Journal of Transport Geography 60 (2017) 155–166

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

A multilevel spatial interaction model of transit flows incorporating spatial and network autocorrelation

Kasper Kerkman [a,*], Karel Martens [a,b], Henk Meurs [a]

[a] Institute for Management Research, Radboud University, P.O. Box 9108, Nijmegen 6500 HK, The Netherlands
[b] Leona Chanin Career Development Chair, Faculty of Architecture and Town Planning, Technion - Israel Institute of Technology, Amado Building, Technion City, 32000 Haifa, Israel

ARTICLE INFO

Article history:
Received 26 April 2016
Received in revised form 15 February 2017
Accepted 24 February 2017
Available online xxxx

Keywords:
Spatial interaction model
Spatial autocorrelation
Public transport
Demand forecasting

ABSTRACT

While it is well known that ignoring spatial dependence often results in misspecification of models, travel demand models almost never account for this phenomenon. The aim of this paper is to empirically demonstrate the importance of accounting for potential spatial dependence between observations in the specification of spatial interaction models. As a case study, we analyze travel flows on the public transport system in an urban region in the Netherlands. We develop five distinct spatial interaction models (SIMs) of increasing complexity, each encompassing a lower and upper level model. At the lower level, the attractiveness of neighborhoods for boarding and alighting is modeled based on spatial and transit supply characteristics. At the upper level, spatial interactions among zones are modeled taking into account competing origins, competing destinations as well as network characteristics. We systematically compare more traditional SIM formulations with a SIM that explicitly accounts for spatial and network autocorrelation. The results show a substantial difference between the former models and the latter, in terms of the estimated total marginal impacts of the different variables and the pattern of the error terms. The results of our study underscore that the failure to incorporate autocorrelation effects in travel models is likely to influence model outcomes, which in turn may have profound implications for the very design of public transport networks in cities and regions.

1. Introduction

The aim of this paper is to demonstrate the importance of accounting for potential spatial dependence between observations in the specification of spatial interaction models. As a case study, we analyze public transport travel flows in an urban region in the Netherlands.

Spatial interaction modeling (SIM) is an approach to explore, analyze and explain flows of people, goods or information over space. It has been used frequently to model migration flows, freight transport flows, trade flows, and the distribution of telephone calls. The advantage of spatial interaction models is that these models can take into account the influence of both spatial characteristics and characteristics of the transport network simultaneously.

Spatial interaction models always form a core component of travel demand models, which are routinely employed around the world as part of the transport planning process (e.g., Bates, 2000). In traditional models, travel demand is forecasted in a four-step process: trip generation, trip distribution, mode choice, and trip assignment. In the second step of these models, the data on trip production and attraction per zone generated in the first step (trip generation) are recombined into trips between origins and destinations, based on the attractiveness of zones and travel impedance between them. In most cases, a SIM is used to execute this step.

The trip distribution step of these travel demand models almost never account for the potential spatial dependence of travel flows, even though it is well known that ignoring this phenomenon often results in misspecification of models, resulting in flawed predictions (LeSage & Pace, 2008). Over the past decades, more advanced spatial interaction models have been developed which do account for spatial dependence (Anselin, 2010). Recently, these models have successfully been applied in studies on patent citation distributions (Fischer and Griffith, 2008), migration flows (LeSage and Pace, 2008), and commodity flows (Chun et al., 2012). To the best of our knowledge, these advanced models have not yet been applied to study traffic flows or incorporated in transport models. This is problematic, especially because transport models are employed around the world to justify often very capital intensive transport investments. Against this background, the goal of this paper is to demonstrate the importance of accounting for spatial interdependence in the specification of transport models. For

* Corresponding author.
E-mail addresses: k.kerkman@fm.ru.nl (K. Kerkman), k.martens@fm.ru.nl (K. Martens), h.meurs@fm.ru.nl (H. Meurs).

http://dx.doi.org/10.1016/j.jtrangeo.2017.02.016
0966-6923/

this purpose, we will analyze the flows of public transport trips in the Arnhem-Nijmegen region, the Netherlands.

The paper is structured as follows. Following this introduction, we will briefly discuss the importance of taking spatial autocorrelation into account in (transport) models on spatial interactions (Section 2). In Section 3, we discuss previous applications of SIMs of transit flows. After this, we specify our models used to analyze transit flows (Section 4). In Section 5, we provide a description of the data for our case study. The modeling results are presented in Section 6. Section 7 gives the conclusions and discusses the implications of the findings.

2. Spatial autocorrelation

Since travel flows are a particular form of spatial interaction, transport models always contain some type of spatial interaction model. Conventional four-step transport models have a submodel related to trip distribution describing the interactions between origins and destinations (Ortuzar and Willumsen, 2001). In the estimation of the distribution function it is assumed that the errors in these models are independent among the zones. Note that errors in destination choice models of the logit-type are often assumed to be independent as well, yielding similar problems as discussed in the remainder of this paper (Ben-Akiva and Lerman, 1985). In this paper such models are outside the scope, since we do not have disaggregated data available. The generalized spatially correlated logit model (SCL) proposed by Bhat and Guo (2004) constitutes a generalized extreme value (GEV) based formulation of the phenomenon of interest in this paper, in case disaggregate data are available. In this model spatial interactions are an important component.

Measurement error occurs when the location of a variable or the value of a variable are observed with imperfect accuracy. The main problem is that the geometric and graphical representation of the location of points, lines or areal boundaries (i.e., a map), gives an imperfect impression of the uncertainty associated with errors in their measurement.

Other spatial errors of measurement have to do with the imperfect way in which data on socio-economic phenomena are recorded and grouped in spatial units of observation (e.g. types of administrative units/zones). This interdependence of location and value in spatial data leads to distinctively spatial characteristics of the errors. These are the common phenomena of spatial dependence and spatial heterogeneity:

- Dependence is mostly due to the existence of spatial spillovers, as a result of a miss-match between the scale of the spatial unit of observation and the phenomenon of interest (e.g., continuous processes represented as points, or processes extending beyond the boundaries of administrative regions).
- Heterogeneity is due to structural differences between locations and leads to different error distributions (e.g., differences in accuracy of census counts between low-income and high-income neighborhoods).

It is well known that many (transport) models [aiming to capture spatial interaction] actually do not take into account the possible spatial dependence between errors. As a result, such models lead to misspecification, since the independence assumption regarding errors is violated (see for example Bolduc et al., 1992, 1989; LeSage and Pace, 2008). The failure to take spatial autocorrelation into account may be due to the high complexity of spatial interaction models that do address this phenomenon. The higher complexity of models employing spatial data compared to time-series data is the result of the fact that, unlike time-series, spatial autocorrelation is multidirectional. An observation of an attribute at one location can be correlated with the value of the same attribute at any different location, and vice versa.

The importance of accounting for spatial dependence obviously depends on the actual strength of the unobserved spatial relationships. More specifically, the size of the bias in models estimating trips or flows depends on the size of the correlation between the error terms and the number of trips. In case these correlations are positive, the parameters in the spatial interaction models will be overestimated.

3. SIMs of transit flows

Although a SIM is generally used for modeling the trip distribution in traditional four-step travel demand models, it has rarely been used to analyze and explain transit flows explicitly. A reason for this might be the fact that, until recently, little reliable and detailed data on transit passenger flows was available. New data sources are coming rapidly available in recent years, especially due to technological developments. This increasing availability of 'big' data is both an opportunity and a challenge for spatial and regional studies (Arribas-Bel, 2014; Kitchin, 2013; Rae and Singleton, 2015). Also in transport research, new data sources are increasingly available and used (Yue et al., 2014). Especially the number of data collection techniques that can capture personal trip trajectories have increased tremendously. Examples of these are GPS trackers, mobile phones, and transit smart cards. In addition to traditional data sources such as paper interviews, travel diaries, and stated preference data, these new data sources bring new opportunities, as they give more complete and detailed information about travel patterns (Tao et al., 2014). We build on these possibilities, by using smart card data of all bus boardings and alightings to estimate SIMs for transit passenger flows.

Only a very limited number of studies are available that have constructed SIMs for public transport travel flows. Goh et al. (2012) and Smith et al. (2012) successfully estimate SIMs for the London rail network and the Seoul Metropolitan Subway system, respectively. Goh et al. (2014) applied a modified gravity model to the passenger flows in the Seoul bus system. They found that the geographical environment had a far larger influence on the usage of the bus system than it had on the subway system. These studies, however, were mainly motivated by analyzing social phenomena and social networks exemplified by the transit passenger flows. As such, they do not provide many insights in how different aspects of transit supply and spatial characteristics influence passenger flows. Moreover, none of these models considered the possible influence of spatial dependence in their modeling approaches.

In the current study, we extend these efforts in a number of ways in order to estimate a more realistic SIM of transit passenger flows, exploring the influence of both spatial and network characteristics at different spatial scales on transit travel flows. Firstly, we model transit flows as a multilevel phenomenon. At the lower level, boardings and alightings are modeled as a function of spatial characteristics and transit supply characteristics at the neighborhood level. This includes population characteristics, densities, employment, transit frequency, transfer options, and bus stop characteristics. At the upper level we model spatial interactions among zones using spatial as well as network characteristics. At this level, we also take spatial competition into account.

Secondly, in our model we explicitly take spatial dependence into account by including spatial and network autocorrelation. We will explore a number of formulations capturing these effects. First, we correct for spatial autocorrelation in the lower level (boardings) model by including information about adjacent neighborhoods in a spatial autoregressive model. Second, at the upper level model (flows), we correct for potential network autocorrelation accounting for the influence flows may have on flows on related OD-pairs. The exact, formal specification of the SIMs is described in the following section.

Note that our SIMs, while accounting for multiple types of autocorrelation, are still static in nature. That is, our models aim to explain the observed static pattern of transit flows, although we are of course well aware that these flows are the result of the activity patterns of individuals, which are dynamic in nature and change in space and time. In this