instruments:				
Past Levels of Reachability				x
Physical Reachability (Peripherality)				x
Distance to Roman Roads				x
Distance to Railway lines in 1870				x
Distance to Casini Roads				x
Observations	1,664	1,664	1,664	1,664
Adjusted R^2	0.544	0.651	0.694	0.615
Over-identification p-value				0.467

First Stage Statistics for:	
log(Reachability by Private Vehicle)	60.486
log(Reachability by Public Transport)	30.659

From the inspection of the columns (1) to (4) of the table, we find that both public transport and private vehicle reachability measures have a significant impact on employment density. Indeed, according to all specifications, if the average of optimal travel times to a particular destination decreases, employment density increases. Interestingly, we identify a similar behavior of reachability estimates in both, density and productivity analyses. In both cases, and when all controls are included, they increase after the introduction of Toulouse sectors and EPCI areas fixed-effects, and after instrumentation techniques are applied. Further, the coefficients for reachability by the public transport network are barely significant before instrumenting and/or introducing local effects. This suggests that the omitted variable problem was biasing the estimates downwards. Further, it suggests that the planner may be using transport infrastructure to better connect deprived areas and to attract employment activity.

On one hand, when the average of optimal travel times by the road network to a particular destination decreases by 10%, local employment density increases by 23.8%. On the other hand, when the average of optimal travel times by the public transport network to a particular destination decreases by 10%, local employment density increases by 37.6%. These results suggest that higher levels of local connectivity impact positively the local level of employment density. Therefore, more secluded areas have lower employment levels. These results provide evidence of an indirect effect of public transport and private vehicle exposure measures on productivity through employment their impact on the distribution of employment, since reachability levels affects employment density, and at the same time employment density affects productivity. By the interaction of both effects we achieve the complete indirect effect of reachability levels on productivity. In our case, we multiply the reachability estimates of our preferred density specification (column 4 of Table 5) and our preferred estimate of the impact of density on productivity (column 5 of Table 3). The interaction of both estimates suggest that if the time to arrive to a particular destination by the road network (public transport network) doubles, productivity decreases by 3.8% (6%) through the impact on employment density.

Table 6. First Stage IV Results for All Endogenous Regressors in Density Analysis.

First Stages Results for Density Analysis

	Dependent variable:	
	log(Reachability by Private Vehicle)	*log(Reachability by Public Transport)*
log(Past Reachability)	-0.136***	0.237***
	(0.024)	(0.071)
log(Peripherality)	1.236***	0.107
	(0.110)	(0.320)
Distance to Historical PT Network	-0.016***	0.010*
	(0.002)	(0.006)
Distance to 1870 Railway	0.001	0.023***
	(0.002)	(0.005)
Distance to Cassini Roads	0.011***	0.002
	(0.002)	(0.005)

Distance to Roman Roads	-0.136***	0.237***
	(0.024)	(0.071)
Time Effects	Yes	Yes
EPCI Effects	Yes	Yes
Sector Effects	Yes	Yes
Observations	1,724	1,664
Adjusted R^2	0.922	0.828
F-Statistic for Instruments	60.486	30.659
F-Statistic	421.861	161.329

5. Conclusions

This paper develops a framework to investigate the sources of productivity variation across small-scale geographical units within the metropolitan area of Toulouse. We focus on two factors: agglomeration and transport exposure measures. On the one hand, we investigate the positive spatial externalities on productivity coming from the concentration of economic activity. On the other hand, we investigate the channels by which transport have an impact on productivity. Firstly, transport extends the geographic scope of agglomeration externalities, by reducing the interaction cost between economic agents placed in different locations. And secondly, transport increases the quantity and the quality of the local workforce. Indeed, firms and employees, and more those with higher skills and more added value, are attracted by the higher connectivity of places with higher levels of transport exposure. Therefore, in this paper we estimate the extent of agglomeration externalities at the urban scale taking into account the impacts of transport exposure on local employment and productivity.

Transport exposure is measured by two continuous accessibility-type indexes: an employment accessibility index and a reachability. The *accessibility to employment* index is measured by the weighted sum of inverse optimal travel times, where the weights are measures of employment density at destination, and the *reachability* index is computed by the average of optimal travel times from all potential origins along the road and public transport network. Both indexes project transport improvements through changes in commuting times, and they measure the local and global levels of transport exposure for each location.

We recover the productivity effects of agglomeration and transport measures by the implementation and estimation of a wage determination model in two stages. The first stage assesses the importance of industrial concentration and employees' characteristics against true productivity differences across zones on the average of local industrial wages. The second stage explains local productivity differences on our local factors of interest: local agglomeration and local transport measures. Finally, we recover the size of the indirect effect of transport exposure on productivity through, first, the estimation of the relationship between local employment density and transport exposure measures, and second, by interacting these estimates with those relating employment density and productivity.

We deal with the endogeneity of agglomeration and transport exposures measures provoked by the omission of unobserved local characteristics and the reverse-causality between agglomeration and transport measures, and productivity. Firstly, to deal with the omitted variable bias we take advantage of the panel nature of our data and we introduce Toulouse sectors and EPCI areas effects into the estimation productivity effects of agglomeration and transport exposure measures. Secondly, to control for the reverse-causality, we perform an instrumental variables approach where the instruments are historical levels of agglomeration and transport exposure measures, the physical-distance-based counterparts of the accessibility and reachability indexes and several indicators of the distance to historical infrastructure, i.e. the distance to the historical public-transport plan during the years 1863 and 1957, the distance to two ancient roads, the roman and Cassini roads, and the distance to the 1870 railway network.

Interestingly, less exhaustive specifications of the model display higher employment density effects and lower transport exposure effects on productivity than specifications with fixed effects and instrumentation techniques. The decline in the size of agglomeration effects reveals that the skills-sorting issue remains an important matter at the metropolitan level, where high skilled employees and high value-added firms may be persistently locating in denser and more accessible municipalities and/or denser and more accessible sectors within the municipality of Toulouse. Also, the underestimation of transport effects suggest that the non-random condition of transport location is an important matter, where the results are in line with a planner that tends to assign transport infrastructure to more secluded areas to stimulate its growth.

Our key findings suggest that both agglomeration and transport exposure have an important role in determining the internal productivity structure of a metropolitan area. According to our preferred specification, where local effects are introduced and instrumentation techniques are applied, we find that if local density is doubled, productivity increases by 1.6%. Further, the effects of transport exposure measures differ for our two modes, private vehicle and public transport. On one hand, both accessibility to employment measures have a positive and significant effect on local productivity. Specifically, when accessibility to employment by the road network increases by 10%, local productivity does by 0.49%. Further, when accessibility to employment by the public transport network increases by 10%, productivity does by 0.56%. On the other hand, we identify the direct effect of reachability levels on local productivity and the indirect effect through its impact on the spatial distribution of employment. In the case of reachability levels by the road network, measured by the average of optimal travel times, they affect productivity directly and indirectly. Firstly, a 10% decrease in the average of optimal travel times by the road network increases productivity by 1.65%. Further, this same decrease makes local employment density to increase by 23.8%. In the case of reachability levels by the public transport network, the effect is only indirect through its impact on employment density. Specifically, if the average of optimal travel times by the public transport network decreases by 10%, density of employment increases by 37.6%, which translates into an impact in local of productivity of a 3.8%.

Concluding, all these results point out that, first, agglomeration externalities are locally happening at the metropolitan level, and second, that private vehicle and public transport exposure levels affects productivity directly and indirectly by locally increasing the *quality* and the *quantity* of the workforce, respectively.

References

Abel, Jaison R., Dey, Ishita, and Gabe, Todd M, 2012. Productivity and the density of human capital. Journal of Regional Science, 52(4):562–586.

Abowd, John M., Kramarz, Francis, and Margolis, David N, 1999: High wage workers and high wage firms. Econometrica, 67(2):251–333.

Ahlfeldt, Gabriel M., Redding, Stephen, Sturm, Daniel M. and Wolf, Nikolaus, 2015. The economics of density: evidence from the Berlin Wall. Econometrica, 83 (6). pp. 2127-2189.

Amiti, Mary and Cameron, Lisa, 2007. Economic geography and wages. Review of Economics and Statistics, 89(1):15–29.

Andersson, Fredrik, Burgess, Simon, and Lane, Julia I, 2007. Cities, matching and the productivity gains of agglomeration. Journal of Urban Economics, 61(1):112–128.

Andini, Monica, de Blasio, Guido, Duranton, Gilles, and Strange, William., 2013. Marshallian labour market pooling: Evidence from Italy. Regional Science and Urban Economics, 43(6):1008–1022.

Angel Garcia-Lopez, Miquel, 2012. Urban spatial structure, suburbanization and transportation in Barcelona. Journal of Urban Economics 72:176–190.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2012. On the road: Transportation infrastructure and economic growth in China. Technical report, NBER Working Paper no. 17897.

Baum-Snow, Nathaniel and Ferreira, Fernando, 2015. Causal inference in urban economics. In Duran- ton, Gilles, Henderson, Vernon, and Strange, Will, editors, Handbook of Urban and Regional Economics, volume 5A. North-Holland, Amsterdam.

Baum-Snow, Nathaniel and Pavan, Ronni, 2012. Understanding the city size wage gap. Review of Economic Studies, 79(1):88–127.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang, 2017. Roads, railroads an decentralization of Chinese cities. The Review of Economics and Statistics 2017 99:3, 435-448.

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122(2):775–805.

Baum-Snow, Nathaniel and Matthew E. Kahn. 2005. Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000. Brookings-Wharton Papers on Urban Affaires: 2005 1(4):147–197.

Behrens, Kristian, Duranton, Gilles, and Robert-Nicoud, Frederic, 2014. Productive cities: Sorting, selection, and agglomeration. Journal of Political Economy, 122(3):507–553.

Billings, Stephen B, 2011. Estimating the value of a new transit option. Regional Science and Urban Economics 41(6):525–536.

Brinkman, Jeffrey C., 2016. Congestion, agglomeration, and the structure of cities. Journal of Urban Economics, vol. 94(C),13-31.

Chandra, Amitabh and Eric Thompson, 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Regional Science and Urban Economics 30(4):457– 490.

Ciccone, Antonio, 2002. Agglomeration effects in Europe. European Economic Review, 46(2):213–227.

Ciccone, Antonio and Hall, Robert E, 1996. Productivity and the density of economic activity. American Economic Review, 86(1):54–70.

Ciccone, Antonio and Peri, Giovanni, 2006. Identifying human capital externalities: Theory with an application to US cities. Review of Economic Studies, 73(2):381–412.

Cingano, Federico and Schivardi, Fabiano. Identifying the sources of local productivity growth, 2004. Journal of the European Economic Association, 2(4):720–742.

Combes, Pierre-Philippe, 2000. Economic structure and local growth: France, 1984–1993. Journal of Urban Economics, 47(3):329–355.

Combes, Pierre-Philippe, 2011. The empirics of economic geography: How to draw policy implications? Review of World Economics, 147(3):567–592.

Combes, Pierre-Philippe and Duranton, Gilles, 2006. Labor pooling, labor poaching, and spatial clustering. Regional Science and Urban Economics, 36(1):1–28.

Combes, Pierre-Philippe and Lafourcade, Miren, 2005. Transport costs: Measures, determinants, and regional policy implications for France. Journal of Economic Geography, 5(3):319–349.

Combes, Pierre-Philippe and Lafourcade, Miren, 2011. Competition, market access and economic geography: Structural estimation and predictions for France. Regional Science and Urban Economics, 41:508–524.

Combes, Pierre-Philippe, Magnac, Thierry, and Robin, Jean-Marc, 2004. The dynamics of local employment in France. Journal of Urban Economics, 56(2):217–243.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent, 2008a. Spatial wage disparities: Sorting matters! Journal of Urban Economics, 63(2):723–742.

Combes, Pierre-Philippe, Mayer, Thierry, and Thisse, Jacques-François, 2008b. Economic Geography: The integration of Regions and Nations. Princeton University Press, New Jersey.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, and Roux, Sébastien, 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. NBER Chapters,in: Agglomeration Economics, pages 15-66 National Bureau of Economic Research, Inc.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent. The identification of agglomeration economies. Journal of Economic Geography, 11(2):253–266, 2011.

Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent, 2012a . The costs of agglomeration: Land prices in French cities. Discussion Paper 9240, Centre for Economic Policy Research.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego, and Roux, Sébastien, 2012b. The productivity advantages of large markets: Distinguishing agglomeration from firm selection. Econometrica, 80(6):2543–2594.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, and Roux, Sébastien, 2012c. Sorting and local wage and skill distributions in France. Regional Science and Urban Economics, 42(6): 913–930.

Combes, Pierre-Philippe, D́emurger, Sylvie, and Li, Shi, 2013. Urbanization and migration externalities in China. Discussion Paper 9352, Centre for Economic Policy Research.

De La Roca, Jorge and Puga, Diego, 2017. Learning by working in big cities. *The Review of Economic Studies*, 106–142.

Desmet, Klaus and Fafchamps, Marcel, 2005. Changes in the spatial concentration of employment across US counties: A sectoral analysis 1972-2000. Journal of Economic Geography, 5(3):261–284.

Desmet, Klaus and Esteban Rossi-Hansberg, 2013. Urban accounting and welfare. *American Economic Review* .

Di Addario, Sabrina and Patacchini, Eleonora, 2008. Wages and the city. Evidence from Italy. Labour Economics, 15(5):1040–1061.

Donaldson, D. and R. Hornbeck, 2016. Railroads and American economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2): 99–858.

Donaldson, David, 2013. Railroads of the Raj: Estimating the impact of transportation infrastructure. American Economic Review.

Duranton, Gilles and Puga, Diego, 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. American Economic Review, 91(5):1454–1477.

Duranton, Gilles, 2015. Roads and trade in Columbia. Processed, University of Pennsylvania.

Duranton, Gilles, Peter Morrow, and Matthew A. Turner. 2014. Roads and trade: Evidence from the US. *The Review of Economic Studies*, 81(2):681–724.

Duranton, Gilles and Matthew A. Turner, 2011. The fundamental law of road congestion: Evidence from US cities. *American Economic Review* 101(6):2616–2652.

Duranton, Gilles and Matthew A. Turner.,2012. Urban growth and transportation. *Review of Economic Studies* 79(4):1407–1440.

Fujita, Masahisa, Krugman, Paul R., and Venables, Anthony J, 1999. The Spatial Economy: Cities, Regions, and International Trade. MIT Press, Cambridge.

Garcia-Lopez, Miquel-Angel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. Suburbanization and highways: When the Romans, the Bourbons and the first cars still shape Spanish cities. Journal of Urban Economics, Elsevier, vol. 85(C), pages 52-67

Garcia-López, Miquel-Àngel and Hémet, Camille and Viladecans-Marsal, Elisabet, How Does Transportation Shape Intrametropolitan Growth? An Answer from the Regional Express Rail (November 2017). Journal of Regional Science, Vol. 57, Issue 5, pp. 758-780, 2017

Gibbons, Stephen and Stephen Machin. 2005. Valuing rail access using transport innovations. *Journal of Urban Economics* 57(1):148–169.

Glaeser, Edward L. and Kerr, William R, 2009. Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? Journal of Economics and Management Strategy, 18(3):623–663.

Glaeser, Edward L. and Mare, David C, 2001. Cities and skills. Journal of Labor Economics, 19(2): 316–342.

Glaeser, Edward L., Kallal, Heidi, Scheinkman, José A., and Schleifer, Andrei, 1992. Growth in cities. Journal of Political Economy, 100(6):1126–1152.

Glaeser, Edward L., Resseger, Matt, and Tobio, Kristina, 2009. Inequality in cities. Journal of Regional Science, 49(4):617–646.

Glaeser, Edward L., Kerr, William R., and Ponzetto, Giacomo A.M, 2010a . Clusters of entrepreneurship. Journal of Urban Economics, 67(1):150–168.

Glaeser, Edward L., Rosenthal, Stuart S., and Strange, William C, 2010b . Urban economics and entrepreneurship. Journal of Urban Economics, 67(1):1–14.

Graham, Daniel J. Variable returns to agglomeration and the effect of road congestion. Journal of Urban Economics, 62(1):103–120, 2007.

Graham, Daniel J, 2009. Identifying urbanisation and localisation externalities in manufacturing and service industries. Papers in Regional Science, 88(1):63–84.

Graham, Daniel J., Melo, Patricia S., Jiwattanakulpaisarn, Piyapong, and Noland, Robert B, 2010. Testing for causality between productivity and agglomeration economies. Journal of Regional Science, 50(5):935–951.

Lucas, Robert E., Jr. and Esteban Rossi-Hansberg. 2002. On the internal structure of cities. *Econometrica* 70(4):1445–1476.

Melo, Patricia C., Graham, Daniel J., and Noland, Robert B, 2009. A meta-analysis of estimates of urban agglomeration economies. Regional Science and Urban Economics, 39(3):332–342.

Micucci, Giacinto and Giacinto, Valter di, 2009. The producer service sector in Italy: Long-term growth and its local determinants. Spatial Economic Analysis, 4(4):391–425.

Mion, Giordano, 2004. Spatial externalities and empirical analysis: The case of Italy. Journal of Urban Economics, 56(1):97–118.

Redding, Stephen and Sturm, Daniel, 2008. The costs of remoteness: Evidence from German division and reunification. American Economic Review, 98(5):1766–1797.

Redding, Stephen and Venables, Anthony J, 2004. Economic geography and international inequality. Journal of International Economics, 62(1):63–82.

Rice, Patricia, Venables, Anthony J., and Patacchini, Eleonora, 2006. Spatial determinants of productivity: Analysis for the regions of Great Britain. Regional Science and Urban Economics, 36(6): 727–752.

Rosenthal, Stuart S. and Strange, William C, 2001. The determinants of agglomeration. Journal of Urban Economics, 50(2):191–229.

Rossi-Hansberg, Esteban, 2004. Optimal urban land use and zoning. *Review of Economic Dynamics* 7:69–106.

Thierry Mayer, Florian Mayneris and Loriane Py, 2017. The impact of urban enterprise zones on establishments' location decisions: Evidence from French ZFUs. *Journal of Economic Geography*, 17(4):709–752.

Appendix A. Instruments Construction: Past levels of agglomeration and transport exposure measures.

This appendix provides additional information on the computation of past levels of agglomeration and transport exposure measures. In the following appendix, i.e. Appendix B, we provide a description of the rest of instruments used in this study.

In order to compute past levels of agglomeration and transport exposure measures, we need disaggregated data on previous employment levels and disaggregated data on previous optimal travel times between every pair of zones within the study. The farthest year for which we have access to both sources of data is 2004. From the *Agence d´Urbanisme et Aménagement de Toulouse* (AUAT), we have the geo-coordinates of all establishments operating in Toulouse during that year. Then, from the *Déclaration Annuelle de Données Sociales* (DADS), we identify the number of employees working on the establishments and aggregate them by zone. In this manner, employment levels for 2004 are obtained for all zones of the study.

However, we have not direct access to the distribution of optimal travel times for this year. Yet, we can predict them, since we know in detail the structure of the public transport network in 2004, displayed by Figure A.1. Specifically, our predictive approach makes use of observed optimal travel times for the years 2013 and 2015, and regress them on observed characteristics of the public transport network in those two years, i.e. we regress the observed distribution of optimal travel times in 2013 and 2015 on observed variables related to the structure of the public transport network for those two years.

Then, we interact the estimated coefficients of this regression with the observed variables of the public transport network in 2004, obtaining the predicted distribution of optimal travel times by the public transport network in 2004. Finally, the predicted levels are introduced in the computation of past transport exposure levels.

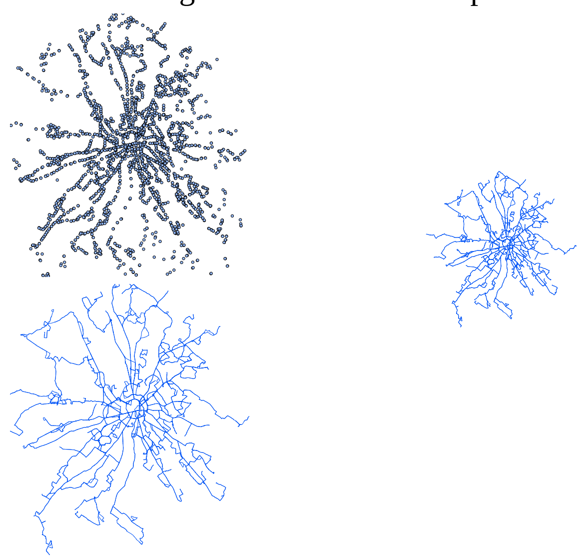
Unfortunately, this predictive approach can only be done for the prediction of optimal travel times by the public transport network, and not for the road network. The reason is that road infrastructure has not changed much in the last years, and consequently, infrastructure variation is not a good predictor of changes in road travel times. In other words, we have not enough variability in observed road characteristics to provide consistent predictions. Further, we lack access to other sources of road travel times variation that we could use to predict previous road travel times, as for instance, road congestion levels or road quality improvements. Thus, instruments identifying past levels of transport exposure are only computed for the public transport network. Nevertheless, we include other instruments that are more oriented to instrument road exposure measures, like several distance indicators of distances to ancient road infrastructures.

In order to predict optimal travel times by the public transport network in 2004, we first estimate a prediction regression of optimal travel times in 2013 and 2015 on public transport network characteristics of origins and destinations in those two years. Then, we recover the estimated coefficients and interact them with 2004 observed variables. The product is a vector of optimal predicted travel times in 2004 between all pairs of origins and destinations. The estimated prediction regression is the following:

$$TT_{t,od} = a + b_1 X_{t,o} + b_2 X_{t,d} + g_1 long_o + g_2 lat_o + d_1 long_d + d_2 lat_d + qDist_{od} + e_{t,od}$$

where $TT_{t,od}$ is the optimal travel time by the public transport network for our two times *t* available, i.e. 2013 and 2015, between every origin *o* and destination *d*, and $X_{t,o}$ and $X_{t,d}$ are vectors of public transport network characteristics at time *t* in origin *o* and destination *d*, respectively. Then, controls on zones´ location and the Euclidean distances between each pair are also included: $long_o$ and $long_d$, and lat_o and lat_d, are the longitudes and latitudes of the geographic coordinates of origin *o* and destination *d*, respectively; and $Dist_{od}$ is the physical distance between origin *o* and destination *d*. The predicted vectors of coefficients $\left\{\hat{b}_1,\hat{b}_2,\hat{g}_1,\hat{g}_2,\hat{d}_1,\hat{d}_2,\hat{q}\right\}$ is interacted with the vectors of observed variables $\left\{X_{2004,o},X_{2004,d},long_o,long_d,lat_o,lat_d,Dist_{od}\right\}$, evaluated in 2004. The result of this interaction is a vector of predicted optimal travel times for 2004, $\hat{T}T_{2004,od}$. This vector of predicted optimal travel times is used in the estimation of past levels of transport exposure measures, i.e. employment accessibility and reachability levels.

Figure A.1: Public Transport in 2004.



Source: Own elaboration from data provided by the *Agence d´Urbanisme et Amenagement de Toulouse (AUAT).*

Table A1 displays the results of the prediction regression. We consider three public transport modes: *bus*, *metro* and *tram*. Therefore, all variables are computed for these three modes and evaluated at both, the origin and the destination zones. The network characteristics for each origin-destination pair included in the regression are: (i) a categorical variable specifying the closest public transport mode, (ii) the distance to each of the modal networks, (iii) the local length of the modal network, and (iv) the total number of lines and stations locally per mode. Also, and together with the geographical position controls and Euclidean distances between nodes, we include the area and the area covered by the water for all origins and destinations.

The results suggest that the group of variables chosen for the predictive exercise is relevant and explain almost the 60% of the variation of optimal travel times by the public transport network.

Further, all the variables are highly significant and the F-Static is high, evidencing an efficient predictive power of the regression.

Table A1. Prediction regression for the distribution of travel times in 2004.

Prediction Regression for Public Transport Travel Times in 2004		
	Dependent variable: Travel Times in 2013 and 2015.	
	Coefficient	**St. Deviation**
Closest Mode Origin: Metro	-0.558***	(0.205)
Closest Mode Origin: Tram	4.149***	(0.320)
Dist. Bus Network at Origin	0.001***	(0.00002)
Dist. Metro Network at Origin	0.002***	(0.00002)
Dist. Tram Network at Origin	-0.002***	(0.00002)
Length Bus Network at Origin	0.0001***	(0.00001)
Length Metro Network at Origin	-0.001***	(0.0001)
Length Tram Network at Origin	-0.001***	(0.0001)
Total Bus Lines at Origin	-0.339***	(0.017)
Total Metro Lines at Origin	-3.473***	(0.159)
Total Tram Lines at Origin	-3.009***	(0.248)
Total Bus Stations at Origin	-0.102***	(0.004)
Total Metro Stations at Origin	0.419***	(0.037)
Total Tram Stations at Origin	0.070	(0.052)
Area Origin	0.00000***	(0.000)
Area Water Origin	-0.033***	(0.0001)
Closest Mode Destination: Metro	-1.352***	(0.204)
Closest Mode Destination: Tram	2.366***	(0.318)
Dist. Bus Network at Destination	0.0002***	(0.00002)
Dist. Metro Network at Destination	0.001***	(0.00002)
Dist. Tram Network at Destination	0.0002***	(0.00002)
Length Bus Network at Destination	0.00003***	(0.00001)
Length Metro Network at Destination	0.0002*	(0.0001)
Length Tram Network at Destination	-0.001***	(0.0001)
Total Bus Lines at Destination	-0.207***	(0.017)
Total Metro Lines at Destination	-1.144***	(0.158)
Total Tram Lines at Destination	1.800***	(0.246)
Total Bus Stations at Destination	-0.196***	(0.004)
Total Metro Stations at Destination	0.291***	(0.037)
Total Tram Stations at Destination	0.340***	(0.051)
Area Destination	0.00000***	(0.000)
Area Water Destination	-0.032***	(0.0001)
Physical Distance	0.002***	(0.00001)
Origin Longitude	0.001***	(0.00000)
Origin Latitude	-0.0004***	(0.00000)

Destination Longitude	0.0002***	(0.00000)
Destination Latitude	-0.0003***	(0.00000)
Time Fixed Effects	Yes	
Observations	1,365,848	
R^2	0.591	
Adjusted R^2	0.591	
F Statistic	51,838.690***	

Appendix B. Instrument Analysis.

In this appendix we explore the nine instruments proposed for our four endogenous regressors, i.e. the local density of employment, the accessibility to employment through the public transport and the road networks, and the reachability levels by those two networks computed by the average of optimal travel times from any origin to a particular destination. Likewise, we remind that the instrumental candidates are past levels of local employment density computed for 2004, past transport exposure measures computed for 2004, two synthetic indexes of reachability and employment accessibility computed with physical distances instead of optimal travel times, and finally, the distance to four historical networks, i.e. the roman roads, the Cassini roads, the 1870 railway and the public transport network from 1863 to 1957.

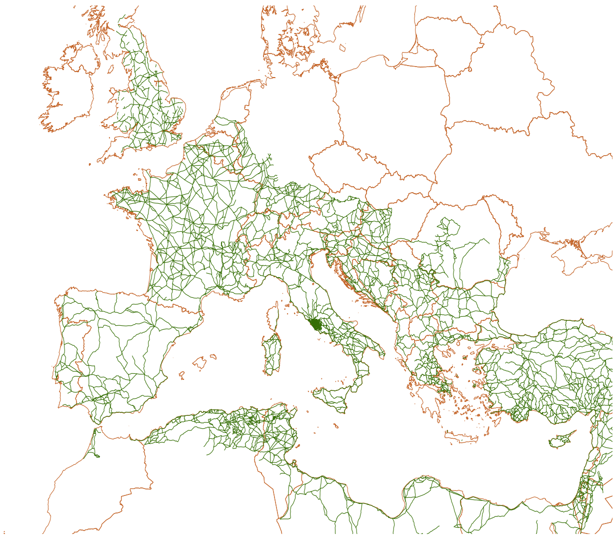
As a first group of candidates, we consider past levels of employment density and employment accessibility, computed for 2004. Following Ciccone and Hall (1996), they were the first one to argue that endogeneity may be caused by “contemporaneous” local shocks. Therefore, and considering that these shocks did not have any effect on the distribution of the employment in the past, we can instrument employment density between 2013 and 2015 by lagged employment variables. This strategy rests on the hypothesis that employment agglomeration in the past is not related to modern differences in productivity. Same logic is applied to the instrumental candidate measuring past levels of employment accessibility. Indeed, past levels of reachability are correlated with actual levels, but they do not have a direct impact on nowadays productivity.

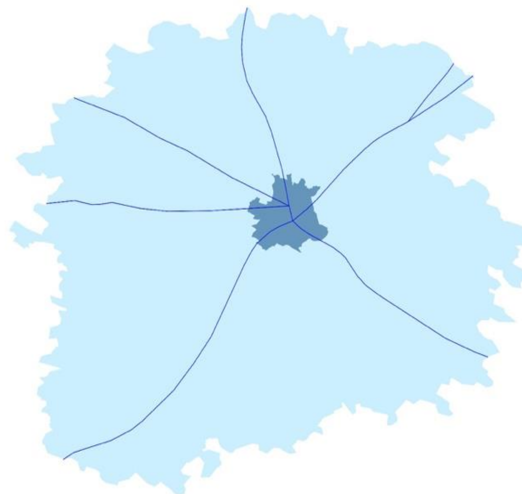
The second group of candidates is composed by the two synthetic indexes of employment accessibility and reachability, where instead of using optimal travel times, we introduce physical distances between nodes. The rationale behind is the following: geographical proximity alone should not have an independent impact on productivity. Physical proximity does not affect productivity independently, it is instead the interaction between economic agents what generates productivity gains. This interaction would not be possible without a network connecting the nodes.

Finally, the third and last group comprises a set of distance indicators to several types of historical infrastructure. Indeed, it has been frequently pointed out in the literature that modern networks are built following the routes traced in the past. It is easier and cheaper to build new transportation infrastructure by improving existing one, or nearby it. Therefore, past networks of infrastructure shapes current ones. For this reason, historical infrastructure indicators are useful instruments for current levels of transport exposure measures since they better satisfy the exogeneity condition, and more specifically the exclusion restriction, than more modern infrastructure indicators. Certainly, while the relevance of historical instruments decreases with time, their exogeneity increases with it.

Within this group, the first candidate considered is the distance to the closest roman road. The main Roman roads passing through Toulouse (Tolosa at the time) were built around 118 BC. As a whole, our first candidate, the Roman network around Toulouse, was based on 352 km of roads and it is represented on the right of Figure B.1, together with the overall ancient Roman road network, on the left.

Figure B.1: Roman Roads

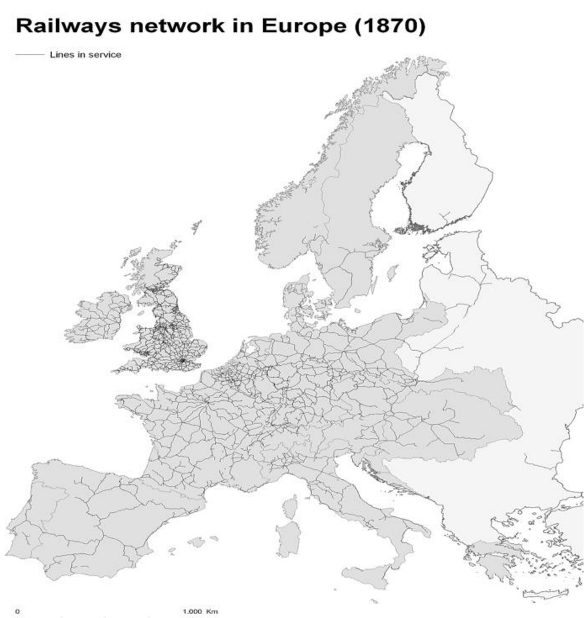


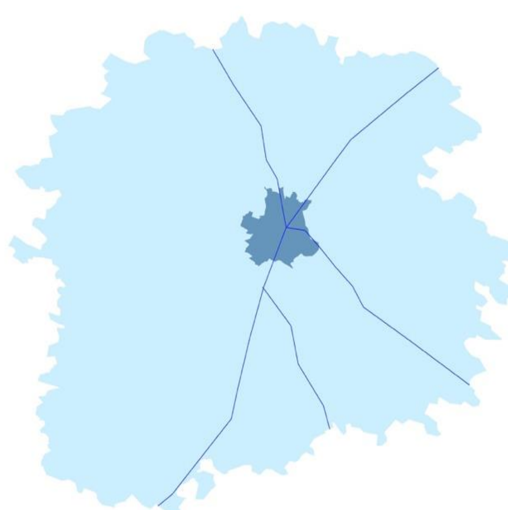


Source: Own elaboration based on the Digital Atlas of Roman and Medieval Civilizations.

The second candidate is the railroad network as it existed in 1870 in the region around Toulouse, based on 321 km of railroad lines. Certainly, the effect of trains and railways on employment and productivity outcomes is out of the scope of this paper, as we focus on roads and public transport infrastructure. However, we believe that historical railway structures can be useful instruments as they should not only affect current railway patterns, but also the current structure of road and public transport infrastructure. The 1870 railway network is displayed by Figure B.2, as well as the map of railways network in Europe (Martí-Henneberg, 2013).

Figure B.2: Railway Network in 1870.

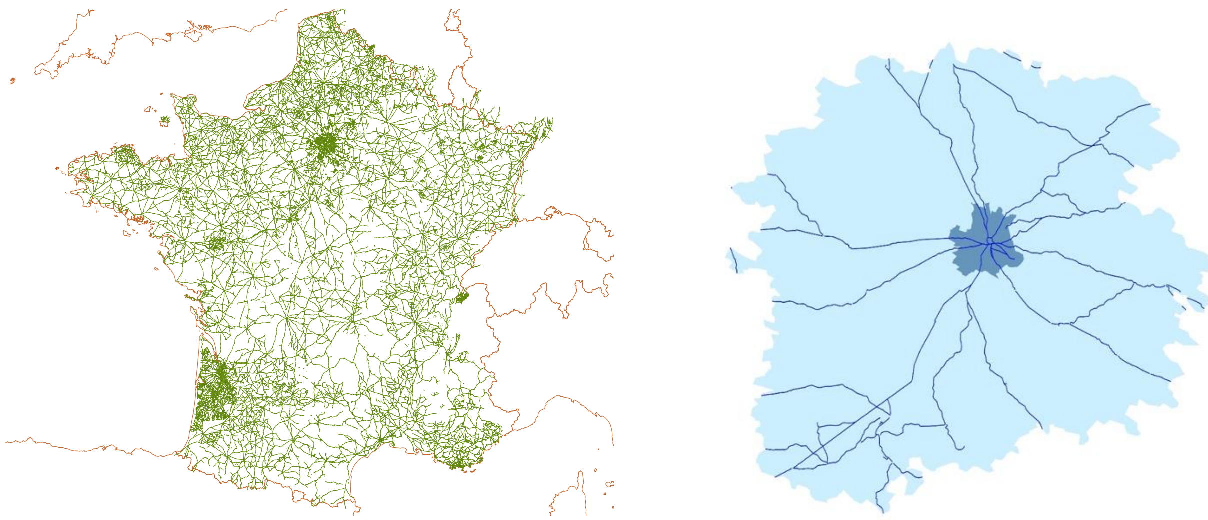




Source: Own elaboration based on Martí-Henneberg (2013) maps. [21]

The third candidate is the distance to the network of Cassini roads. In 1747 César-François Cassini de Thury was formally commissioned by Louis XV to draw the entire map of France. Due to financial difficulties, the Revolution and regime changes, the constitution of this map was delayed and it was releas in 1815. Therefore, the Cassini roads corresponds to the road infrastructure existing in France within the years 1747 and 1815. As a whole, the Cassini network along the metropolitan area of Toulouse was based on 1.064 km of roads. Figure B.3 presents the network of Cassini roads in the Metropolitan Area of Toulouse, and the full digitalized map of the Cassini network in France.

Figure B.3: Cassini Roads.



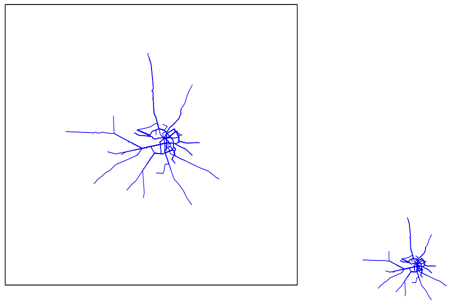
Source: Own elaboration based on the Geo Historical Data Research Project from Harvard University. [22]

The forth and last candidate corresponds to the historical public transport network of Toulouse for the years 1863 to 1957. The network evolved considerably between those two dates. It started as a "omnibus à impériale", where four-wheels vehicles were driven by horses around the city. This omnibus was substituted in 1887 by a railway network where vehicles, although still moved by horses, were circulating only along the rails. Finally, from 1913 onwards, the network was electrified, looking very much alike than the tramways that exist nowadays. The network is represented in Figure B.4.

Figure B.4: Historical Public Transport Network Toulouse - 1863 to 1957.

[21] https://www.sciencedirect.com/science/article/abs/pii/S0966692312002517

[22] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28674



Source: Data Toulouse Métropole. [23]

[23] https://data.toulouse-metropole.fr/explore/dataset/transports-en-commun-a-toulouse-entre-1863-et-1957-terminus/information/

Appendix C. Evaluation of a transport investment: Toulouse third metro line.

[Introduction of the project. Date of opening, exact route, purpose…]

From the results obtained in Section 5, we identify that transport exposure measures significantly affects the productivity structure of a metropolitan area in two ways. Firstly, directly, through the change of reachability levels by the road network, and secondly, indirectly, through the change in reachability levels by the public transport network which leads to productivity changes by modifying the metropolitan distribution of employment density.

In this appendix, we apply these results to compute the productivity gains of a real infrastructure change: the construction of a third metro line in Toulouse that will start operating in 2022. Figure C.1 displays the trajectory of this new line in Toulouse Metropolitan Area.

Figure C.1: Course of the Third Metro Line.

Not surprisingly, even if the infrastructure change evaluated only concerns the public transport network, travel times by the road network are also adjusted. This is explained by the following reasoning: when a transport investment takes place in any of the networks operating in a metropolitan area, commuters usage of all networks is affected, being the demand for all modes adjusted. In our case, the construction of a third metro line in the Toulouse public transport network changes the way commuters use both, public transport services and their private vehicles. Figure C.2 and Figure C.3 display the change (in percentual levels) of public transport and private vehicle reachability measures, respectively, after the construction of the new metro line. Both maps show that zones crossed by the trajectory of the new metro line, displayed in Figure C.1, host the most intense decreases on average travel times. Further, the intensity of those decreases diminish the farther we go from the infrastructure change. Importantly, both figures disclose that the decrease in public transport travel times is quantitatively stronger than the decrease in private vehicle travel times.

Regarding the direct effect on productivity of road reachability, from Section 5 we know that when the average travel time by private vehicle decreases by 1%, productivity increases by

0.289%. Therefore, we can compute the percentual change in the productivity of each zone by multiplying the percentual change in local levels of private vehicle reachability after the metro line construction, by an elasticity of -0.289,

$$E_{Productivity|PV\ Reachability} = \frac{\%\Delta\, Productivity}{\%\Delta\, PV\ Reachability} \Leftrightarrow$$

$$\begin{aligned} \%\Delta\, Productivity &= E_{Productivity|PV\ Reachability} * \%\Delta\, PV\ Reachability \\ &= (-0.289) * \%\Delta\, PV\ Reachability \end{aligned}$$

The distribution of the effects is displayed in Figure C.4. Further, the average increase on productivity after the construction of the third metro line is 0.89%.

Regarding the indirect effect of public transport reachability on productivity through employment density, first, from Section 5, we learn that there is a significant effect of public transport reachability on employment density, represented in the following expression.

$$E_{Density|PT\ Reachability} = \frac{\%\Delta\, Density}{\%\Delta\, PT\ Reachability} \Leftrightarrow$$

$$\%\Delta\, Density = E_{Density|PT\ Reachability} * \%\Delta\, PT\ Reachability$$

Specifically, when the average time it takes to arrive to a particular destination by the public transport network increases by 1%, density levels decrease by 2.955%. Then, and also from Section 5, we know that when density increases by 1%, productivity do so by 0.01%. Therefore,

$$E_{Productivity|Density} = \frac{\%\Delta\, Productivity}{\%\Delta\, Density} \Leftrightarrow$$

$$\begin{aligned} \%\Delta\, Productivity &= E_{Productivity|Density} * \%\Delta\, Density \\ &= E_{Productivity|Density} * E_{Density|PT\ Reachability} * \%\Delta\, PT\ Reachability \\ &= 0.01 * (-2{,}955) * \%\Delta\, PT\ Reachability \end{aligned}$$

This means that the full indirect effect of public transport reachability through density on productivity is equal to (0.01 * (-2.955) =) -0.02955%. The distribution of this change is displayed by Figure C.5. Further, the average indirect effect of public transport reachability levels on productivity is equal to 0.13%.

Finally, and according to our results, we can conclude that the construction of the third metro line in Toulouse is beneficial for metropolitan productivity. While the change in private vehicle reachability levels increases productivity by 0.89%, on average, public transport reachability improvements do so by 0.13%.

Figure C.2: Public Transport Reachability Change after the Third Metro Line.

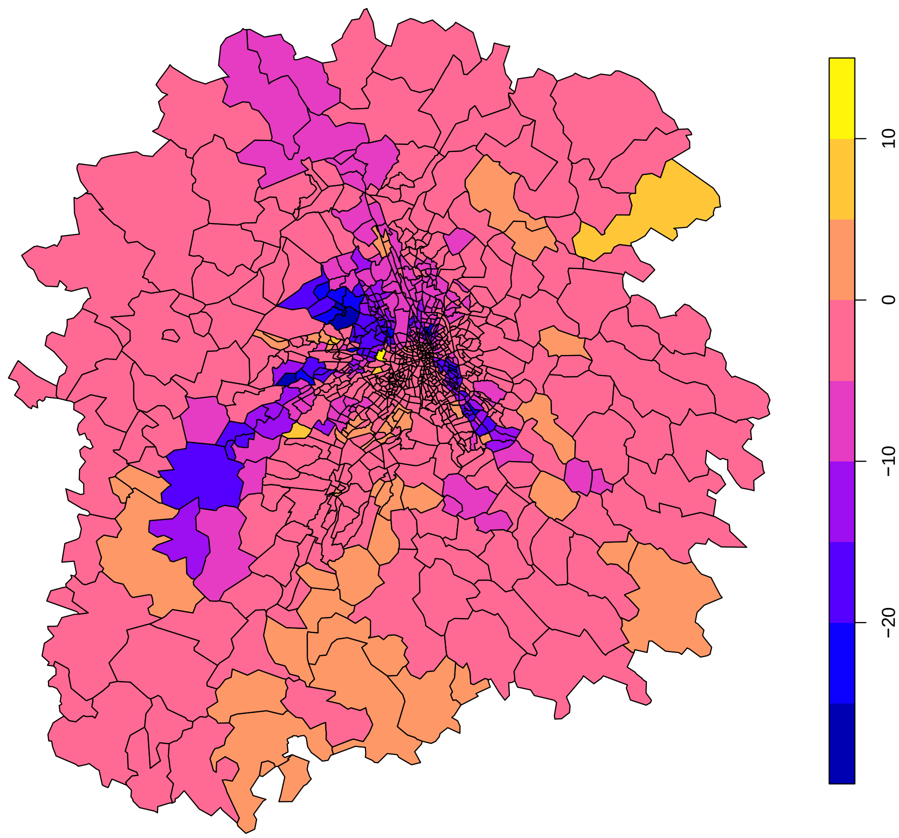


Figure C.3: Private Vehicle Reachability Change after the Third Metro Line.

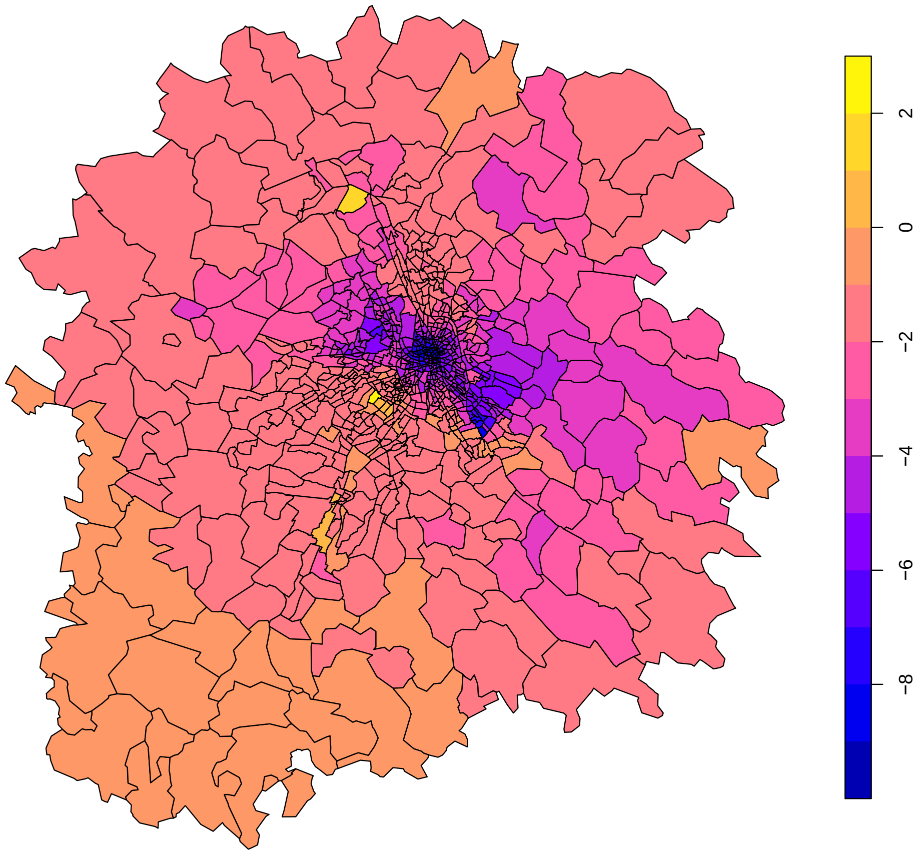


Figure C.4: Productivity Change after the Third Metro Line: Private Vehicle Reachability.

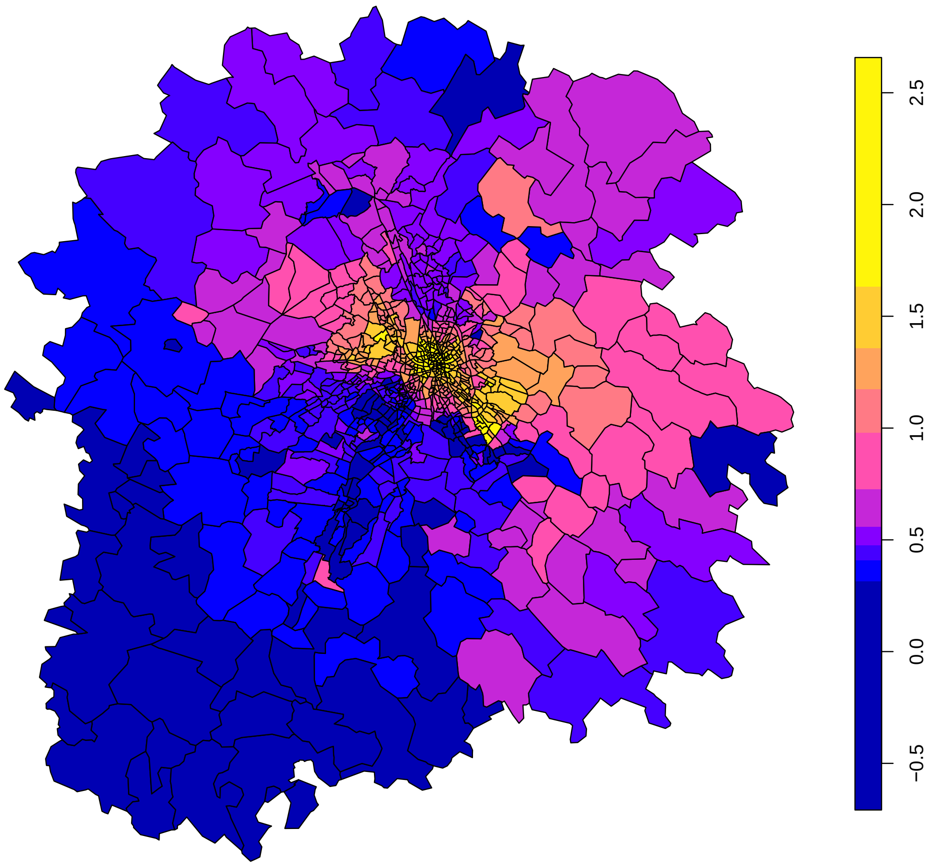


Figure C.5: Productivity Change after the Third Metro Line: Public Transport Reachability.