Transportation Research Part F 58 (2018) 500–510

Contents lists available at ScienceDirect

Transportation Research Part F

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

Modelling passenger waiting time using large-scale automatic fare collection data: An Australian case study

Ahmad Tavassoli [a,*], Mahmoud Mesbah [a,b], Ameneh Shobeirinejad [c]

[a] *School of Civil Engineering, The University of Queensland, Brisbane, Australia*
[b] *Department of Civil and Environmental Engineering, Amirkabir University, Tehran, Iran*
[c] *School of Information Communication Technology, Griffith University, Brisbane, Australia*

ARTICLE INFO

Article history:
Received 7 October 2017
Received in revised form 5 May 2018
Accepted 25 June 2018

ABSTRACT

Passenger waiting time at transit stops is an important component of overall travel time and is perceived to be less desirable than in-vehicle travel time or access time. Therefore, an accurate model to estimate waiting time is necessary to better plan for transit and to improve patronage. The majority of previous studies on waiting time have either made very limiting assumptions on the arrival distribution of passengers or lacked a large-scale and high-quality dataset. The smartcard fare collection system in South-East Queensland, Australia, has provided the opportunity of very large-scale and highly accurate data on passenger boarding and alighting times and locations. In this research, all 130,000 daily rail passengers in all 145 stations of a network are considered. First a methodology is developed to match each individual passenger with the most likely rail service he/she boarded. Then, a hazard-based duration modelling approach is adapted to model passenger waiting time as a function of a variety of factors that influence waiting time. Log-logistic accelerated failure time (AFT) models are inferred to be appropriate among the models tested. The results indicate that: (a) the waiting time can be predicted accurately at various confidence levels; (b) the waiting time at all network stations can be predicted with a single model; and (c) a wide range of influencing parameters are statistically significant in the model, which can be categorized to temporal, infrastructure and operation, demographics, and trip characteristics parameters. The results of this study can be used for demand estimation, operational analysis, transit scheduling, and network design through an understanding of the effects of influential variables on waiting time.

1. Introduction

Increasing the share of transit is the objective of many transport authorities as a means to reduce traffic congestion, emissions, and increase transport systems' efficiency. To this aim, a deep understanding of the components of this mode is essential. Considering a transit journey, waiting time is perceived more onerous than in-vehicle travel time and access time (Wardman, 2004; Ceder, 2016). Waiting time has been studied extensively in the literature and is used in many transit models as a key variable affecting travel time, travel time reliability, and use of transit from a passenger's perspective (Cats and Gkioulou, 2014). The waiting time models in the literature mainly identify two extremes for the spectrum of passengers.

* Corresponding author at: School of Civil Engineering, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia.

E-mail address: a.tavassoli@uq.edu.au (A. Tavassoli).

https://doi.org/10.1016/j.trf.2018.06.037
1369-8478/

When the headways are long, the majority of passengers use the schedule to minimize their waiting time, however, depending on the service reliability, they leave a margin to arrive at the station/stop before the scheduled service time. When the headways are short, most passengers arrive randomly as they can catch a service at any time. At any given station/stop there is a mixture of both groups of passengers. This fact makes the prediction of waiting time challenging.

The literature is reviewed in greater detail in the next section. Despite the importance of waiting time, lack of large-scale and high-quality data has limited the previous models to accurately estimate waiting time. The theoretical models reviewed in the next section were developed with few specific assumptions or the empirical models were calibrated with small or incomplete datasets. With the availability of the smartcard fare payment, a new horizon is now available to develop robust models validated by high-quality data on a large-scale. A unique dataset from TransLink, the transit authority in South-East Queensland (SEQ), Australia, is used in this study. The smartcard fare payment system (called GoCard), requires passengers of all types to 'touch on' when entering a station or boarding a transit vehicle as well as 'touch off' when exiting a station or alighting a transit vehicle. Because of the strong fare incentives, close to 85% of transit passengers use a GoCard. In this study, all transactions of all rail passengers across the network during a complete day are included for analysis. This is equal to approximately 130,000 passengers in 145 stations during one day.

This study aims to model the passenger waiting time using the above large-scale database. After testing various theoretical distributions, a hazard-based duration model is proposed to estimate the waiting time. The prediction power of the model is tested by empirical observations. Furthermore, the effect of a wide range of independent variables is tested in the model.

2. Literature review

Waiting time has been identified as a significant parameter since the 1970 s. Huddart (1973) study, based on data collected in London, UK, showed that passengers arrive randomly when the headway is short and target a specific service when the headways are long. Ceder and Marguier (1985) derived waiting time distribution based on the variation of some assumptions, such as no queuing to board a transit vehicle and a constant rate of passenger arrivals. In addition to headway, waiting time is also affected by a range of other variables. Many studies have investigated the effect of reliability on passenger waiting time (Turnquist, 1978; Bowman and Turnquist, 1981; Currie and Csikos, 2007; Luethi et al., 2007). Frumin and Zhao (2012) focused on stops with heterogeneous services (more than one service pattern). The proposed study of this research considers this factor as the GoCard data source is rich enough to identify the destination and therefore, the study can include services connecting the origin and destination in headway calculations. The effect of weather on waiting time has been investigated by Lam, Lam, Morrall, and Morrall (1982) and Mishalani, Mccord, and Wirtz (2006). Cats and Gkioulou (2014) examined the effect of information and reliability on waiting time.

In terms of the scale of data used, usually smaller scale studies have been undertaken on bus and light rail systems in which an interview or direct observation was used (Salek and Machemehl, 1999; Zahir, Matsui, & Fujita, 2000; Fan and Machemehl, 2009; Iseki and Taylor, 2010; Psarros, Kepaptsoglou, & Karlaftis, 2011; Bouzir et al., 2014). Larger-scale studies such as those in Brisbane (Tavassoli, Mesbah, & Hickman, 2017), London (Frumin and Zhao, 2012) and Melbourne (Csikos and Currie, 2008) have used a heavy rail data source. Note, that in the former study only a part of the network (London Overground) was included and in the latter, the arrival distribution was focused as a time series rather than the waiting time itself.

A hazard-based duration modelling approach is suitable for dealing with duration data that are positive and can be censored and time-varying (Bhat and Pinjar, 2008). This hazard-based approach is common in many disciplines including biomedical, social sciences, and engineering (Hensher and Mannering, 1994). In the transport field, this method has been applied over the last two decades in modelling of sometime-related events including safety, traffic studies, vehicle ownership, and activity based models. Examples include time between planning and execution of an activity (Bhat and Pinjar, 2008), residential location (Rashidi, Auld, & Mohammadian, 2012), duration of shopping activity (Bhat, 1996), length of traffic delay (Mannering, Kim, Barfield, & Ng, 1994), incident duration (Tavassoli Hojati, Ferreira, Washington, Charles, & Shobeirinejad, 2014), the analysis of urban travel time (Anastasopoulos et al., 2012) and congestion duration for rail passenger flow (Shi et al., 2016).

Many of the previous studies on waiting time estimation cannot be generalized to other cases because the research is based on: (1) small sample size of up to several hundred passengers; (2) incomplete or poor-quality waiting time data; and (3) even though there are a few proposed models with rather a high efficiency, their applicability to other locations and times are not verified for consistency. This means some of the variables could be significant in one location but not significant at another location.

Therefore, an analysis of passenger waiting time on a large-scale is carried out here to validate the factors arising in the literature, to identify other potential factors that might influence the waiting time, and to better understand patronage behavior in the Australian context.

3. Model development

Since waiting time varies from one passenger to another, a probabilistic method can be utilized to model waiting time. The results of probabilistic methods provide useful insights for selecting an appropriate strategy in transit operations and