Next train to the polycentric city: The effect of railroads on subcenter formation

Miquel-Àngel Garcia-López[*†]
Universitat Autònoma de Barcelona and *Institut d'Economia de Barcelona*

Camille Hémet[*‡]
Paris School of Economics and *Institut d'Economia de Barcelona* and *CEPR*

Elisabet Viladecans-Marsal[*§]
Universitat de Barcelona and *Institut d'Economia de Barcelona* and *CEPR*

July 2017

Abstract: Recent evidence reveals that transportation's improvements within metropolitan areas have a clear effect on population and job decentralization processes. Yet, very little has been said on how these improvements affect the spatial organization of the economic activity in the suburbs. This paper analyses the effects of transportation's changes on employment subcenters formation. Using data from metropolitan Paris between 1968 and 2010, we first show that the spatial pattern of job decentralization is reinforcing the polycentric nature of Paris: the number of subcenters grew from 21 in 1968 to 35 in 2010 and the employment growth was very intense within them. Second, our main contribution is to show that the new rail transit clearly affects the emergence of subcenters: not only does the presence of a rail station increase the probability of a suburban municipality of becoming (part of) a subcenter by 5 to 10%, but a 10% increase in municipality proximity to a suburban station is found to increase its chance to be part of a subcenter by 3 to 5%.

Key words: urban spatial structure, decentralization, subcenters, polycentric city, transportation
JEL classification: R11, R12, R14, R4, O2

*We are very grateful to Gilles Duranton, Laurent Gobillon, Miren Lafourcade, Daniel P. McMillen, Henry Overman, Jacques-Francçois Thisse, Matthew Turner, and Jean-Claude Prager, as well as seminar and conference participants and two anonymous referees for their comments and suggestions. We also want to thank Corentin Trévien and Thierry Mayer for their very helpful GIS maps and data on the transportation networks.Financial support from the Société du Grand Paris, the Ministerio de Ciencia e Innovación (research projects ECO2010-20718 and ECO614-52999-R (M.A. Garcia-López), and ECO2010-16934 and ECO2013-41310-R (E. Viladecans-Marsal)), Generalitat de Catalunya (research projects 2014SGR1326 (M.A. Garcia-López) and 2014SGR420 (E. Viladecans-Marsal)), and the "Xarxa de Referència d'R+D+I en Economia Aplicada' is gratefully acknowledged.

†Corresponding author. Department of Applied Economics, Universitat Autònoma de Barcelona, Edifici B, Facultat d'Economia i Empresa, 08193 Cerdanyola del Vallès, Spain (e-mail: miquelangel.garcia@uab.cat; phone: +34 93 581 4584; website: http://gent.uab.cat/miquelangelgarcialopez).

‡Ecole Normale Supérieure (PSE), 48 Boulevard Jourdan, 75014 Paris, France (e-mail: camille.hemet@psemail.eu; phone: +33 (0)1 43 13 62 25; website: http://sites.google.com/site/camillehemet).

§Department of Economics, Universitat de Barcelona, Avinguda Diagonal 690, 08034 Barcelona, Spain (e-mail: eviladecans@ub.edu; phone: +34 93 403 4825; website: http://sites.google.com/site/eviladecans).

1. Introduction

Over the past fifty years, the Paris metropolitan area has been undergoing an unprecedented process of employment decentralization, whereby the employment share of the central city (Paris) fell from more than 45% in 1968 to less than a third today. In the meantime, national and regional governments have dedicated huge amount of money to improve public transportation in the area, with particular attention paid to the rail transit network. Suggestive evidence of this investment is provided by the Regional Express Rail (*Réseau Express Régional* in French, RER henceforth), a new and more efficient regional railway network that started operating in late 1960s and represented about 600 kilometers of rails in 2010. The main goal of this paper is thus to investigate the role played by the improvement of rail transit between 1968 and 2010 on the emergence of employment subcenters in Paris.

We organize our investigation in two parts. Departing from a previous work in which we found that Paris' job decentralization and, in particular, its intrametropolitan growth is shaped by the rail transit (see Garcia-López, Hémet, and Viladecans-Marsal, Forthcoming), we first briefly describe metropolitan Paris and its rail transit, and analyze the spatial pattern of job decentralization: Is Paris decentralization diffuse or clustered around subcenters? McMillen's non-parametric approach (McMillen, 2001) allows us to identify employment subcenters in all six census years. The number of subcenters grew from 21 in 1968 to 35 in 2010, some municipalities emerging as (part of) subcenters over the period while others were dropped out during the process. More importantly, our analysis reveals that employment growth in the subcenters during this period was very intense, both in absolute and relative terms. As a result, it seems clear that the spatial pattern of job decentralization in Paris is reinforcing the polycentric nature of its urban spatial structure.

Second, we investigate the role played by transportation on the emergence of those employment subcenters: Does rail transit cause subcenter formation? Our results show that the answer is 'yes': (1) the presence of a rail station increases the probability of a suburban municipality of being (part of a) subcenter by 5%, and (2) a 10% increase in municipality proximity to a suburban station causes about a 3% increase in its probability of being (part of) a subcenter. These results are robust to subcenter size and definition, and we only find that the effects are heterogeneous in terms of the type of rail: the suburban train and the Regional Express Rail (RER). While the effects for suburban train are similar to the average results mentioned above, the effect of the RER are much higher: the presence of a RER station increases the probability of being (part of) a subcenter by 10%, and the corresponding effect of getting 10% closer to a RER station amounts to 5%.

Our investigation contributes to the literature in three ways. As far as we know, we are the first to simultaneously study employment decentralization and subcenter formation in a very long time period. As Duranton and Puga (2015) point out, very little is known about the details

of the spatial patterns of decentralized employment[1] and the frontier of knowledge was defined 15 years ago by Glaeser and Kahn (2001) and McMillen and Smith (2003). Using data at the county level for 335 US cities, Glaeser and Kahn (2001) show that job decentralization between 1950 and 1990 was mainly diffuse. On the contrary, McMillen and Smith (2003) use 1990 data at the Transportation Analysis Zone level to identify employment subcenters in 62 US cities and find that they are mainly polycentric[2]. In our paper, we analyze trends in job decentralization in a non-US city, Paris, and track the emergence of its employment subcenters from 1968 to 2010 using data at a fine spatial scale, the municipality. Our results reveal that the recent spatial pattern of job decentralization in Paris have reinforced its polycentric spatial configuration that was already apparent in 1968.

Second, this paper is the first to empirically study the role played by transportation on subcenter formation and thus the first to provide empirical evidence that supports theoretical models of urban spatial structure. As is well known, transportation plays a crucial role in the spatial distribution of residences and firms within cities. In the classical monocentric city model, transportation (accessibility) is the main factor that determines urban land use (Duranton and Puga, 2015). In non-monocentric models, the emergence of subcenters (and their number) depends on the interplay between agglomeration economies, transportation and population (Fujita and Ogawa, 1982, Helsley and Sullivan, 1991, Henderson and Mitra, 1996, Henderson and Slade, 1996, Berliant, Peng, and Wang, 2002, Lucas and Rossi-Hansberg, 2002, Anas and Kim, 1996, Berliant and Wang, 2008). From an empirical point of view, McMillen and Smith (2003) are the only ones to explore the connection between transportation and subcenters. Because their work is restricted to 1990, they can only focus on the number of subcenters and not on subcenter formation. Furthermore, since the number of subcenters is arguably determined simultaneously with transportation (and metropolitan population), this paper provides an interesting description of the data but not an estimate of causal effects as noted by Duranton and Puga (2015). In our paper, we study the causal effect of transportation on subcenter formation by using decennial census data from 1968 to 2010 to track transportation improvements and the emergence of subcenters in metropolitan Paris. We follow Duranton and Turner (2012) and address endogeneity concerns relying on Instrumental Variables (IV) techniques with historical instruments built on the 19th century railroads (1870). Our results confirm this causality.

A final contribution of our research is further our understanding of the role of transportation infrastructure on shaping cities. At the intermetropolitan level, we know that transportation fosters urban growth (Duranton and Turner, 2012), population suburbanization (Baum-Snow, 2007, Baum-Snow, Brandt, Henderson, Turner, and Zhang, 2015, Garcia-López, Holl, and Viladecans-Marsal, 2015), employment decentralization (Baum-Snow, 2010, 2014), and modify

[1]Most papers characterize the intrametropolitan location of employment by identifying subcenters and/or estimating density functions: while Giuliano and Small (1991), McDonald and Prather (1994) and McMillen (2001, 2004) show the existence of employment subcenters in some US cities, Glaeser and Kahn (2004) consider that job decentralization is mainly diffuse and directly related with cars. Others study the determinants of intrametropolitan growth (e.g. Bollinger and Ihlanfeldt, 1997, 2003, Garcia-López and Muñiz, 2013) and of growth in (within) subcenters (e.g. Giuliano and Small, 1999, Giuliano, Redfearn, Agarwal, and He, 2012).

[2]Similar results are obtained by Arribas-Bel and Sanz-Gracia (2014) using local indicators of spatial association (LISA) to identify subcenters in 359 MSAs with census tract data for 1990, 2000 and 2010.

zoning (Garcia-López, Solé-Ollé, and Viladecans-Marsal, 2015). At the intrametropolitan level, transportation affects employment and population densities (McMillen and McDonald, 1998, Redfearn, 2009*a*, Garcia-López, 2010, 2012, Brooks and Lutz, 2016), local growth (Garcia-López, 2012, Mayer and Trévien, 2015, Garcia-López *et al.*, Forthcoming), the housing market and the property values (Baum-Snow and Kahn, 2000, Bowes and Ihlanfeldt, 2001, McMillen and McDonald, 2004, Baum-Snow and Kahn, 2005, Gibbons and Machin, 2005, Ryan, 2005, Redfearn, 2009*b*, McMillen and Redfearn, 2010, Diao, Zhu, and Zhu, 2016, Li, Yang, Qin, and Chonabayashi, 2016), land use (Hurst and West, 2014, Schuetz, 2015), air pollution (Chen and Whalley, 2012) or cities' gentrification (Grube-Cavers and Patterson, 2015). Our paper distinguishes itself from this large literature by showing that transportation also influences the spatial pattern of decentralized employment within cities by fostering the emergence of employment subcenters, a topic that remains largely unexplored, to the best of our knowledge.

The paper is organized as follows. Section 2 describes the main features of job location and railroad changes in the Paris metropolitan area over the past fifty years. Section 3 explores the pattern of this process, revealing that is has been clustered rather than diffused. Section 4 analyses the influence of railroad transit on the employment subcenters formation and finally Section 5, summarizes and concludes.

2. The Paris metropolitan area

France is organized into *régions*, administrative units with a governing body elected by direct vote (the Regional Council). Regions are divided into *départements*, also with a government (the General Council of the Department). Departments are made up of *communes* (municipalities), the lowest level of the administrative division governed by the City Council. Some big municipalities are subdivided into *arrondissements*. In terms of planning, the national government sets the legal framework, but it does not establish a national spatial plan. On the contrary, regions and municipalities are responsible of planning policies: the former prepare the general strategic plan (for example, infrastructure plans) and the latter develop local land use plans and approve building permits. While in the past, regional plans provided only non-binding guidelines and strategies, after November 2014 local land use plans need to adopt the logic of the regional plan (OECD, 2017).

This study focuses on the Paris metropolitan area, a French administrative region known as *Ile de France*. It is the densest and most populated metropolitan area in France, with 9.8 inhabitants per hectare in 2010 for a total of 11,786,234 inhabitants and 1,200,000 hectares. It is also the region with the highest employment density, with a total of 5,668,902 jobs in 2010 (21.6% of French employment). From a sector point of view, 0.2% of metropolitan jobs are in the agricultural sector, 5.2% in the construction sector, 8.4% in industry and the remaining 86.2% in the tertiary sector (trade, services, public administration, education, among others) (Garcia-López *et al.*, Forthcoming).

Metropolitan Paris is composed of its CBD, the city of Paris (with 20 *arrondissements*), and 1,280 suburban municipalities. The average land size of the CBD is 526 ha (with a standard deviation

of 435 ha). For suburban municipalities, the average size is 724 ha (466 ha SD) for the ones located less than 25 km from Paris (average distance to the CBD), and it is 1,146 ha (953 ha SD) for those located beyond the 25 km. In our empirical strategy we take into account these differences in land size by using density variables (as dependent or control variables).

Relying on detailed population and employment data at the municipal level from six census waves (1968, 1975, 1982, 1990, 1999 and 2010), we are able to track the evolution of the urban spatial structure of the Paris metropolitan area over the past forty years. We use municipal data because census tract data are only available from 1990[3].

We also use precise transportation data, provided by Mayer and Trévien (2015) and the Institut d'Aménagement et d'Urbanisme (IAU), to characterize the changes in the area's public transportation over the period.

2.1 *Employment decentralization in Paris*

Garcia-López *et al.* (Forthcoming) show that metropolitan Paris is undergoing a process of employment decentralization in which its central city (Paris) loses jobs in favor of suburban locations. In particular, the number of jobs in central Paris declined by 7.1% between 1968 and 2010. Over the same period, while the metropolitan area as a whole grew by about one third, the share jobs located in central Paris dropped from 45.3 to 31.7%. As a result, this evolution of jobs' location reveals that metropolitan Paris is decentralizing both in absolute and relative terms.

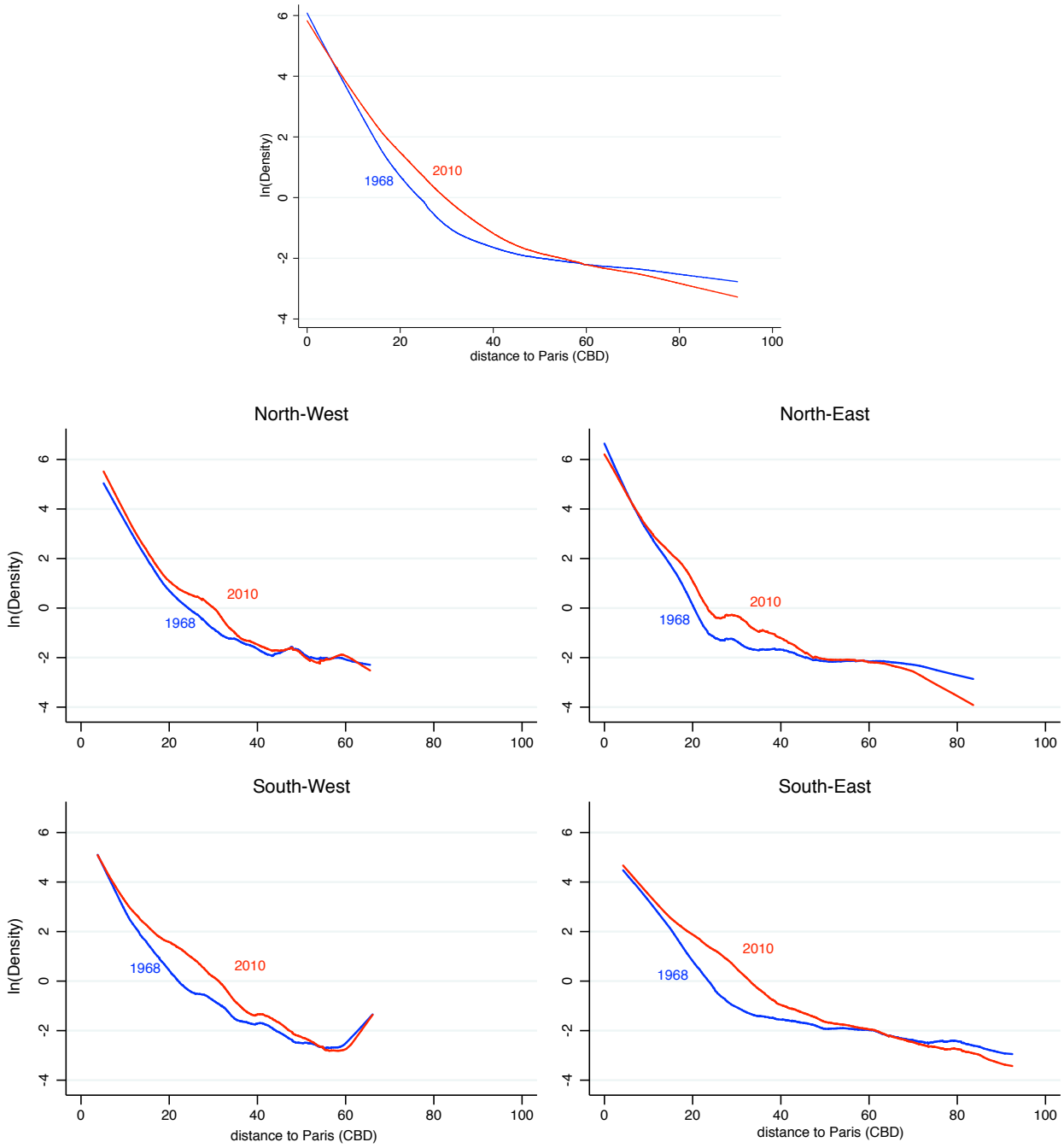
Additional evidence of this decentralization process comes from estimating the traditional monocentric density function for each census year. Using municipal data, we regress the log of employment density on the distance to the center of the CBD (Paris). In order to take non-linearities into account, our estimations are based on a non-parametric method known as Locally Weighted Regression (LWR), with a bandwidth of 0.5[4] (McMillen, 2001)[5]. The LWR density estimates drawn in Figure 1 are for the whole city (top graph) and for four city quadrants. These graphs clearly illustrate the decentralization process between 1968 (blue line) and 2010 (red line). Indeed, between these two dates, employment density in the most central municipalities (0 to 10 km from the CBD) decreased (top and North-East graphs), did not change (South-West graph) or slightly increased (North-West and South-East graphs), while it increased greatly in municipalities located between 10 and 60 km from the CBD (all five graphs), in line with a (relative and/or an absolute) decentralization process. We can also note a reduction in employment density for the most peripheral (mostly rural) municipalities (more than 60 km from the CBD). This, combined with the increase observed for the suburban municipalities indicate the emergence of new suburban subcenters and reinforcement of existing ones around kilometers 30 (all five graphs), 40 (North-East and South-West graphs), and 50 and 60 (North-West graph).

[3]With a total area of 1,200,000 hectares and 1,300 observations, Paris is closer to Los Angeles with 1,146 tracts and 910,000 ha (see Giuliano and Small, 1991) than to Chicago with 14,290 tracts and 920,000 ha (see McMillen and McDonald, 1998). As McMillen (2001) points out, subcenter identification (in Section 3) depends on the size of the unit of observation: large observations may produce fewer subcenter sites than more disaggregated data.

[4]A window size of 0.5 means that the nearest 50% of observations are weighted.

[5]Results are similar when using window sizes of 0.1, 0.3, 0.8 and 1.0 and are available upon request.

Figure 1: Employment decentralization in metropolitan Paris, 1968–2010



Note: Density estimates based on LWR with a window size of 0.5.

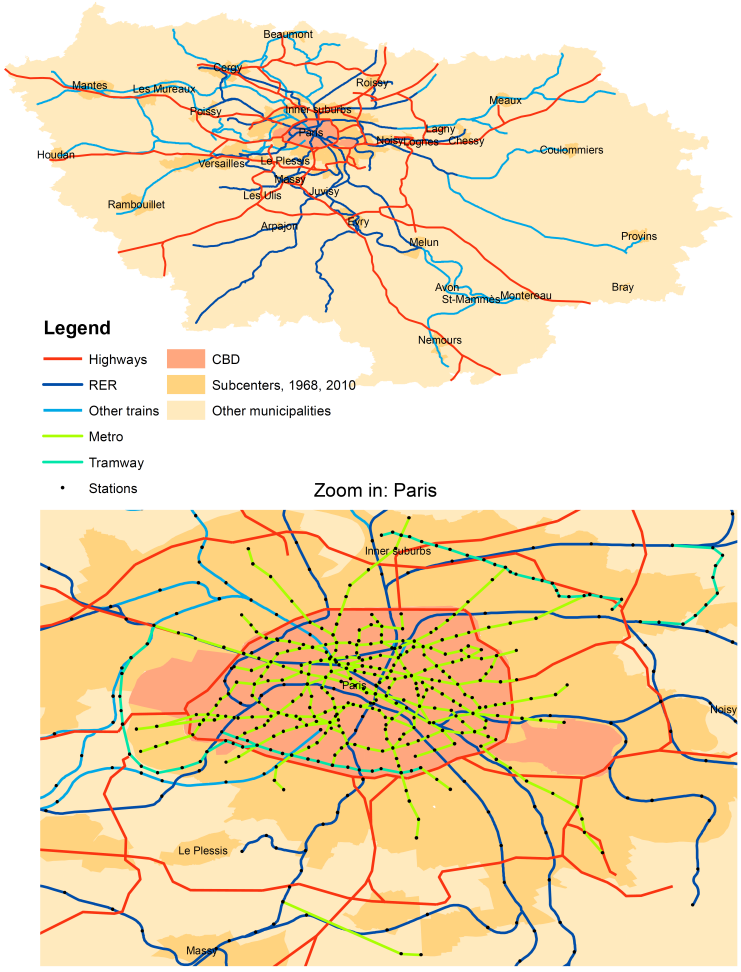
2.2 *Rail transit in Paris*

As documented with more details in Garcia-López *et al.* (Forthcoming), the public transportation infrastructure of the Paris metropolitan area has dramatically improved since the 1960s and today is mostly based on a railroad network made of more than 1,600 km of lines and including four

network types, as illustrated in Figure 2. First, a suburban train (henceforth *train)* connecting Paris to the rest of the metropolitan area (suburbs as well as some of the most remote rural municipalities), that underwent substantial improvement over the 1960s.

Second, the Paris region is endowed with a regional express network (*Réseau Express Régional* in French, RER henceforth) which started operating in late 1960s. Like the train, the RER connects Paris to the suburbs, but for a shorter maximum distance of about 30 km. Most of the RER lines follow the train lines and were designed to improve the existing train network. An important distinction between the train and RER networks is that the latter has connections within Paris. This means the RER enables passengers to commute from one part of the Paris Metropolitan Area to another, going through Paris, but without having to switch to another train to cross the city. This represents a clear improvement to regional transit overall. As a whole, the RER network increased its number of lines from 1 to 5, its total length from 39 to 587 km, its number of stations from 22 to 243, and its number of municipalities with stations from 16 to 167 between 1975 and 2010.

Figure 2: Transportation infrastructures in metropolitan Paris, 2010



In addition to these regional railroad networks, Paris is characterized by a very dense subway system (*métro* henceforth), which started in 1900, and is mainly connecting areas within Paris. Between 1968 and 2010, the métro network kept expanding, such that a few métro stations are now located beyond Paris, in the immediate outskirts of the CBD.

The city of Paris and its closest suburban area (the first ring of municipalities out of Paris) also enjoy a tramway network. This fourth network is much more recent, dating back from the beginning of the 1990s, and is still expanding.

Finally, it is important to note that the origins of these rail transit networks can be traced back to the 19th century. In the empirical strategy of the following sections, we exploit this link to correct for the potential biases related to the endogenous location of rail stations and subcenter formation. In particular, we rely on IV techniques with two historical instruments as sources of exogenous variation: the distance to the nearest 1870 railroad line and a dummy variable indicating whether a given municipality was crossed by a railroad line in 1870. In Appendix A we extensively document and discuss the validity of these 1870 rail variables as instruments for the location of modern railroad stations.

3. Is Paris decentralization diffuse or clustered around subcenters?

After establishing that the Paris metropolitan area went through a job decentralization process between 1968 and 2010, we now want to characterize the spatial pattern of this process: Does decentralization follow a *polycentric* spatial pattern, reinforcing existing secondary centers (subcenters) and/or fostering the emergence of new ones? Or does it rather reflect a *dispersed* spatial pattern, in which suburban land is occupied by low-density settlements? To answer these questions, we first identify subcenters for each census year between 1968 and 2010 before analyzing the evolution of employment inside these subcenters vs. non-central locations between 1968 and 2010.

3.1 Identifying and characterizing subcenters

The empirical literature has proposed alternative procedures to identify subcenters. The most used are those based on density and employment thresholds (Giuliano and Small, 1991, Giuliano, Redfearn, Agarwal, Li, and Zhuang, 2007, Muñiz, Garcia-López, and Galindo, 2008) and on employment and/or population density peaks (McDonald, 1987, McDonald and Prather, 1994, Craig and Ng, 2001, MuÑiz, Galindo, and Garcia-López, 2003, Redfearn, 2007, Garcia-López, 2010, Craig, Kohlhase, and Perdue, 2016). Despite using different methods[6], they all agree that an employment subcenter is a place with a significantly larger employment density than nearby locations that has a significant effect on the overall spatial distribution of jobs.

We identify employment subcenters using the method first developed by McDonald and Prather (1994) and improved by McMillen (2001) because it is consistent with theory and, in

[6] Although in general they identify different sets of subcenters in the same city (for instance, see Redfearn (2007) in Los Angeles when using his own methodology than when relying on Giuliano and Small (1991) and McMillen (2001)'s procedures), there are examples in which different methods lead to similar sets of subcenters (see for instance Craig *et al.* (2016) in Houston when using McMillen (2001) and Redfearn (2007)'s strategies),

particular, based on (deviations from) the monocentric model: The main idea is to estimate densities following a monocentric spatial pattern; the predicted densities obtained are then substracted from the corresponding real densities; and the positive and statistically significant residuals are finally selected.

While McDonald and Prather (1994) estimate by Ordinary Least Squares (OLS) a two dimensional density function (log of employment density on the distance to CBD), as in Figure 1, McMillen (2001) suggests estimating a three-dimensional density function (log of employment density on north-south and east-west distances to CBD) with a Locally Weighted Regression (LWR). Both improvements allow to take into account geographical differences, which, in terms of the spatial pattern of densities, can occur in any direction from the CBD (e.g. steeper density gradients on the north side than on the south side of the city). They additionally allow to define any type of monocentric spatial pattern: concave, convex or linear.

We therefore estimate the following employment density function:

$$\begin{aligned} \ln(\text{Employment density}) = \gamma_0 + \gamma_1 \times \text{north-south distance to CBD} \\ + \gamma_2 \times \text{east-west distance to CBD} \end{aligned} \tag{1}$$

where density is measured as jobs per hectare, and distances are in kilometers. The CBD is defined as the 20 *arrondissements* that make up the city of Paris. Distance to CBD is the distance to the centroid of the 4th *arrondissement* (corresponding to the town-hall of Paris).

Since our estimates are based on LWR, we need to define a bandwidth. As McMillen (2001) points out, this is a critical choice because we need a monocentric benchmark. We experiment with alternative window sizes ranging from 1% to 9% and from 10% to 90% (see Table B.1 in Appendix B). After visual inspection, we find that the first monocentric spatial structure appears when the nearest 50% observations are included in each local regression. Interestingly, this is the value used by McMillen (2001) for some US cities. We also experimented with a selection rule based on the Akaike information criterion. However, the selected window size (7%) was clearly related to a polycentric spatial structure (see and compare Tables B.1 and B.2 in Appendix B).

Second, for each site we compute the residual as the difference between the log of real employment density and the estimated log of employment density. We then select those that are significantly positive, according to their own standard errors that can vary over space (McMillen, 2001). We use two critical thresholds, 1.96 and 1.64, that are associated with a 5% and a 10% significance level, respectively.

Finally, we group the selected sites in subcenters when they are contiguous. We use a "queen" criterion for contiguity: two sites (municipalities) are contiguous if they share at least one point in their boundaries. See McMillen (2001, 2003) and Garcia-López (2010) for further details on this procedure.

This methodology enables us to identify subcenters for each census year between 1968 and 2010, which are described in Table 1. For each year, we report two figures, corresponding to subcenters identified using positive residuals significant at the 5 and at the 10% level, respectively. From Panel A, we can see that the number of subcenters identified at the 5% level (respectively, at the 10% level) increased from 20 to 26 (respectively from 21 to 35) between 1968 and 2010. Panel B

reveals that these subcenters hosted 1,756,000 jobs in 2010 (respectively 1,979,000), corresponding to an increase of about 600,000 jobs since 1968 (respectively 560,000). We can also observe that subcenters are heterogeneous in terms of size, with an increasing number of large subcenters (more than 20,000 jobs), in line with the decentralization process: in 1968, between 15 and 20% of subcenters hosted more than 20,000 jobs, contrasting with a range of 42 to 50% in 2010. Panel C shows the share of metropolitan employment in the CBD and the subcenters. While the CBD share decreased from 45% in 1968 to 32% in 2010, subcenters concentrated about a third of all city jobs during the studied period. Finally, we want to emphasize that the number of subcenters and the number of jobs in the subcenters do not differ much whether the subcenters are identified using positive residuals a the 5% level or at the 10% level. We will henceforth use the subcenters identified at the 10% level in our analysis.

Table 1: Employment subcenters in metropolitan Paris, 1968–2010

	1968		1975		1982		1990		1999		2010	
Resid. significant at:	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Panel A: Number of subcenters												
All subcenters	20	21	26	27	30	32	29	34	31	34	26	35
$\geq$ 10,000 jobs	7	8	13	14	18	16	19	21	19	19	19	23
$\geq$ 20,000 jobs	3	4	8	7	9	12	13	13	13	12	13	15
Panel B: Jobs ('000) in subcenters												
All subcenters	1,152	1,419	1,319	1,667	1,237	1,665	1,373	1,788	1,441	1,782	1,756	1,979
$\geq$ 10,000 jobs	1,088	1,350	1,254	1,601	1,169	1,582	1,326	1,720	1,369	1,693	1,714	1,909
$\geq$ 20,000 jobs	1,022	1,290	1,184	1,501	1,040	1,513	1,252	1,618	1,281	1,602	1,624	1,788
Panel C: Metropolitan shares in the CBD and subcenters												
CBD (Paris)	45.3%	45.3%	41.0%	41.0%	38.4%	38.4%	35.8%	35.8%	31.8%	31.8%	31.7%	31.7%
All subcenters	26.9%	33.2%	28.2%	35.7%	26.3%	35.4%	27.0%	35.2%	28.6%	35.3%	31.0%	34.9%
$\geq$ 10,000 jobs	25.4%	31.6%	26.8%	34.2%	24.8%	33.6%	26.1%	33.9%	27.1%	33.6%	30.2%	33.7%
$\geq$ 20,000 jobs	23.9%	30.2%	25.3%	32.1%	22.1%	32.2%	24.7%	31.9%	25.4%	31.8%	28.6%	31.5%

Note: LWR estimates use a window size of 0.5 (i.e., the nearest of the 50% observations).

Table 2 further describes the identified subcenters (at the 10% level) regarding the number of municipalities they encompass: All of them in Panel A, with more than 10,000 jobs in Panel B, and with more than 20,000 jobs in Panel C. We can first notice that the number of municipalities that form a subcenter of their own or that are part of a subcenter is quite stable, from 88 (73, 67) in 1968 to 89 (76, 65) in 2010. We can also observe a certain stability in the composition of subcenters: among all the municipalities belonging to a subcenter, 57 (42, 38) are constantly identified as a subcenter (or as part of a subcenter) over time. For the case of Los Angeles, Redfearn (2009*a*) also finds evidence of the persistence of employment subcenters during the 20th century. The remaining municipalities may have emerged as part of a subcenter at some point, or, alternatively, stopped being considered as such. The last two lines of each Panel also report subcenters' surface and show that 40 to 50% of land was associated with constantly identified subcenters in 2010. Furthermore, total subcenters' land size increased by 20% between 1968 and

2010, despite their constant number. This indicates that the *new* subcenters are spatially larger.

Table 2: Municipalities and land in employment subcenters in metropolitan Paris, 1968–2010

	1968	1975	1982	1990	1999	2010
Panel A: All subcenters						
Number of municipalities						
All municipalities	88	97	93	95	95	89
Emerging as (part of) subcenters	–	18	7	11	12	5
Disappearing as (part of) subcenters	–	9	11	9	12	11
Always in subcenters	57	57	57	57	57	57
Hectares of land						
All municipalities	50,048	58,122	59,159	61,947	67,055	60,245
Always in subcenters	31,140	31,140	31,140	31,140	31,140	31,140
Panel B: Subcenters $\geq$ 10,000 jobs						
Number of municipalities						
All municipalities	73	81	75	80	78	76
Always in subcenters	42	42	42	42	42	42
Hectares of land						
All municipalities	41,038	49,086	48,998	54,296	57,582	51,514
Always in subcenters	21,814	21,814	21,814	21,814	21,814	21,814
Panel C: Subcenters $\geq$ 20,000 jobs						
Number of municipalities						
All municipalities	67	73	69	68	69	65
Always in subcenters	38	38	38	38	38	38
Hectares of land						
All municipalities	35,047	40,834	43,401	47,102	45,492	42,458
Always in subcenters	17,613	17,613	17,613	17,613	17,613	17,613

Note: Employment subcenters identified using McMillen (2001)'s method with a LWR window size of 50%, and for positive residuals significant at the 10% level.

3.2 *The spatial pattern of decentralization*

In order to determine the spatial pattern of employment decentralization, we now compare the evolution of employment inside and outside employment centers over the period of interest.

Table 3 displays the number and the percentage of jobs located in each type of area for all census years (columns 1 to 6), and the corresponding variation between 1968 and 2010 (column 7). In Panel A, the figures correspond to the 20 *arrondissements* of Paris, the CBD. Panel B shows figures for different classifications of subcenters. The *All subcenters* category refers to subcenters identified at a given date, some of them appearing or disappearing as subcenters between two waves. In other words, the geography is not constant as the municipalities included in this category differ between two points in time (as shown in Table 2 Panel A). By contrast, the *Always subcenters* and the *Subcenters in 2010 (not all years)* categories consider constant geography: the former by including the 57 municipalities that are identified as a subcenter (or part of one) in all six years, the latter by including the 32 municipalities identified as (part of) a subcenter in

2010, but not in all previous years. In Panel C, we assess the importance of employment centers by jointly considering the CBD and the *All subcenters* categories. Finally, figures for non-central municipalities are in Panel D. Similar to Panel B, geography is not constant in the *All non-central municipalities* category, and it is constant in the *Always non-central* and the *Non-central in 2010 (not all years)* categories.

Table 3: The spatial pattern of decentralized employment in metropolitan Paris, 1968–2010

	1968 [1]	1975 [2]	1982 [3]	1990 [4]	1999 [5]	2010 [6]	1968–2010 [7]
Panel A: CBD							
Paris (20 *arrondissements*)	1,936	1,918	1,808	1,815	1,601	1,798	-138 (-7.1%)
Metropolitan share	45.3%	41.0%	38.4%	35.8%	31.8%	31.7%	
Panel B: Subcenters							
All subcenters	1,419	1,667	1,665	1,788	1,782	1,979	560 (39.5%)
Metropolitan share	33.2%	35.7%	35.4%	35.2%	35.3%	34.9%	
Always subcenters	1,028	1,136	1,112	1,168	1,183	1,372	344 (33.5%)
Metropolitan share	24.0%	24.3%	23.6%	23.0%	23.5%	24.2%	
Subcenters in 2010 (not all years)	104	221	316	438	506	606	502 (482.7%)
Metropolitan share	2.4%	4.7%	6.7%	8.6%	10.0%	10.7%	
Panel C: All employment centers (A+B)							
Paris + All subcenters	3,355	3,585	3,473	3,603	3,383	3,777	422 (12.6%)
Metropolitan share	78.4%	76.7%	73.8%	71.0%	67.1%	66.6%	
Panel D: Non-central municipalities							
All non-central municipalities	992	1,090	1,232	1,472	1,659	1,893	901 (90.8%)
Metropolitan share	21.6%	23.3%	26.2%	29.1%	32.9%	33.4%	
Always non-central	754	880	961	1,136	1,260	1,383	629 (83.4%)
Metropolitan share	17.6%	18.8%	20.4%	22.4%	25.0%	24.4%	
Non-central in 2010 (not all years)	455	518	508	519	493	510	55 (12.1%)
Metropolitan share	10.6%	11.1%	10.8%	10.2%	9.8%	9.0%	

Note: Employment values are thousands of jobs. Growth rates are in parentheses. Employment subcenters identified using McMillen (2001)'s method with a LWR window size of 50%, and for positive residuals significant at a 10% level.

The figures reported in Table 3 confirm the decentralization process commented in Section 2: the CBD lost 138,000 jobs (-7%) between 1968 and 2010 and reduced its metropolitan share of jobs from 45% to 31% (Panel A).

The decentralization process and the overall employment growth experienced by the city benefited subcenters (Panel B). Despite always concentrating around a third of all employment in the Paris metropolitan area, their jobs grew from 1.5 to 2 millions (39%) in the studied period. If we consider constant geographies, we can see that employment growth took place in the *Always subcenters* category, with 344,000 new jobs (33%) and, in particular, in those municipalities that ended up as subcenters in 2010 (the *Subcenters in 2010 (not all years)* category), with 500,000 new jobs (483%).

Similarly, both processes also benefited non-central municipalities (Panel D), which grew in 900,000 jobs (91%). While most growth took place in the *Always non-central* category (629,000 jobs, 83%), municipalities that ended up as non-central in 2010 only grew in 55,000 jobs (12%).

Finally, the importance of the polycentric nature of Paris is clear when we jointly consider all employment centers (Panel C) and compare with non-central municipalities (Panel D). In this sense, despite the reduction in their metropolitan share between 1968 and 2010, employment centers still concentrated 67% of jobs in 2010. On the contrary, non-central municipalities represented around one fifth of all jobs in the metropolitan area in 1968, but up to one third in 2010. In other words, while 3,8 millions of jobs were concentrated in 90 municipalities (89 subcenters + Paris), 1,8 millions of jobs were located in 1,191 municipalities in 2010.

In summary, the increasing number of employment subcenters, their high employment growth rates and the fact that they concentrate a third of metropolitan jobs allow us to conclude that job decentralization in Paris is clustered around subcenters. We have to admit though that a small part of job decentralization is diffuse.

4. Does rail transit cause subcenter formation?

After analyzing the job decentralization process in the Paris metropolitan area, we now turn to the most important part of this paper, where we contribute to the literature by establishing that rail transit causes subcenter formation. To answer to this key question, we proceed in two steps. We first investigate whether the existence of a rail station in a suburban municipality increases the probability that this municipality becomes (part of) a subcenter. Then, we examine whether proximity to rail stations also increases the likelihood of becoming (part of) a subcenter, even when the station is not built on the municipal ground.

In both steps, our empirical strategy consists in regressing the probability that a municipality becomes a subcenter on a *rail station variable*. In Section 4.1, where we explore the role of the existence of a rail station, this variable indicates whether there is a station within the administrative boundaries of the municipality or the number of stations and lines in the municipality. Alternatively, in Section 4.2, this variable measures the distance between a municipality and the closest station.

We control for characteristics related to Paris urban spatial structure of metropolitan such the distance to the CBD, and employment and population densities. Also for municipal geography with altitude, index of terrain ruggedness, and elevation range variables. History variables are the population levels between 1962 and year t-1 and dummy variables for municipalities (1) that were Roman settlements (based on DARMC[7] maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), and (3) between the 16th and the 19th centuries (based on Bairoch, 1988), and (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps). Socioeconomic variables are the unemployment rate, the

[7]The Digital Atlas of Roman and Medieval Civilizations (DARMC) is a website with free GIS maps for the Roman and medieval worlds (see darmc.harvard.edu/icb/icb.do).

shares of employment in manufacturing, in construction, and in services, the share of executives and professionals, and the share of population with university degree.

The general equation, that we estimate using probits, can thus be expressed as follows:

$$\begin{aligned}\text{Prob}(\text{subcenter}_{i,t}) = \delta_0 &+ \delta_1 \times \text{Rail station variable}_t \\ &+ \delta_2 \times \ln(\text{densities}_{i,t}) + \delta_3 \times \ln(\text{distance to CBD}_i) \\ &+ \sum_i (\delta_{4,i} \times \text{geography}_i) + \sum_i (\delta_{5,i} \times \text{history}_i) \\ &+ \sum_i (\delta_{6,i} \times \text{socioeconomy}_{i,t})\end{aligned} \tag{2}$$

In order to correct for the potential biases related to the endogenous location of rail stations, we use a dummy variable indicating whether a given municipality was crossed by a rail (a train line) in 1870 (Section 4.1) and the distance to the nearest 1870 railroad line (Section 4.2) and as instruments, as documented and discussed in Appendix A. However, the use of these historical instruments comes with a caveat: since they are time invariant, we can not estimate using panel data techniques. As a result, we pool all observations together, irrespective of the year, and include year fixed-effects in our regressions. Since only suburban trains operated in 1968 and RER started operating between late 1960s and mid 1970s, we show results for the 1968–2010 and the 1975–2010 periods. As a robustness, we also run some cross-sectional regressions.

4.1 Do rail stations lead to subcenters?

In order to establish whether the existence of a rail station in a suburban municipality increases the probability that this municipality becomes (part of) a subcenter, we estimate Eq. (2) using the 1,280 suburban municipalities (excluding the 20 *arrondissements* of Paris).

The marginal effects of the corresponding (second-stage) results are displayed in Table 4. Panel A shows results when estimating by Ordinary Least Squares (OLS) and Panel B their Two-Stage Least Squares (TSLS) counterparts. Columns 1 to 7 presents results estimated using all suburban municipalities. In columns 1 and 2, the station variable represents the number of lines times station, which counts the total number of lines having a stop in a municipality (it can be seen as a weighted count of the number of stations). The station variable then simply counts the number of stations in columns 3 and 4, and indicates the existence of a station in columns 5 to 7.[8] We will restrict our comments on the specifications that includes all control variables (densities, distance to CBD, geography, history and socioeconomy), after noting that the marginal effects are significantly reduced in the conditional regressions[9].

[8]Therefore, if a municipality has two stations, with n_1 lines stopping in one station and n_2 lines in the other one, the "number of lines-stations" variable takes a value of $n_1 + n_2$, the "number of stations" variable takes a value of 2, and the dummy variable is equal to 1.

[9]The importance of these control variables can be noticed in the evolution of the pseudo R^2: without controls, its value is below 20%; with them, it is close to 90%. The variables driving the big jump are employment density and population density. They increase pseudo R^2 to 70%, something expected because the identification of subcenters use a job density criterion. Adding the distance to CBD, the explanatory power increases to 85%. Finally, historic, geographic and socioeconomic variables only add around 3 percent points. However, remember that these variables are included because our historical instruments need to fulfill the exclusion restriction (see Appendix A). More detailed gradual results are available upon request.

While OLS conditional results in Panel A columns 2 and 4 show non-significant zero marginal effects, TSLS result in Panel B column 4 indicates that an additional station increases the probability that the municipality becomes (part of) a subcenter by 2.8%. This effect is exactly the same as that of having an additional line stopping in the municipality (Panel B column 2). This does not come as a surprise given that most of the suburban municipalities are only crossed by one train line, so that the number of lines-stations is actually very close to the number of stations.

Table 4: The effect of rail stations on subcenter formation, (IV) Probit Marginal effects

Dependent var.:	Probability of being (part of) a subcenter									
	All suburban municipalities							Without always subcenters		
Variable type:	Number of lines-stations		Number of stations		Dummy=1 for municipality with stat			Dummy=1 for municipality with stat		
Period/year:	68-10 [1]	68-10 [2]	68-10 [3]	68-10 [4]	68-10 [5]	68-10 [6]	75-10 [7]	75-10 [8]	1975 [9]	2010 [10]
Panel A: OLS results										
Station variable	0.068[a]	-0.000	0.069[a]	-0.000	0.141[a]	0.003	0.004	0.004	0.010	0.010[c]
	(0.007)	(0.001)	(0.007)	(0.001)	(0.013)	(0.004)	(0.005)	(0.004)	(0.007)	(0.006)
Pseudo R^2	0.19	0.88	0.18	0.88	0.17	0.88	0.87	0.81	0.88	0.85
Panel B: TSLS results										
Station variable	0.087[a]	0.027[b]	0.092[a]	0.028[b]	0.142[a]	0.040[c]	0.047[c]	0.053[b]	0.035[c]	0.075[c]
	(0.024)	(0.012)	(0.025)	(0.013)	(0.046)	(0.022)	(0.025)	(0.022)	(0.021)	(0.045)
F-S statistic	73.24	26.99	71.68	25.34	127.12	41.11	39.41	38.04	33.81	31.63
Instrument:	Dummy=1 if municipality is crossed by a 1870 rail									
ln(Densities)	N	Y	N	Y	N	Y	Y	Y	Y	Y
ln(Dist to CBD)	N	Y	N	Y	N	Y	Y	Y	Y	Y
Geography	N	Y	N	Y	N	Y	Y	Y	Y	Y
History	N	Y	N	Y	N	Y	Y	Y	Y	Y
Socioeconomy	N	Y	N	Y	N	Y	Y	Y	Y	Y
Observations:	7680	7680	7680	7680	7680	7680	6400	6115	1223	1223
	(1280 suburban municipalities × 6 census years)						(1280×5)	(1223×5)		

Notes: Regressions in columns 1 to 8 include year effects. Robust standard errors and are in parentheses (and are clustered by municipality in regressions in columns 1 to 8). [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

Regarding the existence of a station, we estimate the same non-significant zero OLS effect (Panel A column 6), and a slightly larger TSLS effect, around 4% over the whole period (Panel B column 6). This TSLS effect increases to 4.7% when we focus on the 1975-2010 period (Panel B column 7), suggesting either that the effect is delayed in time, or that the transportation system built after 1975 (mostly the RER) explains a larger part of the overall effect.

In order to dig further into this time variation, we focus on the 1975-2010 period in columns 8 to 10, taking the municipalities systematically identified as subcenters out of the sample. While most of the estimated OLS marginal effects remain non-significant, the TSLS estimated effect over the period jumps to 5.3% (Panel B column 8) confirming the idea of an reinforced effect in the most recent period (the effect goes from 3.5% in 1975 to 7.5% in 2010 (Panel B columns 9 and 10), but the difference is not significant).

We check the robustness of our results in Appendix C. In Table C.1 Panel A we show that estimates are robust to subcenter size and definition: the effect is always between 3.5% and 4.4%, whether we focus on subcenters with more or less than 50,000 jobs (columns 1 and 2), and whether we rely on subcenters identified using the 5% criterion (columns 3 and 4) (instead of 10% in the main results). In Table C.2 Panel A we test the validity of our identification strategy by following Chandra and Thompson (2000) and dropping some observations. Since the 1870 railroad network was probably planned to serve the most important municipalities during the 19th centuries, we first drop municipalities that were important. We do not have population data at the municipality level for these years, as a result we use our historical dummy variables that signal the most important towns through history. That is, we drop municipalities that were Roman settlements and/or major towns during the 10th and 19th centuries and/or with a monastery built between the 12th and 16th centuries (columns 1 and 2). Alternatively, we also drop observations of municipalities with a rail station built during the 19th century (columns 3 and 4). In both cases, results still show a significant and positive effect of having a rail station on the probability of becoming (part of) a subcenter.

We now refine our results by investigating the train station effect, looking alternatively at two different train types: suburban train vs. RER. The corresponding results are displayed in columns 1 to 4 of Table 5 for the former train type, and in columns 5 to 8 for the latter. Again, OLS and TSLS results are in Panels A and B, respectively. Over the 1968-2010 period, we estimate that the presence of a suburban train station in a municipality (column 2) increases the likelihood that it becomes a subcenter by 0.8% (OLS Panel A) and 3.4% (TSLS Panel B). As before, this effect slightly increases (to 0.9-3.9%) when we focus on the 1975-2010 period (column 3), and to 0.7-4.4% once we exclude municipalities that are always identified as a subcenter (column 4).

On the other hand, OLS and TSLS results for the RER (columns 5 to 8) show significant but opposite effects: while all conditional OLS marginal effects are negative, TSLS marginal effects are positive. In particular, the TSLS results for the RER reveal that this particular type of train has a much stronger impact on subcenter formation than the suburban train. The existence of a RER station is indeed found to increase the probability of becoming (part of) a subcenter by 14% over the 1968-2010 period (Panel B column 6), an effect about four times as large as for suburban trains. Interestingly, looking at the later period (after 1975) does not show a significantly different effect (13.5%, Panel B column 7).

The difference between our TSLS estimates and their OLS counterparts suggests that construction of railroads in Paris is endogenous. Why? It may be due to classical measurement error, but, since similar TSLS-OLS differences are found when using different measures of railroads (number of lines-stations vs. number of stations vs. dummy for municipalities with stations), we rule out this possibility.

It may also be due to a negative correlation between the existence of a station and the error term because of missing variables or reverse causation. Despite controlling for geography, population history, socioeconomy and year fixed-effects, the possibility remains that the TSLS-OLS differences could be explained by a missing variable such as the municipal land use regulations, which could be associated with a higher probability of becoming (part of) a subcenter and with

not having a railroad station. Alternatively, it may be that conditional on controls, less dense non-central municipalities on average experience positive shocks and receive an station. Although not reported for reasons of space, first-stage results confirm this through a negative estimated marginal effect of -0.011 for the log of job density (suburban train probits) and of -0.025 for the log of population density (RER probits).

Table 5: The effect of train and RER stations on subcenter formation, (IV) Probit Marginal effects

Dependent var.:	Probability of being (part of) a subcenter							
	Train stations				RER stations			
	All suburban municipalities			Without always sub	All suburban municipalities			Without always sub
Period:	68-10	68-10	75-10	75-10	68-10	68-10	75-10	75-10
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: OLS results								
Station dummy	0.100[a]	0.008[b]	0.009[b]	0.007[b]	0.123[a]	-0.012[a]	-0.012[a]	-0.006[b]
	(0.014)	(0.004)	(0.004)	(0.003)	(0.017)	(0.003)	(0.004)	(0.003)
Pseudo R^2	0.06	0.88	0.87	0.81	0.04	0.88	0.88	0.81
Panel B: TSLS results								
Station dummy	0.175[a]	0.034[c]	0.039[c]	0.044[a]	0.680[a]	0.140[a]	0.135[a]	0.102[a]
	(0.049)	(0.020)	(0.021)	(0.017)	(0.191)	(0.046)	(0.036)	(0.034)
F-S statistic	114.12	44.56	43.62	36.83	29.43	10.74	9.91	11.21
Instrument:	Dummy=1 if municipality is crossed by a 1870 railroad line							
ln(Densities)	N	Y	Y	Y	N	Y	Y	Y
ln(Dist to CBD)	N	Y	Y	Y	N	Y	Y	Y
Geography	N	Y	Y	Y	N	Y	Y	Y
History	N	Y	Y	Y	N	Y	Y	Y
Socioeconomy	N	Y	Y	Y	N	Y	Y	Y
Observations:	7680	7680	6400	6115	7680	7680	6400	6115
	(1280 × 6 years)		(1280×5)	(1223×5)	(1280 × 6 years)		(1280×5)	(1223×5)

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

4.2 Does proximity to rail stations lead to subcenters?

We now want to examine the effect of the distance to a train station on subcenter formation. This presents a double advantage: it enables us to measure the spatial effect of the presence of a train station, and allows to consider the effect of a train station on municipalities that do not possess any. In other words, we investigate whether the effect of a rail station can go beyond the boundaries of the municipality where the station is located.

To this aim, we now use the distance (in log) of a municipality's centroid to the closest train station as the *train station variable*. Table 6 reports OLS and TSLS results in Panels A and B, respectively. Columns 1 to 4 present results for all railroad stations. In columns 5 to 8, we focus

on suburban train stations. Finally, we analyze the effect of proximity to RER stations in columns 9 to 11.

Table 6: The effect of rail proximity on subcenter formation, (IV) Probit Marginal effects

Dependent var.:	Probability of being (part of) a subcenter										
	All suburban stations				Train stations				RER stations		
	All suburban muni			Without muni-stat	All suburban muni			Without muni-stat	All muni		Without muni-stat
Period:	68-10	68-10	75-10	75-10	68-10	68-10	75-10	75-10	75-10	75-10	75-10
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
PanelA: OLS results											
ln(Distance)	-0.080[a]	-0.006[c]	-0.008[b]	-0.003	-0.065[a]	-0.008[a]	-0.009[a]	-0.007[a]	-0.049[a]	0.011[a]	0.007[a]
	(0.007)	(0.003)	(0.003)	(0.003)	(0.007)	(0.002)	(0.003)	(0.001)	(0.005)	(0.002)	(0.001)
Pseudo R^2	0.26	0.88	0.88	0.94	0.14	0.88	0.88	0.95	0.13	0.89	0.96
Panel B: TSLS results											
ln(Distance)	-0.092[a]	-0.024[a]	-0.027[a]	-0.027[a]	-0.105[a]	-0.024[a]	-0.027[a]	-0.030[b]	-0.127[a]	-0.042[b]	-0.049[a]
	(0.014)	(0.007)	(0.007)	(0.007)	(0.016)	(0.009)	(0.009)	(0.012)	(0.017)	(0.019)	(0.019)
F-S statistic	203.20	82.82	82.44	39.92	171.24	43.08	42.40	19.25	87.80	21.28	12.92
Instrument:	ln(Distance to the nearest 1870 railroad line)										
ln(Densities)	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
ln(Dist to CBD)	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Geography	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
History	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Socioeconomy	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Observations:	7680	7680	6400	4885	7680	7680	6400	4885	6400	6400	4885
	(1280×6)		(1280×5)	(977×5)	(1280×6)		(1280×5)	(977×5)	(1280×5)		(977×5)

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

The main conclusion to be drawn from the TSLS results (Panel B) is that train stations have extended spatial effects: being closer to a station increases the probability to be (part of) a subcenter, even for municipalities without any station. As in the previous section, we find a very similar effect of the proximity to any type of train or to suburban train alone. In this case, getting closer to a station by one kilometer increases the probability of becoming a subcenter by 2.4% (columns 2 and 6), or by 2.7% considering the 1975-2010 period (columns 3 and 7). On the other hand, the effect of being one kilometer closer to an RER station is estimated at 4.2% (column 10). We obtain our preferred results when we restrict our sample to the 977 suburban municipalities that do not have any station within their boundaries (columns 4, 8 and 11). This effect, of 3% for suburban trains and 4.9% for RER, confirms the spatial effect of railroad stations.

As previously the larger TSLS estimates (vs. OLS) suggest the endogeneity of railroad construction in Paris and could be associated to a missing variable (local land use regulations) and to reverse causation (positive shocks to less dense non-central municipalities).

Finally, we check that our results are robust to subcenter size (more or less than 50,000 jobs) and definition (subcenters identified using the 5% threshold instead of 10%), and, by following Chandra and Thompson (2000), to a more restrictive sample after dropping observations for historical towns and municipalities crossed by railroads during the 19th century. These tests are reported in Tables C.1 Panel B and C.2 Panel B in Appendix C.

To summarize the results of this section, two main points can be highlighted. First, railroad stations do play a role in the subcenter formation process, and this effect is spatially lagged: the existence of a train station increases the probability of becoming part of a subcenter by 4 to 5%, and decreases at a rate of about 3% per kilometer. Second, the RER is the type of train having the most important effect, with a direct effect of around 14% for municipalities with a station, and a spatial decay of about 5% per kilometer.

5. Conclusions

In this paper we investigate the effect of railroad construction on the emergence of employment subcenters in metropolitan Paris between 1968 and 2010. Because of the potential endogeneity problem of railroad provision, we rely on IV estimations that use historical instruments built on the 19th century railroad network: a dummy for municipalities crossed by 1870 railroads, and the (log) distance to the nearest 1870 railroad.

After briefly describing Paris' job location main characteristics and its rail transit, we focus on the suburbs to study the spatial pattern of Paris' job decentralization. We identify employment subcenters and, despite some municipalities emerge as (part of) subcenters whereas others were dropped, the number of subcenters grew from 21 in 1968 to 35 in 2010. Since employment growth in these subcenters was very intense over the period, we conclude that employment decentralization in Paris is more clustered around subcenters than diffuse, thus reinforcing the polycentric nature of the city.

Finally, we investigate whether railroads cause the emergence of employment subcenters in Paris and our results confirm the causal effect. On average, the probability of a suburban municipality of being (part of a) subcenter increases by 5% if the municipality has a rail station within its boundaries, or by 3% if, although not having a rail station, municipality proximity to a rail station increases by 10%. Results for the RER confirm its effect is stronger: a 10% and a 5% increase in the probability of becoming (part of a) subcenter if the municipality has a RER station or if municipality proximity to a RER station increases by 10%, respectively.

The contribution of the paper is relevant because, as far as we know, it provides the first empirical evidence on the causal effect of transportation (railroad) on subcenter formation. Furthermore, these new results are useful for urban planners facing and dealing the consequences of urban growth and, in particular, of population suburbanization, employment decentralization and urban sprawl: while these phenomena might reduce city's agglomeration economies (Glaeser and Kahn, 2004), the emergence of employment subcenters can potentially compensate and even overcome these loses by offering new agglomeration economies and avoiding CBD's congestion costs (McMillen, 2004).

References

Anas, Alex and Ikki Kim. 1996. General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration. *Journal of Urban Economics* 40(2): 232–256.

Arribas-Bel, Daniel and Fernando Sanz-Gracia. 2014. The validity of the monocentric city model in a polycentric age: US metropolitan areas in 1990, 2000 and 2010. *Urban Geography* 35(7): 980–997.

Bairoch, Paul. 1988. *Cities and Economic Development: From the Dawn of History to the Present*. University of Chicago Press.

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

Baum-Snow, Nathaniel. 2010. Changes in transportation infrastructure and commuting patterns in us metropolitan areas 1960-2000. *American Economic Review Papers & Proceedings* 100: 378–382.

Baum-Snow, Nathaniel. 2014. Urban transport expansions, employment decentralization, and the spatial scope of agglomeration economies. *Mimeo* .

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2015. Roads, railroads and decentralization of chinese cities. *Mimeo* .

Baum-Snow, Nathaniel and Matthew E. Kahn. 2000. The effects of new public projects to expand urban rail transit. *Journal of Public Economics* 77(2): 241–263.

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 Affairs* : 147–206.

Berliant, Marcus, Shin-Kun Peng, and Ping Wang. 2002. Production Externalities and Urban Configuration. *Journal of Economic Theory* 104(2): 275–303.

Berliant, Marcus and Ping Wang. 2008. Urban growth and subcenter formation: A trolley ride from the Staples Center to Disneyland and the Rose Bowl. *Journal of Urban Economics* 63(2): 679–693.

Bollinger, Christopher and Keith Ihlanfeldt. 1997. The impact of rapid rail transit on economic development: The case of atlanta's marta. *Journal of Urban Economics* 42(2): 179–204.

Bollinger, Christopher and Keith Ihlanfeldt. 2003. The intraurban spatial distribution of employment: which government interventions make a difference? *Journal of Urban Economics* 53(3): 396–412.

Bowes, David R. and Keith R. Ihlanfeldt. 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics* 50(1): 1 – 25.

Brooks, Leah and Byron Lutz. 2016. Vestiges of transit: Urban persitence at a micro scale. Working paper.

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.

Chen, Yihsu and Alexander Whalley. 2012. Green infrastructure: The effects of urban rail transit on air quality. *American Economic Journal: Economic Policy* 4(1): 58–97.

Craig, Steven and Pin T. Ng. 2001. Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *Journal of Urban Economics* 49(1): 100–120.

Craig, Steven G., Janet E. Kohlhase, and Adam W. Perdue. 2016. Empirical polycentricity: The complex relationship between employment centers. *Journal of Regional Science* 56(1): 25–52.

Diao, Mi, Yi Zhu, and Jiren Zhu. 2016. Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Studies* Forthcoming.

Duranton, Gilles and Diego Puga. 2015. Urban land use. In Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics Volume 5*. Amsterdam: North-Holland, 467–560.

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

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2): 161–196.

Garcia-López, Miquel-Àngel. 2010. Population suburbanization in Barcelona, 1991-2005: Is its spatial structure changing? *Journal of Housing Economics* 19(2): 119–132.

Garcia-López, Miquel-Angel. 2012. Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics* 72(2–3): 176–190.

Garcia-López, Miquel-Àngel, Camille Hémet, and Elisabet Viladecans-Marsal. Forthcoming. How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail. *Journal of Regional Science* .

Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics* 85(0): 52 – 67.

Garcia-López, Miquel-Àngel and Ivan Muñiz. 2013. Urban spatial structure, agglomeration economies, and economic growth in barcelona: An intra-metropolitan perspective. *Papers in Regional Science* 92(3): 515–534.

Garcia-López, Miquel-Àngel, Albert Solé-Ollé, and Elisabet Viladecans-Marsal. 2015. Does zoning follow highways? *Regional Science and Urban Economics* 53(C): 148–155.

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

Giuliano, Genevieve, Christian L. Redfearn, Ajay Agarwal, and Syvia He. 2012. Network accessibility and employment centers. *Urban Studies* 49(1): 77–95.

Giuliano, Genevieve, Christian L. Redfearn, Ajay Agarwal, Chen Li, and Duan Zhuang. 2007. Employment concentrations in Los Angeles, 1980-2000. *Environment and Planning A* 39(12): 2935–2957.

Giuliano, Genevieve and Kenneth A. Small. 1991. Subcenters in the Los Angeles region. *Regional Science and Urban Economics* 21(2): 163 – 182.

Giuliano, Genevieve and Kenneth A. Small. 1999. The determinants of growth of subcenters. *Journal of Transport Geography* 7: 189 – 201.

Glaeser, Edward L. and Matthew E. Kahn. 2001. Decentralized employment and the transformation of the American city. In W.G. Gale and J.R. Pack (eds.) *Brookings-Wharton Papers on Urban Affairs: 2001*. Brookings Institution Press, 1–47.

Glaeser, Edward L. and Matthew E. Kahn. 2004. Sprawl and urban growth. In J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics Volume 4*. Amsterdam: North-Holland, 2481–2527.

Grube-Cavers, Annelise and Zachary Patterson. 2015. Urban rapid rail transit and gentrification in canadian urban centres: A survival analysis approach. *Urban Studies* 52(1): 178–194.

Helsley, Robert W. and Arthur M. Sullivan. 1991. Urban subcenter formation. *Regional Science and Urban Economics* 21(2): 255–275.

Henderson, Vernon and Arindam Mitra. 1996. The new urban landscape: Developers and edge cities. *Regional Science and Urban Economics* 26(6): 613–643.

Henderson, Vernon and Eric Slade. 1996. Development games in non-monocentric cities. *Journal of Urban Economics* 34(2): 207–229.

Hurst, Needham B. and Sarah E. West. 2014. Public transit and urban redevelopment: The effect of light rail transit on land use in Minneapolis, Minnesota. *Regional Science and Urban Economics* 46(C): 57–72.

Li, Shanjun, Jun Yang, Ping Qin, and Shun Chonabayashi. 2016. Wheels of fortune: Subway expansion and property values in Beijing. *Journal of Regional Science* Forthcoming.

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

Martí-Henneberg, Jordi. 2013. European integration and national models for railway networks (1840–2010). *Journal of Transport Geography* 26(0): 126–138.

Mayer, Thierry and Corentin Trévien. 2015. The impacts of urban public transportation. Evidence from the Paris metropolitan area. Working Paper g2015/03, INSEE.

McDonald, John F. 1987. The identification of urban employment subcenters. *Journal of Urban Economics* 21(2): 242–258.

McDonald, John F. and Paul J. Prather. 1994. Suburban employment centers: The case of Chicago. *Urban Studies* 31(2): 201–218.

McMillen, Daniel P. 2001. Nonparametric employment subcenter identification. *Journal of Urban Economics* 50(3): 448–473.

McMillen, Daniel P. 2003. Identifying sub-centres using contiguity matrices. *Urban Studies* 40(1): 57–69.

McMillen, Daniel P. 2004. Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas. *Journal of Regional Science* 44(2): 225–244.

McMillen, Daniel P. and John McDonald. 2004. Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983-1999. *Real Estate Economics* 32(3): 463–486.

McMillen, Daniel P. and John F. McDonald. 1998. Suburban Subcenters and Employment Density in Metropolitan Chicago. *Journal of Urban Economics* 43(2): 157–180.

McMillen, Daniel P. and Christian L. Redfearn. 2010. Estimation And Hypothesis Testing For Nonparametric Hedonic House Price Functions. *Journal of Regional Science* 50(3): 712–733.

McMillen, Daniel P. and Stefani C. Smith. 2003. The number of subcenters in large urban areas. *Journal of Urban Economics* 53(3): 321–338.

Michaels, Guy. 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics* 90(4): 683–701.

MuÑiz, Ivan, Anna Galindo, and Miquel-Àngel Garcia-López. 2003. Cubic spline population density functions and satellite city delimitation: The case of Barcelona. *Urban Studies* 40(7): 1303–1321.

Muñiz, Ivan, Miquel-Àngel Garcia-López, and Anna Galindo. 2008. The effect of employment sub-centres on population density in Barcelona. *Urban Studies* 45(3): 627–649.

OECD. 2017. *The governance of land use in France: Case studies of Clermont-Ferrand and Nantes Saint-Nazaire*. OECD Publishing, Paris.

Redfearn, Christian L. 2007. The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics* 61(3): 519–541.

Redfearn, Christian L. 2009*a*. Persistence in urban form: The long-run durability of employment centers in metropolitan areas. *Regional Science and Urban Economics* 39(2): 224–232.

Redfearn, Christian L. 2009*b*. How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. *Regional Science and Urban Economics* 39(3): 297–306.

Ryan, Sherry. 2005. The value of access to highways and light rail transit: Evidence for industrial and office firms. *Urban Studies* 42(4): 751–764.

Schuetz, Jenny. 2015. Do rail transit stations encourage neighbourhood retail activity? *Urban Studies* 52(14): 2699–2723.

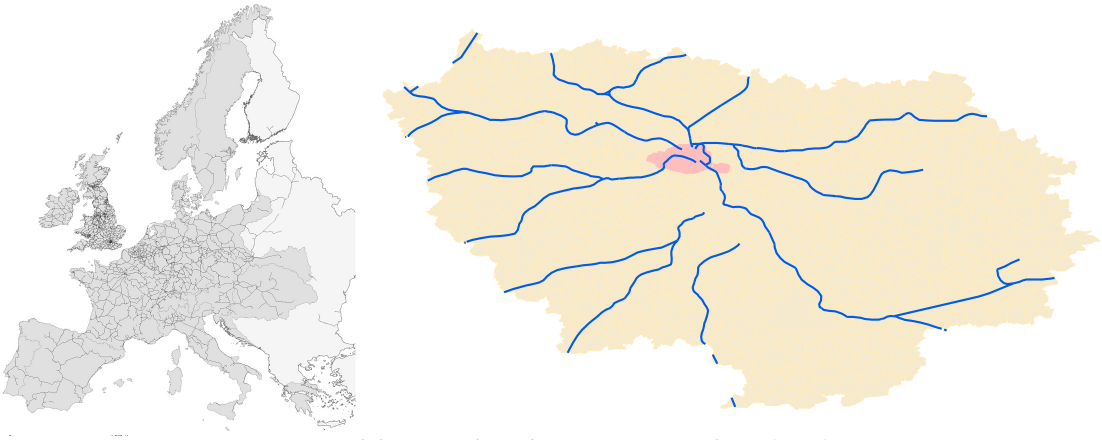
Appendix A. Rail transit in Paris: Past and present

One of the main purposes of this paper is to evaluate whether and to what extent transportation has fostered the emergence of employment subcenters in metropoliran Paris. However, first we need to deal with an identification issue because transportation and its improvements are not placed randomly. On the contrary, they are endogenous to employment and/or population location and growth. Planners may for instance decide to improve the connection of deprived areas in order to boost their economic activity or attract population. In order to address this issue, we adopt an instrumental variable approach in which some variables, named instruments, are used as sources of exogenous variation for our transportation endogenous variables.

Recent literature highlights the advantages in terms of exogeneity and relevance of using 'historical' and 'planned' instruments. For instance, Baum-Snow (2007), Michaels (2008) and Duranton and Turner (2012) use the 1947 plan of the interstate highway system as an instrument for modern highways in the US, and Duranton and Turner (2012) additionally rely on the 1898 railroad network. Garcia-López (2012) uses the ancient Roman roads, and the 19th century main road and railroad networks as instruments for highways and railroads in metropolitan Barcelona. Finally, Garcia-López *et al.* (2015) use the ancient Roman roads and the 1760 Bourbon roads (post routes) to instrument current highways in Spain.

Following the above mentioned literature, we instrument modern railroads in metropolitan Paris with a historical instrument, the 1870 railroad network. The first French railroads were built at the beginning of the 19th century, but slightly later than in the UK due to Napoleon wars: the first line connecting Paris to a city located 18 km away (Saint-Germain) was not opened before 1837. In 1870, the railroad network was based on 698 km of railroad lines. Due to the high levels of centralization in France, it had a star-shaped form centered around Paris (Figure A.1).

Figure A.1: The 1870 railroads



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

Is the 1870 rail network a valid instrument?

As above mentioned, the fact that modern roads and railroads were built following ancient infrastructures has already been argued and used in the literature. Common sense would suggest that in France as well, past infrastructures shape current ones due to practical reasons: it is easier and cheaper to build new transportation infrastructures as an improvement of old ones for instance, or close to them (Duranton and Turner, 2012). We now test empirically the credibility of this assumption in the context of the metropolitan area of Paris. To do so, we conditially regress our endogenous rail variables on their historical counterparts and some control variables:

$$\begin{aligned} \text{2010 Rail transit variable} = \alpha_0 + \alpha_1 \times \text{1870 rail transit variable} \\ + \sum_i (\alpha_{2,i} \times \text{control variables}) \end{aligned} \tag{A.1}$$

It is important to point out the importance of the control variables, in particular geography and history. Although ancient transportation infrastructures may be exogenous because of the length of time since they were built, the significant changes undergone by society and economy in the intervening years, and, in particular, because neither of them were built to anticipate employment and population changes in a distant future; it is also true that other factors such as the geography are likely to have influenced the construction and location of both ancient and modern transportation infrastructures for obvious reasons related to the feasibility and convenience of infrastructure building. From this point of view, it is crucial to include geographic characteristics such as altitude, index of terrain ruggedness, and elevation range as controls to comply with the exogeneity condition.

On the other hand, it is equally important to control for the historical context, since it may explain both the presence of former infrastructure and the economic importance of present-days municipalities. In order to fulfill the exclusion restriction, and because there are no historical employment and population data at the municipal level prior to 1962 and 1968, we control for history by including the population level in 1962 and dummy variables indicating (1) whether municipalities were Roman settlements, (2) whether they used to be major towns between the 10th and the 15th centuries and (3) between the 16th and the 19th centuries, and (4) whether they had a monastery built between the 12th and 16th centuries. These dummy variables come from the Digital Atlas of Roman and Medieval Civilizations, with the exception of the major cities of the 16th to 19th centuries which are identified in Bairoch (1988). To put it differently, we assume *conditional* exogeneity of the proposed instruments, as suggested by (Duranton and Turner, 2012).

Finally, it is important to notice that, to satisfy the exclusion restriction, our endogenous variables and potential instruments are not exactly built in the same way. While the former refer to stations (dummy for municipalities with modern station/s, distance to the nearest modern station) the latter refer to railroad lines (dummy for municipalities crossed by 1870 railroad line/s, distance to the nearest 1870 railroad line). Furthermore, in order to support the validity of our identification strategy we run some robustness checks in Appendix C Table C.2. Following Chandra and Thompson (2000), we exclude historical towns and municipalities with

1870 stations/lines and restrict our attention to municipalities that were 'accidentally' crossed by 1870 railroads and did not received 1870 stations.

Regarding the relevance of our potential instruments, Table A.1 shows results for versions of Eq. (A.1) in which we analyze the relationship between modern and past railroads in terms of the presence of stations and proximity to them. In particular, in Panel A, we study whether suburban municipalities crossed by a 1870 rail receive a rail stations. In all cases (pooled vs. cross-section regressions in columns 1 and 2-3, all railroads vs. train and RER regressions in columns 1 and 4-5) we find significant and positive coefficients for the presence of 1870 rails. In Panel B, we estimate the effect of municipality proximity to 1870 rail on the municipality proximity to the nearest modern rail. Conditional on control variables, estimated coefficients for the 1870 distance variable are positive and highly significant. As a whole, results in Table A.1 clearly show that historical rails matter for modern rail construction and location.

Table A.1: Modern rail transit as a function of past rail transit, OLS estimates

Panel A: Rail stations						**Panel B:** Proximity to rail stations			
Dependent var.:	Dummy=1 if muni with station					Dependent var.:	ln(Dist to nearest station)		
Rail type:	Rail			Train	RER		Rail	Train	RER
Period/year:	75-10	1975	2010	75-10	75-10	Period:	75-10	75-10	75-10
	[1]	[2]	[3]	[4]	[5]		[6]	[7]	[8]
Dummy=1 if crossed	0.179[a]	0.177[a]	0.168[a]	0.173[a]	0.031[a]	ln(Distance to	0.180[a]	0.121[a]	0.093[a]
by 1870 rail	(0.029)	(0.030)	(0.030)	(0.028)	(0.009)	nearest 1870 rail)	(0.029)	(0.028)	(0.026)
Adjusted R^2	0.37	0.30	0.39	0.24	0.24	Adjusted R^2	0.54	0.42	0.71
F-S statistic	38.04	33.81	31.63	36.83	11.21	F-S statistic	39.92	19.25	12.92
ln(Densities)	Y	Y	Y	Y	Y	ln(Densities)	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y	Y	ln(Dist to CBD)	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Geography	Y	Y	Y
History	Y	Y	Y	Y	Y	History	Y	Y	Y
Socioeconomy	Y	Y	Y	Y	Y	Socioeconomy	Y	Y	Y
Observations:	6115	1223	1223	6115	6115	Observations:	4885	4885	4885
	(1223×5)			(1223×5)			(977 muni × 5 years)		

Notes: Pooled regressions in Columns 1 and 4 to 8 include year effects. Cross section regressions in Columns 2 and 3 include a constant. Columns 1 to 3, Columns 4 and 5, and Columns 6 to 8 show first-stage results for regressions in Table 4 Columns 8 to 10, Table 5 Columns 4 and 8, and Table 6 Columns 4, 8 and 11, respectively. Robust standard errors are clustered by municipality and are in parentheses. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

Appendix B. LWR and urban spatial structure in metropolitan Paris, 1968–2010

Table B.1: Employment spatial structure and LWR: A benchmark to identify subcenters

	1968	1975	1982	1990	1999	2010
1% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
3% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
5% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
7% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
9% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
10% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
30% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
50% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric
70% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric
90% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric

Table B.2: Employment spatial structure and LWR: Akaike information criterion

	1968	1975	1982	1990	1999	2010
1% LWR Akaike inf. crit.	746	767	770	778	774	782
3% LWR Akaike inf. crit.	397	433	446	462	460	479
5% LWR Akaike inf. crit.	346	386	402	423	422	446
7% LWR Akaike inf. crit.	336	377	393	417	417	444
9% LWR Akaike inf. crit.	340	380	396	421	421	450
10% LWR Akaike inf. crit.	345	386	401	426	426	456
30% LWR Akaike inf. crit.	573	598	593	605	590	631
50% LWR Akaike inf. crit.	817	837	820	828	797	843
70% LWR Akaike inf. crit.	1081	1112	1094	1109	1067	1120
90% LWR Akaike inf. crit.	1284	1331	1318	1346	1301	1364

Appendix C. Does rail transit cause subcenter formation? Robustness checks

Table C.1: The effect of rail on subcenter formation, IV Probit - Marginal effects: Robustness to subcenter size and significance

Panel A: The effect of rail stations

Dependent var.:	Probability of being subcenter			
	Subcenter jobs		5% residuals	
	≥50,000	<50,000	All obs	Without alw-sub
Period:	75-10	75-10	75-10	75-10
	[1]	[2]	[3]	[4]
Station dummy	0.036[c]	0.039[c]	0.035[c]	0.044[b]
	(0.020)	(0.023)	(0.021)	(0.021)
F-S statistic	34.20	42.76	38.95	38.29
ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y
History	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y
Observations:	6214	6117	6400	6195
Instrument:	Dummy=1 if crossed by a 1870 rail			

Panel B: The effect of proximity to rail stations

Dependent var.:	Probability of being subcenter			
	Subcenter jobs		5% residuals	
	≥50,000	<50,000	All obs	Without alw-sub
Period:	75-10	75-10	75-10	75-10
	[5]	[6]	[7]	[8]
ln(Distance)	-0.015[a]	-0.017[a]	-0.013[c]	-0.011[c]
	(0.005)	(0.006)	(0.008)	(0.006)
F-S statistic	81.35	80.63	80.41	84.55
ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y
History	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y
Observations:	6214	6117	6400	6195
Instrument:	ln(Dist to the nearest 1870 rail)			

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.

Table C.2: The effect of rail on subcenter formation, IV Probit - Marginal effects: Robustness to identification strategy

Panel A: The effect of rail stations

Dependent var.:	Probability of being subcenter			
	No historic towns		No 1870 stations	
	All obs.	Without alw-sub	All obs	Without alw-sub
Period:	75-10	75-10	75-10	75-10
	[1]	[2]	[3]	[4]
Station dummy	0.067[a]	0.046[b]	0.049[b]	0.037[c]
	(0.023)	(0.022)	(0.023)	(0.021)
F-S statistic	34.55	35.13	8.53	10.64
ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y
Lagged pop	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y
Observations:	6185	5955	5385	5245
Instrument:	Dummy=1 if crossed by a 1870 rail			

Panel B: The effect of proximity to rail stations

Dependent var.:	Probability of being subcenter			
	No historic towns		No 1870 lines	
	Without mun-stat	Without alw-sub	Without mun-stat	Without alw-sub
Period:	75-10	75-10	75-10	75-10
	[5]	[6]	[7]	[8]
ln(Distance)	-0.023[a]	-0.010[c]	-0.025[a]	-0.020[a]
	(0.006)	(0.005)	(0.004)	(0.004)
F-S statistic	52.85	50.69	95.90	90.45
ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y
Lagged pop	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y
Observations:	4780	4675	4110	3945
Instrument:	ln(Dist to the nearest 1870 rail)			

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. [a], [b], and [c] indicates significant at 1, 5, and 10 percent level, respectively.