Transportation Research Part B 95 (2017) 105–125

Contents lists available at ScienceDirect

Transportation Research Part B

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

Crowding cost estimation with large scale smart card and vehicle location data

Daniel Hörcher*, Daniel J. Graham, Richard J. Anderson

Railway and Transport Strategy Centre, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom

ARTICLE INFO

Article history:
Received 8 June 2016
Revised 27 October 2016
Accepted 28 October 2016
Available online 9 November 2016

Keywords:
Public transport
Crowding
Revealed preference
Smart card data
AVL data

ABSTRACT

Crowding discomfort is an external cost of public transport trips imposed on fellow passengers that has to be measured in order to derive optimal supply-side decisions. This paper presents a comprehensive method to estimate the user cost of crowding in terms of the equivalent travel time loss, in a revealed preference route choice framework. Using automated demand and train location data we control for fluctuations in crowding conditions on the entire length of a metro journey, including variations in the density of standing passengers and the probability of finding a seat. The estimated standing penalty is 26.5% of the uncrowded value of in-vehicle travel time. An additional passenger per square metre on average adds 11.9% to the travel time multiplier. These results are in line with earlier revealed preference values, and suggest that stated choice methods may overestimate the user cost of crowding. As a side-product, and an important input of the route choice analysis, we derive a novel passenger-to-train assignment method to recover the daily crowding and standing probability pattern in the metro network.

1. Introduction

Public transport services are congestible: as the number of users increases relative to the available capacity, the attractiveness of the service drops and customers become less willing to use it or pay the same fare. This reduction in willingness to pay can be considered as a user cost, just like the value of time that passengers have to allocate for travelling. However, the cost of crowding is much more difficult to measure than other consumption externalities in transport, e.g. the time lost in road congestion. In the latter case there is a physical relationship between the density of traffic and the average speed of vehicles. In public transport travel time is almost unaffected by the density of passengers[1]. The user cost of crowding appears in the form of discomfort caused by the physical proximity of fellow travellers[2], less personal space and limited access to certain amenities of the vehicle, such as preferred seats, fresh air, handle bars, or quick access to doors.

The main planning and policy areas where crowding cost measurements are utilised are demand modelling, investment appraisal as well as service quality and tariff optimisation. Considering that the cost of discomfort in very high crowding may reach the uncrowded travel time cost of the same trip (Whelan and Crockett, 2009), congestion has been shown to be

* Corresponding author.
E-mail address: d.horcher@imperial.ac.uk (D. Hörcher).

[1] Increased boarding and alighting times and other congestion-related delay factors are partly or fully compensated by shorter headways and less waiting time in peak periods. The actual travel time that passengers experience may remain independent of crowding, especially in case of high capacity urban rail systems.

[2] For a detailed review on the psychological aspects of crowding discomfort see Evans and Wener (2007) and Thomas (2009).

http://dx.doi.org/10.1016/j.trb.2016.10.015

an important factor of demand, and therefore congestion-relieving investments do provide important welfare benefits for society (Cats et al., 2016; Haywood and Koning, 2015; Prud'homme et al., 2012). Crowding costs have important implications on the optimality of supply-side decisions too (de Palma et al., 2015; Tirachini et al., 2013). As crowding is a consumption externality, the optimal fare should reflect the marginal external cost that an additional trip imposes on fellow passengers, when capacity cannot be adjusted. Thus, adapting methods originally developed for road congestion pricing in public transport is clearly a relevant challenge on the research agenda. Given the rising interest in understanding crowding-related problems in public transport planning and policy, it is crucial to improve the empirical methods that we apply to quantify the discomfort that crowding causes.

Decades after the appearance of the first speed-flow functions, advances in discrete choice modelling finally made the quantitative measurement of crowding disutilities possible. Wardman and Whelan (2011) and Li and Hensher (2011) provide comprehensive reviews of the evolution of crowding cost estimation. The vast majority of crowding disutility measurements were performed using stated preference (SP) techniques. Revealed preference experiments are less widely used by researchers,[3] mainly because it is difficult to collect data and control for all travel attributes in real choice situations, or there is insufficient variation between alternatives in crowding levels. Our goal is to illustrate that surveying and SP methods are no longer needed to measure the cost of crowding when such data are not available, and a comprehensive revealed preference experiment can be conducted instead using automated fare collection (AFC) and vehicle location (AVL) data.

The revealed preference approach that we propose here provides a number of advantages compared to traditional SP methods. Most importantly, data collection is cheaper, faster, more reliable and reproducible. Also, limitations in sample size can be relaxed significantly. Choices are realistic in the sense that we observe how passengers behave in their everyday life, under ordinary conditions. Additionally, the in-vehicle travel environment often varies during a public transport journey – in our RP framework these variations can be controlled for, while in a survey it is hard to describe and explain them for respondents. In general, the range of retrievable choice attributes is wider, given the rich information content of AFC and related datasets.

However, to enjoy these advantages we have to cope with multiple challenges:

1. We have to recover passengers' in-vehicle crowding experience. This can be solved by merging AFC and AVL data in a passenger-to-train assignment process.
2. The actual route chosen is not recorded in the data, it has to be inferred. We present two alternative methods to identify route choices.
3. Crowding varies across different segments (links or inter-station legs) of alternative routes. On the other hand, what we observe is a choice between two routes, i.e. two sets of links. Therefore we have to find a way to aggregate non-additive link-level attributes in the choice model.
4. In reality passengers are not fully informed about travel conditions on alternative routes, especially not prior to the trip; their decisions are based on expectations, that we have to model with reasonable assumptions.

We address all these challenges in the paper. The ultimate goal of the experiment is to estimate a crowding-dependent travel time multiplier as a function of crowding density and the probability of standing.

The paper is structured as follows. Section 2 reviews the earlier literature of crowding-related empirical analyses. We devote separate sub-sections for passenger-to-train assignment techniques, as well as the evolution of stated preference crowding cost estimation models and recent contributions with revealed preference experiments. Section 3 is the backbone of this paper: after a data description we present the passenger-to-train assignment methodology, which is a key input for the subsequent behavioural analysis. Section 3.3 details the discrete choice model, and the way in which route choices and route attributes are inferred from the data. Finally, Section 4 delivers the estimation results and Section 5 concludes.

2. Previous literature

The main contribution of this paper is a revealed preference crowding cost estimation experiment. However, in order to analyse customers' decisions we have to recover the travel conditions that they experienced and expected to experience on competing routes – most importantly, the density of in-vehicle crowding. Therefore the passenger-to-train assignment method, discussed in Section 3.2, is a crucial preparatory part of the analysis. In Section 2.1 we review previous attempts to merge smart card and metro vehicle location datasets, while Section 2.2 investigates the literature of crowing cost estimation.

2.1. Inferring in-vehicle crowding

Smart card data (SCD) refers to the information retrieved from automated fare collection systems widely used in major urban public transport networks. Electronic ticketing gained popularity because of its obvious advantages versus paper tickets in the form of a faster, cheaper, safer, more transparent and more comfortable mean of tariff enforcement. However, as Bagchi and White (2005) envisioned in an early contribution, SCD is an important information source for travel behaviour

[3] Some of the innovative exceptions include Kroes et al. (2013) and Tirachini et al. (2016).