The Journal of Choice Modelling 18 (2016) 1–17

Contents lists available at ScienceDirect

The Journal of Choice Modelling

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

Location choice with longitudinal WiFi data

Antonin Danalet*, Loïc Tinguely, Matthieu de Lapparent, Michel Bierlaire

École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR), Switzerland

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form
11 April 2016
Accepted 12 April 2016

Keywords:
Location choice
Panel data
Pedestrians
Dynamic model
Initial conditions problem

ABSTRACT

Location-aware data collection technologies provide new insights about location choices. Only a few dynamic models of location choice exist in scientific literature. To our knowledge, none of them correct for serial correlation. In this paper, we model choice of catering locations on a campus using WiFi traces. We use the Wooldridge (2005) correction method that deals with the initial values problem and related endogeneity bias in estimation. Cross-validation, price elasticity and simulation of a scenario predicting the opening of a new catering location are presented. Predicted market shares of the new catering location correspond to point-of-sale data of the first week of opening.

1. Introduction

Properly modeling destination choices is important in order to understand travel behavior and travel demand, both at the urban scale and in pedestrian facilities. In transportation, destination choice modeling is often used by local and national authorities for planning future infrastructures and policies (e.g., Fox et al., 2014) and for the planning and design of multimodal transport hubs (Hoogendoorn and Bovy, 2004). In tourism, the choice of destinations is important for analyzing demand for holidays locations (e.g., Yang et al., 2013) and for the management of pedestrian flows in museums (Yoshimura et al., 2014) and in parks (O'Connor et al., 2005). In all of these contexts, destination choice models commonly infer on the relevant factors that influence the decisions and allow the testing of policies when building new infrastructures or optimizing current ones. Demand management strategies can be evaluated.

Most of destination choice models rely on cross-sectional data (e.g., BenAkiva and Lerman, 1985; Zhu and Timmermans, 2011; Scott and He, 2012; Kalakou et al., 2014). As they are collected at one point in time, the related frameworks of analysis are static. As stated by Hsaio (2003), "a longitudinal, or panel, data set is one that follows a given sample of individuals over time, and thus provides multiple observations on each individual in the sample". Panel data are difficult and expensive to collect using standard survey techniques (Yang and Timmermans, 2015), and sometimes nonexistent, e.g., for the analysis of induced traffic at an aggregate level (Weis and Axhausen, 2009). In absence of actual panel data, pseudo-panel data are constructed by grouping individuals from cross-sectional data into cohorts and by considering behavior of cohorts as individuals (Deaton, 1985; Weis and Axhausen, 2009; McDonald, 2015). However, actual panel data from new technologies are more and more used (Carrion et al., 2014; Kazagli et al., 2014). Network traces (e.g., WiFi traces or cell tower data) are

* Corresponding author.
E-mail addresses: antonin.danalet@alumni.epfl.ch (A. Danalet), loic.tinguely@alumni.epfl.ch (L. Tinguely), matthieu.delapparent@epfl.ch (M.d. Lapparent), michel.bierlaire@epfl.ch (M. Bierlaire).

http://dx.doi.org/10.1016/j.jocm.2016.04.003
1755-5345/

increasingly available for location choices (see Section 2.2). Compared to traditional surveys, network traces follow individuals over longer periods (see Section 2.1). This makes it possible to collect sequences of activity locations covering several days, weeks or months. Location choice models must be adapted to use these data.

In this paper, we model dynamics of location choices for catering. We make best use of panel data by taking into account state dependence and serial correlation. We solve the initial values problem and related endogeneity bias in estimation using Wooldridge's (2005) correction method. Accounting for panel data nature in location choices has never been treated in the literature before. It allows us to correct for serial correlation, while understanding people's habits in their decision process. The methodology is applied to sequences of catering locations on a campus collected using WiFi access points (Danalet et al., 2014).

The rest of the paper is organized as follows. Review of literature is presented in Section 2. We detail methodology in Section 3. Our case study is discussed in Section 4. It also includes cross-validation and forecasting. Conclusions are drawn in Section 5.

2. Literature review

2.1. From diary surveys to location-aware technologies

One recent trend in travel demand modeling is resorting to location-aware technologies (Chen and Yang, 2014; Danalet et al., 2014; Miller, 2014; Carrel et al., 2015). Traditionally, disaggregate data of revealed preferences about activity and travel patterns are collected from diary surveys, where people describe 1 or 2 past days (Ettema, 1996; Carrel et al., 2015). The largest panel surveys include a six-week period for 317 participants (Axhausen et al., 2002), a six-week period for 261 participants (Axhausen et al., 2007) and a twelve-week period for 71 participants (Schlich, 2004). Most long-term surveys cover a maximum of 7 days and are not panel data (Ortúzar et al., 2011; Carrel et al., 2015). Location-aware technologies improve the quality of surveying. For instance, combination of GPS devices carried by respondents with standard recall questionnaires makes for easier implementation of longitudinal surveys (Frignani et al., 2010; Yang and Timmermans, 2015). Recall methods can also be directly implemented on mobile devices (Rindfüser et al., 2003; Cottrill et al., 2013).

Location-aware technologies can also be used alone. It can be from the communication infrastructure side, such as cell tower traces or WiFi access points traces (Bekhor et al., 2013; Calabrese et al., 2013; Danalet et al., 2014). It can also be from the individuals' devices (Etter et al., 2012; Buisson, 2014; Chen and Yang, 2014; Carrel et al., 2015). Etter et al. (2012) show that it is possible to predict up to 60% of next visited places from passive smartphone data.

2.2. Location choice

Location choice models are common in studies of urban transportation policies and planning. BenAkiva and Lerman (1985) mention three of them, for the Paris region and Maceio, Brazil. Such models have been applied to the choice of location for grocery shopping (Timmermans, 1996; Dellaert et al., 1998; Fox et al., 2004; Scott and He, 2012). They also relate to other applications: choice of a departure airport (Furuichi and Koppelman, 1994), the choice of a hospital for patients by general practitioner (primary care physicians) (Whynes et al., 1996), the choice of tourist destinations (Woodside and Lysonski, 1989; Um and Crompton, 1990; Eymann and Ronning, 1997; Oppermann, 2000; Seddighi and Theocharous, 2002; Bigano et al., 2006; Chi and Qu, 2008; Gössling et al., 2012; Yang et al., 2013) and in particular recreational outdoor facilities (Fesenmaier, 1988; Scarpa and Thiene, 2005; Thiene and Scarpa, 2009), the choice of migrants (Fotheringham, 1986) and the optimal allocation of charging stations for electric vehicles (He et al., 2013).

Regarding pedestrians, Zhu and Timmermans (2011) propose heuristic rules pertaining to bounded rationality. They compare them with random utility maximization discrete choice models. The models are validated on the same sample used for estimation. Cross validation is not carried out. Ton (2014) studies route and location choice in train stations based on tracking and counting data. Counting data come from infrared scanners and tracking data come from WiFi and Bluetooth scanners. Count data are used to model pedestrians without smartphones. The choice is between locations for a given activity type, e.g., which coffee shop knowing that the individual is visiting one. Kalakou et al. (2014) apply a similar approach for location choice for a given activity type in an airport.

2.2.1. Attributes of the choice of a location

The main attributes in location choices in urban context are travel time, travel cost and distance (Cambridge Systematics Europe, 1984; BenAkiva and Lerman, 1985; Whynes et al., 1996). Other variables are used: park-finding time, parking cost, type of neighborhood, and the number of different services (banks, post offices, medical facilities, offices, shops, etc.) in the zone (Cambridge Systematics Europe, 1984; BenAkiva and Lerman, 1985). Another typical attribute is the *size* in the context of aggregation of alternatives (see Section 2.2.2). It represents the number of elemental alternatives in the considered aggregate alternatives (subsets of the choice set). The interpretation of this attribute is complicated, since it absorbs both the preference for a large set of destinations compared to a small one and the correlation between destinations in the set. The expected sign is opposite in the two situations (Frejinger and Bierlaire, 2007). In shop patronage, the main attributes are the retail floor space, the accessibility and the price (Arnold et al., 1983; Scott and He, 2012). Other attributes include parking