CASPT 2018 Extended Abstract

A METHODOLOGY FOR CORRECTING SMARTCARD TRIP MATRICES BY FARE EVASION

Luis-Angel Cantillo, Juan Carlos Muñoz, Sebastián Raveau, Paula Iglesias, Sebastián Tamblay.

Abstract It is increasingly common for public transport systems to incorporate Automatic Vehicle Location (AVL) and Automated Fare Collection (AFC) technologies as a source to improve their operation and planning. The information collected from these sources can be used to infer an Origin-Destination trip matrix which becomes a capstone for the system. However, trips or trip stages, evaded by users (not registered by AFC), are not included in these matrices, and therefore can hinder the proper planning and operation of the system. The present work presents a methodology for correcting the public transport travel matrix for evasion, starting from an estimated stop-based evasion rate, and information from an Origin-Destination survey and smartcard transactions. An optimization model is used to build the corrected OD matrix, which minimizes the error between observed evasion rates and evaded rate predicted at each stop, while mimicking the travel structure of the Origin-Destination survey and the registered transactions. Finally, a case study applied to the integrated public transport system of Santiago is presented, where the average tax evasion rate for 2017 in bus trip legs is estimated at close to 30%.

Keywords: Public Transport, Fare Evasion, O-D Matrices, Passive Data.

Luis-Angel Cantilllo
Pontificia Universidad Católica de Chile
Santiago, Chile
lacantillo@uc.cl

Juan Carlos Muñoz
Pontificia Universidad Católica de Chile
Santiago, Chile
jcm@ing.puc.cl

Sebastián Raveau
Pontificia Universidad Católica de Chile
Santiago, Chile
sraveau@ing.puc.cl

Paula Iglesias
DICTUC
Santiago, Chile
piglesias@dictuc.cl

Sebastián Tamblay
Pontificia Universidad Católica de Chile
Santiago, Chile
stm@ing.puc.cl

1 Introduction

Cards with AFC technology are a great opportunity to "track" public transport users and build matrices describing the trips between origin/destination pairs in the system. For systems where passengers tap-in their cards in every trip-leg, but are not requested to tap-out, building these matrices is harder because the destination of every trip-leg is not immediately available from the data. Trepanier et al (2007), Munizaga & Palma (2012) (refined in Devillaine et al., 2012, and Munizaga et al., 2014) and Gordon et al. (2013) elaborate methodologies to estimate the destination of each trip-leg, using the information AFC and AVL of the buses. This way, it is also possible to link trip-legs into full trips and infer the O-D matrix of validated trips in a public transport system. But, what happens with trips that evade their fare or enter without validating because they are in possession of a transit pass? In this paper we address the problem of correcting the origin destination matrices to incorporate an estimation of the trips being evaded. This process must recognize that these passengers may not choose their routes to reach their destinations in the same way that passengers paying their fare do (e.g. avoiding metro, where evading could be considerably harder).

In different transport systems throughout the world, due to their structure, operation and enforcement mechanisms, it is quite common that some passengers evade paying their fare when entering the system either in one or all legs of their trip. Gallegos et al. (2016) provide a methodology for correcting the origin destination travel matrix by fare evasion. The authors assume that they have not only the information from passive smartcards data, but also an OD survey for the city from where they can infer the true structure of the trips in the city. Thus, they assume that evading users choose their route in the same way as paying users. In our paper we propose a methodology that considers different route choices between both user types. This is especially relevant in systems where payment enforcement is very different in different parts of the system. For example in Santiago fare evasion is almost null in Metro stations, but in buses it has reached an average of over 30% citywide.

2 Problem Description and Model Formulation

2.1 Problem description

In the literature reviewed, there is consensus that there are different forms of evasion, which generate distinct types of distortion in the OD matrix (Gallegos et al., 2016). Delbosc & Currie (2016) categorize fare evaders for the city of Melbourne (Australia) as "deliberate" and "involuntary". The first type is characterized by being repeat offenders and presenting a lower risk aversion; while the second type corresponds users which are not systematic recidivists, usually evading under certain circumstances of personal or system negligence, especially when the form of payment is complicated.

In cases in which the public transport system has fare integration, many trips are composed of more than one leg. A traveller that effectively evades the fare must evade the validation of the card in every leg. Table 1 characterizes the “types of fare evasion” depending on the category of evader and the type of the trip.

Table 1 Types of evasion according to category of fare evader and trip.

Type of trip	User behavior	
	Deliberate evader	**circumstantial evader**
1 leg	Complete fare evasion	Complete fare evasion
2 or mor legs	Complete fare evasion	Parcial or complete fare evasion

Thus, a trip is classified as "Complete fare evasion" when none of its stages is validated by the traveller, that is, the entire trip is not registered in the transaction database of the smart cards. "Partial evasion" occurs when at least one of the legs of the trip is registered in the database, while some others are not. Partial evasion creates a distortion in the real origin or destination of the trip, and therefore in the resulting matrix based on smartcard data.

2.2 Model formulation

The proposed method constructs an O-D matrix of fare evaders, using an optimization model; so that the matrix obtained is consistent with zoning fare evasion rates, travel structure of the O-D survey and behaviour models (route choice) for users. Fig. 1 shows a schema of the matrix correction methodology, proposed in this research.

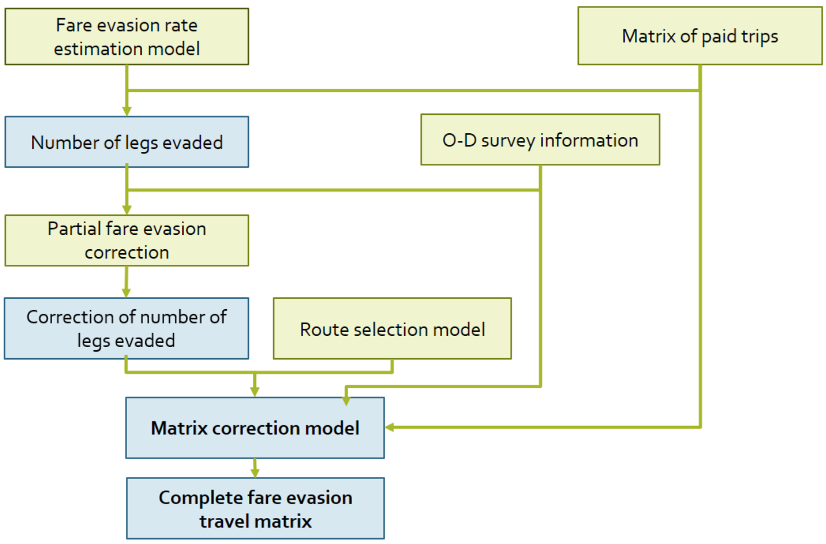

Fig. 1 Schema of the matrix correction methodology.

In general, the matrix correction model is an optimization model with a quadratic objective function and a linear domain. Its objective function is to minimize the error between the number of stages "observed" and those resulting from the model, for each zone. The above, defined a prior zoning.

The optimization model has a quadratic objective function: to minimize the difference between the number of stages "observed" and those resulting from the model, for each zone. The domain of the model is linear, thus the problem in convex and has a unique solution Each zone in the model can be associated to a stop or a group of stops.

The model considers the following notation.

Set indices:
i, j = Zones of origin and destination of the trip.
k = Bus stop.
r = Route.
or, d = Municipality of origin and destination of the trip.
g = Grouping of bus stops.

Parameters:
E_g = Trip stages evaded in the group's stops g.
P_{ij}^r = Ratio of bus trips between the pair of zones (i, j) that use the r route.
α_k^r = Binary parameter that takes the value of 1 if the bus stop k belongs to the group g.
τ_k^g = Parámetro binario que toma el valor de 1 si el paradero k pertenece al grupo g.
T_{ij}^{NE} = Number of bus trips, validated between the pair of zones ij.
NPE^{EOD} = Average number of bus trip stages.
T_{od}^{EOD} = Number of bus trips, between the municipalities o and d, of the O-D survey.
β_i^o = Binary parameter that takes the value of 1 if zone i is in the municipality o.
δ_j^d = Binary parameter that takes the value of 1 if zone j is in the municipality d.
ε, μ, ϕ, σ, γ, λ = Slack parameters of the constraints.

Variables:
$\overline{E_g}$ = Number of stages evaded in bus stops, which are part of group g, according to the built matrix.
T_{ij}^{E} = Number of evaders traveling between the par ij zones

The optimization model is:

$$\underset{T_{ij}^{E}}{MIN} \sum_{g} \left(E_g - \overline{E_g} \right)^2$$

$S.A.$

$$\overline{E_g} = \sum_{i,j,r,k} T_{ij}^{E} * P_{ij}^{r} * \alpha_k^{r} * \tau_k^{g} \qquad \forall g \qquad R1$$

$$NPE^{EOD} - \varepsilon \leq \frac{\sum_{g} \left(\overline{E_g} + \sum_{i,j,r,k} T_{ij}^{NE} * P_{ij}^{r} * \alpha_k^{r} * \tau_k^{g} \right)}{\sum_{i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right)} \leq NPE^{EOD} + \varepsilon \qquad R2$$

$$\frac{\sum_{d} T_{od}^{EOD}}{\sum_{o,d} T_{od}^{EOD}} - \phi \leq \frac{\sum_{d,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}}{\sum_{o,d,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}} \leq \frac{\sum_{d} T_{od}^{EOD}}{\sum_{o,d} T_{od}^{EOD}} + \phi \qquad \forall o \qquad R3$$

$$\frac{\sum_{o} T_{od}^{EOD}}{\sum_{o,d} T_{od}^{EOD}} - \sigma \leq \frac{\sum_{o,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}}{\sum_{o,d,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}} \leq \frac{\sum_{o} T_{od}^{EOD}}{\sum_{o,d} T_{od}^{EOD}} + \sigma \qquad \forall d \qquad R4$$

$$\frac{T_{od}^{EOD}}{\sum_{d} T_{od}^{EOD}} - \gamma \leq \frac{\sum_{i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}}{\sum_{d,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}} \leq \frac{T_{od}^{EOD}}{\sum_{d} T_{od}^{EOD}} + \gamma \qquad \forall o,d \qquad R5$$

$$\frac{T_{od}^{EOD}}{\sum_{o} T_{od}^{EOD}} - \lambda \leq \frac{\sum_{i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}}{\sum_{o,i,j} \left(T_{ij}^{E} + T_{ij}^{NE} \right) * \beta_i^{o} * \delta_j^{d}} \leq \frac{T_{od}^{EOD}}{\sum_{o} T_{od}^{EOD}} + \lambda \qquad \forall o,d \qquad R6$$

$$T_{ij}^{E} \geq 0 \qquad \forall i,j \qquad R7a$$

$$T_{ii}^{E} = 0 \qquad \forall i,j \qquad R7b$$

3 Experiments and results

3.1 Problem case study

The methodology proposed in this research was applied to Transantiago, the public transport system of Santiago (Chile). Due to the size of the city (seven million inhabitants), we grouped the over 11,000 stops into 1203 zones, triggering an optimization problem of 1,448,412 variables, subject to 5965 restraints.

As inputs for this model, we used the bus fare evasion measurement database, developed by the Metropolitan Public Transportation Board (DICTUC, 2012), the most recent O-D survey for Santiago, and the validated travel matrix resulting from the Munizaga & Palma (2012) methodology, built from smartcard transactions.
Finally, we incorporate a stochastic route choice model between pairs of zones, that was developed by Raveau et al (2017).

3.2 Results

Table 2 and Table 3 summarize the results from the case study. Interestingly, the results suggest that evaded trips have fewer legs than paid trips, which may be intuitive, but was not requested to the model.

Table 2 Results of fare evasion rates by macrozones

Macrozone	Leg fare evasion rate (observed)	Leg fare evasion rate (Modelled)	Trip fare evasion rate (Modelled)
Downtown	19.3%	26.8%	30.2%
North	28.5%	32.6%	39.5%
East	15.5%	21.3%	26.9%
West	26.9%	27.3%	34.8%
South	25.6%	26.0%	34.1%
Southeastern	27.3%	26.3%	34.7%
Total	**24.5%**	**26.5%**	**33.7%**

Table 3 Matrix of fare evasion rates between pair of macrozones.

Trip fare evasión rate	**Downtown**	**North**	**East**	**West**	**South**	**Southeastern**	**Total**
Downtown	8.3%	13.4%	37.1%	38.0%	63.1%	48.7%	**30.2%**
North	5.7%	47.1%	10.9%	44.8%	67.6%	84.3%	**39.5%**
East	9.1%	31.4%	25.3%	51.2%	32.8%	49.2%	**26.9%**
West	14.2%	40.6%	17.6%	40.5%	47.7%	82.6%	**34.8%**
South	10.1%	51.5%	12.1%	38.2%	40.7%	45.8%	**34.1%**
Southeastern	27.2%	61.4%	20.1%	37.6%	31.8%	40.3%	**34.7%**
Total	**12.4%**	**42.8%**	**23.5%**	**41.0%**	**42.9%**	**47.0%**	**33.7%**

References

- Delbosc, A., Currie, G. (2016) Cluster analysis of fare evasion behaviours in Melbourne, Australia. Transport Policy 50, 29-36.
- Devillaine, F., Munizaga, M., And Trepanier, M. (2012). Detection of activities of public transport users by analyzing smart card data. Transportation Research Record: Journal of the Transportation Research Board, 2276 (1), 48-55.

- DICTUC (2012) Mediciones de perfil de carga con evasión en la ciudad de Santiago. Estudio realizado por DICTUC para la Coordinación Transantiago.
- Eliason, S. R. (1993). Maximum likelihood estimation: Logic and practice (Vol. 96). Sage Publications.
- Gallegos, N. (2016) Construcción de factores de corrección por evasión para la matriz origen-destino de viajes en transporte público en Santiago obtenida a partir de transacciones Bip!, Tesis para optar al grado de Magíster en Ciencias de la Ingeniería, mención Transporte, Universidad de Chile.
- Gordon, J.B., H.N. Koutsopoulos, N.H.M. Wilson, and J.P. Attanucci. "Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data." Transportation Research Record: Journal of the Transportation Research Board, No. 2343, pp 17-24
- Ipsos (2013) Encuesta Origen Destino de viajes 2013. Realizado por Ipsos para Metro S.A.
- Munizaga, M., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies, 24, 9-18.
- M Trépanier, N Tranchant, R Chapleau (2007) Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems 11 (1), 1-14.
- Munizaga, M., Devillaine, F., Navarrete, C. And Silva, D. (2014). Validating travel behavior estimated from smartcard data. Transportation Research Part C: Emerging Technologies, 44, 70-79.
- Raveau, S., Muñoz, J.C., Prato, C., Soto, A., Tamblay, S., & Iglesias, P. (2017). A Behavioural Planning Tool for Modelling Public Transport Systems. Transit Data 2017.
- Troncoso, R., & de Grange, L. (2017). Fare evasion in public transport: A time series approach. Transportation Research Part A: Policy and Practice, 100, 311-318.