SMART CARD DATA AND PLANNING FOR PUBLIC TRANSPORT

A case study from South East Queensland, Australia

Barbara T.H. Yen

Introduction

Brisbane, Australia, has a large multimodal public transport system. Like other cities in the world, Brisbane uses a smart card fare system. Public transport smart card data are widely used to measure passengers’ travel patterns, such as intermodal transfer behaviour (Bagchi & White, 2004; Jang, 2010; Seaborn et al., 2009; Yen et al., 2018b), mode choice (Chakirov & Erath, 2011; Yen et al., 2018a), passengers’ waiting time (Jang, 2010; Ingvardson et al., 2018) and travel pattern variability (Morency et al., 2007). However, only a few studies pay attention to public transport service supply or the interaction between public transport service supply and demand. Trépanier et al. (2009) measured public transport performance by using smart card data to produce operational indicators, such as network supply indicators (e.g., vehicle-kilometres) and passenger service indicators (e.g., average trip length). Eom et al. (2015) used data mining techniques on smart card data in Seoul to analyse the performance of public transport services. With their developed data mining logic, they identified metro routes which had low levels of service (LOSs).

In most studies, passengers’ travel behaviour, such as arriving time at a station, has been viewed as static behaviour. However, passengers’ travel behaviour is heavily dependent on public transport reliability and travel information (e.g., the timetable) (Cats & Gkioulou, 2017). Thus, the research question addressed by this case study chapter is: How do passengers behave when there is a fixed timetable? Do passengers coordinate their arrivals to coincide with the scheduled service so as to minimise the waiting time or do passengers just turn up and wait for the service? It is widely assumed that passengers’ expected waiting time is half the service headway under the assumption of the random arrival of passengers (Fan & Machemehl, 2002) and uniform headways (Bowman & Turnquist, 1981). However, longer headways might gradually change passengers’ arrival pattern from random to coordinated passenger arrivals or “aware” arrivals, because passengers would have motivations to learn the schedule (Bowman & Turnquist, 1981). On the other hand, if public transport services are frequent enough, passengers might have “turn up and go” behaviour, since they no longer need to match their arrival time with the timetable (Clifton
et al., 2018). This chapter examines the evidence to understand how passengers behave when faced with a fixed timetable of service.

In order to investigate the impact of public transport service supply on passengers’ arrival time and thus waiting time patterns, this chapter evaluates passengers’ waiting time patterns against public transport service frequency (i.e., headway) using South East Queensland (SEQ), Australia, as the case study. SEQ has a zone-based fare system with no penalty for a multimodal transfer within the same zone, although a full fee is charged for an interzone transfer within a single journey (Yen et al., 2018a). Passengers can use one smart card, the Go-card, for all public transport modes, including heavy rail, bus, ferry and light rail. Unlike many highly populated urban areas but consistent with Australian experience, SEQ is a car-dominated area, with 76% private vehicle mode share (Department of Transport and Main Roads, 2018). Exploring public transport users’ behaviour, such as waiting time, in this chapter has operational and policy implications for public transport systems operating in low to medium population density areas. To simplify the question, this chapter focuses on heavy rail passengers only because the data from smart cards is more robust in this context.

The chapter is structured as follows. The next section describes the SEQ transport study area and summarises the heavy rail system which is the context of this case study chapter. Then, the methodology and data used are described. The penultimate section provides the results together with interpretation. The chapter concludes with a discussion and suggestions for further research on using smart card data for service planning.

**Case study area**

*The South East Queensland rail system*

This section presents a brief outline of the rail system in SEQ. The SEQ region of Australia, which includes Brisbane, the Sunshine Coast and the Gold Coast, has merged into a 200-kilometre-long city (Spearritt, 2009). SEQ’s public transport system is a network of heavy rail, light rail, buses and ferries. Figure 36.1 shows the rail network, including both heavy rail and light rail. For a larger multimodal public transport network in SEQ, transfers during journeys are inevitable but, because of the zone-based fare system, there are no transfer penalties across all public transport modes.

SEQ’s heavy rail network of over 200 kilometres connects the Sunshine Coast and Gold Coast to Brisbane. The heavy rail network is radial and centre oriented, with 11 lines and 154 heavy rail stations. The heavy rail network has relatively low ridership due to the low service frequency, as discussed subsequently.

*South East Queensland smart card data: Go-card data*

Go-card is a region-wide smart card which implements a zone-based fare scheme in which passengers use one card for all public transport modes. Fares are automatically collected based on the zones travelled. All Go-card passengers must tag on when boarding and tag off when alighting from buses or ferries or when entering and leaving a heavy rail or light rail station. Go-card readers are installed on board vehicles or on platforms. Go-card data provide origin and destination data records for each transaction or tap (Soltani et al., 2015). SEQ has four types of Go-cards: adult, child, senior and concession – the latter three types have a 50% fare reduction as compared to the adult card. Due to privacy issues, the Go-card transaction data provided by TransLink for analysis do not identify card types.
Figure 36.1  SEQ’s rail network

Source: TransLink, 2020
This study obtained travel behaviour by Go cardholders with spatial, temporal, operational and trip characteristics of the passenger trips and journeys. As Brisbane has a simple rail network, only the busiest heavy rail lines are selected for analysis: Beenleigh/Ferny Grove Line (north-west to south-east line in Figure 36.1) and Airport/Gold Coast Line (east to south-east line in Figure 36.1). This chapter uses a cross-sectional slice of Go-card transaction data over two months (November and December 2014) with 605,958 data points (i.e., tap-ons), provided by TransLink.

Method and data processing

This chapter evaluates passengers’ waiting time patterns, which can be measured by the time difference between passengers’ arriving time at a station (i.e., the Go-card tap-on time) and the departure time of the next heavy rail service at that station. Usually the mean or expected passenger waiting time is assumed to be equal to half the headway, with the assumption of deterministic headway (i.e., the variance of headway is 0) and random arrival of passengers (i.e., passenger arrivals follow the Poisson process) (Mishalani et al., 2006). However, this might not be the case in practice. From the passengers’ point of view, there are several factors that might influence their arrival pattern, such as real-time information on public transport services, the on-time rate and passenger waiting time perceptions. For the passengers who are aware of the service schedule, they might shift from random arrival to coordinated arrival patterns (Bowman & Turnquist, 1981). To test the assumption that the passenger average waiting time is equal to half the headway (i.e., random arrival pattern), the research question is rephrased as: Is the actual proportion of passengers waiting up to half the headway equal to 0.5? The analysis uses a $z$-test as identified in the following equations:

$$Z = \frac{(p - \pi)}{(s / \sqrt{n})}$$  \hspace{1cm} \text{Equation 36.1}

$$s^2 = \pi(1 - \pi)$$  \hspace{1cm} \text{Equation 36.2}

where $p$ is the actual proportion of passengers waiting up to half the headway; $\pi$ is the expected proportion waiting, up to half the headway ($=0.5$); $s$ is the standard deviation of the sample; and $n$ is the sample size. With the setting of the $z$-test, we have the following hypothesis:

$H_0$: Actual proportion of passengers waiting up to half the headway is equal to 0.5.

$H_1$: Actual proportion of passengers waiting up to half the headway is larger than 0.5.

The $z$-test follows a normal distribution and is best used for a larger sample size (greater than 30): a condition which is met in this analysis, as there are 605,958 tap-ons in the dataset.

Heavy rail headways and data processing

The timetable for heavy rail in SEQ has three different headways: 7, 15 and 30 minutes. Service frequencies are generally 30 minutes off-peak in the outer suburbs and 15 minutes (or better) in the inner suburbs due to extra peak hour services. Service frequency is higher in the peak period (e.g., 7–9 am), in particular for inner suburbs that are close to the Brisbane central business district (CBD). Each of the 51 heavy rail stations on the two lines is identified as having...
7, 15 or 30-minute headway for the different time periods. Passengers are assumed to take the next service after they tap on with their Go card. In this case, each passenger's waiting time is the difference between the arriving time at the station (Go-card tap-on time) and the departing time of the next rail service at the station. According to each passenger's arriving time, the actual proportion of passengers waiting every minute can be determined. For example, for a 7-minute service headway, there is only 0–6 minutes waiting time, and the proportion waiting for each minute can be derived from the time difference between the Go-card tap-on data and the heavy rail service timetable. The average waiting time is expected to be 3.5 minutes, with the expected proportion waiting up to half the headway (=0.5). A \( z \)-test is then conducted for a 7-minute service headway for each minute (i.e., 0–6 minutes), with the results identifying whether \( (H_0) \) is accepted. This data processing logic is repeated for all three headway types.

Results

Although there are three different service headways for the Beenleigh/Ferny Grove and Airport/Gold Coast rail lines (7, 15 and 30 minutes), Table 36.1 shows only the actual count of Go-card tap-ons for those passengers who have 0 to 6 minutes waiting time for the 7-minute headway, but the process is similar for the other two headways. The cumulative count is the cumulative number of passengers arriving from 0 minutes onwards. The expected cumulative count is the average cumulative count of the total count divided by the number of possible arrival minutes (=19,857/7) from 0 minutes. The actual proportion is equal to cumulative count (i.e., row 2) divided by total count (i.e., row 4) for every minute due to the random arrival assumption. Using the same concept, the expected proportion is equal to the expected cumulative count (i.e., row 3) divided by total count for every minute (which is the smallest analysis unit). The \( z \) score and sample standard deviation can be calculated using Equations 36.1 and 36.2. Therefore, when waiting time is 3 minutes, the expected proportion of passengers waiting up to half the headway is 0.5714 (shaded in pale grey in Table 36.1), which is the closest proportion to 0.5. The actual proportion for 3 minutes is 0.0681. The \( z \) score is −143.3267, which rejects the null hypothesis that the actual proportion waiting up to half the headway is equal to 0.5. Table 36.1 shows that most of the passengers wait 5 minutes (shaded in darker grey in Table 36.1), which is also the median waiting time.

<table>
<thead>
<tr>
<th>Items</th>
<th>Waiting time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(1) Frequency</td>
<td></td>
</tr>
<tr>
<td>(2) Cumulative count</td>
<td></td>
</tr>
<tr>
<td>(3) Expected cumulative count</td>
<td>2836.7</td>
</tr>
<tr>
<td>(4) Total count</td>
<td>19857.0</td>
</tr>
<tr>
<td>(5) Actual proportion</td>
<td>0.0001</td>
</tr>
<tr>
<td>(6) Expected proportion</td>
<td>0.1429</td>
</tr>
<tr>
<td>(7) Sample standard deviation</td>
<td>0.0025</td>
</tr>
<tr>
<td>(8) ( z ) score</td>
<td>−57.5080</td>
</tr>
</tbody>
</table>

Table 36.1 Wait time proportion for 7-minute headway for heavy rail services
Following the same calculation process, Table 36.2 shows the arrival pattern and median waiting time for each service headway. In terms of passenger arrival patterns, the percentage of passengers who arrive within half the headway is significantly lower than 50% for all headways. This suggests that passengers do not arrive randomly and prefer to arrive early. According to the median waiting time, passengers have significantly longer waiting time than half the headway.

### Conclusions and discussion

This chapter investigated passengers’ waiting time patterns and whether waiting time is equal to the expected waiting time of half the heavy rail service headway. The analysis used a z-test to verify actual waiting time patterns (shown as the actual proportion in Table 36.1) and expected waiting time pattern (shown as the expected proportion in Table 36.1). The results offer some useful insights into how a public transport service characteristic of service frequency influences passengers’ arrival pattern and contributes to the travel demand management literature, as well as the smartcard literature on measuring waiting time.

The z-test results for all three headways were rejected at the 5% significance level, indicating that passengers’ actual waiting time is not half the headway. Table 36.2 shows that the proportion of passengers who arrive within half the headway for all three headways is smaller than 50%, with median waiting time of 5 minutes for the 7-minute headway, 10 minutes for the 15-minute headway, and 19 minutes for the 30-minute headway. This implies that passengers are waiting longer than the theoretically assumed half the headway.

In SEQ, heavy rail is considered on time with the following definition:

“All services, including peak and off-peak, are measured as on-time when they arrive within three minutes and 59 seconds of their scheduled time, or 5 minutes and 59 seconds on inter-urban services – Gold Coast, Rosewood and Sunshine Coast” (Queensl & Rail, 2020). With the possibility that the heavy rail service might arrive early and not wanting to miss the service, passengers appear to have adjusted their arriving time at the station as shown in the case study. It should be noted that this case study analysis includes the Gold Coast, which has a longer on-time buffer of 5 minutes and 59 seconds. If the longest buffer time is deducted from the median waiting time for the 15 and 30 minute services, the passengers’ waiting time is very close to half the headway for the two longer service headways. This result indicates that passengers’ perceived waiting time is shorter than actual waiting time at a heavy rail station. The perceived shorter waiting time could potentially be translated into a risk of missing the service; thus, passengers tend to arrive earlier. In SEQ, at the time of data collection in 2014, there was a real-time information machine at each station but no other channels, such as a mobile phone.
app, that usually provide timetable schedules but not real-time information. A real-time passenger information system could potentially reduce the perceived uncertainty when providing accurate information that might increase passenger satisfaction and, ultimately, increase ridership for heavy rail in SEQ.

These preliminary results suggest areas for future research. In the case study, passengers’ waiting time for heavy rail in SEQ is much longer than expected, but the impacts of other variables are not considered, such as the time period (peak or off-peak), user frequency and therefore familiarity with the services, and geographic area. The Go-card data could be further segmented and analysed by time period (e.g., peak and off-peak time periods), user frequency (e.g., number of trips per week) or trip purpose (e.g., for commuting or other purposes) to better measure and understand waiting time variations. Since the inner CBD has a higher service frequency, segmentation into the CBD area and outer suburbs might provide different waiting time patterns that are worth exploring. Passengers making trips that have both origins and destinations in the inner CBD might have different waiting time patterns that, for example, might equal half the service headway. In this case, timetable improvements or the provision of a real-time information system might be more important for passengers and stations in outer suburbs to improve public transport reliability and satisfaction and thus public transport patronage in car-dominated regions.

As a final note, as this chapter was written before the announcement of the COVID-19 pandemic, future research must take account of the way in which public transport modes have been claimed as high-risk modes. Indeed, a similar study with updated smart card data would be useful to understand the impact of this public health incident on public transport ridership and operations.

Acknowledgements

This study was partially sponsored by the Ministry of Science and Technology, Republic of China (Taiwan), under contract number 108–2218-E-009–061 and 109–2221-E-009–047-MY3.

References


