Transportation Research Part B 71 (2015) 85–99

Contents lists available at ScienceDirect

Transportation Research Part B

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

Determining optimal frequency and vehicle capacity for public transit routes: A generalized newsvendor model

Avi Herbon, Yuval Hadas *

Department of Management, Bar-Ilan University, Israel

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form 21 October 2014
Accepted 22 October 2014
Available online 17 November 2014

Keywords:
Public transit
Frequency
Vehicle size
Non-linear optimization
Newsvendor model

ABSTRACT

The level of service on public transit routes is very much affected by the frequency and vehicle capacity. The combined values of these variables contribute to the costs associated with route operations as well as the costs associated with passenger comfort, such as waiting and overcrowding. The new approach to the problem that we introduce combines both passenger and operator costs within a generalized newsvendor model. From the passenger perspective, waiting and overcrowding costs are used; from the operator's perspective, the costs are related to vehicle size, empty seats, and lost sales. Maximal passenger average waiting time as well as maximal vehicle capacity are considered as constraints that are imposed by the regulator to assure a minimal public transit service level or in order to comply with other regulatory considerations. The advantages of the newsvendor model are that (a) costs are treated as shortages (overcrowding) and surpluses (empty seats); (b) the model presents simultaneous optimal results for both frequency and vehicle size; (c) an efficient and fast algorithm is developed; and (d) the model assumes stochastic demand, and is not restricted to a specific distribution. We demonstrate the usefulness of the model through a case study and sensitivity analysis.

1. Introduction

Public transit is a capital-intensive system that operates within a dynamic environment. One of the key factors in this environment is passenger demand that is basically uncertain and depends on spatial and temporal attributes such as the location of the stations along the route and service frequency during the day. The awareness of public transit passengers for higher service standards are increasing. Characterized by adequate timetables, high frequencies and available seats, this awareness make the problem of determining route frequencies and vehicle capacity an appealing problem both to researchers and practitioners. Analyzing the monetary properties (resulting from frequencies and vehicle capacity) of the demand side (Jansson, 1980;Mohring, 1972) is an important issue that provides the authorities and operators with the means to better design competitive fare structures.

A common practice in public transit planning is to determine the frequency of service based on accumulated hourly passenger counts; average travel time; vehicle capacity; desired occupancy (load standard); and the minimum frequency permitted according to the time of day (Ceder, 2007). Criteria for quality of public transport can also include vehicle and transfer comfort at the terminal; regularity of service; service coverage; frequency level; and crowding (Pel et al., 2014;Tirachini et al., 2013; Wardman and Whelan, 2011). From the operator's perspective, on the other hand, the objective is to make as

---

* Corresponding author.
E-mail addresses: avher@bezeqint.net (A. Herbon), yuval.hadas@biu.ac.il (Y. Hadas).

http://dx.doi.org/10.1016/j.trb.2014.10.007
0191-2615/© 2014 Elsevier Ltd. All rights reserved.

much profit as possible (Guihaire and Hao, 2008). While the perspectives may seem to differ, decisions taken by the operator that do not consider consumer criteria will deteriorate profits due to competition and possible decline of passenger demand. In this paper, we consider the operator's monetary perspective as the focal objective. In order to realize his monetary perspective, the operator must achieve appropriate service levels and frequencies. A suitable frequency setting should provide sufficiently regular service to satisfy the users and sufficiently sparse service to reduce the required departures and thereby the operator's costs (Guihaire and Hao, 2008).

The emergence of automated identification technology can assist in coping with the above problem. Automatic vehicle location (AVL) and automatic passenger counting (APC) are well-known technologies that can track the location of vehicles en route and collect the number of alighting\boarding passengers at each stop (Tétreault and El-Geneidy, 2010). According to Hadas and Shnaiderman (2012), the data acquired can be used to analyze as well as to enhance the performance of public transit systems and to support the development of advanced models. Using AVL technology, it is possible to accurately forecast the estimated arrival times of buses and to implement bus-holding strategies to coordinate transfers (Dessouky et al., 2003). Strathman et al. (2002) analyzed the operating performance for Tri-Met, the transit provider for the metropolitan area of Portland, Oregon. They demonstrated how AVL–APC data can be utilized to monitor and evaluate service performance in relation to adopted standards. Tétreault and El-Geneidy (2010) used AVL and APC data to select stops and estimate run times for a new service that will run parallel to a heavily used bus route in Montreal, Canada.

According to Desaulniers and Hickman (2007) summary of PT planning optimization models, the problem of setting frequencies can be approached both at the network level and for a single route. At the route level, the primary goal is to select frequencies that maximize passenger service subject to various constraints, such as overall fleet size, route capacity in relation to demand and policies regarding minimum desirable frequencies.

The transit network frequency setting problem (TNFSP) was mathematically addressed by Salzborn (1972). His work was based on determining frequencies given passenger arrival rates with the objective of decreasing fleet size (peak period) and passenger waiting time (off-peak period). The model assumes deterministic travel time and demand. Furthermore, bus capacity (overcrowding) is not a constraint. Schéele (1980) proposed a nonlinear model (solved by a non-optimal algorithm) for bus network vehicle allocation and frequency setting by minimizing the total travel time given passenger trip assignment. The model assumes deterministic demand and overcrowding is not allowed. Another mathematical approach is suggested by Furth and Wilson (1982). They used a relaxation of a non-linear model for maximizing ridership benefit and waiting time saving benefit under bounded fleet size, headway and budget. Gao et al. (2004) proposed a heuristic solution based on sensitivity analysis designed to optimize frequencies settings. The algorithm is designed to help transit planners to adjust an existing transit network to evolutions in demand and various other parameters. Constantin and Florian (1995) developed a non-linear, non-convex mixed integer programming model for the TNFSP with the goal of minimizing the total expected travel and waiting time of passengers under fleet size constraints. The TNFSP was also addressed heuristically. Han and Wilson (1982) considered the problem as one of allocating vehicles among the routes of the network. They proposed a two-stage heuristic to reach their objective, minimizing the maximum "occupancy level" at the maximum load point for each route. Chowdhury and I-Jy Chien (2002) considered transfer coordination for intermodal transit networks by optimizing both headways and slack times. A mathematical programming model is first developed and then a procedure is presented that first optimizes headways without taking coordination into account. It then optimizes slack times in the context of intermodal transit.

All the above mentioned TNFSP models share several properties. First, the objective functions are time base. Secondly, stochastic demand is not considered, hence overcrowding (if addressed) does not lead to fail to board passengers. Finally, operators under-utilizing their vehicle is not considered.

Another approach for setting frequencies stems from microeconomics. This is a broader approach since it considers in detail the costs of both passengers and operators.

Mohring (1972) developed a model that considers optimal frequency and stop spacing as a means of minimizing bus company operating costs and user costs (walking, waiting, and riding). The model assumes deterministic demand and does not takes into account overcrowding (bus capacity is realized from frequency and demand). Jansson (1980) further investigated the above mentioned model (the square root formula) for peak and off-peak periods, each having different deterministic demand patterns. An optimal frequency and bus size model was introduced as well. More recently, Jara-Dfaz and Geschwender (2003) extensively reviewed the evolvement over the years of the models that Mohring and Jansson developed. Specifically, they investigated the effect of vehicle size on operating costs and passenger crowding that resulted from delayed boarding. The model assumes deterministic demand that is uniformly distributed. Jara-Díaz et al. (2008) address the goal of maximizing social benefits (considering both users and operators) in the case of inelastic demand and where the cost function includes waiting time and travel time. They found the optimal levels of frequency and vehicle size on a public transport corridor. This model, like that of Jara-Dfaz and Geschwender, does not treat overcrowding. A variation of the model developed by Delle Site and Filippi (1998) adds operational tactics in the form of short-turns in addition to setting vehicle size and frequencies. This model also assumes that all passengers are able to board. A better realization of overcrowding was introduced by Oldfield and Bly (1988) in the form of an infinite waiting time component. Tirachini and Hensher (2011) increased bus capacity (spare capacity) to absorb random variation in demand. From the above it is clear that a more detailed analysis of the effect of passenger demand on capacity should be sought. When demand is lower than capacity (crowding), there is a more positive effect on passengers since boarding time decreases (Jara-Dfaz and Geschwender, 2003) and perceived time decreases (Wardman and Whelan, 2011), (Pel et al., 2014). Furthermore, crowdedness has an effect on wellbeing,