Does the built environment shape commuting? The case of Lyon (France)

Charles Raux, Ayana Lamatkhanova, Lény Grassot

WORKING PAPERS DU LAET

Does the built environment shape commuting? The case of Lyon (France)

Charles RAUX
Ayana LAMATKHANOVA
Lény GRASSOT

Is built environment the most influential factor on travel behavior when compared to individual socioeconomic characteristics? This paper extends the empirical knowledge by providing and comparing quantitative estimates of these various effects on both commuting distance and mode choice in a European city spatial context, while using up-to-date and novel methodology. Eight indicators of built and social environment are identified in order to characterize clusters of residential locations, giving a rich view of spatial and social diversity of locations. To disentangle the causal effects of residential self selection and built environment, both sample selection and specific matching preprocessing (õcoarsened exact matchingö, a novel approach in the field) are implemented. Regarding commuting distance, the true effect of built and social environment appears modest with an increase of distance in the range of 10-20%. It comes behind individual socioeconomic characteristics such as car availability and qualification. Regarding commuting mode choice, again the true effect of built and social environment is modest, with a nearly 20%pt increase of car share and around 10%pt decrease or public transport share for the most prominent effects, and it comes behind car availability. These results suggest the primary importance of influencing directly car use, if not car ownership, in the European context, while trying to modify the built environment would provide only limited results.

Keywords: built environment, commuting, mode choice, distance travelled, Lyon, France

www.laet.science.fr

Does the built environment shape commuting? The case of Lyon (France)

Charles RAUX

Univ Lyon, CNRS, LAET, F-69007, LYON, France

Ayana LAMATKHANOVA

Univ Lyon, CNRS, LAET, F-69007, LYON, France

Lény GRASSOT

Univ Lyon, CNRS, LAET, F-69007, LYON, France

Does the built environment shape commuting? The case of Lyon (France)

Charles Raux[a*], Ayana Lamatkhanova[a], Lény Grassot[a]
[a] Univ Lyon, CNRS, LAET, F-69007, LYON, France

* corresponding author: email: charles.raux@laet.msh-lse.fr,
address: LAET, MSH, 14 av. Berthelot, 69363 Lyon Cedex 07, France.
Phone: +33472726403, Fax: +33472726448

WP Raux Lamatkhanova Grassot 2020 v2.docx - 4/30/2021 3:16:00 PM

Abstract

Is built environment the most influential factor on travel behavior when compared to individual socioeconomic characteristics? This paper extends the empirical knowledge by providing and comparing quantitative estimates of these various effects on both commuting distance and mode choice in a European city spatial context, while using up-to-date and novel methodology. Eight indicators of built and social environment are identified in order to characterize clusters of residential locations, giving a rich view of spatial and social diversity of locations. To disentangle the causal effects of residential self selection and built environment, both sample selection and specific matching preprocessing ("coarsened exact matching", a novel approach in the field) are implemented. Regarding commuting distance, the true effect of built and social environment appears modest with an increase of distance in the range of 10-20%. It comes behind individual socioeconomic characteristics such as car availability and qualification. Regarding commuting mode choice, again the true effect of built and social environment is modest, with a nearly 20%pt increase of car share and around 10%pt decrease or public transport share for the most prominent effects, and it comes behind car availability. These results suggest the primary importance of influencing directly car use, if not car ownership, in the European context, while trying to modify the built environment would provide only limited results.

Keywords: built environment; commuting; mode choice; distance travelled; Lyon; France

L'environnement construit façonne-t-il les navettes domicile-travail ? Le cas de Lyon

Résumé

L'environnement construit est-il le facteur le plus influent sur les comportements de mobilité quotidienne quand on le compare aux caractéristiques socio-économiques individuelles ? Cet article propose des estimations quantitatives de ces différents effets à la fois sur les distances parcourues et le choix de mode de transport pour le navettage domicile-travail dans un contexte urbain européen, tout en utilisant une méthodologie des plus récentes. Huit indicateurs de l'environnement construit et social sont identifiés afin de construire une typologie des localisations résidentielles, laquelle donne une représentation enrichie de la

diversité spatiale et sociale des localisations. Afin de démêler les influences de l'auto-sélection résidentielle et de l'environnement construit, nous mettons en œuvre la méthode de sélection d'échantillon et une nouvelle méthode d'appariement (« coarsened extact matching »). L'effet contrôlé de l'environnement construit sur les distances de navettage apparaît modeste car la croissance de distance qui peut être lui imputée en comparant la banlieue ou le périurbain au centre va de 10 à 20%. Cette influence est moindre que celle de caractéristiques individuelles telles la disponibilité d'une voiture et la qualification professionnelle. L'effet de l'environnement construit sur le choix du mode de déplacement est également modeste avec une hausse de 20%pt de la part de la voiture et une baisse de 10%pt de la part des transports publics. Cet effet vient derrière celui de la disponibilité d'une voiture. Ces résultats montrent l'importance primordiale d'influencer directement l'usage de la voiture, si ce n'est sa possession, dans un contexte européen, et relativisent l'effet de l'aménagement sur la réduction de l'usage de l'automobile en zone urbaine.

Mots-clés : environnement urbain ; navettes domicile-travail ; choix modal ; distance ; Lyon

INTRODUCTION

Urban sprawl is often viewed as an undesirable development type because of its various harmful consequences regarding the air quality and climate change (through vehicle miles travelled), traffic congestion, physical health and obesity, and public service costs (Ewing and Hamidi, 2015). At least two decades of controversy on the respective advantages of compactness and sprawl have fostered numerous studies, particularly in the transport and land-use literature. On the specific topic of climate change mitigation, on the basis of a literature review, Desjardins (2011) concluded that planning's contribution could only be modest. However, the debate is still lively as shown by the controversy reopened in the Journal of the American Planning Association following Stevens (2017) paper. The main issue at stake is the potential to moderate vehicular travel intensity in urban areas through changes in the built environment.

This paper analyzes and estimates the respective scales of the influence of built environment (BE) and socioeconomic characteristics (SEC) on travel behavior (TB) in an European context, the Lyon urban area. BE has originally been defined through the famous 3Ds of built environment: Density, Diversity, and Design, initially coined by Cervero and Kockelman (1997) and since widely used in North American studies. 3D has grown to 5D with the addition of Destination accessibility and Distance to transit (Ewing and Cervero, 2001). We extend these dimensions to social ones, reflecting additional factors which affect residential location choices. To this adds the critical concern of residential self-selection (RSS) which would mediate the causal relationship between BE and TB (Guan et al, 2019).

Due to data availability we focus our analysis on commuting behavior of working people. Our research questions are basically the following. Does built environment have the most influential effect on commuting behavior (commuting distance and mode choice) when compared to individual socioeconomic characteristics? How can these effects be computed?

Eight indicators of built and social environment have been identified in order to characterize residential locations: they give a richer view of spatial and social diversity of locations than a unique indicator such as distance from the agglomeration centre or density. RSS is tackled through socio-economic variables attached to the individual and related to the most relevant attributes for residential location choice. RSS is controlled through two approaches, the sample selection approach for commuting distance and the matching approach for mode choice. Regarding matching preprocessing, the use of "coarsened exact matching" is novel in the transport and land use field. We extend the empirical knowledge by providing and comparing quantitative estimates of the various effects, whether BE or individual socio-economic characteristics, on both commuting distance and mode choice in a European spatial context.

In the first section we give an overview of the literature. The second section sets out the data elaboration and the statistical methods. The third section shows the results while the fourth one discusses them and the fifth one concludes.

1   LITERATURE REVIEW

The literature on the relationships between built environment and travel behavior is "vast but messy" (Handy, 2017). Moreover, the literature points at the important issue of residential self selection. Finally, the methodological aspects of this issue are reviewed.

1.1 The relationships between built environment and travel behavior

Based on a synthesis of 50 empirical studies Ewing and Cervero (2001) found that trip lengths are primarily a function of the BE and secondarily of socioeconomics (SEC). Mode choices depend on both but probably more on SEC. Pursuing with a meta-analysis of more than 200 studies Ewing and Cervero (2010) found that vehicle-miles of travel (VMT) is most strongly related to destination accessibility, that walking is related to land use diversity, intersection density and the number of destinations within walking distance. Bus and train use are related to proximity to transit and street network design. Population and job densities are only weakly associated with travel behavior (TB) when the other 4Ds are controlled. Both reviews included studies comparing various areas, whether metropolitan areas, districts or neighborhoods in U.S. regions.

By pooling household travel and BE data from 15 various regions in the USA (regional household travel surveys), Ewing et al (2015) found that socioeconomic influence is the strongest even if BE and transit service also influence household travel decisions. Among the 5D the strongest ones are diversity, design and destination accessibility, while the weakest is density.

In a more recent meta-analysis (based upon a meta-regression of 37 studies) focusing on VMT and controlling for self-selection (see below), Stevens (2017) found that compact development (mainly the density and distance to downtown) does make people drive less but concluded that the impact is "fairly small". These conclusions reopened the controversy on the effects of BE on travel (see the comments in the same and following issues of JAPA[1]).

However, the previous results are mostly based upon empirical case studies in the USA. On the European side, characterized by denser cities and higher walking and public transport use in general, the literature is more recent. Based on longitudinal analyses of the German nation-wide travel survey, focusing on mode choice, Scheiner (2010) concludes that the increase in car use is mostly explained by the increase in car ownership while BE appears to have an impact on car use for car owners. In The Netherlands, focusing on commuting, Susilo and Maat (2007) show with longitudinal analyses of the Dutch national travel survey that commuting distance (whether by car or rail) is barely influenced by urban forms and travel accessibility but more by job locations in line with individual profiles. In the UK Aditjandra et al (2012) applied structural equation modeling to a panel of respondents who reported residential relocation in the metropolitan area of Tyne and Wear. They found that neighborhood characteristics do influence TB after controlling for residential self-selection. With cross-sectional data on commuting behavior in the Lisbon metropolitan area (Portugal), controlling for self-selection, de Abreu e Silva (2014) found that attitudinal variables significantly influence TB but do not cancel the effects of land-use patterns on this behavior. In Dublin (Ireland) Humphreys and Ahern (2019) observed residential self-selection but not to such an extent that it is more important than land-use factors in determining modal-split characteristics.

Empirical studies on the link between BE and TB are lacking in France, except for two. Pouyanne (2005) explored the interactions between urban forms and travel patterns inside the metropolitan area of Bordeaux. Aguilera (2005) analyzed the increase in commuting distances in the three metropolitan areas of Paris, Lyon and Marseille. However, both studies were conducted at an aggregate zonal level inside the urban areas, and not individual level. The debate is lively: it is generally admitted that density is a strong determinant of mobility pattern and that making cities (again) compact will help to reduce car dependence. This position is of course criticized by some authors (see e.g. Charmes, 2010) who point that the issue at stake is not continuous sprawl but rather scattering of locations.

---

[1] https://www.tandfonline.com/toc/rjpa20/83/1

1.2   The issue of residential self selection

An important issue arises in the causal mechanism linking BE and TB, that is to say the residential self-selection (RSS). RSS hypothesizes that residents may choose a BE that is consistent with their attitudes (AT) and preferences regarding travel modes or land use configuration (for instance moving to a dense and diversified neighborhood in order to be able to walk to various amenities): this would bias the link evidenced between BE and TB and, according to the majority of the literature, overestimates the influence of BE.

Empirical findings regarding RSS are documented through 38 studies (most of them from North America, five from UK, Germany and The Netherlands) in Cao et al (2009) and some of the recent studies previously quoted also address this issue. Cao et al noticed that the effect of BE is statistically significant after taking account of RSS but that RSS's influence is rarely estimated. Mokhtarian and Cao (2008) reviewed the methodologies used in these studies and recommended usage of longitudinal methods. Mokhtarian and van Herick (2016) reviewed ten studies (all comparing locally neighborhoods in various U.S. regions) quantifying the RSS effect and evidencing a great variation of values across studies.

Cao and Chatman (2016) underlined that travel preference is only one of the factors influencing residential choices and that the resulting effect of BE and RSS on TB may come from various influences. These include elasticities of travel to BE, the structure of preferences in the population, choice factors other than BE (i.e. social environment or housing type) and the supply of different development types. This explains why the influence of BE on TB can be over or underestimated. In their review of the recent literature on RSS Guan et al (2019) discussed the issue of attitudes measurement. They underlined that beyond travel attitudes and preferences a multitude of subjective factors may contribute to RSS, including life style, values, norms and so on. Moreover, it is not clear whether attitudes are exogenous or endogenous (i.e. influenced by both BE and TB) since there are probably feedback effects from BE and TB to AT.

Moreover, the RSS issue based on travel attitudes is controversial. For instance Ettema and Nieuwenhuis (2017) found in a study on households who relocated recently in three Dutch transit-oriented development locations near The Hague that travel attitude "is insufficient to fully reflect self-selection process". Even Wang and Lin (2019) found no evidence of RSS in their case study in Beijing. In the two Norwegian metropolitan areas of Oslo and Stavanger Wolday et al (2019) found no strong empirical basis in support of controlling for travel-based residential self-selection. For Naess (2014) the RSS issue based on travel attitudes is like a "tempest in a teapot". Following the discussion of this paper by van Wee and Boarnet (2014) it can be concluded that: a) RSS based on socio-economic variables should be distinguished from RSS based on attitudes; b) there is a consensus that the former is important to be included; c) the role of attitudes is not fully understood yet; and d) there is a need to estimate the respective influence of BE and RSS in TB.

Regarding RSS based on socio-economic variables, a recent survey of the literature on residential location choice models (Schirmer et al, 2014) shows that the most significant attributes of location are: a) regarding BE: built density, proximity to transport networks, open space and green area, land use mix (residential and commercial); b) regarding point of interests (POI): education, retail, recreation and sport, transportation facilities, urban amenities; c) for socioeconomic environment: population density, household types (size, working or not, families or not), origin or race, incomes, density of jobs, school quality; and d) for employment accessibility: generally a gravity measure.

1.3   Matching and sample selection approaches

When trying to disentangle the causal effect of BE and RSS a randomized experiment would be ideal. However, such kind of experiment is generally infeasible in social sciences and

the scholar must generally rely on observational data (Winship and Morgan, 1999). These authors introduced the issue thanks to a conceptual framework distinguishing a "control" group (in our case it could be those living in the city centre) and a "treatment" group (those living in the suburb). In order to compare the outcome (for instance commuting distance) between the "treatment group" and the "control group", the bias resulting from a potential correlation between the outcome variable and the treatment assignment needs to be eliminated. Winship and Morgan reviewed and discussed strengths and weaknesses of various statistical and econometric methods available, that is, regression methods for cross-sectional data – including selection models, matching on propensity score and instrumental variables – and methods for longitudinal data. Mokhtarian and van Herick (2016) outlined and discussed the basics of propensity-score matching and sample selection approaches.

According to Ho et al (2007) preprocessing of data with matching has the interesting potential to remove some of the biases referred to above. By matching pairs of units that are (nearly) identical in observable covariates, one unit assigned to the treated group and the other to the control group, one mimics a randomized experiment. This method is said to reduce "imbalance" between the two groups. In our case for example observable covariates would include individual socioeconomic characteristics (SEC) such as age, gender and so on, and matching would imply to find pairs of individual with similar SEC, one living in the city centre (control), the other in the suburb (treated). Among matching methods, propensity score matching (PSM) (Rosenbaum and Rubin, 1983) is commonly applied. PSM has been applied in transport and land use studies in several studies (Cao et al, 2010; Cao, 2010; Cao and Fan, 2012; Cao and Schoner, 2014).

However, the matching approach assumes that all the selection is on the observables (Ho et al, 2007; Winship and Morgan, 1999), i.e. the probability of being assigned to the treatment condition is only a function of the observed variables. This assumption may not be valid if unobserved variables (such as attitudes or preferences) could be correlated simultaneously with the probability of choosing to live e.g. in the suburbs and the travel behaviour outcome. In this case the sample selection approach (Heckmann, 1978) is more appropriate since it controls also on the unobservables. It combines a selection (probit) model with two outcome models (linear regressions), one for each group (control and treated). However, as recalled by Bushway et al (2007) some methodological pitfalls must be avoided in the application of this approach. The main one is the importance of "exclusion restrictions" which means that some variables must be included in the selection equation (they must have a power to predict selection) while they should not be in the outcome equation (and they should not have a strong power of prediction on the outcome).

The sample selection model will be applied to a continuous outcome such as commuting distance. Regarding commuting mode choice the relevant outcome model is a discrete choice model (multinomial logit). However, there is no simple analog of the Heckman method for discrete choice models (Bushway et al, 2007). Thus for commuting mode choice the matching approach will be applied before the logit estimation.

However, most matching methods (including PSM) in real applications need repeated trials to obtain a better balance between the two groups because their application in a single use does not guarantee imbalance reduction and hence reduction in bias (Iacus et al, 2012). Coarsened Exact Matching (CEM) is a matching method which helps to avoid these pitfalls (for theoretical aspects see Iacus et al, 2011; for details about the advantages of CEM over many matching methods see Iacus et al, 2012). Measuring imbalance between the empirical distribution of pretreatment covariates of the treated and control group is done with an L1 distance (also called Manhattan distance). For that the covariates need to be discretized (for continuous variables) or recoded (for discrete variables). Through this coding variables' values are "coarsened" (i.e. grouped) so that they have the same meaning according to the researcher

assumptions. This is the case for instance for age or household size. Some variables are already coarsened and taken as they are, like gender, or may be recoded like qualification level which is coarsened along main diploma stages. Hence in the following we apply CEM preprocessing.

2 METHODOLOGY

In order to implement the treatment/control methodology we need to define the substance of grouping. Since we are studying the influence of BE on TB it is justified to base the grouping on spatial areas. Thus our empirical strategy consists in:
- first, elaborating "treatment" and "control" groups: this will be done via clustering residential locations;
- second, selecting the variables at the individual level which capture the self-selection phenomenon and will be used with both sample selection and matching (CEM) approaches;
- third, estimate the models, compute the treatment (BE) effects and compare them with individual socioeconomic effects.

First, we set out the elaboration of our data before detailing in a second subsection the statistical methods implemented.

2.1 Data collection and elaboration

Our study focuses on Lyon area, given the availability of data and our specific knowledge of this area acquired through various previous studies. Various delimitations exist in this area, as elaborated by the INSEE. The "urban unit" perimeter is based on the physical continuity of development (it aggregates communes with buildings less than 200 meters from each other), with 127 communes in the case of Lyon in 2020. The "urban area" perimeter is based on a functional approach since it aggregates to the urban unit the communes of which at least 40% of resident working people work in the urban area (these communes are the "periurban ring"). There are overall 498 communes in the Lyon "urban area". Our aim is to cut ourselves from either the "physical" or the "functional" perimeter by choosing a quadrilateral around the Lyon area centre, gathering communes both from the urban unit and the periurban ring. Thus our study area is a 50 x 30 km rectangle around the city of Lyon (France). It contains 123 "communes" (including the Lyon city with its nine "arrondissements" or districts distinguished), of which 20 belong to the periurban ring, with approximately 1.5 million inhabitants including around 750,000 working people (see Figure 1 hereafter). This map shows the main transport infrastructures (divided highways and semi-heavy public transport – subway and tramways), the administrative boundaries of the "communes" (or municipalities) and their density of inhabitants.

2.1.1 Data on commuting

Ideally we would have based our analysis on a mobility survey related to overall daily mobility of the whole population (both households and their individual members), as measured for instance in conventional household travel surveys (HTS). However, the sampling of the HTS in Lyon covers mainly the dense urban area and is parsimonious in low density outskirts. This is why we rely upon another survey focused on (work) commuting only which has the advantage of a much higher sampling rate and more variability in the representation of various locations in the urban region. Therefore we use an extract from the French national census about the population holding a job (INSEE MOBPRO 2013). This institutional source has a high sampling rate (around 27%) and provides dozens of individual information including the usual home-work transportation mode.

Focusing the analysis on individual home-work mobility is specifically relevant regarding our research questions. Indeed work activity has some mandatory aspects such as activity scheduling and location which influence travel mode choice and distance travelled from home. These mandatory aspects do not apply to other activities such as shopping, personal matters or leisure, where scheduling and location are freer. This is why it is relevant to perform a separate analysis for commuting behavior. Moreover, as shown by Salon (2015) the influence of BE on VMT depends on travel purpose.

However, it should be noted that we only know some characteristics of the commute, that is to say the "usual" commuting mode and the municipalities of residence and workplace, but not the frequency (i.e. daily or not). Travel mode is distinguished as [Private car], [Public transportation], [Two-wheeled vehicle (bike or motorcycle, not distinguished till now in the census)], [Walking]. Distance between home and workplace is then measured between centroids of the respective communes (computed with IGN shape file GEOFLA) while the intra-commune distance is measured as $d_{ii} = 0.5 * \sqrt{area_i}$ where $area_i$ is the surface of the commune i.

Furthermore, this data source contains some socio-demographic variables related to the commuter (see below). Overall statistics on commuting in the sample (N=201,355 individuals) are given in Table 1.

Table 1: Commuting distance statistics and mode share on the overall sample

Distance (km)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.758	2.258	4.797	6.82	9.876	54.424

Travel mode

walking	2_wheels	car	public_transport	total
26 444	9 451	109 308	56 132	201 335
13%	5%	54%	28%	100%

Figure 1: Map of the Lyon study area

The legend of the figure contains the following:

- Studied area
- Selected municipalities
- Subway & Tram (SYTRAL, 2015)
- Motorway (IGN, Route500, 2016)
- 1-9 Arrondissements of Lyon

Density (inhab/km²)
- 48 - 171
- 171 - 277
- 277 - 395
- 395 - 542
- 542 - 946
- 946 - 1849
- 1849 - 9529

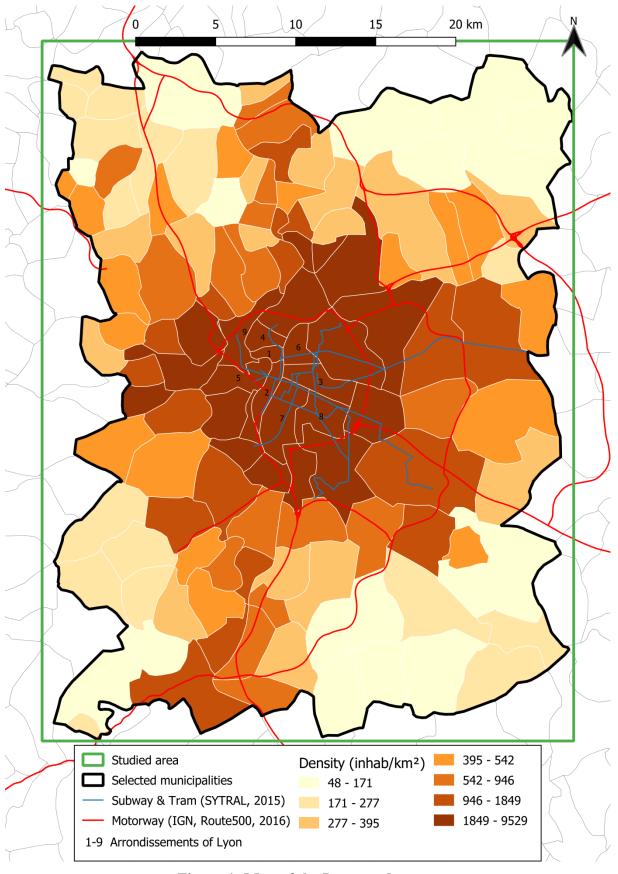### 2.1.2 Covariates for clustering communes for treatment and control groups

The "commune" is the smallest administrative subdivision in France. Since most of our land use data are available at this spatial level, this will be the unit of analysis for grouping areas. Thus our measure of "treatment" is based on a typology of communes. Our aim is to elaborate a set of two or more groups of communes across the study area. When arranged by pairs these groups will play the role of "control" and "treatment" for the individuals living within these groups of communes.

Since our study area is heterogeneous from a built environment and social perspective it would be irrelevant to base this typology only on a distance to a city center. Locations characterization is multidimensional. This is why we use clustering. The choice of variables for clustering results from a compromise between what is indicated by the literature and data availability in our study area (see Table 2). Some of these variables address the potential impact of BE on commuting behavior (Ewing and Cervero, 2001, 2010) and include the Ds as typically measured (Ewing et al, 2015). Commuting distance is impacted by density and jobs accessibility (see "Destination accessibility" in Table 2) but also with the proximity to transport networks ("Distance to transit" – subway or tramway – and "Distance to main roads") which may favor longer distance commute through the speed they offer. Commuting mode may be influenced by these proximities to transport networks but also by density and the possibility to walk ("Design", i.e. "walkability" index). "Diversity" reflects the mix of land use through points of interest. To this we add two variables measuring the social environment of the location, one relating to population ethnic origin, the other to income (as pointed in the literature review above; Schirmer et al, 2014). The eight variables are recapped in Table 2.

Table 2: Built and social environment dimensions

Dimensions	Description
Density	Population density of the commune: it is the ratio of population to commune area (in square kilometers)
Distance to transit	Distance to transit network: the distance from the centre of the commune to the nearest heavy public transportation station (subway or tramway)
Distance to main roads	Distance to major road network (divided highways): the distance from the centre of the commune to the nearest road interchange
Design	"Walkability" in the commune (derived from Frank et al, 2010): maximal number of 3-way intersections within a floating radius of 1 kilometer in the commune
Diversity	POI in each commune are available with the "Base Permanente des Equipements" (BPE, 2015). This database records the presence of equipments and services categorized in 27 types covering public services, banks and other commercial services, retail, education, health services and so on. We measure the intensity of POI as the presence, in the commune, of at least one equipment for each of the 27 categories. Thus the variable may vary from 0 (no equipment at all) to 27 (all 27 categories of equipments are present in the commune)
Destination accessibility	Accessibility to jobs: according to Hansen's indicator of accessibility (1959) it is computed for each commune i as $\sum_k \frac{Jobs_k}{d_{ik}^2}$ where $Jobs_k$ is the number of jobs in commune k and d_{ik} is the distance between the commune i and the commune k, with $d_{ii} = 0.5 * \sqrt{area_i}$
Social environment	Percentage of immigrants in the commune (those who are born as foreigners outside France and living in France, whether still foreigners or naturalized French)
Social environment	Median of the "equivalised disposable income" in the commune: this is the total income of a household, after tax and other deductions, that is available for spending or saving, divided by the number of household members each weighted according to the OECD equivalence scale (FiLoSoFi, 2015)

Note that by clustering communes of residential location of the individuals, the transit accessibility of the workplace (destination) seems to be ignored. However, by including the "Distance to transit" in the commune of residence – at the origin of the commute – the potential of using transit to access to workplace is dealt with.

Distances are first computed as Euclidean distances starting from the commune centre defined as the location of the town hall (source Geofla). Then a correction factor of 1.3 is applied to take account of average network design in urban areas. Descriptive statistics of the eight variables are given in Table 3.

Table 3: Descriptive statistics of variables used for clustering communes (N=123)

	Unit	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Coded into
Density	inhab. / km²	51	307	560	1 790	1 412	19 745	Log
Immigrants	%	1%	4%	6%	8%	10%	29%	Log
Median Inc.	euros / year	15 006	23 333	25 297	25 210	27 476	37 287	-
Jobs access.	-	577	2 354	4 136	8 896	9 029	73 360	Log
POI 27	number	4	12	15	16	21	27	-
Walkability	within 1km radius	34	111	172	209	284	510	Squareroot
Dist. roads	meters	495	1 720	2 520	3 032	4 224	7 743	Squareroot
Dist. transit	meters	570	8 532	15 998	16 234	23 651	40 925	Squareroot

As shown in Table 3, given their dispersion some variables are transformed in order to approximate a normal distribution: percentage of immigrants, population density and jobs accessibility are transformed into logarithms; distances to transit and main roads, and walkability index are transformed into square root. Even after this transformation the level of correlations between them is high (see Appendix A): density is highly correlated with jobs accessibility, POI 27, walkability and distance to transit; jobs accessibility is highly correlated with POI 27, walkability and distance to transit. This justifies using clustering in order to synthesize these various BE indicators.

2.1.3 Choice of covariates for individual self-selection of residential location

According to the previous literature survey of the most significant location attributes influencing residential location choice, we can identify the main dimensions of *individual* characteristics which could capture the preferences in location choice: these include life cycle features (age and family size), wealth through income level, social profile, immigrant or not and preference for car mobility.

As suggested in the literature review coarsened exact matching (CEM) preprocessing is applied (Iacus et al, 2012). Measuring imbalance between the empirical distribution of pretreatment covariates of the treated and control group is done with an L1 distance (also called Manhattan distance). For that the covariates need to be discretized (for continuous variables) or recoded (for categorical variables). Through this coding variables' values are "coarsened" (i.e. grouped) so that they have the same meaning according to the researcher assumptions. This is the case for instance for age or household size. Some variables are already coarsened and taken as they are, like gender, or may be recoded like qualification level which is coarsened along main diploma stages.

Again we have to make a compromise with the availability of variables in the MOBPRO database. We capture lifecycle stage through individual's age (discretized in classes) and household size. Income is unfortunately not available, so we approximate it through the housing tenure status (owner, social housing tenant, other tenant) and the qualification level (the highest diploma obtained by the individual). Person origin is measured through the status of immigrant or not. Preference for car mobility is captured through the car availability for the individual: we divide the number of cars available in the household by the number of working people and discretize this ratio (zero, less than one, one or more). We also have the variable "Living in a house" which could have been a proxy for preference for a house. However, this variable can be considered as partly a consequence of "treatment" and it should not be included in the matching process (Ho et al, 2007). Statistics over the whole sample are given in Table 4.

The same set of variables is used for the selection equation in the sample selection approach (i.e. the set of Z variables, see below).

Table 4: Descriptive statistics of individual covariates used in selection

Household size

	1	2	3	4	5+	total
	41 276	56 833	41 226	39 494	22 506	201 335
	21%	28%	20%	20%	11%	100%

Age

	(0,24]	(24,34]	(34,44]	(44,54]	(54,100]	total
	22 751	58 571	50 753	44 641	24 619	201 335
	11%	29%	25%	22%	12%	100%

Housing tenure of the household

other tenant	owner	social_housing_tenant	total
74 855	92 455	34 025	201 335
37%	46%	17%	100%

Car availability (ratio of cars per working people in the household)

zero	less than one	one or more	total
28 304	56 934	116 097	201 335
14%	28%	58%	100%

Immigrant

no	yes	total
173 616	27 719	201 335
86%	14%	100%

Qualification

GCSE (age 15)	Technical school certificate	High school diploma	Degree	total
29 067	34 288	37 216	100 764	201 335
14%	17%	18%	50%	100%

2.1.4 *Choice of covariates for outcome equations*

A subset of these variables is selected for outcome equations, the same for commuting distance and mode choice. These are age, gender, housing tenure, qualification and car availability. They form the set of X variables in the outcome equations (see below).

"Household size" and "Immigrant" individual status are included in the Z set and not the X set. These two variables are hypothesized to have a strong influence on residential area choice but not on commuting distance, nor on commute mode choice. Thus the condition of "exclusion restriction" is respected.

2.2 Statistical methods

In this section the methods for computation of treatment effects are presented, first with the sample selection approach for commuting distance, second with the matching approach for mode choice. Third, the computation of the influence of BE due to RSS is presented.

2.2.1 *Sample selection approach for commuting distance*

Following Heckman et al (2001) and Mokhtarian and van Herick (2016) we consider the model of potential outcomes

$$Y(1) = X\beta_1 + U_1 \qquad\qquad Y(0) = X\beta_0 + U_0 \qquad\qquad D^* = Z\theta + U_D$$

The first two equations are outcome equations and D^* is a latent variable which generates $D(Z)$ the observed treatment decision with two values, 1 (treated state) and 0 (untreated state).

The outcome equation may be rewritten

$$Y = DY(1) + (1 - D)Y(0)$$

The average treatment effect for a randomly chosen individual is then

$$ATE = \bar{x}(\beta_1 - \beta_0) \qquad (3)$$

where $\bar{x}$ is the vector of sample means of the X variables for the whole sample.

With the assumption of jointly normally distributed errors, the average treatment on the treated is

$$TT(x, z, D[z] = 1) = \bar{x}(\beta_1 - \beta_0) + (\rho_1\sigma_1 - \rho_0\sigma_0)\frac{\varphi(z\theta)}{\emptyset(z\theta)}$$
$$= \bar{x}(\beta_1 - \beta_0) + (\lambda_1 - \lambda_0)\frac{\varphi(z\theta)}{\emptyset(z\theta)} \qquad (4)$$

where σ_1 and σ_0 are variance parameters, ρ_1 and ρ_0 are correlation parameters, φ is the standard normal probability density function and $\emptyset$ the standard normal cumulative density function.

In the sample selection approach ATE is the "true effect" of BE on any random individual, when self-selection is controlled (Mokhtarian and van Herick, 2016). TT is the total effect of BE on treated group. According to equations (3) and (4) TT can be decomposed in two components: the first one is the ATE for treated individuals which we denote $ATEt$; the second one is the influence of self-selection which we denote $RSSt$.

$$TT = ATEt + RSSt \qquad (5) \qquad \text{and} \qquad ATEt = \bar{x_t}(\beta_1 - \beta_0) \qquad (6)$$

where $\bar{x_t}$ is the vector of sample means of the X variables for the treated group only[2].

This model, also referred to as the "mover-stayer" model, is estimated in two steps, first the selection (probit) model to estimate the coefficients of Z and then the outcome equations by linear regressions, with Limdep 11 (Econometric Software Inc.). The full information maximum likelihood estimation has been dismissed because of convergence issues. The two-step method is more robust (Bushway et al, 2007), although standard errors are large.

The effect of BE can be compared to the effects of individual situation which is measured here by socioeconomic characteristics (SEC, selected previously). The marginal effects of these covariates are computed as follows. For a categorical covariate each category (except an arbitrary base category) is binary: the marginal effect of this category (e.g. being female) compared to the base case (here being male) can be computed as

$$\frac{y_1 - y_0}{y_0} = \exp(\beta) - 1$$

where y_1 and y_0 are the values of the dependent variable when the category is 1 (resp. zero) and β is the estimated coefficient of the category.

---

[2] We corrected the procedure of computation of treatment effects indicated in Limdep 11 documentation (section E56.2.6).

2.2.2 Matching approach and logit for mode choice

Following Ho et al (2007) the average treatment effect (ATE) is the expected difference in outcome measured over the whole population

$$ATE = \frac{1}{n}\sum_{i=1}^{n} E[Y_i(1) - Y_i(0)|X_i] \qquad (1)$$

where i denotes the n individuals, Y_i is the potential outcome (e.g. car use) and X_i is the set of individual covariates.

Another quantity of interest is the "average treatment of the treated" (TT) measured on *matched* individuals in the treatment group, which is expressed as

$$TT = \frac{1}{\sum_{i=1}^{n} T_i}\sum_{i=1}^{n} T_i E[Y_i(1) - Y_i(0)|X_i] \qquad (2)$$

where $T_i = 1$ for treated individuals, zero otherwise

In the matching approach ATE is the total observed influence of BE when self-selection is *not* controlled for. TT is the "true effect" of BE on treated group, that is to say when self-selection is controlled.

Regarding commuting mode choice the relevant outcome model is a discrete choice model (multinomial logit). The CEM matching approach implemented in R (R Core Team, 2019) is applied. The logit estimations are performed with Nlogit 6 (Econometric Software Inc.)

Regarding commute mode choice, the effects of individual SEC can also be estimated. Logit models of commute mode choice are estimated for treatment group of each of the six pairs of control/treatment in the matching approach. Only individual characteristics are available (with no alternative specific attributes such as time, cost and so on). They do not vary across alternatives, thus we make them alternative specific. We also include commuting distance (in logarithm) in order to control its effect on mode choice.

In order to compute the marginal effects of the various dummy covariates on modal shares (e.g. setting gender as male and then female and compare the results) we use the simulation capability of Nlogit (see Hensher et al, 2015). The changes in percent share come from simulations computing the predicted choice probability times the number of observations in the simulated sample (they are reported only when at least one mode-specific coefficient of the covariate is significant at the 95% level).

2.2.3 Computing the influence of BE due to self-selection

Some authors (e.g. Mokhtarian and van Herick, 2016) have explicitly computed a quantity named "built environment proportion" (BEP), defined as the ratio of the true BE effect (when RSS is controlled for) to the total effect of BE. While the numerator is ATE in the sample selection approach, the restriction of the denominator to TT is inappropriate according to Mokhtarian and van Herick (2016). Inverting the labeling of control and treatment would produce different results. It seems more natural to include in the denominator not only TT but also TUT, that is a quantity similar to TT computed on the untreated individuals, in order to obtain the total effect. However, as shown by van Herick and Mokhtarian (2019) this BEP measured in the sample selection approach is equal to 1.0[3]. Moreover, different methods for measuring the BEP on the same set of data yield substantially different results (van Herick and Mokhtarian, 2020).

Despite this limitation we compute this ratio for commuting distance and also the influence of BE due to self-selection that is to say $(1 - BEP) * 100$ in order to compare our

---

[3] This is empirically confirmed by our own computations of BEP.

results with those found in the literature (basically those listed in Mokhtarian and van Herick, 2016).

Regarding the matching approach, the BEP is computed as the ratio of TT to ATE, following again Mokhtarian and van Herick (2016).

3   RESULTS

We first show the typology of communes followed by the effects computed on the commuting distance with the sampling selection approach and then on the commute mode choice with the matching approach.

3.1   Typology of communes

The sample consists of 123 communes, the commune of Lyon being divided into its nine districts ("arrondissements"). The eight variables (see Table 3, some of them transformed) are centered and scaled. A hierarchical clustering (not shown here) indicates a relevant number of clusters around four. We then perform a kmeans clustering technique based on the algorithm of Hartigan and Wong (1979) implemented in R. Clustering is repeated with five random starts with a number of clusters going from three to five. Finally a four clusters typology is retained with a ratio of between sum of squares to total sum of squares of 65.5 % and sizes of 23, 23, 42 and 35 observations (communes). Summary statistics per cluster are given in Table 5 while detailed statistics are given in Appendix B.

Table 5: Summary statistics of the four clusters of communes*

	Immigr.	Median inc.	Density	Jobs access	Dist. Transit	Dist. Motorways	Walkab.	POI 27	N
Unit	% in commune	euros per year**	1000 inhab. / km²	**	km	km	**		
1. "Urban centre"	16%	76	6.7	100	3.1	1.8	100	24	23
2 "Suburb"	11%	81	1.3	23	13.8	2.2	72	19	23
3 "Periurban"	6%	100	0.7	16	16.5	3.2	49	15	42
4 "Deep periurban"	4%	95	0.2	6	26.2	4.2	26	10	35
All	8%	90	1.8	30	16.2	3.0	56	16	123

* statistics per commune averaged over the cluster
** for the ease of interpretation of clusters regarding income, jobs accessibility and walkability index, the respective figures are coded in base 100, 100 being the maximum mean observed in one of the clusters
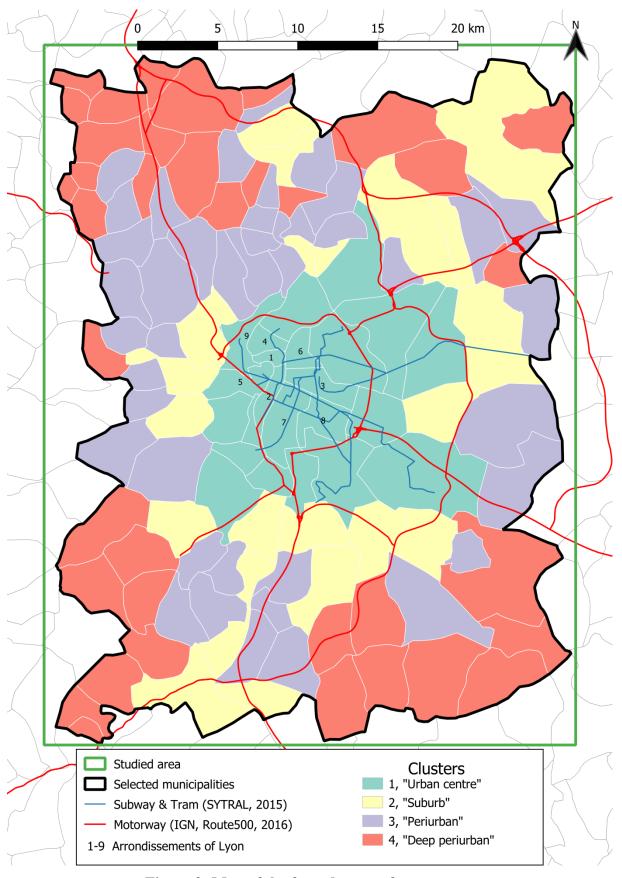
These clusters are described in Appendix B and shown on the Lyon area map in Figure 2. We give a "nickname" to each cluster, summing up their profile: "Urban centre", "Suburb", "Periurban" and "Deep periurban" respectively. Roughly speaking the communes of the three outer clusters ("Suburb", "Periurban" and then "Deep periurban") are respectively located at greater distance from the city centre, a monocentric feature which was expected. However, while the communes of cluster 1 "Urban centre" are contiguous this is not the case for the three other clusters. This reflects the polycentric feature of the conurbation with a spatial and social diversity of communes regarding the eight covariates at stake in the clustering.

16

Figure 2: Map of the four clusters of communes

Table 6 shows the distribution of individuals according to their residence in the four clusters of communes.

Table 6: Distribution of individuals across the four clusters of communes

Cluster	1	2	3	4	total
Number of individuals	151 532	27 809	16 373	5 621	201 335

Since there are four clusters the pairing of clusters in control and treatment groups yields six pairs. The control group is systematically the cluster with the highest level of average density in the pair, the other being the treatment group.

In the following the treatment effect is analyzed, first for commuting distance then for commute mode choice.

3.2 Commuting distance with the sample selection approach

Table 7 gives the number of individuals and the mean commuting distance (one way) in each cluster.

Table 7: Commuting distance in each cluster

Clusters	1	2	3	4	Total
N obs.	151 532	27 809	16 373	5 621	201 335
Mean distance	5.63	9.69	10.57	13.89	6.82

The detailed results for the six series of sample selection models involving for each pair of clusters a binomial probit model (selection) and two linear regression models (one for treated and one for control group) are given in Appendix C. Table 8 below recapitulates the various treatment effects which are computed from model results.

Table 8: Treatment effects on commuting distance (sample selection)

	Control	Treatment	ATE = BE true effect (km)	ATE / mean distance in control (all)	TT = BE overall effect (km)	(1-BEP) = influence of BE due to RSS	ATEt = BE true effect on the treated (km)	ATEt / mean distance in treatment
Pair 1-2	Urban centre	Suburb						
N obs.	151 532	27 809						
Distance	5.63	9.69	1.73	31%	2.28	24%	1.71	18%
Pair 1-3	Urban centre	Periurban						
N obs.	151 532	16 373						
Distance	5.63	10.57	2.25	40%	2.63	14%	2.29	22%
Pair 1-4	Urban centre	Deep periurban						
N obs.	151 532	5 621						
Distance	5.63	13.89	2.11	37%	5.83	64%	2.28	16%
Pair 2-3	Suburb	Periurban						
N obs.	27 809	16 373						
Distance	9.69	10.57	1.96	20%	1.42	-38%	1.94	18%
Pair 2-4	Suburb	Deep periurban						
N obs.	27 809	5 621						
Distance	9.69	13.89	1.70	18%	1.95	12%	1.70	12%
Pair 3-4	Periurban	Deep periurban						
N obs.	16 373	5 621						
Distance	10.57	13.89	29.75	281%	31.39	5%	29.20	210%

ATE, which is the true BE effect, is an estimate of the increase in daily commuting distance (one way) for a randomly selected commuter. For instance (see 4th column of Table 8), for pair 1-2 a commuter is expected to increase his or her one way commuting distance by 1.73

km on average when living in cluster 2 ("Suburb") compared to living in cluster 1 ("Urban centre").

TT is an estimate of the increase of daily commuting distance for a commuter living in an outer cluster (treatment) when compared with a commuter living in the inner cluster (control), all other covariates being constant. It aggregates the true effect of BE and the effect of residential self selection. TT varies between 1.42 km (pair 2-3) and 5.83 km (pair 1-4).

ATEt is the true effect of BE on the treated. It is mostly similar to ATE, but computed on individuals belonging to the treated group. The ratios of ATEt to mean distance in the treated group show that the influence of BE on commuting distance for those individuals amounts to about 10-20% of this distance.

The outlier results of ATE and ATT for pair 3-4 lead us to assess the robustness of the estimation to varying sets of covariates. One of the two "exclusion restriction" covariates is "Immigrant" which is significant in the (probit) selection equation in the initial estimation. If we exclude this variable from the estimation we obtain a huge value for ATE expressed in km (the constant for the least squares regression model of treated cases goes around 16, where the dependent distance variable is coded in logarithm). The same kind of variation occurs, as a test, for pair 1-2, where both "exclusion restriction" covariates are significant in the initial probit model: for instance, excluding "Immigrant" covariate produces an ATE of 4.76 km and a TT of 2.03 km. We come back to this point in the discussion section.

The marginal effects of individual SEC are computed and can be compared to the BE effect for the treated groups (ATEt). They are shown in Table 9 and illustrated in Figure 3.

Table 9: Marginal effects on commuting distance (sample selection, treated)

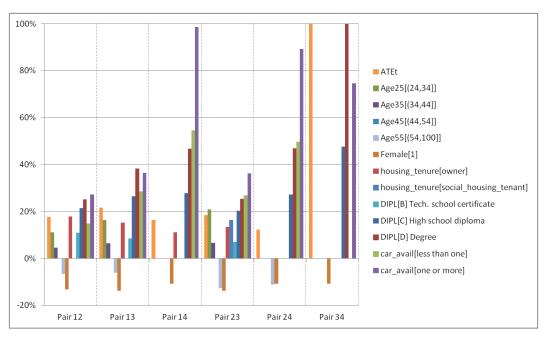|	Pair 1-2	Pair 1-3	Pair 1-4	Pair 2-3	Pair 2-4	Pair 3-4
ATEt	18%	22%	16%	18%	12%	210%
Age25[(24,34]]	11%	16%		21%		
Age35[(34,44]]	5%	7%		7%		
Age45[(44,54]]						
Age55[(54,100]]	-7%	-6%		-13%	-11%	
Female[1]	-13%	-14%	-11%	-14%	-11%	-11%
housing_tenure[owner]	18%	15%	11%	13%		
housing_tenure[social_housing_tenant]				16%		
DIPL[B] Tech. school certificate	11%	9%		7%		
DIPL[C] High school diploma	22%	27%	28%	20%	27%	48%
DIPL[D] Degree	25%	38%	47%	25%	47%	105%
car_avail[less than one]	15%	29%	55%	27%	50%	
car_avail[one or more]	27%	36%	99%	36%	89%	75%

Figure 3: Marginal effects on commuting distance (sample selection, treated)

3.3 Commute mode choice with the matching approach

Table 10 gives the shares of commuting travel modes in each cluster.

Table 10: Commuting mode shares in each cluster

	walking	two-wheels	car	public transport	total	N commuters
cluster 1 "Urban centre"	14.8%	5.4%	46.8%	33.1%	100.0%	151 532
cluster 2 "Suburb"	8.3%	2.8%	74.1%	14.9%	100.0%	27 809
cluster 3 "Periurban"	7.9%	2.8%	80.2%	9.0%	100.0%	16 373
cluster 4 "Deep periurban"	8.0%	1.7%	83.4%	6.8%	100.0%	5 621

As explained previously the matching approach is applied before the analysis of "treatment effect" on commuting mode choice. The CEM algorithm implemented in R is used (Iacus et al, 2009). This algorithm performs exact matching of individuals on the six individual covariates selected above, by first sorting all the individuals in strata, each of which has identical values for the six covariates. These are household size, age, housing tenure, car availability, immigrant status and qualification (see Table 4). Then individuals from one group who cannot be matched with anyone of the other group are discarded. Since we have enough data we opt to a k-to-k solution which gives the same number of treated and control units within each strata. Table 11 below shows the result of k-to-k matching for each of the six pairs. The percentage of matched individuals in each treatment group is close to 100%.

20

Table 11: CEM k-to-k matching for the six pairs of clusters

	Control	Treated	Control	Treated
	Cluster 1	Cluster 2	Cluster 2	Cluster 3
All	151 532	27 809	27 809	16 373
Matched	27 804	27 804	15 402	15 402
% matched in treated group		100.0%		94.1%
	Cluster 1	Cluster 3	Cluster 2	Cluster 4
All	151 532	16 373	27 809	5 621
Matched	16 368	16 368	5 601	5 601
% matched in treated group		100.0%		99.6%
	Cluster 1	Cluster 4	Cluster 3	Cluster 4
All	151 532	5 621	16 373	5 621
Matched	5 620	5 620	5 528	5 528
% matched in treated group		100.0%		98.3%

Table 12 gives for each pair of clusters the number of matched observations (equal for the control and the treatment group given the k2k matching) and the shares of commuting travel modes in each group of matched individuals. Then it gives the TT (the difference of shares of the four travel modes in percentage point between the two matched groups) which is the true BE effect, ATE the observed difference in travel mode shares between the two initial groups (all individuals, including not matched), the ratio of TT to ATE (which is the BEP) and the influence of BE due to RSS.

Table 12: Treatment effects on commuting mode shares for treatment groups (matched)

	Control (matched)	Treatment (matched)	TT (shares) = BE true effect	ATE = observed difference (all)	TT / ATE = BEP	(1-BEP) = influence of BE due to RSS
Pair 1-2	Urban centre	Suburb	%pt	%pt		
N	27 804	27 804				
walking	13.5%	8.3%	-5.2%	-6.5%	81%	19%
2wheels	5.4%	2.8%	-2.7%	-2.6%	103%	-3%
car	57.0%	74.1%	17.1%	27.3%	63%	37%
public_trans	24.1%	14.9%	-9.2%	-18.2%	50%	50%
total	100.0%	100.0%				
Pair 1-3	Urban centre	Periurban				
N	16 368	16 368				
walking	12.8%	7.9%	-4.9%	-6.9%	71%	29%
2wheels	6.0%	2.8%	-3.2%	-2.5%	126%	-26%
car	61.3%	80.3%	19.0%	33.5%	57%	43%
public_trans	19.9%	9.0%	-10.9%	-24.1%	45%	55%
total	100.0%	100.0%				
Pair 1-4	Urban centre	Deep periurban				
N	5 620	5 620				
walking	11.7%	8.0%	-3.7%	-6.7%	55%	45%
2wheels	5.3%	1.7%	-3.6%	-3.6%	98%	2%
car	66.1%	83.4%	17.3%	36.6%	47%	53%
public_trans	16.8%	6.8%	-10.0%	-26.3%	38%	62%
total	100.0%	100.0%				
Pair 2-3	Suburb	Periurban				
N	15 402	15 402				
walking	8.2%	8.1%	-0.2%	-0.4%	45%	55%
2wheels	2.8%	2.8%	0.1%	0.1%	81%	19%
car	76.5%	80.1%	3.7%	6.2%	59%	41%
public_trans	12.5%	9.0%	-3.5%	-5.8%	61%	39%
total	100.0%	100.0%				
Pair 2-4	Suburb	Deep periurban				
N	5 601	5 601				
walking	9.0%	8.0%	-1.0%	-0.3%	338%	-238%
2wheels	2.7%	1.7%	-1.0%	-1.0%	93%	7%
car	78.2%	83.5%	5.3%	9.3%	57%	43%
public_trans	10.2%	6.8%	-3.3%	-8.0%	42%	58%
total	100.0%	100.0%				
Pair 3-4	Periurban	Deep periurban				
N	5 528	5 528				
walking	7.9%	7.9%	0.0%	0.1%	16%	84%
2wheels	3.2%	1.8%	-1.4%	-1.1%	127%	-27%
car	81.0%	83.5%	2.5%	3.2%	78%	22%
public_trans	7.9%	6.9%	-1.1%	-2.2%	50%	50%
total	100.0%	100.0%				

Besides the treatment effect in the treated group, the effects of individual SEC covariates can also be estimated. The results of the six logit mode choice models and the marginal covariate effects are given in Appendix D. They are represented as six bar charts in Figure 4.

For instance, the first bar chart (top left) shows for Pair 1-2 the change in percentage point for each of the four modes shares induced by a change in dummy covariates. A change in car equipment level "Cmore1" from zero to one (there is at least one car per working people in the household) induces a 52%pt increase of car share in commuting and a 39%pt (resp. 8%pt) decrease in public transport (resp. walking) share. These levels of impacts of car equipment are estimated independently of other covariates and can be compared to the "treatment of the

treated" effect (TT, already shown in Table 12), i.e. the BE effect for those living in cluster 2 ("Suburb"), when compared to those living in cluster 1 ("Urban centre").
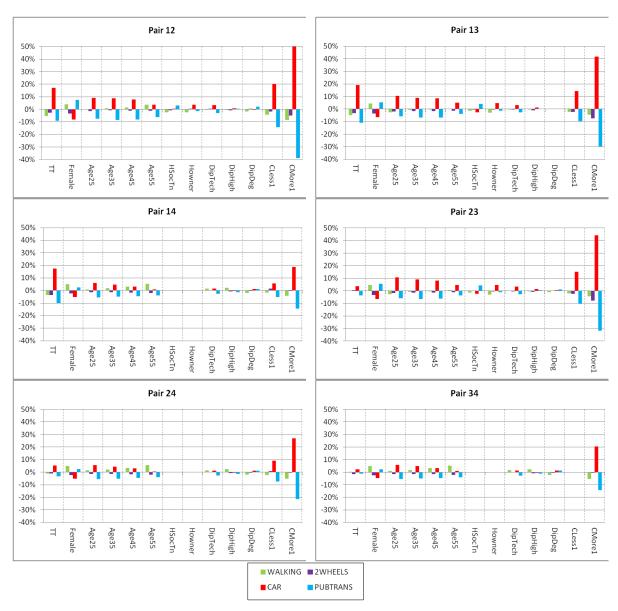

Figure 4: Marginal effects of various covariates on commuting modes shares

4   DISCUSSION

The results are commented regarding first commuting distance, second mode choice. Then they are put in perspective with the empirical literature. Finally, the limitations of this study are discussed.

4.1   Commuting distance

Regarding commuting distance, with the sample selection approach, the true BE effect measured by the ATE is shown as meaningful (see Table 8), especially when comparing "Urban centre" to "Suburb" (an increase of 31%), "Periurban" (40%) or "Deep periurban" (37%), if we except the outlier result of pair 3-4 (see below). The true BE effect on the treated (ATEt), i.e. on those living in outer clusters, is also meaningful but lower, and modest since the increase of the mean distance in the treated group goes from 12% to 22%.

ATE is lower than TT in general, evidencing some self-selection. The influence of BE due to RSS has large variations: it goes from 12% for the pair comparing "Periurban" to "Suburb" to 64% for the pair comparing "Deep periurban" to "Urban centre" (excluding the outlier results of pair 3-4, see below). However, we find a case of reverse influence of self-selection, that is for pair "Suburb" – "Periurban" (Pair 2-3 in Table 8). This can be referred to the concept of residential neighborhood type dissonance, where a person is considered dissonant if his or her current residential neighborhood does not coincide with his or her preferences regarding land use or travel (Schwanen and Mokhtarian, 2005; for a recent discussion of this issue see Guan et al, 2019). However, as pointed by Cao and Chatman (2016) travel preference is only one of the factors influencing residential choices, which could explain the lower overall effect of BE (including RSS) when compared to its true effect.

When compared with individual socioeconomic characteristics, the true effect of BE on the treated (ATEt, see Table 9 and Figure 3) comes never first, but after qualification level and car availability (at least a car in the household). When treated cluster is 2 "Suburb" or 3 "Periurban" the effect of car availability (one or more per commuter) amounts to 27%-36% increase of commuting distance. It is much higher when treated cluster 4 "Deep periurban" is considered, with a range of 80-100% (that is a doubling of commuting distance).

A limitation regarding the application of the sample selection approach is shown through the outlier results of ATE and ATT for the pair "Periurban" – "Deep periurban" (Pair 3-4 in Table 8). We assessed the robustness of the estimation to varying sets of covariates and found large variations in the estimation of the treatment effects. This point was also raised by Zhou and Kockelman (2008). Such results may indicate limits in the specification of the model where other more relevant variables might have been included (depending on their availability). Overall one cannot exclude the potential for "omitted variables". This could also explain why there are such large variations of treatment effects measured in the empirical literature.

4.2   Commute mode choice

First, the largest TT effects (the BE true effects in the matching approach) are observed for the three pairs where cluster 1 ("Urban centre") is the control group (see Table 12). For these pairs the BE produces an increase of car share between 17.1 and 19.0 %pt, a decrease of both public transport share between 9.2 and 10.9 %pt and walking share around 5 %pt. This is meaningful while not high. Moreover, regarding again these three pairs, the influence of BE due to RSS is also meaningful: from 19 to 45% for walking, 27 to 53% for car and 50 to 62% for public transport.

Regarding the three outer clusters, that is to say cluster 2 ("Suburb"), 3 ("Periurban") and 4 ("Deep periurban"), the variations in modal shares between them is much lower, and thus the TT-BE effect is significantly lower.

When comparing TT-BE effect with individual socioeconomic characteristics, the overall picture is that for each pair of control/treatment groups the impact of car availability "CMore1" (at least one car per working people in the household) is the highest over all other effects (see Figure 4), with a range of 19%-52%pt. The level of car availability "CLess1" (non zero but less than one car per working people) comes second for pair 1-2, 2-3 and 2-4, with an overall range of 6%-20%pt. The next highest effect is the TT-BE effect. Then come the socioeconomic covariates with gender and age categories as the most prominent effects.

Moreover, for the last three pairs (2-3, 2-4 and 3-4) comparing "Suburbs", "Periurban" and "Deep periurban", while car availability remains as the highest effect, the TT-BE effect comes this time after individual socioeconomic covariates effects (such as gender or age).

Overall, the BE has a meaningful effect in favoring car share to the detriment of public transport and walking, when moving from an urban centre characterised by density of jobs and high supply of public transport. However, car availability effect is in the forefront, and far above

the BE effect when considering full car availability. The BE effect comes second for the three pairs comparing "Urban centre" to the three outer clusters.

4.3 From the point of view of the empirical literature

Our results are in line with Stevens (2017) conclusion since we find a moderate impact of BE on commuting distance (10-20%). Specifically regarding distance our results are opposite to those of Ewing and Cervero (2001) since we show that SEC are primarily influencing commuting distance, before BE.

Regarding mode choice, our results are partly in line with Scheiner (2010) in the case of Germany, who shows the critical effect of car equipment whatever the distances travelled but also the strong effect of BE, reducing car use in cities opposed to suburbs and rural areas.

Regarding the relative scales of treatment effects, two studies using the sample selection approach can be referred to. Zhou and Kockelman (2008), when comparing Suburb/Rural (treated) to CBD/Urban (control) in Austin (Texas), obtained 58% for the ratio of ATE to TT. From Cao (2009) comparing four traditional (control) and four suburban neighborhoods (treatment) in Northern California the ratio ATE/TT (25.8/33.8) can be computed as 76%. Our ratio ATE/TT (which can be inferred from Table 8) goes from 36% to 88% (excluding the outlier result of pair 3-4 and the reverse result for pair 2-3). Thus our results are broadly in line with the two previous studies in an US context.

Finally, regarding the influence of BE due to RSS our range of values between 12% and 64% computed for commuting distance with the sample selection approach is comparable to the review by Mokhtarian and Van Herick (2016) with a range between 10 and 42%.

4.4 Limitations of this study

A first limitation would be that using car availability as an exogenous variable in a model linking BE and TB would induce a simultaneity bias (a particular form of endogeneity bias as identified by Mokhtarian and Cao, 2008). Indeed, car ownership is playing a mediating role between BE and car use, as shown by Van Acker and Witlox (2010), Cao et al (2007) or Ding et al (2017) through structural equation modeling. BE and SEC influence car ownership, thus indirectly travel choices, while BE also influences directly travel choices. However, both our matching and sample selection approaches include car availability in the selection covariates and thus control the mediating role of car ownership. By this way we believe having minimized this bias risk.

A second limitation is inherent to the material which is available. The first is that travel attitudes are not considered, basically because such data are not available in our study. However, by implementing the sample selection approach such "unobservables" are controlled. Regarding mode choice the matching approach reduces the biases from self-selection although not as thoroughly as the sample selection approach. Yet preference for car commuting is partially captured in the matching through the car equipment variable.

Third, the effect of each built attribute cannot be analysed separately, due to correlation between most of them. This reflects the current spatial structure of the city and its past urban policy (e.g. transit infrastructure is supplied where density is high). With clustering we think to have overcome this issue. Moreover, the clustering we propose is a progress when compared to studies which only consider the distance to the city centre.

Fourth, using only commuting travel data may limits the scope of the results since the BE can influence other trips like e.g. shopping or leisure trips.

A further limitation is that the clustering of communes obtained with covariates measured at the level of the commune is subject to the Modifiable Areal Unit Problem (Openshaw, 1984). This issue is exacerbated by the fact that data are till now available only at the commune level which is an especially rough division of geographic space. Associated to

the MAUP are both a scale effect, given the communes surfaces – which impacts also the measurement of home-work distance, and a zone effect which depends on the boundaries of communes. Alternative aggregations whether at finer levels or with different boundaries would probably give different outcomes in clustering. Such alternative was not possible in our case study given the dependency on administrative data given at the commune level. This issue is also raised by Guan et al (2019), for instance on the possibility that people may value regional accessibility (relevant for job access in our application to commuting) more than neighbourhood-level accessibility in residential choice. This issue is left open. However, the application of our methodology to other urban areas would help to test the robustness of the results regarding the relative effects of BE and individual socioeconomic characteristics in various spatial contexts.

5   CONCLUSION

The links between built and social environment on the one hand and commuting behavior on the other hand have been explored with a control of residential self selection, thanks to publicly available data. This makes this study reproducible in other areas with similar data availability.

Regarding commuting distance, the true effect of built and social environment appears modest with an increase in the range of 10-20% of this distance when comparing e.g. "urban centre" and "periurban". This comes behind some individual socioeconomic characteristics such as car availability and qualification. Car availability produces an increase of commuting distance in the range of 27%-36% when moving to suburb or periurban and even as high as 100% in the deep periurban. Regarding commuting mode choice, again the true effect of built and social environment is modest, with a nearly 20%pt increase of car share and around 10%pt decrease or public transport share for the most prominent effects. It comes behind car availability, especially when there is at least one car per working people in the household with an increase in car share ranging from 20%pt to 50%pt. However, these results are valid subject to the limitations raised previously, i.e. measurement issues and MAUP.

When it comes to policy implications, in the European context, these results suggest the importance of influencing directly car use, if not car ownership, while trying to modify the built environment would provide only limited results, at least from the point of view of commuting behavior. The policy measures are known, whether parking policy, carpooling with high occupancy reserved lanes, restriction of road capacity devoted to car traffic or even road or congestion pricing. In cities where housing is expensive and the property market rather seized up, trying to bring workers homes closer to their jobs locations would look rather vain.

One perspective of this study would be to expand the analysis to other European urban areas with comparable available data.

ACKNOWLEDGEMENTS

REFERENCES

Aditjandra, P.T., Cao, X., Mulley, C., 2012. Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. *Transportation Research Part A* 46 (2012) 22–32

Aguilera, A., 2005. Growth in commuting distances in French polycentric metropolitan areas: Paris, Lyon and Marseille. *Urban Studies*, 42(9), pp. 1537-1547.

BPE, 2015. Base permanente des équipements (BPE) - 2015, INSEE.

Bushway, S., Johnson, B.D., Slocum, L.A., 2007. Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *Journal of quantitative criminology* 23 (2), 151-178.

Cao, X. J., Schoner, J. (2014). The influence of light rail transit on transit use: An exploration of station area residents along the Hiawatha line in Minneapolis. *Transportation Research Part A: Policy and Practice*, 59, 134-143.

Cao, X., 2010. Exploring Causal Effects of Neighborhood Type on Walking Behavior Using Stratification on the Propensity Score. *Environment and Planning A: Economy and Space*, 42(2), 487–504. https://doi.org/10.1068/a4269

Cao, X., Chatman, D., 2016. How will smart growth land-use policies affect travel? A theoretical discussion on the importance of residential sorting. *Environment and Planning B: Planning and Design* 2016, volume 43, pages 58 –73.

Cao, X.J., 2009. Disentangling the influence of neighborhood type and self-selection on driving behavior: an application of sample selection model. *Transportation* (2009) 36:207–222.

Cao, X.J., Fan, Y., 2012. Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B: Planning and Design* 2012, volume 39, pages 459 – 470

Cao, X.J., Mokhtarian, P.L., Handy, S.L., 2007. Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation* (2007) 34:535–556. DOI 10.1007/s11116-007-9132-x

Cao, X.J., Mokhtarian, P.L., Handy, S.L., 2009. Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, 29:3, 359-395, DOI: 10.1080/01441640802539195

Cao, X.J., Xu, Z., Fan, Y., 2010. Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments. *Transportation Research Part A* 44 (2010) 797–805.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, Diversity, and Design. *Transportation Research Part D*, Vol. 2, No. 3, pp. 199-219.

Charmes, E. (coord.), 2010, La densification en débat, *Etudes foncières*, n° 145, mai-juin, p. 19-38

de Abreu e Silva, J., 2014. Spatial self-selection in land-use–travel behavior interactions: Accounting simultaneously for attitudes and socioeconomic characteristics. *The Journal of Transport and Land Use*, 7(2), 63–84.

Desjardins, X., 2011. The contribution of urban and regional planning to climate change mitigation. Cybergeo: European Journal of Geography https://doi.org/10.4000/cybergeo.23531 (in French).

Ding, C., Wang, D., Liu, C., Zhang, Y., Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A* 100 (2017) 65–80.

Ettema, D., Nieuwenhuis, R., 2017. Residential self-selection and travel behaviour: What are the effects of attitudes, reasons for location choice and the built environment? *Journal of Transport Geography* 59 (2017) 146–155

Ewing, R., Cervero, R., 2001. Travel and the Built Environment: A Synthesis. *Transportation Research Record*, Vol. 1780, pp. 87-114.

Ewing, R., Cervero, R., 2010. Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, 11 May 2010, 30 p. +

Ewing, R., Hamidi, S., 2015. Compactness versus Sprawl: A Review of Recent Evidence from the United States. *Journal of Planning Literature,* 2015, Vol. 30(4) 413-432.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M., Joyce, A., Kircher, J., Greene, W., 2015. Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Studies* 2015, Vol. 52(13) 2330–2348

FiLoSoFi, 2015. Fichier Localisé Social et Fiscal (FiLoSoFi) - Année 2015. https://www.insee.fr/fr/statistiques/3560121

Frank L D, Sallis J F, Saelens B E, Leary L, Cain K, Conway T L, Hess P M, 2010. The development of a walkability index: application to the Neighborhood Quality of Life Study. *Br J Sports Med* 2010; 44: 924–933

Guan, X., Wang, D. Cao, X.J., 2019. The role of residential self-selection in land use-travel research: a review of recent findings. *Transport Reviews*, DOI: 10.1080/01441647.2019.1692965

Handy, S., 2017. Thoughts on the Meaning of Mark Stevens's Meta-Analysis. *Journal of the American Planning Association*, 83:1, 26-28, DOI:10.1080/01944363.2016.1246379

Hansen, W.G., 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 25:2, 73-76. DOI: 10.1080/01944365908978307

Hartigan, J. A., Wong, M. A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1(1979), pp. 100-108

Heckman JJ. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46:931–61.

Heckman, J.J., Tobias, J.L., Vytlacil, E.J., 2001. Four parameters of interest in the evaluation of social programs. *Southern Economic Journal* 68(2), 210–223.

Hensher, D.A., Rose, J.M., Greene, W.H., 2015. *Applied Choice Analysis*. Second Edition. Cambridge University Press.

Ho, D.E., Imai, K., King, G., Stuart, E.A., 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* (2007) 15:199–236

Humphreys, J., Ahern, A., 2019. Is travel based residential self-selection a significant influence in modal choice and household location decisions? *Transport Policy* 75 (2019) 150–160

Iacus S. M., King, G., Porro, G., 2009. cem: Software for Coarsened Exact Matching. *Journal of Statistical Software*, June 2009, Volume 30, Issue 9.

Iacus S. M., King, G., Porro, G., 2011. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, March 2011, Vol. 106, No. 493, Theory and Methods, pp. 345-361.

Iacus S. M., King, G., Porro, G., 2012. Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20, 1, pp. 1-24.

IGN BD TOPO, 2011. Descriptif de contenu (Version 2.1). Institut Géographique National. Octobre 2011.

INSEE MOBPRO, 2013. Mobilités professionnelles des individus : déplacements commune de résidence / commune de travail en 2013. Recensement général de la population - Fichiers détail. Parution juin 2016. www.insee.fr

Mokhtarian, P.L., Cao, X.J., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B* 42 (2008) 204–228.

Mokhtarian, P.L., van Herick, D., 2016. Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches. *The Journal of Transport and Land Use*, Vol. 9, No 1 [2016] pp. 9–28 http://dx.doi.org/10.5198/jtlu.2016.788

Næss, P., 2014. Tempest in a teapot: The exaggerated problem of transport-related residential self-selection as a source of error in empirical studies. *The Journal of Transport and Land Use*, vol. 7 n° 3 [2014] pp. 57–79 doi: 10.5198/jtly.v7i3.491

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books

Pouyanne, G., 2005. Urban Form and Travel Patterns. An application to the metropolitan area of Bordeaux. *Canadian Journal of Regional Science*, vol 28:1, 19-47

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rosenbaum, P.R., Rubin D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.

Salon, D., 2015. Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. *Research in Transportation Economics* 52 (2015) 34-45.

Scheiner, J., 2010. Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002. *Journal of Transport Geography* 18 (2010) 75–84

Schirmer, P.M., Eggermond, M.A.B., Axhausen, K.W., 2014. The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*. Vol. 7 No. 2 [2014] pp. 3–21

Schwanen, T., Mokhtarian, P.L., 2005. What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D* 10 (2005) 127–151.

Stevens, M.R., 2017. Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, 83:1, 7-18, DOI:10.1080/01944363.2016.1240044

Susilo, Y. O., Maat, K., 2007. The influence of built environment to the trends in commuting journeys in the Netherlands. *Transportation* (2007) 34:589–609. DOI 10.1007/s11116-007-9129-5

Van Acker, V., Witlox, F., 2010. Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography* 18 (2010) 65–74

van Herick, D., Mokhtarian, P.L., 2019. Quantifying the share of total apparent effect that is genuinely due to a treatment: a comparison of methods. *Under peer review*; available from the authors.

van Herick, D., Mokhtarian, P.L., 2020. How much does the method matter? An empirical comparison of ways to quantify the influence of residential self-selection. *Travel Behaviour and Society* 18 (2020) 68–82.

Van Wee, B., Boarnet M.G., 2014. Reaction to the paper "Tempest in a Teapot: The exaggerated problem of transport-related residential self-selection as a source of error in empirical studies". *The Journal of Transport and Land Use*, vol. 7 n° 3 [2014] pp. 81–86 doi: 10.5198/jtlu.v7i3.491

Wang D., Lin, T., 2019. Built environment, travel behavior, and residential self-selection: a study based on panel data from Beijing, China. *Transportation* (2019) 46:51–74. DOI 10.1007/s11116-017-9783-1

Winship, C., Morgan, S.L., 1999. The estimation of causal effects from observational data. *Annual Review of Sociology* 25: 659-706.

Wolday, F., Næss, P., Cao, X.J., 2019. Travel-based residential self-selection: A qualitatively improved understanding from Norway. *Cities* 87 (2019) 87–102

Zhou, B., Kockelman, K/M., 2008. Self-Selection in Home Choice. Use of Treatment Effects in Evaluating Relationship Between Built Environment and Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2077, pp. 54–61. DOI: 10.3141/2077-08

Appendix A

Table A.1: Correlations between covariates of communes*

	Density	Immigrants	Median Inc.	Jobs access.	POI 27	Walkability	Dist. Roads	Dist. Transit
Density	1.00							
Immigrants	0.71	1.00						
Median Inc.	-0.40	-0.69	1.00					
Jobs access.	0.92	0.67	-0.24	1.00				
POI 27	0.82	0.77	-0.47	0.81	1.00			
Walkability	0.74	0.73	-0.47	0.73	0.88	1.00		
Dist. Roads	-0.45	-0.54	0.31	-0.46	-0.48	-0.52	1.00	
Dist. Transit	-0.84	-0.68	0.31	-0.93	-0.79	-0.73	0.42	1.00

* Density, Immigrants and Jobs access. are in logarithms, Walkability, Dist. roads and Dist. transit in square root.

Appendix B

Table B.1: Detailed descriptive statistics of the four clusters

	Immigrants	Median inc.	Density	Jobs access.	Dist. Transit	Dist. Motorways	Walkab.	POI 27	N obs.
Cluster 1									23
Min.	9%	15 006	1 465	8 321	570	708	149	17	
1st Qu.	10%	19 105	2 609	14 187	1 419	1 236	304	23	
Median	16%	20 210	4 165	17 847	2 959	1 569	385	25	
Mean	16%	21 145	6 738	29 300	3 063	1 778	371	24	
3rd Qu.	18%	24 114	9 571	40 368	4 294	2 305	436	26	
Max.	29%	29 548	19 745	73 360	8 375	3 461	510	27	
St. dev.*	0.61	0.95	0.65	0.67	0.39	0.60	0.59	0.44	
Cluster 2									23
Min.	7%	16 579	179	1 675	4 926	495	130	14	
1st Qu.	8%	20 865	802	3 798	8 151	1 604	207	18	
Median	10%	22 834	1 098	5 452	10 070	2 077	269	20	
Mean	11%	22 638	1 281	6 733	13 785	2 213	268	19	
3rd Qu.	13%	24 258	1 678	9 012	19 621	2 410	301	22	
Max.	20%	27 015	2 795	18 858	27 892	5 389	447	26	
St. dev.*	0.48	0.64	0.55	0.56	0.72	0.82	0.63	0.55	
Cluster 3									42
Min.	3%	23 293	209	2 193	8 040	1 197	70	10	
1st Qu.	5%	25 273	427	3 225	12 990	1 745	135	14	
Median	5%	27 476	521	4 191	15 500	3 163	170	15	
Mean	6%	27 890	657	4 796	16 464	3 196	182	15	
3rd Qu.	6%	29 842	784	5 514	19 550	4 206	228	17	
Max.	10%	37 287	2 279	11 818	28 827	6 402	324	23	
St. dev.*	0.44	0.81	0.38	0.41	0.45	0.91	0.57	0.47	
Cluster 4									35
Min.	1%	23 039	51	577	14 206	1 128	34	4	
1st Qu.	3%	25 085	161	1 204	19 393	2 787	74	9	
Median	4%	26 128	213	1 719	26 164	4 265	95	11	
Mean	4%	26 357	233	1 829	26 220	4 197	98	10	
3rd Qu.	5%	27 499	292	2 314	31 176	5 077	115	12	
Max.	9%	31 207	494	4 136	40 925	7 743	171	16	
St. dev.*	0.59	0.46	0.38	0.49	0.53	0.88	0.44	0.53	

* the standard deviations are computed on centered and scaled values of variables

Description of the clusters

Cluster 1 ("Urban centre") groups together 23 communes with a high population density (6,700 inhab./km² on average), including the two main communes (the nine districts of Lyon and Villeurbanne) and the greatest in size adjoining communes. The immigrant level is the highest (16% compared to overall average of 8%). The median disposable income is 76% of the maximum (observed in cluster 3). These communes benefit from the maximum jobs accessibility and on average nearly the whole set of equipments (24 compared to POI 27 categories). They have the best walkability (maximum index) and their center is close to main road interchanges (less than 2 km on average) and heavy transit network (3.1 km on average with a minimum of 570 m).

Cluster 2 ("Suburb") groups together 23 communes with one fifth of the average density of cluster 1 and includes average sized communes located mostly in the second ring of suburbs. The immigrant level is also above the average (11%) and the median disposable income is 81% of the maximum. These communes have a markedly lower jobs accessibility (23% of the maximum) but a good set of equipments (19 on average). They have a good walkability (72% of the maximum) and their center is close to main road interchanges (2 km) but remote from heavy transit network (nearly 14 km on average).

Cluster 3 ("Periurban") groups together 42 communes with low density (one tenth of cluster 1) and includes average sized communes beyond the second ring of suburbs. The immigrant level is lower than the average (6%) and the median disposable income is at the maximum. These communes have a low jobs accessibility (16% of the maximum) and a moderate set of equipments (15 on average). They have a moderate walkability (49% of the maximum) and their center is close to main road interchanges (about 3 km) but very remote from heavy transit network (more than 16 km on average).

Cluster 4 ("Deep periurban") groups together 35 communes with very low density (one thirtieth of cluster 1) and includes mainly little sized communes in the peri-urban. The immigrant level is the lowest (4%) and median disposable income is near the maximum. These communes have a very low accessibility to jobs (6% of the maximum) and a poor set of equipments (10 on average). They have a poor walkability (26% of the maximum) and their center is reasonably close to main road interchanges (about 4 km) but very remote from heavy transit network (more than 26 km on average).

Table C.1: Sample selection models for commuting distance

Binomial Probit Model	Pair 1-2	Pair 1-3	Pair 1-4	Pair 2-3	Pair 2-4	Pair 3-4
TREATED	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant	-1.80039***	-2.54928***	-2.97154***	-.97056***	-1.64472***	-.78312***
AGE25	-.02451*	-.16356***	-.09119***	-.15173***	-0.04485	.12817***
AGE35	-.03483**	-.06828***	-0.0077	-0.01688	.07486**	.11733***
AGE45	-0.01594	.08996***	.07552***	.14745***	.15290***	0.0171
AGE55	-.05480***	.15505***	0.04325	.26435***	.15134***	-.10244**
TENANT_S	-0.00335	-.32016***	-.45616***	-.34416***	-.49607***	-.24980***
OWNER	.28490***	.41248***	.49248***	.17561***	.33995***	.20638***
LESS1	.37401***	.42247***	.40931***	.17717***	.20222***	0.07153
MORE1	.78410***	.94352***	.87784***	.30633***	.33674***	0.0869
IMMIGRAN	-.21450***	-.44164***	-.46446***	-.31725***	-.37089***	-.07866*
NBHHMEMB	.10339***	.15564***	.14350***	.04696***	.04788***	0.00104
DIPL_B	.03769***	.07599***	.08882***	.04208*	.06129**	0.02215
DIPL_C	-.12479***	-0.00687	-.08199***	.11995***	0.00848	-.12201***
DIPL_D	-.34032***	-.12664***	-.28117***	.21735***	-0.02675	-.26711***
Log likelihood	-71803.72	-46535.53	-20851.35	-27955.16	-14375.14	-12347.22
McFadden Pseudo R2	0.072	0.133	0.140	0.040	0.051	0.012
N	179 341	167 905	157 153	44 182	33 430	21 994
Sample Selection Model	TREATED = 1					
LDISTANC	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant	1.53128***	1.64843***	1.37813***	2.01583***	1.65719***	3.74343**
AGE25	.10617***	.15163***	0.03865	.18885***	0.04589	-0.12151
AGE35	.04531**	.06315**	-0.01513	.06424**	-0.02353	-0.16361
AGE45	0.00281	0.0423	0.00421	0.00345	-0.00769	-0.01224
AGE55	-.06810***	-.06286**	-.11179*	-.13462***	-.11768**	0.02953
FEMALE	-.14076***	-.14744***	-.11376***	-.14783***	-.11339***	-.11351***
TENANT_S	-0.01797	0.0518	0.01735	.15138***	0.07232	0.35503
OWNER	.16416***	.14163***	.10526*	.12640***	0.06285	-0.16941
LESS1	.13923***	.25188***	.43573***	.23820***	.40346***	.33220*
MORE1	.24104***	.31063***	.68639***	.30893***	.63822***	.55739***
DIPL_B	.10386***	.08195***	.07330*	.06861**	0.06279	0.0347
DIPL_C	.19486***	.23548***	.24491***	.18471***	.24130***	.39013***
DIPL_D	.22460***	.32372***	.38299***	.22555***	.38450***	.71814***
LAMBDA	-0.00221	-.14380***	0.01382	-.51446***	-0.11828	-1.70591
Adjusted R-squared	0.041	0.049	0.051	0.050	0.051	0.052
N	27 809	16 373	5 621	16 373	5 621	5 621
Sample Selection Model	TREATED = 0					
LDISTANC	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant	1.19900***	1.21895***	1.21683***	1.46788***	1.50200***	0.42065
AGE25	.03040***	.03661***	.03360***	.12370***	.11135***	-0.00087
AGE35	-.05543***	-.05503***	-.05753***	.04776**	.04071**	-0.05476
AGE45	-.11771***	-.12553***	-.12415***	-0.00968	-0.00638	0.03787
AGE55	-.20552***	-.21860***	-.21012***	-.08916***	-.07392***	0.04768
FEMALE	-.14015***	-.14012***	-.14011***	-.14065***	-.14072***	-.14835***
TENANT_S	.14071***	.15315***	.15769***	0.01116	0.00749	0.25805
OWNER	.14058***	.13267***	.12493***	.14170***	.13455***	-0.01045
LESS1	.14147***	.15104***	.15596***	.12159***	.12897***	.26328**
MORE1	.28587***	.29488***	.30072***	.21216***	.22225***	.35798***
DIPL_B	.03021***	.02813***	.02516***	.09642***	.09645***	0.05928
DIPL_C	-0.00059	-0.01032	-0.00614	.17937***	.19114***	.36027***
DIPL_D	-.02389***	-.04170***	-.03240***	.19788***	.22289***	.58614***
LAMBDA	-.20363***	-.23368***	-.47412***	-.19152**	-.21334*	-2.2405
Adjusted R-squared	0.048	0.048	0.048	0.041	0.041	0.049
N	151 532	151 532	151 532	27 809	27 809	16 373

***, **, * : significance at 1%, 5%, 10% level.

Table D.1: Mode choice logit models (matching, treated group)

	Pair 1-2	Pair 1-3	Pair 1-4	Pair 2-3	Pair 2-4	Pair 3-4
ASC-walk	2.48944***	1.92841***	0.71059	1.97753***	.85641*	.96566*
Dist-walk	-2.14922***	-2.26641***	-2.17565***	-2.26581***	-2.17885***	-2.21790***
F-walk	.75136***	.84817***	1.00277***	.87294***	1.00189***	.96883***
AG25-walk	-0.15773	-.64897***	0.14854	-.66841***	0.22733	0.19437
AG35-walk	0.00743	-.24384*	0.3237	-.25236*	0.37842	0.32792
AG45-walk	0.12933	-0.17852	.56683**	-0.18233	.60789**	.61139**
AG55-walk	.48561***	0.02208	.91278***	-0.00473	.96032***	.91884***
HT-walk	-.43243***	-0.22614	-0.16667	-.25877*	-0.2323	-0.24036
HO-walk	-.45257***	-.57219***	-0.25692	-.58833***	-0.24687	-0.27017
DDB-walk	-0.11866	-0.04154	0.27215	-0.05821	.29183*	.30616*
DDC-walk	-0.09069	-0.0732	.41072**	-0.08563	.45176**	.44675**
DDD-walk	-.29025***	-.19886*	-.46045**	-.24460**	-.43267**	-.45015**
CL-walk	-1.35015***	-.67347***	-0.36035	-.68471***	-0.57161	-0.55355
CM-walk	-2.18389***	-1.31546***	-.89899**	-1.34143***	-1.11910***	-1.11127***
ASC-2whl	.35212*	0.4422	-1.96183*	.53099*	-1.71412	-1.50118
Dist-2whl	-.30153***	-.16791***	-.20028*	-.14904***	-.20826*	-.21021*
F-2whl	-1.32657***	-1.28430***	-1.69879***	-1.29847***	-1.67987***	-1.69706***
AG25-2whl	-.71733***	-1.01897***	-1.18789***	-1.02517***	-1.25099***	-1.15475***
AG35-2whl	-.52307***	-.75441***	-1.03215***	-.75063***	-1.03066***	-1.00712***
AG45-2whl	-.56085***	-.73012***	-1.06481***	-.69412***	-1.07191***	-1.04183***
AG55-2whl	-.50337***	-.65482***	-2.30529***	-.60949***	-2.31208***	-2.30188***
HT-2whl	-.32336***	-0.12204	-0.11673	-0.10011	-0.03739	-0.08968
HO-2whl	-0.04275	-.20089*	0.23962	-.20500*	0.31671	0.23996
DDB-2whl	0.00079	-0.22417	0.01238	-0.25748	0.00802	-0.02651
DDC-2whl	-.29510**	-.42636**	-0.45203	-.41916**	-0.44608	-0.49171
DDD-2whl	0.08977	-0.0698	-0.2876	-0.07242	-0.32236	-0.34373
C-2whl-2whl	-1.39653***	-1.38268***	0.64467	-1.50158***	0.33167	0.26158
CM-2whl	-2.43501***	-2.29598***	-0.20562	-2.43332***	-0.47774	-0.61378
ASC-pubtr	.31771***	-.68208***	-2.73047***	-.56767**	-2.31223***	-2.83317***
Dist-pubtr	.45556***	.62126***	.86040***	.63043***	.89302***	.90280***
F-pubtr	.69146***	.72017***	.37542***	.72757***	.37973***	.36021***
AG25-pubtr	-.79504***	-1.00178***	-1.14681***	-1.00131***	-1.16141***	-1.15241***
AG35-pubtr	-.87938***	-1.02584***	-.92652***	-1.01518***	-.93303***	-.93506***
AG45-pubtr	-.83358***	-1.01250***	-.83736***	-.95286***	-.84742***	-.87499***
AG55-pubtr	-.64041***	-.61240***	-.80472***	-.57782***	-.80773***	-.81695***
HT-pubtr	.22757***	.46515***	0.26799	.46052***	0.29337	0.38938
HO-pubtr	-.16337***	-.20163***	0.09691	-.19295**	0.11693	0.09951
DDB-pubtr	-.30691***	-.37916***	-.51486**	-.41088***	-.51932**	-.50657**
DDC-pubtr	0.02712	0.00125	-0.22671	-0.00802	-0.22115	-0.20154
DDD-pubtr	.17867***	0.13446	0.19933	0.13995	0.20093	0.21106
CL-pubtr	-1.87751***	-1.77334***	-1.03806**	-1.91576***	-1.57006***	-1.04031*
CM-pubtr	-2.98264***	-2.63799***	-1.59461***	-2.80945***	-2.11849***	-1.59936***
N obs.	27 804	16368	5620	15402	5601	5528
L-likelihood	-18162.183	-8959.5282	-2579.8742	-8438.4188	-2562.0589	-2528.3836
R2Adj	0.1947	0.2116	0.2415	0.2138	0.243	0.2439

***, **, * : significance at 1%, 5%, 10% level.

Table D.2: Covariates effects on mode shares (matching, treated group)

	Pair 1-2				Pair 1-3				Pair 1-4			
	WLK	2WH	CAR	PBT	WLK	2WH	CAR	PBT	WLK	2WH	CAR	PBT
TT	-5%pt	-3%pt	17%pt	-9%pt	-5%pt	-3%pt	19%pt	-11%pt	-4%pt	-4%pt	17%pt	-10%pt
Female	4%pt	-4%pt	-8%pt	8%pt	4%pt	-3%pt	-6%pt	6%pt	5%pt	-2%pt	-5%pt	2%pt
Age25	0%pt	-1%pt	9%pt	-8%pt	-3%pt	-2%pt	10%pt	-6%pt	1%pt	-1%pt	6%pt	-5%pt
Age35	1%pt	-1%pt	9%pt	-9%pt	-1%pt	-2%pt	9%pt	-7%pt	2%pt	-1%pt	5%pt	-5%pt
Age45	2%pt	-1%pt	8%pt	-8%pt	0%pt	-2%pt	9%pt	-7%pt	3%pt	-2%pt	3%pt	-5%pt
Age55	4%pt	-1%pt	4%pt	-6%pt	0%pt	-1%pt	5%pt	-4%pt	5%pt	-2%pt	1%pt	-4%pt
HSocTn	-2%pt	-1%pt	0%pt	3%pt	-1%pt	0%pt	-3%pt	4%pt				
Howner	-2%pt	0%pt	4%pt	-1%pt	-3%pt	0%pt	5%pt	-1%pt				
DipTech	0%pt	0%pt	3%pt	-3%pt	0%pt	0%pt	3%pt	-3%pt	1%pt	0%pt	1%pt	-3%pt
DipHigh	0%pt	-1%pt	1%pt	0%pt	0%pt	-1%pt	1%pt	0%pt	2%pt	-1%pt	0%pt	-1%pt
DipDeg	-2%pt	0%pt	-1%pt	2%pt					-2%pt	0%pt	1%pt	1%pt
CLess1	-4%pt	-2%pt	20%pt	-14%pt	-2%pt	-2%pt	14%pt	-10%pt	-2%pt	1%pt	6%pt	-5%pt
CMore1	-8%pt	-5%pt	52%pt	-39%pt	-5%pt	-7%pt	42%pt	-30%pt	-4%pt	0%pt	19%pt	-14%pt
	Pair 2-3				Pair 2-4				Pair 3-4			
	WLK	2WH	CAR	PBT	WLK	2WH	CAR	PBT	WLK	2WH	CAR	PBT
TT	0%pt	0%pt	4%pt	-4%pt	-1%pt	-1%pt	5%pt	-3%pt	0%pt	-1%pt	2%pt	-1%pt
Female	5%pt	-3%pt	-7%pt	6%pt	5%pt	-2%pt	-5%pt	2%pt	5%pt	-2%pt	-5%pt	2%pt
Age25	-3%pt	-2%pt	11%pt	-6%pt	1%pt	-1%pt	6%pt	-5%pt	1%pt	-1%pt	6%pt	-5%pt
Age35	-1%pt	-2%pt	9%pt	-7%pt	2%pt	-1%pt	4%pt	-5%pt	2%pt	-1%pt	5%pt	-5%pt
Age45	-1%pt	-1%pt	8%pt	-6%pt	3%pt	-2%pt	3%pt	-5%pt	3%pt	-2%pt	3%pt	-5%pt
Age55	0%pt	-1%pt	5%pt	-4%pt	5%pt	-2%pt	1%pt	-4%pt	5%pt	-2%pt	1%pt	-4%pt
HSocTn	-1%pt	0%pt	-2%pt	4%pt								
Howner	-3%pt	0%pt	5%pt	-1%pt								
DipTech	0%pt	-1%pt	3%pt	-3%pt	2%pt	0%pt	1%pt	-3%pt	2%pt	0%pt	1%pt	-3%pt
DipHigh	0%pt	-1%pt	1%pt	0%pt	2%pt	-1%pt	0%pt	-1%pt	2%pt	-1%pt	0%pt	-1%pt
DipDeg	-1%pt	0%pt	0%pt	1%pt	-2%pt	0%pt	1%pt	1%pt	-2%pt	-1%pt	1%pt	1%pt
CLess1	-2%pt	-3%pt	15%pt	-10%pt	-2%pt	1%pt	9%pt	-8%pt				
CMore1	-4%pt	-8%pt	44%pt	-32%pt	-5%pt	0%pt	27%pt	-21%pt	-5%pt	-1%pt	20%pt	-14%pt