Do passengers consider common lines? Observing and understanding route choice behaviour of public transport passengers from smart card data

Jacqueline Arriagada*[a] , Marcela Munizaga [a], Angelo Guevara [a] , Carlo G. Prato [b]

[a] División Ingeniería de Transporte, Departamento de Ingeniería Civil, Universidad de Chile, Blanco Encalada 2002, Santiago, Chile

[b] School of Civil Engineering, The University of Queensland
St Lucia 4072, Brisbane, Australia

Understanding route choice in public transport systems is relevant to improve their design via the understanding and prediction of the preferences of passengers when choosing their routes. A typical approach to route choice modelling consists of a two-stage approach: (i) the definition of the consideration set, namely a subset of attractive alternatives for the individual, and (ii) the modelling of the route choice given the consideration set (Prato, 2009). The identification of the alternatives in the consideration set requires the definition of alternative routes. According to De Cea & Fernández (1993), an alternative route is any path that a transit user can use to reach the destination and it can be identified by a sequence of nodes or route sections, where the latter is a route portion between two consecutive nodes. Noticeably, existing route choice models do not consider whether multiple lines can be treated as belonging to a unique alternative. In this study, we propose a model able to identify whether multiple attractive lines are indeed considered as the same alternative route by the traveller.

Several studies, mostly in the transit assignment area, assume that a passenger travelling over a route will consider a set of attractive lines at each transit stop and will board the first bus belonging to that set. This attractive set is defined as common lines (Chiriqui & Robillard, 1975), which is defined as the group of lines that minimizes the total expected travel time. Accordingly, it seems plausible to treat the alternatives belonging to common lines as a unique one (Spiess & Florian, 1989). In our study, we examine whether passengers are using common lines and we develop and test a route choice model in public transport by considering common lines usage.

For this study, we use automatic fare collection Smart Card (SC) data from the public transport system in Santiago, Chile. The data comprise SC data from 20 working days on morning peak periods for passengers who travelled for 15 or more days. Accordingly, these are panel data with multiple choice situations for the same passenger. For the analysis of common lines, we identified 66,249 route sections and on which we used the definition of common lines in three ways: (i) the standard definition (Chriqui & Robillard, 1975), where travel time and waiting time are assumed to have the same value, (ii) adding a waiting time penalty equal to two, and (iii) adding a waiting time penalty equal to three. We explored the last two cases because there is evidence in the public transport route choice literature that the rates of substitution between travel time and waiting time is different from 1, varying between 0.9 and 2.5. In the observed network, there is a large percentage of route sections where all the lines that serve it are common lines. The exact figure will vary depending on the consideration of waiting time to define common lines: (i) 97.6%, (ii) 99.3% and (iii) 99.7%. Regarding users' behaviour, more than 99% of the trips were made on a line within common lines and less than 0.6% of trips were made on a line that does not minimize the expected travel time (possibly because of the destination or the high waiting time). This means that passengers are mainly traveling in lines that minimize travel time and it is necessary to identify a waiting time penalty in the definition of common lines.

In order to analyze whether significant differences exist between observed and expected trip distributions according to Spiess & Florian (1989), we performed the Pearson's chi-squared test and found that passengers were using common lines as a unique alternative in 80% of the route sections, which in turn means that they were not using common lines as a unique alternative in 20% of the route sections. We also calculated the coefficient of concentration (Glasser, 1962) that measures the stickiness of each passenger to a specific route. It is expected that, in the presence of common lines, passengers do not stick to a specific route as they perceive different routes as the same alternative. We found that 10% of passengers stick to one route even when there are common lines.

Our data analysis suggested the necessity of accounting for the consideration of common lines within a route choice model. We proposed a latent class model to capture the heterogeneity between passengers regarding the consideration of common lines. The class membership probability is modelled by using the Bernoulli distribution, where the probability of belonging to a class is based on an indicator of the differences between observed and expected trip distributions for each passenger. The classes' route choices are modelled by using two Path Size Correction Logit models, with one class having consideration sets where common lines are part of the same unique alternative and the second-class having consideration sets where each elemental route is an alternative. For both models, the consideration sets were formed by combining historical choices observed for 50,921 origin-destination pair of zones, where the origin and destination zones were defined as a 300-meter radio area around the origin and destination stop.

Model estimates show that the two class models capture the effect of travel time, waiting time, walking time and number of transfers with the expected negative signs, and the path size terms capture the overlap between routes with the expected positive sign, which is consistent with the theoretical derivation of the path size correction. The rates of substitution suggest that passengers who consider aggregate alternatives have higher waiting time, walking time and transfer penalties (in particular for bus to bus or metro to bus) than passengers who consider elemental alternatives.

References

Chriqui, C., & Robillard, P. (1975). Common Bus Lines. *Transportation Science*, *9*(2), 115–121.

De Cea, J., & Fernández, E. (1993). Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. *Transportation Science*, *27*(2), 133–147.

Glasser, G. J. (1962). *Variance Formulas for the Mean Difference and Coefficient of Concentration*. *Journal of the American Statistical Association,* 57, 648-654.

Nassir, N., & Hickman, M. (2017). Statistical Inference of Transit Passenger Boarding Strategies from Farecard Data. *Transportation Research Record*, *2652*, 1–22.

Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100.

Spiess, H., & Florian, M. (1989). Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, *23*(2), 83–102.