Transportation Research Part A 84 (2016) 44–54

Contents lists available at ScienceDirect

Transportation Research Part A

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

A zonal inference model based on observed smart-card transactions for Santiago de Chile

Sebastián Tamblay *, Patricia Galilea, Paula Iglesias, Sebastián Raveau, Juan Carlos Muñoz

Pontificia Universidad Católica de Chile, Department of Transport Engineering and Logistics, Vicuña Mackenna 4860, Macul, Santiago 7820436, Chile

ARTICLE INFO

Article history:
Available online 19 November 2015

Keywords:
Public transport
Origin–destination matrix
Smartcard data
Zoning system

ABSTRACT

The collection of origin–destination data for a city is an important but often costly task. This way, there is a need to develop more efficient and inexpensive methods of collecting information about citizens' travel patterns. In this line, this paper presents a generic methodology that allows to infer the origin and destination zones for an observed trip between two public transport stops (*i.e.*, bus stops or metro stations) using socio-economic, land use, and network information. The proposed zonal inference model follows a disaggregated Logit approach including size variables. The model enables the estimation of a zonal origin–destination matrix for a city, if trip information passively collected by a smart-card payment system is available (in form of a stop-to-stop matrix). The methodology is applied to the Santiago de Chile's morning peak period, with the purpose of serving as input for a public transport planning computational tool. To estimate the model, information was gathered from different sources and processed into a unified framework; data included a survey conducted at public transport stops, land use information, and a stop-to-stop trip matrix. Additionally, a zonal system with 1176 zones was constructed for the city, including the definition of its access links and associated distances. Our results shows that, *ceteris paribus*, zones with high numbers of housing units have higher probabilities of being the origin of a morning peak trip. Likewise, health facilities, educational, residential, commercial, and offices centres have significant attraction powers during this period. In this sense, our model manages to capture the expected effects of land use on trip generation and attraction. This study has numerous policy implications, as the information obtained can be used to predict the impacts of changes in the public transport network (such as extending routes, relocating their stops, designing new routes or changing the fare structure). Further research is needed to improve the zonal inference formulation and origin–destination matrix estimation, mainly by including better cost measures, and dealing with survey and data limitations.

1. Introduction

According to Ortúzar and Willumsen (2011), one of the key elements for correct urban transport planning is the knowledge of a reliable origin–destination (O–D) matrix of the city of interest. However, in fast paced changing cities with detailed zoning systems this information might be hard, costly, and time consuming to acquire and keep up to date. Furthermore,

* Corresponding author. Tel.: +56 (2) 2354 4270; fax: +56 (2) 2354 0281.
E-mail addresses: sjtambla@uc.cl (S. Tamblay), pgalilea@ing.puc.cl (P. Galilea), piglesias@dictuc.cl (P. Iglesias), sraveau@uc.cl (S. Raveau), jcm@ing.puc.cl (J.C. Muñoz).

http://dx.doi.org/10.1016/j.tra.2015.10.007

even by spending large amounts of money in household activity and travel surveys, there is no guarantee that accurate O–D matrices can be obtained if no additional information is included (Cools et al., 2010). The latter evidences the need to develop more efficient and inexpensive methods of collecting information about citizens' travel patterns.

The last decade has seen an increasing trend in the implementation of information technologies such as Automated Fare Collection and Automatic Vehicle Location Systems, which provide very valuable information about trip patterns. In systems where passengers are forced to tap-in and out in every vehicle (bus or train), a full stop-to-stop (bus stops or metro stations) trip matrix would be readily available. In systems where passengers only tap-in, then the alighting stop of every trip leg must be estimated, as has been done for Santiago (Munizaga and Palma, 2012) and London (Gordon et al., 2013). However, this information is not enough to estimate the impact of changes in the transit system that could change the stop in which users start or end their trips (e.g. extending routes, relocating their stops, designing new routes or changing the fare structure). To study this kind of changes, a zonal O–D matrix that captures full trips (including access and egress) is needed.

Particularly, Santiago de Chile is a vast and unsteady city when it comes to its transport network, which could be benefited from better demand data. In 2007, the city went through a complete restructuration of its transport system, named Transantiago, which completely redesigned the network, changed the operators, and implemented new technologies, amongst other important changes (Muñoz and Gschwender, 2008). Additionally, Santiago's network changes have continued through time, with numerous service modifications and real-estate developments during this new Transantiago era.

In this line, this paper presents a methodology that allows us to estimate a zonal origin–destination matrix for a city, given a stop-to-stop trip matrix. Our proposed model infers a probability for the zone of origin and destination for each trip between two stops, using a Logit model. Model inputs are socio-economic, land use, and public transport information, for any city's period of interest. The methodology is then applied to estimate the model for Santiago de Chile's morning peak period.

The remainder of this paper is structured as follows. Section 2 provides the theoretical basis of our model, presenting its structure along with the proposed estimation procedure. Section 3 introduces and describes the available data for the case of Santiago de Chile, relating it to the zonal inference model when appropriate. The implementation of the proposed methodology to estimate and validate the model can be found on Section 4. Then, Section 5 presents the estimation results and discussion of the zonal inference model. Finally, in Section 6 we draw our conclusions, present future research guidelines, and provide policy applications of this study.

2. Theoretical model

As mentioned, the aim of this study is to develop a methodology for estimating a zonal O–D matrix for a public transport system, without the need of applying large scale origin–destination surveys and instead using passive information. This is achieved through the zonal inference model formulated in this Section, which assigns a probability for the zone of origin and destination for each entry in a stop-to-stop trip matrix. In other words, the model infers the access and egress legs for each trip between two public transport stops (which could correspond to different available public modes, such as bus or metro), considering socio-economic and land use information of the zones, as well as network information. Once the model is formulated, an estimation procedure using survey data and a Maximum Likelihood approach is proposed.

2.1. Zonal inference model

The proposed zonal inference model allows us to obtain the probability that an observed trip using public transport stops k and l (as their boarding and alighting points, respectively), was originated in zone i and had zone j as its destination. This probability is defined as $Prob(ij/kl)$.

In order to obtain these probabilities for each bus stop, metro station, and zone, we follow Daly (1982) and formulate this particular gravitational model:

$$T_{ij}^{kl} = A_i B_j f_{ij}^{kl} \quad \forall (i,j)_{\wedge}(k,l) \tag{1}$$

Subject to the constraints:

$$\sum_m \sum_n T_{mn}^{kl} = T_{kl} \quad \forall (k,l) \tag{2}$$

A_i and B_j are measures of attractiveness (for generating and attracting trips, respectively) of zone i and j, related to the population and land use information for these zones; T_{ij}^{kl} is the (unknown) number of trips made from zone i to zone j that use stops k and l for the initial boarding and final alighting, respectively; f_{ij}^{kl} is an inverse measure of the "cost" of choosing those stops when travelling from zone i to zone j; T_{kl} is defined as the observed trips between public transport stops k and l, which is assumed as known for every k and l (see Section 3 for more details); lastly, sets m and n are the zones of influence for stops k and l, respectively (i.e., zones within walking distance from each public transport stop).

This way, for a given trip using stops k and l (as their initial boarding and final alighting points, respectively), the probability that its origin was zone i and its destination zone j can be expressed as: