International Journal of Sustainable Transportation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ujst20

Impacts of transit-oriented development on car use over a 10-year period in Porto, Portugal: From macro- to micro-analysis

Anna Ibraeva, Gonçalo Homem de Almeida Correia, Cecília Silva & António Pais Antunes

To cite this article: Anna Ibraeva, Gonçalo Homem de Almeida Correia, Cecília Silva & António Pais Antunes (2023) Impacts of transit-oriented development on car use over a 10-year period in Porto, Portugal: From macro- to micro-analysis, International Journal of Sustainable Transportation, 17:11, 1195-1206, DOI: 10.1080/15568318.2022.2160284

To link to this article: https://doi.org/10.1080/15568318.2022.2160284

Published online: 28 Dec 2022.

Submit your article to this journal

Article views: 388

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ujst20

INTERNATIONAL JOURNAL OF SUSTAINABLE TRANSPORTATION
2023, VOL. 17, NO. 11, 1195–1206
https://doi.org/10.1080/15568318.2022.2160284

Taylor & Francis
Taylor & Francis Group

Check for updates

Impacts of transit-oriented development on car use over a 10-year period in Porto, Portugal: From macro- to micro-analysis

Anna Ibraeva[a], Gonçalo Homem de Almeida Correia[b], Cecília Silva[c], and António Pais Antunes[a]

[a]Department of Civil Engineering, University of Coimbra, CITTA, Coimbra, Portugal; [b]Faculty of Civil Engineering and Geosciences, Department of Transport & Planning, Delft University of Technology, Delft, The Netherlands; [c]Faculty of Engineering, University of Porto, CITTA, Porto, Portugal

ABSTRACT

Transit-oriented development (TOD), an urban planning concept that aims to promote sustainable transport modes, has been actively studied in recent years, especially in relation to car trips reduction. However, longitudinal studies on the matter are still rare. In this paper, we analyze the effects of the implementation of a new metro system after its first 10 years of operation focusing on how changes in the number of car trips were influenced by station type – TOD, transit-adjacent development (TAD) and park-and-ride (P&R). Specifically, we perform a before/after analysis of the impact of metro implementation in the Porto area (Portugal) both at a macro scale (civil parish) and at a micro scale (census tract) to analyze the overall effect of metro and more detailed effects only detectable at the micro scale. Census-tract data enables a comprehensive analysis of spatial spillover impacts from different stations, comparing the extent (i.e., the distance range within which the effect of station proximity is noticeable) and the magnitude (the reduction in the number of car trips) of the spillover in each case. Both direct and indirect metro impacts are already visible at the macro scale, yet at the micro scale the magnitude of the spillover effects varies depending on station type. The effects of TOD stations on the reduction of car trips are the strongest across the different station types and are felt up to a 2 km distance from a station, while TAD and P&R effects are weaker and do not reach beyond 1.2 km.

ARTICLE HISTORY
Received 21 March 2022
Revised 11 October 2022
Accepted 14 December 2022

KEYWORDS
Car trip reduction; longitudinal approach; metro systems; spatial regression models; spillover effects; transit-oriented development

1. Introduction

Public transport (transit) systems are expected to play a major role in the progression toward a more sustainable urban mobility, as is explicitly recognized in the UN's Sustainable Development Goals and in the EU's Urban Mobility Framework. In this context, station areas are of critical importance, since their design and environment, together with the level of transport service they provide, can be decisive to increase transit ridership and reduce car use. Transit-oriented development (TOD) is a manifold concept for station area organization that intends to address current urban challenges like heavy traffic and associated pollution by providing dense, mixed-use, and lively neighborhoods in proximity to transit stations. A neighborhood centered around a station with a dense street network that facilitates access to transit and to a variety of local services is expected to discourage the use of cars. However, due to the complexity of modern cities, many factors such as socio-demographics, residential location choice and built environment characteristics at the destination may constrain the success of TOD. In the last ten years, accompanying the implementation of TOD projects in many countries around the globe, numerous attempts have been made to analyze and evaluate TOD's contribution to the reduction of car use and the switch to sustainable transport modes (Ibraeva et al., 2020).

Notwithstanding the major progress made in this research area, several issues remain understudied. As discussed in detail in the next section of this paper, studies that analyze the mode choice impacts of TOD projects using comprehensive region- or city-wide information, notably at the census-block or census-tract level, are seemingly unavailable. Indeed, most works have been developed at the individual, household or neighborhood scale, being typically limited in terms of sample size (being based on survey data) and/or geographical scope (being based on specific preselected station areas). Also, few works have focused on the mode choice impacts of different types of station areas (TOD or non-TOD) and on how these impacts spill over to adjacent areas. Finally, a relatively small number of studies rely on a longitudinal design, despite the advantages they offer compared to a cross-sectional approach.

In this paper, we address all the understudied issues identified above in a study of the changes in the number of car trips for work/study over a 10-year period in the Porto region (Portugal) following the implementation between 2002 and 2011 of Metro do Porto—a new metro system with TOD features. For this purpose, we resort to first difference estimator spatial regression models, thus accounting

CONTACT Anna Ibraeva ibraevanna@student.uc.pt Department of Civil Engineering, University of Coimbra, CITTA, Polo II, Coimbra, 3030-788, Portugal.
© 2022 Taylor & Francis Group, LLC

for the time and space dimensions of change. Our primary data source is the official census data for the years 2001 and 2011 (i.e., before and immediately after metro implementation). Since no other major infrastructure investment occurred during this period in the Porto region, our study design can be considered "a natural experiment".

Two levels of analysis are considered in our study: macro and micro. At the macro level, we rely on data from the 120 (civil) parishes of the Porto region. Based on this information, it is already possible to get relevant insights into the impact of Metro do Porto on the number of car trips, and compare the conclusions with the ones obtained in other studies. However, it is not possible to properly analyze the performance and spillover effects of the different station area types identified in the Metro do Porto network depending on the surrounding environment: TOD, transit-adjacent development (TAD) and park and ride (P&R). This is only viable based on micro-level data, e.g., data from the census tracts of the region (up to 77 in the same parish). It should be noted that the boundaries of parishes, the smallest political-administrative units in Portugal, are very stable over time. Instead, census tracts change considerably across censuses, and sometimes in complex manners (in Portugal as well as in many other countries including the United States; see Logan et al., 2016). Hence, while a longitudinal parish-level study can be performed without substantial data preparation efforts, census-tract-level studies are extremely demanding in this regard. This certainly contributes to explain why longitudinal studies at this level of detail are so scarce.

The remainder of the paper is structured as follows. In the next section, we provide an overview of relevant related literature. The case study (Porto region) addressed in the paper is presented afterwards. Then, we describe the modeling approach applied in this study, justify the use of first-difference estimator regression models, and explain the different spatial specifications considered in the macro- and micro-level analyses. The following section specifies all the variables used in these analyses. Next, we present and discuss the results of the regression models (parish and census-tract levels) and assess models' validity. Finally, the concluding section summarizes the paper and offers some final remarks.

2. Related literature

The city and metropolitan growth trends observed in many parts of the world are motivating an increasing interest in the optimal organization of urban settlements as a potential solution to widespread problems of congestion and pollution (Newman & Kenworthy, 1988, 1996). In this setting, the concept of Transit-Oriented Development, introduced by Calthorpe (1993), has become particularly important, illustrating a type of settlement that is at the same time accessible and self-sufficient. The early works of Cervero (1995) and Cervero and Gorham (1995), exploring the interaction between built environment characteristics and transport mode choice, evidenced overall lower levels of car use in TOD-like neighborhoods than in scattered monofunctional suburbs. Following their publication, a vast body of research on that interaction and, more specifically, on TOD has been developed, whose main findings are summarized and discussed in several literature reviews, including some very recent ones (Ewing & Cervero, 2001; Ewing & Cervero, 2010; Naess, 2015; Ibraeva et al., 2020; Van Wee & Cao, 2022). In this section, we focus on the works deemed most relevant in the context of the present study.

A significant part of the advancements made in TOD impact research resulted from incorporating in the analysis socio-demographic characteristics (Cervero & Radisch, 1996; Cervero, Kockelman, 1997; Handy et al., 2005; Nasri & Zhang, 2014; Park et al., 2018; Nasri & Zhang, 2019), attitudinal factors (Handy et al., 2005; Cervero, 2007; Cao et al., 2009; Cao & Cao, 2014; Zhang, 2014; Cao, 2015), and destination characteristics (Cervero & Radisch, 1996; Cervero, 2007; Pan et al., 2017; Park et al., 2018; Nasri & Zhang, 2019; Choi et al., 2012; Ibraeva et al., 2022). This was accomplished using a wide variety of methodological approaches, including longitudinal research design (Handy et al., 2005; Cao et al., 2007; Cervero & Day, 2008; Brown & Werner, 2008; Olaru & Curtis, 2015; Van de Coevering et al., 2016; Griffiths & Curtis, 2017; Sener et al., 2020; Ibraeva et al., 2021).

Longitudinal TOD studies have received a particular attention in recent years because of the advantages they offer compared to cross-sectional studies. In fact, incorporating various time periods in the analysis allows revealing changes that occur over time, which in the TOD context is extremely beneficial given that modal shifts or changes in the built environment tend to occur slowly. Additionally, compared to cross-sectional studies, which reveal associations, longitudinal studies, under certain conditions (like precedence), provide stronger evidence of causality (Dons et al., 2018; Handy et al., 2008; Van de Coevering et al., 2016). Moreover, through longitudinal studies, it is possible to separate the intervention effects from the preexisting situation if the latter is controlled for. Finally, if a longitudinal dataset contains information about preexisting personal preferences in terms of mode choice, it may also allow controlling for potential self-selection (i.e., when a person's mode choice is defined by long-established favorable attitudes toward a specific transport mode and does not depend on the surrounding built environment or nearby transport facilities).

The relatively few existing longitudinal studies dedicated to travel behavior and TOD generally analyze the transport mode changes occurring after a residential relocation (Handy et al., 2005; Cervero & Day, 2008; Cao et al., 2007; Van de Coevering et al., 2016; Cao & Ermagun, 2017; Huang et al., 2019) or after an infrastructure/development improvement like a new station opening or the requalification of a station area (Griffiths & Curtis, 2017; Olaru & Curtis, 2015; Brown & Werner, 2008; Sener et al., 2020).

Such studies, as well as the numerous cross-sectional studies performed to date, were developed at the individual or the household scale (e.g., Bardaka & Hersey, 2019; Chatman, 2013; Cervero & Radisch, 1996; Handy et al., 2005; Van Acker et al., 2007; Cervero & Day, 2008; Olaru &

Curtis, 2015; Cao et al., 2007; Brown & Werner, 2008; Van de Coevering et al., 2016), or at a neighborhood scale (e.g., Cervero & Gorham, 1995; Cervero & Kockelman, 1997; Cervero, 2007; Griffiths & Curtis, 2017; Nasri & Zhang, 2014). Therefore, the results reported are rather local and concern particular neighborhoods, while the idea of TOD as an urban planning concept is not limited to the design of neighborhoods but also concerns the spatial arrangement of an entire city. In contrast, region-wide studies on the matter (such as Loo et al., 2010; Sung & Oh, 2011; Choi et al., 2012; Pan et al., 2017; Park et al., 2018; Nasri & Zhang, 2019; and Ibraeva et al., 2021, 2022) are relatively rare, despite expanding the area of analysis may provide stronger evidence and, owing to greater variability, disclose phenomena that may not manifest themselves in a neighborhood-specific study.

Existing evidence from region-wide analyses generally supports the findings revealed by site-specific studies. For instance, greater residential density was found to be associated with greater transit usage in New York (Loo et al., 2010), Hong Kong (Loo et al., 2010), Seoul (Sung & Oh, 2011; Choi et al., 2012), Shanghai (Pan et al., 2017) and Washington, D.C. (Nasri & Zhang, 2019). Besides, transit use was positively related to mixed land uses in New York and Hong Kong (Loo et al., 2010), although in Seoul this was only true when the residential use was also present in the land use mix (Sung & Oh, 2011). Considering street organization, the density of four-way intersections in station areas in Seoul was linked to greater levels of transit patronage (Sung & Oh, 2011). The effect on car use appears to be the opposite, as results from eight different metropolitan areas in the USA revealed that the vehicle-miles traveled decrease as intersection density increases (Park et al., 2018). In the same direction, in Washington the probability of car use was found to decrease with decreases in average block size (Nasri & Zhang, 2019). On the other hand, it was observed that transit ridership increases with greater number of bus stops/lines and higher transit frequencies (Sung & Oh, 2011), trip length and station age (Pan et al., 2017). Interestingly, Park et al. (2018) noted that car trips decrease in proximity to light-rail transit stations as opposed to commuter heavy-rail stations.

It is important to add that the areas of analysis in these city-wide studies were limited to 500- to 800-meter station buffers, and station effects beyond those distances might have been overlooked. Moreover, different station areas were in general analyzed all together, i.e., without differentiating between various station types. While this approach may be justified in cases of extremely dense and rather homogeneous Asian megacities, in other contexts it may not be appropriate, as "about 60% of rail transit stations in the U.S. do not have the required features to be called TOD, rather they are more like TADs" (Nasri & Zhang, 2019). By contrast to TOD, TAD stations are located in the proximity but peripherally to a settlement, being poorly connected with it and lacking service, commercial and/or recreational facilities in the immediate station area. This motivated Kamruzzaman et al. (2015) to distinguish between TOD and TAD ("the evil twin of TOD") station areas in an assessment of commuting mode choice behavior in Brisbane, evidencing that TOD residents used transit and active modes significantly more than those living in TADs. Similarly, Nasri and Zhang (2019) differentiated between TOD and TAD stations at both trip ends and reported that living and/or working in TOD areas increases the probability of using non-motorized modes.

Besides TOD and TAD, another very widespread station type in many cities is Park & Ride (P&R). Introduced in remote, low-density areas, P&R stations are expected to relieve congestion in central areas by encouraging drivers from suburban neighborhoods to leave their car at a station-adjacent parking lot and to continue their journey by transit. This should allow to increase accessibility from the suburbs to the city center by easing congestion levels, re-purpose parking lots in the center due to decreased demand, and reduce the expenses of users on fuel and vehicle maintenance. However, some effects of P&R appear to be controversial. For example, in Oxford, UK, 56% of those who started to use a P&R facility to commute to the city center used car before, while the rest used to rely on buses, and in York these figures were 70% and 30%, respectively (Parkhurst, 1995), indicating that a considerable share of P&R users came from sustainable transport modes. Besides, Walton and Sunseri (2010), based on studies for Auckland and Wellington, New Zealand, point out that parking availability in the immediate station area encourages driving to the station among those who would otherwise walk or cycle. A similar issue was observed in Brisbane, where a five-year before/after analysis evidenced that a 17% increase in the P&R capacity coincided with an approximately 11% increase in direct trips to the destination by car, and that an increase in P&R trips was registered in neighborhoods located within walking distance from a station (Kimpton et al., 2020).

The present study substantially extends the work described in Ibraeva et al. (2021), also focusing on the introduction of metro in the Porto region and relying mostly on census data. In that earlier paper, a difference-in-differences model was applied to estimate at the macro level (parishes) what could have happened in terms of car use if metro had not been implemented. In the current paper, we estimate through a first difference estimator model how changes in socio-demographic, built environment and metro supply are related to changes in car use. Moreover, these changes are addressed not only at the macro level but also, and more importantly, at the micro level (census tracts), thus making it possible to properly analyze the performance and spillover effects of station areas of different types. This makes the two papers considerably different.

3. Porto region and Metro do Porto

The region covered by the Metro do Porto system, used as case study in this paper, is comprised by seven municipalities: Gondomar, Maia, Matosinhos, Porto, Póvoa de Varzim, Vila do Conde and Vila Nova de Gaia (Figure 1, left). With a total area of about 718 km^2 and a population of

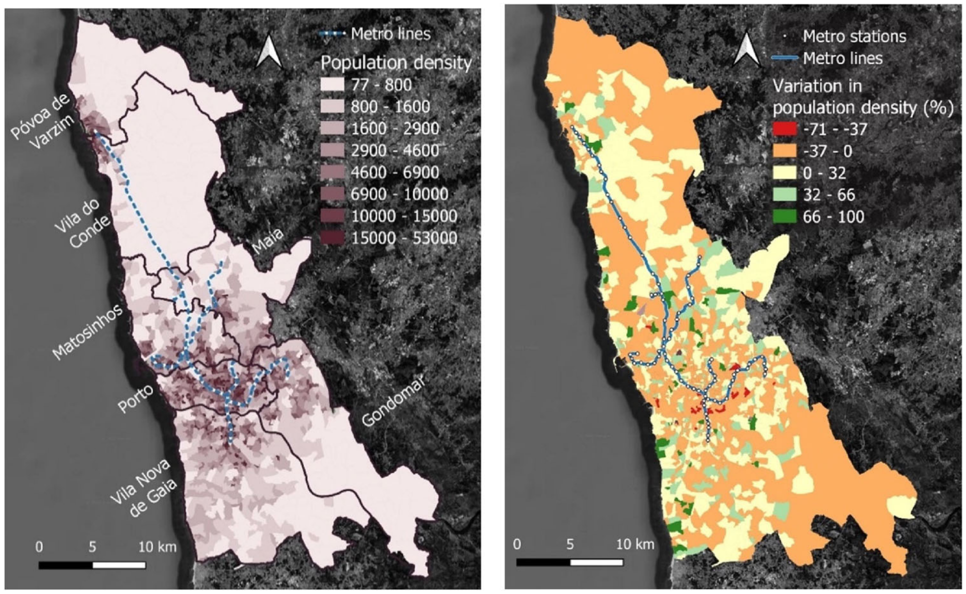

Figure 1. Municipalities of the Porto region, Metro do Porto network and population densities in 2001 (left) and its variation between 2001 and 2011 (right).

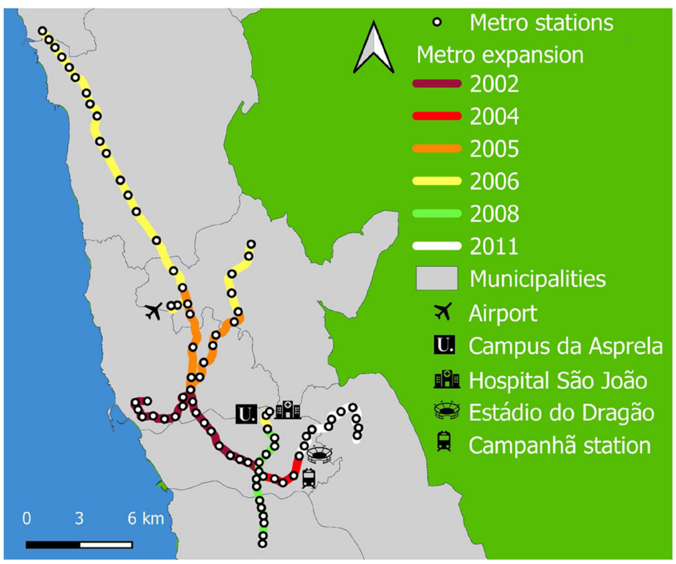

Figure 2. Metro do Porto layout and line opening dates.

roughly 1.2 million, it includes territories with rather distinct characteristics. Porto, the second-largest city in Portugal, hosts main governmental institutions, universities, business headquarters, and numerous traditional commerce establishments. In the adjacent municipalities (Gondomar, Maia, Matosinhos and Vila Nova de Gaia) large industrial sites and big shopping malls are located. The northern municipalities of Póvoa de Varzim and Vila do Conde are mostly rural with agriculture occupying a large part of their territories, yet their coastal areas are quite popular holiday destinations especially in summer months. Although Porto is the regional capital, municipality main towns are far from being mere residential suburbs, also providing employment opportunities in dense and mixed-use built environments. This probably contributes to the long-standing attractiveness of Porto-adjacent municipalities for residential location as Porto's central areas and rural areas of the region were losing population, including in the period of analysis (Figure 1, right).

The decision to launch Metro do Porto was taken in 1994 to tackle the major congestion problems that the region and especially the city of Porto had been suffering since the early 1980s, when car-ownership rates in Portugal started to increase dramatically (from 125 to 335 cars per 1000 residents in 1980–2000, Branco & Ramos, 2003). The layout of the metro network was established in 1996 (Figure 2). The first line opened in 2002, connecting Matosinhos to Porto. This line was extended further to the east to serve Porto's main railway station (Campanhã) and sport facilities (notably, Estádio do Dragão) in 2004. The metro reached Maia in 2005, and, in 2006, Porto airport and the municipalities of Póvoa de Varzim and Vila do Conde. By then, 70% of the planned network was already operating. In 2008, a new line connecting Vila Nova de Gaia to a large university campus (Asprela) and Porto's main hospital (São João) was opened. At last, in 2011, the metro reached Gondomar.

4. Modeling approach

In this section, we present the regression modeling approach adopted in this study to explain the evolution of the number of car trips in the Porto region in the period 2001–2011. First, the general model formulation applied to both the macro- and micro-scale analyses is provided. Second, a more detailed description of the parish and census-tract models is given emphasizing the different spatial specifications implemented in each case.

4.1. General model

An essentially similar approach was used to handle both scales of analysis. The Hausman test, applied to the initial OLS models at the parish and census-tract level, supported the use of fixed-effects models (FEM) in both cases. Detailed

Table 1. Model specification process.

Parish model	Census-tract model
Hausman test (fixed vs random effects)	
chisq = 30.693, df = 8, *p*-value = 0.0002	chisq = 903.95, df = 13, *p*-value < 2.2e-16
Moran's *I* test (presence of spatial autocorrelation)	
Moran's *I* = 0.56, *p*-value < 2.2e-16	Moran's *I* = 0.12, *p*-value = 5.622e-16
Lagrange Multiplier diagnostics	
RLMlag = 19.88, df = 1, *p*-value = 8.248e-06	RLMlag = 0.0093734, df = 1, *p*-value = 0.9229
RLMerr = 0.73192, df = 1, *p*-value = 0.3923	RLMerr = 4.3421, df = 1, *p*-value = 0.03718
Ranges of Variance Inflation Factor (measure of multicollinearity)	
1.0584 – 1.6426	1.0225 – 1.5492
Studentized Breusch-Pagan test (presence of heteroskedasticity)	
BP = 7.7047, df = 8, *p*-value = 0.4628	BP = 42.116, df = 14, *p*-value = 0.0002

information on the results of the test as well as on the whole model specification process is presented in Table 1. It should be noted that, for an analysis with only two time periods, this model is identical to the first difference estimator and takes the following form:

$$\Delta Y_i = \beta_1 \Delta X_{i1} + \beta_2 \Delta X_{i2} + \ldots + \beta_k \Delta X_{ik} + u_i,$$

where ΔY_i is the change in the dependent variable (in our case, the change in the number of car trips per 1000 residents) over the period analyzed (in our case, 2001–2011), ΔX_{i1}, $X_{i2} \ldots X_{ik}$ are the changes in the explanatory variables, and u_i is the idiosyncratic error term. With this model formulation, it is possible to estimate whether/how changes in these variables affect changes in the dependent variable while, at the same time, controlling for omitted unobserved effects that vary between the units but are constant over time.

As both models (especially the census-tract model) accommodate a rather long list of explanatory variables (as shown in Section 5), variance inflation factors were determined to check for potential multicollinearity, but apparently it is not present. Instead, the presence of heteroscedasticity was identified by studentized Breusch-Pagan tests in the census-tract model, therefore robust standard error estimates were calculated for this model when analyzing the significance of explanatory variables.

The application of Moran's I tests (assuming queen contiguity in the spatial weight matrix) detected a significant presence of spatial autocorrelation in parish and census-tract models, requiring the use of a spatial regression approach. While both models are based on the same equation, spatial structures at the macro and micro scale are naturally different (Figure 3). Indeed, parishes have the average area of 6 km^2 (ranging between 0.25 km^2 and 19.43 km^2) and the average population of 9,680 (in 2011), whereas census tracts have the average area of 0.46 km^2 (from 0.008 km^2 to 8 km^2) and the average population of 743. Due to these differences, there are distinct patterns of spatial clustering in each case and they need to be accommodated using different spatial specifications for the parish and census-tract model which are explained next.

4.2. Parish model

Visible spatial clustering and Moran's I statistic (0.56) indicate a quite strong spatial dependence between the parishes. Robust Lagrange Multiplier (RLM) tests, which allow detecting spatial error autocorrelation in the presence of a spatially lagged dependent variable and vice versa (Anselin et al., 1996), were performed to identify the source of spatial dependence. At the macro level, RLM tests pointed to a spatial lag model, and more specifically to a spatial autoregressive (SAR) model, as follows:

$$\Delta Y_i = \rho (I_T \otimes W_N) \Delta y + \beta_1 \Delta X_{i1} + \beta_2 \Delta X_{i2} + \ldots + \beta_k \Delta X_{ik} + u_i,$$

where W_N is the $N \times N$ spatial weights matrix and ρ is the corresponding spatial parameter. The use of a spatially lagged dependent variable is justified when neighboring observations show similar values, which in the context of our study can reflect the natural clustering of urban, suburban or rural parishes. Besides, this specification is appropriate when one expects a value observed in one unit to be influenced by the values of its neighboring units. However, this is hardly the case in our sample since it is difficult to imagine that the increase of the number of car trips in a parish would provoke increases in adjacent parishes. Alternatively, the use of a SAR model is justified in cases when endogenous interactions are possible so that "changes in one region/agent/entity set in motion a sequence of adjustments in (potentially) all regions in the sample such that a new long-run steady state equilibrium arises" (LeSage, 2014). In our case, Metro do Porto could potentially produce an effect in the directly-served parishes that propagated to their first-order neighbors, then to second-order neighbors, and so forth, eventually reaching the whole study area, and, in turn, this global effect might have fed back into the directly-served parishes.

4.3. Census-tract model

Regarding the census-tract model, a specification with spatially correlated errors (SEM) was found to be the most appropriate (as indicated by RLM tests). According to Croissant and Millo (2019), "the spatial error is (…) appropriate when one expects the innovation relative to one observation to influence the outcomes of neighboring ones, as would be the case for an economic shock of some kind to a given region (fully) influencing the relevant dependent variable in that region and also propagating—with distance-decaying intensity—toward nearby ones". Since the implementation of a large-scale infrastructure project (like Metro do Porto) can effectively be considered a "shock", this model

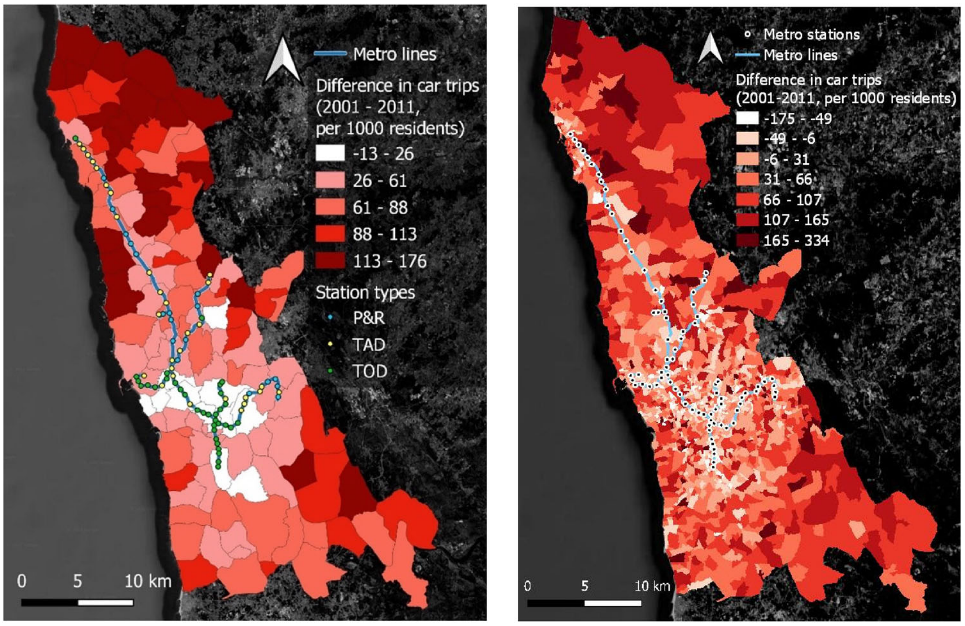

Figure 3. Changes in car trips at the macro (parish) and micro (census tract) level between 2001 and 2011.

specification should properly account for spatial dependence in the error term possibly related to the metro introduction. The SEM model for the census tracts takes the following form:

$$\Delta Y_i = \Delta X_{i1} + \beta_2 \Delta X_{i2} + \ \dots\ + \beta_k \Delta X_{ik} + \ u_i, \ \ u_i = \lambda W u + \varepsilon,$$

where W is the $N \times N$ spatial weights matrix, λ is the corresponding spatial parameter and ε is the error term with $\varepsilon_i \overset{iid}{\sim} N(0,\ \sigma^2)$.

5. Model variables

The variables included in the models were selected considering the land-use/transport model literature and taken into account the data available in the Census 2001 and 2011 provided by Statistics Portugal (INE)—that is, the most recent census data (the final results of Census 2021 will only be known by the end of 2022) . The censuses contain information related to socio-demographic characteristics of the residents, commuting mode choices, travel times and built environment characteristics. Information for both years was aggregated into the same units of analysis forming two panel datasets, one with 120 parishes and the other one with 1,561 census tracts. The distribution of metro trips in 2011 was then mapped to establish a cutoff distance after which the shares of metro ridership become imperceptible (on average, 2 km from a metro station). Observations located beyond the metro's influence buffer were considered as the reference level (control group) in the regression models. Table 2 provides a summary and a brief explanation of all model variables.

For both models considered in this paper, the dependent variable is the change in the number of car trips for work or study per 1000 residents occurred between 2001 and 2011.

Explanatory variables that control for changes in the socio-demographic characteristics are formulated in the same way in both models. These include variables reflecting changes in the population composition by age, level of education (the number of residents with completed secondary studies or higher, serving as a proxy for income level), and unemployment rate.

Variables representing changes in the average trip time reported by the residents for their main work or study trips and in the average metro frequency serving census tracts were also considered in our analyses given their potential influence on transport mode choice.

Regarding the built environment characteristics, the variables considered in the models account for changes in land-use mix and dwelling density. However, the data reflecting changes in land-use mix are not very precise: the land-use mix variable is the share of mostly residential/mostly non-residential buildings as opposed to exclusively residential buildings. Therefore, these data do not provide detailed information about the evolution of the uses present in a building. Nevertheless, they can still differentiate between relatively mixed areas and monofunctional residential-only neighborhoods. To avoid multicollinearity, changes in the residential and building density were replaced by changes in the dwelling density (highly correlated with both residential and building density). Besides, as census-tract areas are rather small, residents likely work/study outside the census tracts of their residence, having to travel through neighboring census tracts to get to their work/study places, so characteristics of these adjacent tracts might also influence the mode choice. For this reason, the spatial lag of the difference in the dwelling density was introduced in the census-tract model.

As can be seen in Table 2, metro-related explanatory variables vary depending on the scale of the model. For the macro-level model, a parish was considered metro-served if a 400-meter buffer from a metro station at least partially overlapped the area of a parish. For the micro-level model, census tracts whose centroids are located within a 400-m buffer and a series

Table 2. Description of model variables.

Variable type	Variable designation	Variable description
Dependent variable	car_trips	difference in the number of car trips per 1000 residents between 2001 and 2011
Explanatory variables common for the parish and census-tract models	dwell_dens	difference in dwelling density between 2001 and 2011
	lu_mix	difference in the share of mostly residential/mostly nonresidential buildings between 2001 and 2011
	plus_65	difference in the number of residents aged 65 or more per 1000 residents between 2001 and 2011
	sec_educ	difference in the number of residents having completed secondary education or more per 1000 residents between 2001 and 2011
	triptime	difference in the average trip time between 2001 and 2011
	unemp	difference in the number of unemployed residents per 1000 residents between 2001 and 2011
	under_13	difference in the number of residents aged 13 or less per 1000 residents between 2001 and 2011
Explanatory variables specific to the parish model	metro	binary variable, equals 1 if a parish is served by metro in 2011 and zero otherwise.
Explanatory variables specific to the census-tract model	dwell_dens_lag	spatial lag of the difference in the dwelling density variable
	PR400/ PR800/ PR1200/PR2000	binary variable, equals 1 if a census-tract centroid is located within a 400-meter buffer/within a range of 400–800-meters/within a range of 800–1200-meters/within a range of 1200–2000-meters from P&R station, and this is the station closest to the centroid.
	TAD400/TAD800/ TAD1200/ TAD2000	binary variable, equals 1 if a census-tract centroid is located within a 400-meter buffer/within a range of 400–800-meters/within a range of 800–1200-meters/within a range of 1200–2000-meters from a TAD station, and this is the station closest to the centroid.
	TOD400/TOD800/ TOD1200/ TOD2000	binary variable, equals 1 if a census-tract centroid is located within a 400-meter buffer/within a range of 400–800-meters/within a range of 800–1200-meters/within a range of 1200–2000-meters from a TOD station, and this is the station closest to the centroid.
	trainhour	average workday number of trains per hour passing on the closest metro station
	up_to400/ up_to800/ up_to1200/ up_to1600/ up_to2000	binary variable, equals 1 if a census-tract centroid is located within a 400-meter buffer/within a range of 400–800-meters/within a range of 800–1200-meters/within a range of 1200–1600-meters/within a range of 1600–2000-meters from the closest metro station.

of consecutive distance ranges (400 m–800 m, 800 m–1200 m, 1200 m–1600 m, 1600 m–2000 m) from the metro stations were identified. The 2-km limit was selected to cover the first-order neighbors of the metro-served parishes as they reported metro trips in 2011 so their inclusion allows evaluating the spillover effect.

To evaluate the effect and spillover magnitude for different station types, a qualitative analysis was performed to identify TOD, TAD and P&R stations. The initial evaluation was based on the visual analysis of station areas using satellite imagery, which allowed to assess street configuration in the area, access conditions and overall stations' insertion in the surrounding environment. More specifically, the aim was to see whether there was a dense street network in each station area and whether the network's configuration facilitated the reach of a station. Additionally, we have evaluated the access conditions considering the existence of sidewalks, lighting and bus stops/parking lots in the area. Finally, using Census and Open Street Map data, station areas were assessed based on the development type (high or low-rise) and the presence of commerce/services. After determining the different station types, network distances between metro stations and census-tract centroids were calculated, to find the closest metro station for each census tract. Census tracts were then grouped into categories reflecting their distance to the respective closest station and its type (e.g., TOD400 is a category meaning that the centroid of a census tract is located within 400 meters from a TOD station).

Considering that the number of operation years of a station can also influence mode choice, station age was initially included in the models, first as a continuous variable, then as a set of categorical values for stations that opened at approximately the same time. However, these were discarded from the final models because they were not statistically significant. A plausible reason for this to happen is because the effect of station age might be captured by the built environment variables, since the older stations are generally located in the most central and dense urban areas.

6. Model results

This section reports the results obtained by the parish and the census-tract models. Concerning the latter we consider two model versions: version I evaluates the overall effect of metro station proximity; and version II additionally evaluates station area type effects. All models were estimated using the "spatialreg" packages in R (Bivand & Piras, 2015). It should be noted that, due to the presence of heteroskedasticity in the census-tract data, robust standard error estimates are provided for the census-tract model (both versions).

6.1. Parish model

The estimation results for the parish model are presented in Table 3. The Nagelkerke pseudo R^2 statistic, analogous in interpretation to the coefficient of determination in OLS, indicates a rather good overall model performance (0.66) for a study related to human behavior and transport mode choices. The autoregressive term (ρ) is statistically significant and positive, meaning that the number of car trips in a spatial unit (parish) increased as the number of car trips in its neighboring units also increased and vice versa.

The impact of metro in the directly-served parishes, as well as metro spillover effects for the first-order neighbors of the directly-served parishes, are visible already at the macro scale. For the directly-served parishes, metro implementation was associated with a decrease of 24 car trips per 1000 residents (over the 10-year inter-census period).

As expected, growth of the unemployment rate and of the number of residents aged 65 or more was inversely related to changes in the number of car trips, while an increase in the number of residents with complete secondary education or higher (a proxy for income level) was associated with an increase in the number of car trips. Considering built environment controls, only the increase in the dwelling density was statistically significant in reducing the number of car trips at a 90% confidence level.

The results displayed in Table 3 were obtained through the model that controls for spatial autocorrelation of the dependent variable, introducing a spatial lag as a covariate. By doing so, it is assumed that the value of the dependent variable in a given spatial unit depends on the values in the neighboring units, but also the value in that unit influences the values in the neighboring units. Faced with this feedback effect, some researchers (LeSage & Pace, 2009; LeSage, 2014; Bivand & Piras, 2015) recommend estimating the direct, indirect and total effects taking into account the value of the spatial autoregressive parameter. The impact measures after correction for the spatial autocorrelation effects are reported in Table 4.

The coefficient for the average direct metro effect remained almost unchanged and similar to the coefficient for the average indirect effect. The average indirect effect coefficient reflects the global spillover from metro implementation for the whole area of analysis: metro introduction in one parish affects the number of car trips in it, which potentially affects as well the number of car trips in the first-order neighbors of the directly-served parishes, then from the first-order neighbors the spillover moves on to subsequent neighbors, eventually covering the whole study area and feeding back to the directly-served parishes. As stated before, introducing a metro station in a parish led to an average decrease of 26 car trips per 1000 residents in that parish, but it also influenced the non-directly served parishes, where the average decrease was almost identical. However, it should be highlighted that this estimation of the global spillover effect (i.e., the effect that affects the whole area, going beyond the neighborhood set of the directly-served units) relies essentially on the spatial autoregressive parameter and the spatial weights matrix (row standardized). The resulting drawback of this model is that, for different variables, the ratio between the direct and the indirect effect is the same (Elhorst, 2014). Acknowledging that the magnitude of the indirect effect may vary depending on the distance to metro (i.e., local spillover effects can be revealed at a micro scale), in the following section (6.2) this issue is addressed in more detail.

Table 3. Estimation results for the parish model.

Variable	Coefficient	Std. error	z-statistic	*p*-value
intercept	30.9507	12.423	2.4914	0.0128*
lu_mix	0.002	0.0343	0.0572	0.9544
under_13	−0.1193	0.1697	−0.7028	0.4822
plus_65	−0.5496	0.1703	−3.2272	0.0013*
sec_educ	0.4193	0.0918	4.5667	<0.0001*
unemp	−0.0558	0.0213	−2.6123	0.0089*
trip_time	0.5014	0.7607	0.6592	0.5097
dwell_dens	−0.0031	0.0017	−1.7849	0.0742.
metro	−24.4215	4.881	−5.0034	<0.0001*
ρ	0.5395	0.078	6.9103	<0.0001*
Nagelkerke pseudo R^2	0.66292			

Note: * $p < 0.05$, $p < 0.1$.

Table 4. Impact measures for the parish model.

Variable	Direct	Indirect	Total
lu_mix	0.0021	0.0021	0.0042
under_13	−0.129	−0.13	−0.259
plus_65	−0.5944*	−0.599*	−1.1935*
sec_educ	0.4535*	0.45709*	0.9106*
unemp	−0.0603*	−0.0608*	−0.1212*
triptime	0.5424	0.5466	1.089
dwell_dens	−0.0034	−0.0034	−0.0068.
metro	−26.4135*	−26.6194*	−53.033*

Note: * $p < 0.05$, $p < 0.1$.

The other coefficients remained largely unchanged, yet it is interesting to note that only the direct and the total effects of the dwelling density were significant (at a 90% confidence level), meaning that, at the parish level, changes in car trips were negatively affected by changes in the density levels of a parish but not by changes in the density levels of its neighbors.

6.2. Census-tract model version I

The estimation results for the first version of the census-tract model, which evaluates the effect of station proximity without distinguishing station area types, are presented in Table 5.

At the micro level, growth in the number of residents aged 13 or less (per 1000 residents) became significant and positively related to the growth of car trips. Indeed, having children is often associated with many additional trips (to day care, sport and leisure facilities, etc.) that are easier to make by car. Surprisingly, growth in the unemployment rate became positively related to the number of car trips in this model. While this effect was not observed in the parish model, it is possible that, in certain census tracts, unemployment increases coincided with increases in car trips, for example, in peripheral rural areas where cars are essential for daily trips and where unemployed could have relocated due to lower cost of life (notably, cheaper housing rents).

Regarding built-environment controls, switching from parish to census-tract analysis shows how some of the factors that might be irrelevant on a macro scale turn out to be significant on a micro scale. In the case of dwelling density, both the variable and its spatial lag were found to negatively affect the number of car trips (the latter only at the 90% confidence level). Interestingly, though the spatial lag

Table 5. Estimation results for census-tract model version I.

Variable	Coefficient	Std. error	*z*-value	*p*-value
intercept	40.2649	6.0834	5.8597	<0.0001*
lu_mix	−0.006	0.0050	−1.2122	0.2255
under_13	0.31	0.046	6.7354	<0.0001*
plus_65	−0.3315	0.0377	−8.7921	<0.0001*
sec_educ	0.5611	0.0333	16.8462	<0.0001*
unemp	0.1041	0.0431	2.4147	0.0157*
triptime	0.7540	0.4594	1.6411	0.1008
dwell_dens	−0.0036	0.0018	−2.0419	0.0412*
lag_ dwell_dens	−0.0063	0.0035	−1.8262	0.0678.
up_to400	−43.5742	3.8063	−11.4479	<0.0001*
up_to800	−31.5188	3.4249	−9.2029	<0.0001*
up_to1200	−23.573	4.0603	−5.8057	<0.0001*
up_to1600	−21.0795	4.2065	−5.0112	<0.0001*
up_to2000	−7.2903	4.6785	−1.5583	0.1192
trainhour	−1.2989	0.1952	−6.6531	<0.0001*
λ	0.1393	0.0403	3.453	<0.0006*
Nagelkerke pseudo R^2	0.5063			
AIC	16273			

Note: * $p < 0.05$, $p < 0.1$.

appeared less significant, spatial lag's coefficient was stronger than the coefficient of the variable itself (-0.006 vs −0.003), meaning that, on a micro scale, densities in the neighboring units might be slightly more important in defining mode choice than densities in the census tract of residence. The land-use mix was not significant, however, as already mentioned, this variable does not provide much detail about existing land uses.

Proximity to a metro station was found to be significant within a 1.6 km range, being inversely related to the number of car trips, with decreasing influence on the dependent variable as distance increases. In fact, for census tracts located within a 400-meter buffer from a metro station, the opening of the station is associated with an average decrease of 43.5 in the number of car trips per 1000 residents, and this coefficient falls by about 10 car trips for each additional 400 meters until a 1200-meter distance from the station. Surprisingly, the coefficients for census tracts located within 800–1200 and 1200–1600-meter buffers are very similar (−23.5 and −21 trips, respectively). Probably, within a 1200-meter buffer from a station, a gradual decline in the influence of metro proximity reflects peoplés willingness to walk to/from the station, while after 1200 meters they prefer to use buses so the influence of distance was weaker for the range 1200–1600 meters. Unfortunately, available data do not allow to check this hypothesis.

Besides distance, the average daily train frequency on the closest station was significant, with each additional train being associated with an average decrease of 1.3 car trips per 1000 residents.

6.3. Census-tract model version II

The estimation results for the second version of the census-tract model, which accounts for station area types, are presented in Table 6. Overall, the coefficients are very similar to those in the previous model version, except for the trip time and train frequency variables: trip time became significant and positively related to the number of car trips, while the coefficient for the train frequency changed from −1.3 to −0.94. Compared to that version, there is a slight increase in Nagelkerke pseudo R^2, as well as a slight decrease in the Akaike Information Criterion (AIC), indicating in both cases a better model fit once station area types are taken into account.

Considering the proximity to different station types, TOD stations reveal the strongest (in terms of distance) spillover effects with a significant negative influence on the number of car trips visible in census tracts located up to 2 km away. The significant influence of TAD is limited to a 1.2 km distance range. Besides, the influence of TOD stations is considerably stronger than that of other station types, especially for census tracts located within 400–800 meters from a station: within this range, TOD stations observed a double decrease in the number of car trips compared to TAD stations (-41 car trips per 1000 residents for TOD versus −21 for TAD). TOD stations' influence decreases with distance: from −52.2 car trips in the immediate station area (up to 400 meters from the station) to −21.5 car trips for census tracts located within 1200–2000 meters, falling on average by 10 car trips every 400 meters. TAD stations on the other hand showed very little variation in coefficients for census tracts located within 400–1200 meters. P&R is the only station area type that does not have a significant negative influence on the number of car trips in the immediate station areas. This probably occurs because P&R immediate station areas are mainly scarcely populated census tracts with scattered private housing, where apparently station proximity does not change much people's habits. However, in some cases, small relatively dense settlements are located within a 400–800-meter buffer from a P&R station and, apparently, for their residents, station openings were important, even if the station area only had a parking lot and no other access options other than car were available. Furthermore, it should be noted that the number

Table 6. Estimation results for census-tract model version II.

Variable	Coefficient	Std. error	*t*-value	*p*-value
intercept	37.5686	6.7562	5.5606	<0.0001*
lu_mix	−0.006	0.005	−1.2016	0.2295
under_13	0.3132	0.0455	6.8854	<0.0001*
plus_65	−0.3345	0.0372	−8.9872	<0.0001*
sec_educ	0.5619	0.0327	17.1644	<0.0001*
unemp	0.1146	0.0429	2.6740	0.0075*
triptime	1.0658	0.4604	2.3147	0.0206*
dwell_dens	−0.0029	0.0018	−1.6633	0.0962.
lag_ dwell_dens	−0.0055	0.0035	−1.5718	0.116
trainhour	−0.9448	0.2094	−4.5110	<0.0001*
TOD400	−52.2267	4.3125	−12.1107	<0.0001*
TAD400	−34.248	6.0367	−5.6733	<0.0001*
PR400	−11.511	10.6667	−1.0792	0.2805
TOD800	−40.9818	3.9828	−10.2897	<0.0001*
TAD800	−20.961	5.4093	−3.8750	0.0001*
PR800	−21.147	7.5964	−2.7838	0.0054*
TOD1200	−28.3823	4.7309	−5.9993	<0.0001*
TAD1200	−20.6092	6.7495	−3.0534	0.0023*
PR1200	−15.9153	9.9968	−1.5920	0.1114
TOD2000	−21.4905	4.3675	−4.9205	<0.0001*
TAD2000	−9.9235	6.842	−1.4504	0.1469
PR2000	−8.5991	7.509	−1.0105	0.3123
λ	0.1239	8.5101	3.0495	0.0023*
Nagelkerke pseudo R^2	0.5141			
AIC	16262			

Note: * p <0.05, p <0.1.

of P&R stations within the Metro do Porto system is small compared to other station types, so the results might differ for networks in which the representativeness of different station types is more balanced.

6.4. Model validation

To assess the internal validity of the models and evaluate their performance on unseen data, a 5-fold cross-validation was performed using the "mlr3" package in R (Lang et al., 2019). For each fold, the dataset was randomly split into test data (20% of observations) and training data (80% of observations), then the model was fit using training data and validated on the test data. Since this procedure requires splitting the sample, it is no longer possible to account for different spatial specifications present in the full dataset, so these estimations were made using classic OLS regression. Once the procedure was applied to all folds, the regression performance measures were averaged.

For the parish model, cross validation resulted in an average $R^2 = 0.58$. Since the OLS regression performed with the full dataset led to an adjusted $R^2 = 0.66$, it can be said that, overall, the model performs relatively well on unseen data, though there is some decrease in the explanatory power. On the other hand, the average Spearman's rank correlation coefficient (measure of correlation between observed responses and predicted responses) was rather high: 0.81.

Considering the census-tract models, the resulting value of $R^2 = 0.49$ (similar for both models), coincided with the adjusted R^2 value for the OLS regression with the full dataset, meaning that this parameter remained stable during cross-validation. Average Spearman's rank correlation coefficient was 0.68, indicating a substantial correlation between the observed values and the predicted responses.

7. Conclusion

Despite the substantial efforts devoted to the assessment of the effects of TOD on travel behavior and, particularly, on the reduction of car trips, research works analyzing those effects over time and space are relatively rare, and are limited in terms of sample size and geographical scope. By applying first difference estimator spatial regression models to a metro system with TOD features spread across a rather vast region using census data, we have addressed its time and space effects at two levels of detail (parishes and census tracts) circumventing such limitations. Moreover, we have analyzed these effects depending on metro station area types at the census-tract level. Considering all these aspects together, the methodology we have applied in our study offers, in our opinion, a significant contribution to the TOD literature, and its application to Metro do Porto delivered solid conclusions on how the implementation of this system leveraged the reduction in car trips in the Porto region.

The positive impact of Metro do Porto with respect to car trip reduction is visible even at the parish level. This goes in line with the findings reported by Ibraeva et al. (2021) where metro-served parishes revealed smaller shares of car use than those which would have been observed in the absence of metro. Similarly, Van de Coevering et al. (2016) note that the further people live from a station the more they tend to use private vehicle over time. Quasi-longitudinal studies from Xi'an (China) and Minneapolis-St. Paul (USA) also show that station proximity has negative influence on car use as it tends to decrease among people who move into station areas (Cao & Ermagun, 2017; Huang et al., 2019). Our results also match the various cross-sectional studies where it was concluded that station proximity (Cervero & Gorham, 1995; Cervero and Arrington, 2008; Nasri & Zhang, 2014; Griffiths and Curtis; 2017; Yin et al., 2019), population density (Cervero & Gorham, 1995; Cervero and Arrington, 2008; Nasri & Zhang, 2014) and elderly population (Handy et al., 2005) are negatively related to car use, while education level (Cervero & Day, 2008) is positively related to it. However, a few studies arrived to different conclusions. For instance, the analysis (of a rather different nature) carried out by Lee and Senior (2013) using 1991–2001 census data from several new light-rail corridors in the UK revealed that the decrease in car use occurred in these corridors was modest and not much different from what was observed in control areas.

Switching to the census-tract level allows detecting with greater precision the effects of station proximity and type, revealing how their magnitude changes with distance. For the period we have examined, such effects can be summarized as follows:

- Overall, the impact of station proximity on car trips reduction is naturally maximum in the immediate station area (−43.6 car trips per 1000 residents in the 400-meter buffer) and falls at a rather constant rate until a 1200-meter distance (with every additional 400 meters from a station the impact diminishes by about 10 trips). Beyond 1600 meters, the impact decreases sharply, becoming practically null.
- The impact of TOD stations within a 400-meter buffer is clearly higher (by 1.5 times) than that of TAD and P&R stations (no significant impact). Until 1200 meters, in the case of TOD stations, this impact decreases very steadily with distance (−10 car trips every 400 meters), and after that stays visible up to 2 km. Instead, the impact of TAD stations remains practically the same between 400 meters and 1200 meters from the station. P&R stations have a significant impact on car trips (about −21 car trips per 1000 residents) only within a distance of 400–800 meters from the station.

These types of findings have important implications for urban planners and practitioners. Since TOD stations reveal greater potential in reducing the number of car trips as compared to other station types, the impact of the introduction of a metro system might be further reinforced by interventions in station areas, and efforts should be made in this direction already at the planning stage. As the decay of the spillover effect from TOD stations is more stable, it might be easier to foresee and plan for their potential outreach.

The presence and magnitude of station's spillover effects suggest that, probably, broader areas should be taken into consideration during the preparation of urban/transport plans for a new station or the evaluation of its effects.

Considering research developments, several issues may be addressed in the future to complete the reported findings. We will mention here three of those issues. First, focusing on the Porto region, due to data limitations it was not possible to recreate in detail the land use mix of the study area. With more refined data on existing land uses, it would be possible to better assess TOD effects and to devise better plans for the expansion of the metro system announced to take place in the next few years. The second issue relates to the availability of Census 2021 data in the near future (end of 2022), which would allow extending the analysis until the recent past. Between 2011 and 2021, practically no investments in infrastructure projects were made in Portugal, as the country was struggling with a severe debt crisis and public spending had to be cut to a strict minimum. In this context, following the short-term post-intervention effects considered in this paper, it would also be possible to look at longer-term effects without the data contamination that would result from changes in the metro and other transport systems. The third issue concerns the geographic scope of our study. Its findings are, naturally, case-specific, and the conclusions might not be the same in other places. To be able to generalize them, more studies are necessary for other metro systems. The methodology we have used has proven to be effective and should not be too difficult to replicate in different geographical contexts, notably in the European Union, since census data is very much the same in every country.

Acknowledgments

The research described in this paper has been carried out in CITTA (Research Center for Territory Transports and Environment) in the framework of the doctoral thesis of Anna Ibraeva (PD/BD/135417/2017). Both the research center and the doctoral thesis are funded by FCT – Fundação para a Ciência e Tecnologia.

ORCID

Anna Ibraeva http://orcid.org/0000-0002-3767-7256
Gonçalo Homem de Almeida Correia http://orcid.org/0000-0002-9785-3135
Cecília Silva http://orcid.org/0000-0003-2868-1840
António Pais Antunes http://orcid.org/0000-0002-8265-9225

References

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77–104. https://doi.org/10.1016/0166-0462(95)02111-6

Bardaka, E., & Hersey, J. (2019). Comparing the travel behavior of affordable and market-rate housing residents in the transit-rich neighborhoods of Denver, CO. *Travel Behaviour and Society*, *15*, 74–87. https://doi.org/10.1016/j.tbs.2019.01.001

Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, *63*(18), 1–36. https://doi.org/10.18637/jss.v063.i18

Branco, J. F., & Ramos, M. J. (2003). *Estrada Viva? Aspectos da Motorização na Sociedade Portuguesa*. Assírio & Alvim.

Brown, B. B., & Werner, C. M. (2008). Before and after a new light rail stop: Resident attitudes, travel behavior, and obesity. *Journal of the American Planning Association*, *75*(1), 5–12. https://doi.org/10.1080/01944360802458013

Calthorpe, P. (1993). The next American metropolis. *Ecology, Community and the American Dream*. Princeton Architectural Press.

Cao, J. (2015). Heterogeneous effects of neighborhood type on commute mode choice: An exploration of residential dissonance in the Twin Cities. *Journal of Transport Geography*, *48*, 188–196. https://doi.org/10.1016/j.jtrangeo.2015.09.010

Cao, J., & Cao, X. (2014). The impacts of LRT, neighbourhood characteristics, and self-selection on auto ownership: Evidence from Minneapolis-St.Paul. *Urban Studies*, *51*(10), 2068–2087. https://doi.org/10.1177/0042098013505887

Cao, J., & Ermagun, A. (2017). Influences of LRT on travel behaviour: A retrospective study on movers in Minneapolis. *Urban Studies*, *54*(11), 2504–2520. https://doi.org/10.1177/0042098016651569

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*, *34*(5), 535–556. https://doi.org/10.1007/s11116-007-9132-x

Cao, X., 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. https://doi.org/10.1080/01441640802539195

Cervero, R. (1995). Sustainable new towns. Stockholm's rail-served satellites. *Cities*, *12*(1), 41–51. https://doi.org/10.1016/0264-2751(95)91864-C

Cervero, R. (2007). Transit-oriented development's ridership bonus: A product of self-selection and public policies. *Environment and Planning A: Economy and Space*, *39*(9), 2068–2085. https://doi.org/10.1068/a38377

Cervero, R., & Day, J. (2008). Suburbanization and transit-oriented development in China. *Transport Policy*, *15*(5), 315–323. https://doi.org/10.1016/j.tranpol.2008.12.011

Cervero, R., & Gorham, R. (1995). Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association*, *61*(2), 210–225. https://doi.org/10.1080/01944369508975634

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Cervero, R., & Radisch, C. (1996). Travel choices in pedestrian versus automobile oriented neighborhoods. *Transport Policy*, *3*(3), 127–141. https://doi.org/10.1016/0967-070X(96)00016-9

Chatman, D. G. (2013). Does TOD need the T? On the importance of factors other than rail access. *Journal of the American Planning Association*, *79*(1), 17–31. https://doi.org/10.1080/01944363.2013.791008

Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of Metro ridership at the station-to-station level in Seoul. *Transportation*, *39*(3), 705–722. https://doi.org/10.1007/s11116-011-9368-3

Croissant, Y., & Millo, G. (2019). *Panel data econometrics with R*. John Wiley & Sons Ltd.

Elhorst, J. P. (2014). Spatial econometrics. *From cross-sectional data to spatial panels*. Springer. https://doi.org/10.1007/978-3-642-40340-8

Dons, E., Rojas-Rueda, D., Anaya-Boig, E., Avila-Palencia, I., Brand, C., Cole-Hunter, T., de Nazelle, A., Eriksson, U., Gaupp-Berghausen, M., Gerike, R., Kahlmeier, S., Laeremans, M., Mueller, N., Nawrot, T., Nieuwenhuijsen, M. J., Orjuela, J. P., Racioppi, F., Raser, E., Standaert, A., Int Panis, L., & Götschi, T. (2018). Transport mode choice and body mass index: Cross-sectional and longitudinal evidence from a European-wide study. *Environment International*, *119*, 109–116. https://doi.org/10.1016/j.envint.2018.06.023

Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(1), 87–114. https://doi.org/10.3141/1780-10

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Griffiths, B., & Curtis, C. (2017). Effectiveness of transit oriented development in reducing car use: Case Study of Subiaco, Western Australia. *Urban Policy and Research*, *35*(4), 391–408. https://doi.org/10.1080/08111146.2017.1311855

Handy, S. L., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427–444. https://doi.org/10.1016/j.trd.2005.05.002

Handy, S., Cao, X., & Mokhtarian, P. (2008). The causal influence of neighborhood design on physical activity within the neighborhood: Evidence from Northern California. *American Journal of Health Promotion: AJHP*, *22*(5), 350–358. https://doi.org/10.4278/ajhp.22.5.350

Huang, X., Cao, J., Yin, J., & Cao, X. (2019). Can metro transit reduce driving? Evidence from Xi'an, China. *Transport Policy*, *81*, 350–359. https://doi.org/10.1016/j.tranpol.2018.03.006

Ibraeva, A., Correia, G. H., A., Silva, C., & Antunes, A. P. (2020). Transit-oriented development: A review of research achievements and challenges. *Transportation Research Part A: Policy and Practice*, *132*, 110–130. https://doi.org/10.1016/j.tra.2019.10.018

Ibraeva, A., Correia, G. H. A., Silva, C., & Antunes, A. P. (2022). Mobility impacts of a new metro system with transit-oriented development features. *Transportation Research Part D: Transport and Environment*, *109*, 103357. https://doi.org/10.1016/j.trd.2022.103357

Ibraeva, A., Van Wee, B., Correia, G. H. A., & Antunes, A. P. (2021). Longitudinal macro-analysis of car-use changes resulting from a TOD-type project: The case of Metro do Porto (Portugal). *Journal of Transport Geography*, *92*, 103036. https://doi.org/10.1016/j.jtrangeo.2021.103036

Kamruzzaman, M., Shatu, F. M., Hine, J., & Turrell, G. (2015). Commuting mode choice in transit oriented development: Disentangling the effects of competitive neighbourhoods, travel attitudes, and self-selection. *Transport Policy*, *42*, 187–196. https://doi.org/10.1016/j.tranpol.2015.06.003

Kimpton, A., Pojani, D., Sipe, N., & Corcoran, J. (2020). Parking Behavior: Park 'n' Ride (PnR) to encourage multimodalism in Brisbane. *Land Use Policy*, *91*, 104304. https://doi.org/10.1016/j.landusepol.2019.104304

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., & Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*, *4*(44), 1903. https://doi.org/10.21105/joss.01903

Logan, J. R., Stults, B. J., & Xu, Z. (2016). Validating population estimates for harmonized census tract data, 2000–2010. *Annals of the American Association of Geographers*, *106*(5), 1013–1029. https://doi.org/10.1080/24694452.2016.1187060

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *Review of Regional Studies*, *44*(1), 13–32. https://doi.org/10.52324/001c.8081

LeSage, J., & Pace, K. P. (2009). *Introduction to spatial econometrics*. Chapman & Hall/CRC.

Loo, B. P. Y., Chen, C., Chan,., & E. T., H. (2010). Rail-based transit-oriented development: Lessons from New York City and Hong Kong. *Landscape and Urban Planning*, *97*(3), 202–212. https://doi.org/10.1016/j.landurbplan.2010.06.002

Naess, P. (2015). Built environment, causality and travel. *Transport Reviews*, *35*(3), 275–291. https://doi.org/10.1080/01441647.2015.1017751

Nasri, A., & Zhang, L. (2014). The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. *Transport Policy*, *32*, 172–179. https://doi.org/10.1016/j.tranpol.2013.12.009

Nasri, A., & Zhang, L. (2019). How urban form characteristics at both trip ends influence mode choice: Evidence from TOD vs. non-TOD zones of the Washington, D.C. metropolitan area. *Sustainability*, *11*(12), 3403. https://doi.org/10.3390/su11123403

Newman, P. W. G., & Kenworthy, J. R. (1988). The transport energy trade-off: Fuel-efficient traffic versus fuel-efficient cities. *Transportation Research Part A: General*, *22*(3), 163–174. https://doi.org/10.1016/0191-2607(88)90034-9

Newman, P. W. G., & Kenworthy, J. R. (1996). The land use-transport connection. An overview. *Land Use Policy*, *13*(1), 1–22. https://doi.org/10.1016/0264-8377(95)00027-5

Olaru, D., & Curtis, C. (2015). Designing TOD precincts: accessibility and travel patterns. *European Journal of Transport and Infrastructure Research*, *15*(1), 6–26. https://doi.org/10.18757/ejtir.2015.15.1.3052

Pan, H., Li, J., Shen, Q., & Shi, C. (2017). What determines rail transit passenger volume? Implications for transit oriented development planning. *Transportation Research Part D: Transport and Environment*, *57*, 52–63. https://doi.org/10.1016/j.trd.2017.09.016

Park, K., Ewing, R., Scheer, B. C., & Tian, G. (2018). The impacts of built environment characteristics of rail station areas on household travel behavior. *Cities*, *74*, 277–283. https://doi.org/10.1016/j.cities.2017.12.015

Parkhurst, G. (1995). Park and ride: could it lead to an increase in car traffic? *Transport Policy*, *2*(1), 15–23. https://doi.org/10.1016/0967-070X(95)93242-Q

Sener, I. P., Lee, K., Durand, C. P., Oluyomi, A. O., & Kohl, H. W. III, (2020). Intention to use light-rail transit in Houston, Texas, United States: Findings from the travel-related activity in neighborhoods study. *International Journal of Sustainable Transportation*, *14*(12), 944–955. https://doi.org/10.1080/15568318.2019.1663962

Van Acker, V., Witlox, F., & Van Wee, B. (2007). The effects of the land use system on travel behavior: A structural equation modeling approach. *Transportation Planning and Technology*, *30*(4), 331–353. DOI: 10.1080/03081060701461675.

Sung, H., & Oh, J.-T. (2011). Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, *28*(1), 70–82. https://doi.org/10.1016/j.cities.2010.09.004

Van de Coevering, P., Maat, K., Kroesen, M., & Van Wee, B. (2016). Causal effects of built environment characteristics on travel behavior: a longitudinal approach. *The European Journal of Transport and Infrastructure Research*, *16*(4), 674–697. https://doi.org/10.18757/ejtir.2016.16.4.3165.

Van Wee, B., & Cao, X. J. (2022). Chapter Four - Residential self-selection in the relationship between the built environment and travel behavior: A literature review and research agenda. *Advances in Transport Policy and Planning*, *9*, 75–94. https://doi.org/10.1016/bs.atpp.2020.08.004

Walton, S., & Sunseri, S. (2010). Factors influencing the decision to drive or walk short distances to public transport facilities. *International Journal of Sustainable Transportation*, *4*(4), 212–226. https://doi.org/10.1080/15568310902927040

Zhang, J. (2014). Revisiting residential self-selection issues A life-oriented approach. *Journal of Transport and Land Use*, *7*(3), 29–45. https://doi.org/10.5198/jtlu.v7i3.460