Investigating Consistency in Passenger Arrivals – Insights from Longitudinal Ticket Validations

Daniel Csikos,
Graham Currie

Institute of Transport Studies, Department of Civil Engineering, Building 60, Monash University, Clayton, Victoria 3800

ABSTRACT

Waiting for public transport is recognised as being more onerous than travel time itself (Wardman 2004). Previous research has established that one category of public transport users ‘target’ their arrival at a time shortly before the service is scheduled. Another group of users arrive randomly, seemingly unaffected by timetable schedules. The time of day plus service characteristics, such as headway and reliability, affect the split of users between these groups.

Little is known about the longitudinal aspects of the non-random arrival behaviour. This paper assesses the day-to-day consistency of transit users’ arrival behaviour by analysing longitudinal ticket validations data for heavy rail services in Melbourne. It makes further distinctions of the public transport users that have up until now been categorised into random or non-random arrival behaviour.

Four archetypal arrival behaviours are derived. Consistency of arrival behaviour is quantified and investigated. Heterogeneity is an overriding feature of public transport commuters’ longitudinal behaviour. Users exhibiting a greater amount of consistency were found to arrive closer to the service’s timetabled time, used fewer scheduled services and travelled earlier in the peak. A systematic difference in arrival behaviour was found for users of a terminus station.

INTRODUCTION

Waiting for public transport is recognised as being more onerous than travel time itself (Wardman 2004). Previous research has established public transport passengers arriving at a stop or station can belong to two main categories (Csikos & Currie 2007, Luethi et al 2007, Jolliffe & Hutchinson 1975, Bowman & Turnquist 1981). One group of passengers ‘target’ their arrival at a time shortly before the timetabled service is scheduled to arrive, gaining the benefit of reduced waiting time. The other group of passengers arrive randomly, seemingly unaffected by timetable schedules. The time of day plus the service characteristics, such as headway and reliability, affect the split of passengers between these groups (Csikos & Currie 2007, Luethi et al 2007). As the headway or reliability decreases, the proportion of public transport users timing their arrivals also decreases. It is suggested that this is due to the diminishing benefit to be gained from reduced waiting time (Bowman & Turnquist 1981). However, the morning commuter period sees a greater proportion of public transport users timing their arrivals, even for short headways such as 5-6 minutes (Csikos & Currie 2007, Luethi et al 2007).

Little is known about the longitudinal aspects of non-random arrival behaviour. The concept of day-to-day consistency of a passenger’s non-random behaviour is largely overlooked in the literature. Insights into the day-to-day consistency of arrival behaviour can contribute to the validation of theoretical models dealing with wait time. It can also improve understanding of transit user segments, an important prerequisite for the development of traveller information (Infopolis 2 1998, Multisystems Inc 2003, Stradling 2002, Warman 2000).

This paper assesses the day-to-day consistency of public transport users’ arrival behaviour by analysing longitudinal electronic ticketing data for heavy rail services in Melbourne. It aims to make further distinctions of the public transport users that have up until now been categorised into random or non-random arrival behaviour.

The paper is structured as follows:

- **Research Background** – presents a review of the research literature relevant to this field
- **Methodology** – describes the aims and approach of the empirical analysis and outlines the issues associated with data collection using electronic ticketing.
- **Results** – outlines the empirical findings.
- **Conclusions** – summarises the main findings of the study including suggestions for further research in this field.

RESEARCH BACKGROUND

‘Conventional wisdom’ suggests that waiting time is one half of headway when services are reliable (Hess et al 2004, Ceder & Marguier 1984, Chang & Hsu 2001, Lam & Morrall 1982). Waiting time is known to increase as headways grow in length (Lam & Morrall 1982, Salek & Machemehl 1999). A number of studies have suggested that passengers arrive randomly when headways are short (e.g. Furth & Muller 2006). At longer headways passengers are thought to time their arrivals to an optimal period before the scheduled arrival time (Hess et al 2004, Ceder & Marguier 1984). The headway threshold between short and long headway waiting behaviours has been suggested as 10 to 12mins for US practice (Furth & Muller 2006). However the same source notes that this can vary with service reliability.

Jolliffe and Hutchinson (1975) explored the different types of passenger behaviour which might be occurring in waiting. Three types were identified. First, a proportion of passengers arrive coincidentally with the vehicle arrival. Second, passengers who target their arrival at an optimal point which minimises their expected wait time. The third type of passenger arrives randomly. Their observation of 10 bus stops for approximately 1 hour over 8 days supported these groupings. More so, they observed that as headways increased, the proportion of non-random passenger arrivals increased. Random arrivals dominate below a short headway threshold between 5 and 10 minutes.

Reliability of transit service has been identified as the first of 10 determinants of service quality for public transport (Morpace International 1999). Poor reliability, like waiting time, is known to be associated with anxiety and stress (Kluger 1998, Wener et al 2003). Reliability has for some time been related to waiting through the work of Newell (1971) and Mohring (1972):

\[ E[W] = 0.5 E[H] (1 + cv^2_H) \]  \hspace{1cm} (1)

Where \( W \) = waiting time and \( cvH \) = the headway coefficient of variation.
Bowman and Turnquist (1981) examined the distinction between random arriving passengers and ‘aware’ passengers who target their arrival times. Their observation of 7 bus stops for approximately 2 hours over 5 to 10 days was conducted to empirically test the proportions of random and ‘aware’ passengers. They estimated that all passengers fell into the ‘aware’ category even though the headways ranged from 5 to 20 minutes. They suggest this is the case because the observations were in the morning peak. The morning peak consists mostly of work trips or other trips repeated regularly where the user has a good opportunity to build ‘awareness’. They speculated that the midday period would not consist entirely of ‘aware’ passengers.

Bowman and Turnquist also focused on the effect that schedule reliability has on the utility function of ‘aware’ passenger arrivals. The utility function was treated as a proxy for arrival probability. The utility function is theorised to reach a peak some point before the scheduled arrival time (Figure 1 - top). It then declines to reflect the penalty of missing an intended service. The suggested utility function for both a 5 minute and a 20 minute headway compared favourably with empirical observations although these were based on limited data. Figure 1 (top) illustrates the case of the 20 minute headway. When comparing utility functions of varying levels of schedule reliability, they theorised that decreasing reliability would have a marked effect on passenger arrival times (Figure 1 - Bottom). When services are reliable (SD=0.5mins) a generally timed arrival pattern would be apparent. For less reliable services (SD=2mins), the peak gets flatter and shifts further away from the scheduled arrival time. The effect of this shift is to sizably increase average waiting times. The number of passengers arriving just after the previous scheduled bus (to the left of Figure 1 bottom) also increases. The authors suggest that the impact of reliability on waiting times has been underestimated on this basis.

Furth and Muller (2006) concur with this view. They adopted an approach using empirical observations of schedule reliability derived from automatic vehicle location (AVL) data. They suggested the concept of ‘optimised waiting time’ which is a user wait time minimisation strategy. This theorises that users become aware of the distribution of (unreliable) arrivals and make allowance for a one in 50 case of missing a late bus by arriving up to the 2 percentile of the distribution of arrivals. They introduce a term called an ‘offset’, which is the time between a public transport user’s arrival and the scheduled departure.

Csikos & Currie (2007) explored the relationship between service reliability and waiting times through an empirical analysis of arrival time distributions for rail services. They used electronic ticketing data to build a database of passenger arrival distributions to undertake the analysis. This provided a large sample of passengers behaviours upon which previous theories could be tested. Some 38,000 individual boardings over 1,470 hours were recorded. Most previous work in this area has been based on small sample manual surveys.

Csikos & Currie confirmed that passengers are ‘aware’ of service timetables in the peak and tend to time arrivals at optimal points before scheduled times. They also confirmed that in the off peak, fewer passengers demonstrate timed arrivals. Whilst it was found that unreliable services generated more random passenger arrivals, the degree to which this occurred was considerably below that suggested in theoretical bus based models. A new phenomenon called ‘late running awareness’ was demonstrated where passenger arrival times are based on knowledge of late running rather than scheduled train times. A high share of passenger arrivals was observed shortly after scheduled trains were due (Figure 2). Further analysis established that in each of these cases a high share of trains were arriving late. Over 70% of trains were at least 3 minutes late in each case (a 3 minute late margin is also indicated in Figure 2). The passenger arrival patterns suggested that passengers were timing arrivals in relation to their experience of late arrival times, not in relation to the scheduled arrival time. Concurrent research by Luethi et al (2007) also observed ‘late running awareness’ type behaviour for short headway commuter services in Zurich.

This ability of public transport users to perceive variability and adjust behaviour accordingly is also described by Bonsall (2004). Whilst all attributes of transport systems are characterised by variability, some patterns of variability are easier to distinguish from others. Bonsall suggests that a traveller’s understanding of variability depends on their experience...
levels, the amount of information at their disposal, and their intellectual curiosity/ability.

In summary, it is clear that public transport service attributes such as headway and reliability affect the arrival behaviour of passengers. This in turn affects the waiting time experienced, an onerous component of a public transport traveller’s journey (Wardman 2004). Two broad categories of arrival behaviour, random and non-random, are now well documented. However, previous investigations have overlooked the longitudinal aspects of non-random arrival behaviour. The day-to-day consistency of a passenger’s non-random behaviour has been missed due to the aggregate nature of the samples. Modern technology involving electronic ticketing and data storage mean that longitudinal analysis can be conducted and sample sizes can be substantially larger using automated systems. Insights into the day-to-day consistency of arrival behaviour can contribute to the validation of theoretical models dealing with wait time. It can also improve understanding of transit user segments, an important prerequisite for the development of traveller information (Infopolis 2 1998, Multisystems Inc 2003, Stradling 2002, Warman 2000).

METHODOLOGY

Analysis aimed to further the work of Csikos & Currie (2007) on passenger arrival profiles at railway stations. By exploring the longitudinal behaviour of habitual commuters, it was the intention to:

1. Identify arrival behaviour archetypes
2. Investigate arrival behaviour consistency
3. Identify transit service characteristics that affect arrival behaviour.

The Csikos & Currie (2007) study resulted in an extensive database derived from electronic ticketing data. Ticket validations which are time stamped on entry to platforms provided a reasonable estimation of rail passenger arrival times. Four weeks of data from May 2006, from first to last service, were recorded. This involved some 38,000 individual boardings over 1,470 hours.

For this research, the focus was on the weekday commuter period (06:00 to 10:00). It is during this period that non-random arrival behaviour is most evident (Csikos & Currie 2007, Luethi et al 2007, Jolliffe & Hutchinson 1975, Bowman & Turnquist 1981), so a longitudinal analysis would gain a more robust insight into arrival behaviour. The four weeks of the sample period provided 20 weekdays (80 hours in total). Periodical ticket serial numbers (monthly and yearly tickets) were the focus for analysis. The ticket serial numbers are anonymous, that is no personal details are attached to them. Tickets are non-transferable between users. It is reasonable to assume that any potential non-compliance with this condition is an insignificant occurrence for periodical tickets used during the commuter period.

The sample involved ticket holders captured for at least 10 weekdays. To observe the complexities of travel behaviour, it is suggested that a sample of at least 2 weeks is required (Schlich & Axhausen 2003). Overall, this longitudinal study samples 1,043 public transport users conducting over 15,000 commuting trips at 7 stations.

The study focuses on users of the metropolitan heavy rail system of Melbourne, with a metropolitan population of approximately 3.6 million. The rail system is a radial network of lines focussing on a CBD including approximately 372 kms of track on 15 lines with 209 stations.

Additional considerations were required for stations where electronic ticketing data could provide a reasonable representation of passenger arrivals. It was important to target stations with principally walk on catchments. Stations with high shares of bus (or tram) access would have passenger arrivals skewed by the arrival characteristics of feeder services. Park-and-ride facilities also have the potential to affect passenger arrivals. Any park-and-ride facilities featured at the targeted stations should be modest in size so as not to contribute more than a minority of the daily commuter numbers.

Only stations with 'side platform' configurations could be targeted. In these stations ticket validators provided a direct link between arrival time and services operating in a single direction. Stations with 'island platform' configurations have ticket validators which could be used by passengers travelling on trains in both directions. Hence it is not possible to link validations to train departures in these cases. This approach linking ticketing data to particular public transport services was used to analyse data in Chicago (Wilson et al 2005).

Seven stations were selected for analysis based on the above criteria. Table 1 provides a summary of the stations and the service levels experienced within the sample period. Data on service reliability was sourced from performance monitoring by the transit authority (Department of Infrastructure Victoria 2006), with all individual scheduled services measured. Table 1 also shows the associated reliability levels measured for each station’s train line. Overall, 4 stations are categorised into a ‘longer headway, more reliable’ group. These stations are served by headways of 15 to 20 minutes and they have unreliability levels of 10% or less in the commuter period. An alternate group of 3 stations (‘shorter headway, less reliable’) are served by headways of around 10 minutes with levels of unreliability up to 30%.

| TABLE 1: Stations Selected for Ticket Data Collection – Associated Characteristics |
|-----------------------------------------|----------------|--------------|----------------|----------------|----------------|
| Group Description                      | Station Name  | Mean Headway | Unreliability | Minutes | Park-and-ride |
|                                       |               | (mins)       |               | to CBD | Spaces        |
| Longer Headway, More Reliable          | Alamein      | 17.18        | 10%           | 6%    | 28            |
|                                       | Croxton       | 15.15        | 9%            | 7%    | 28            |
|                                       | Williamstown | 19.20        | 6%            | 10%   | 27            |
|                                       | Beach        |              |               |       | 67            |
|                                       | Spotswood     | 19.20        | 6%            | 10%   | 19            |
| Shorter Headway, Less Reliable         | Pascoe Vale   | 12.10        | 11%           | 14%   | 27            |
|                                       | Parkdale      | 11.09        | 19%           | 21%   | 45            |
|                                       | Manorhambeena | 11.10        | 20%           | 30%   | 29            |

1Percentage of services 6 or more minutes late at destination, Source (Dept of Infrastructure VIC 2006)
2Terminus Station
RESULTS

Identifying Arrival Behaviour Archetypes

The initial exploration sorted individual train users’ longitudinal arrival behaviour according to their ‘offset’ characteristics (by means of summary statistics such as median and various measures of central tendency). Subsequently, the behaviours could be segmented into 4 visually discernible categories.

Figure 3 illustrates each of these as a cumulative percentage of ‘arrival offsets’. The x-axis represents the ‘arrival offset’ in minutes (in reverse order ending at zero, or the scheduled service time, on the right hand side).

Figure 3 illustrates:

- ‘Like Clockwork’ arrival behaviour (top left in Figure 3) where the train user arrives in a narrow ‘window’ each day. The offset at which they arrive exhibits a strong regularity. The illustrated case has the transit user arriving around 2 – 3 minutes before the scheduled departure. This behaviour is a means of minimising waiting time.

- ‘Consistent within a wider window’ arrival behaviour (top right) occurs where the train user exhibits a slightly weaker regularity. The arrival ‘window’ is wider but the spread of arrivals is evenly distributed. In the illustrated case, the train user arrives between 2 and 7 minutes before the scheduled departure.

- ‘Consistent plus outliers’ arrival behaviour (bottom left) occurs where the train user exhibits regularity most days, but incongruous arrivals on other days. One could speculate that on some days the user’s movement to the train station is interrupted, leading to an arrival later than normal. The illustrated case has the train user ‘normally’ arriving around 3 and 4 minutes before scheduled departure (11 out of 16 days sampled). For 3 of the sample days, however, the arrival offset was unusually fine (near 0 minutes). Another 2 sample days have an arrival offset noticeably longer than normal (at 7 minutes).

- ‘Largely random’ arrival behaviour (bottom right) occurs where the train user’s arrivals are spread throughout the service headway. This behaviour exhibits no significant regularity.

Within the sample of regular commuters, no one arrival behaviour archetype dominated. There was a broadly equal share of arrival behaviours, from Type 1 (‘like clockwork’) through to Type 4 (‘largely random’). Defining consistency of arrival behaviour is largely arbitrary. Whilst Type 1 arrival behaviour as depicted in Figure 3 is undoubtedly consistent, for Type 2 arrival behaviour it is not so obvious. Specifying the bounds of an arrival window that is considered ‘consistent’ is discretionary. Likewise for Type 3 arrival behaviour with regards to the number of outliers (arrivals outside of ‘normal’ arrival offset).

The analysis utilised the difference between a train user’s 25th percentile and 75th percentile arrival offset. This represents the central 50% of the user’s arrivals. It excludes the influence of outliers yet still provides a quantification of arrival consistency. The use of medians (50th percentile) is prevalent in travel behaviour studies (eg, Li et al 2005, Richardson 2006).
Arrival Window and Arrival Offset

Analysis of the train users’ longitudinal arrival behaviour focused on:

1. The spread of times at which they arrived at the station (arrival window).
2. The time between each arrival and the next scheduled service (arrival offset).

Table 2 lists a selection of arrival window and arrival offset summary statistics obtained for each station. For all users, the difference between their 25th percentile arrival time and their 75th percentile arrival time was calculated. For most users, this incorporated multiple services. The third column of the table lists the mean and standard deviation of this measure, as observed at each station.

For the remainder of the summary statistics, users were ranked according to their arrival offset behaviour. For all users, the difference between their 25th percentile arrival offset and their 75th percentile arrival offset was calculated. Using this measure, the users were then divided into quartiles, representing the ‘most consistent’ quartile through to the ‘least consistent’ quartile. Comparisons were made between the ‘most consistent’ quartile and the whole sample. Table 2 summarises this comparison with the median arrival offset and the number of services used. It also compares the median time of arrival during the commuter period (median arrival time).

A number of noteworthy results are found in Table 2:

- With regards to the arrival window (central 50% - 25th to 75th percentile), the mean for all stations is over 10 minutes. The narrowest arrival window of around 12 minutes is observed at Alamein (a longer headway, more reliable station). The broadest arrival window of under 17 minutes is observed at Parkdale (a shorter headway, less reliable station). In all cases, the standard deviation is greater than the mean, reflecting a large amount of variation within the commuter sample. The means and the standard deviations are not dissimilar between the station groupings.

- The median arrival offset of all stations is around 2 or 3 minutes, with the exception of Alamein which has a median arrival offset of over 4 minutes. Alamein is the only terminus station of the sample. This may encourage people to arrive earlier since they know they will be waiting seated on the train not on the platform awaiting the train's arrival. The arrival offset of the consistent users is around 1 or 2 minutes, once again with the exception of Alamein which remains at 4 minutes. It is notable that the consistent users have a finer arrival offset than other users. In all but two cases, the arrival offset is 1 to 2.5 minutes smaller for consistent users. The exceptions are Alamein and Croxton (both longer headway, more reliable stations) where the arrival offset remains the same.

- The sampled train users on average used between 2 and 4 different scheduled services (train scheduled according to timetable). For the longer headway station grouping, an average of around 3 timetabled services was used over the sample period. A slightly fewer number of services were used by consistent users than was used by the full sample. For the shorter headway station grouping, around 4 timetabled services were used over the sample period. About 1 less service was used by consistent users than was used by the full sample.

- The median arrival time (in other words, the middle of the train users’ arrival windows) reflects the typical work start times in the CBD. Of particular interest is the difference between median arrival times of consistent users compared to the full sample. The consistent group of train users typically arrived earlier, with the highest being 22 minutes earlier. The exception to this was Alamein and Croxton (both longer headway, more reliable stations) where the consistent group arrived at the same time as the full sample. For the remaining longer headway stations it was 9 and 15 minutes earlier. The shorter headway stations ranged from 6 to 22 minutes earlier.

<table>
<thead>
<tr>
<th>Group Desc.</th>
<th>Station Name</th>
<th>25th-75th Arrival Window - In minutes (mean) (std dev)</th>
<th>Median Arrival Offset – In minutes (median)</th>
<th>Number of Services Used (mean)</th>
<th>Median Arrival Time (median)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All Users</td>
<td>Most Consistent Quartile</td>
<td>All Users</td>
<td>Most Consistent Quartile</td>
</tr>
<tr>
<td>Longer Headway More Reliable</td>
<td>Alamein</td>
<td>11.8 (22.3)</td>
<td>4.2</td>
<td>4.3</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Croxton</td>
<td>15.6 (17.5)</td>
<td>2.4</td>
<td>2.3</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Williams- Town Beach</td>
<td>13.5 (17.0)</td>
<td>2.3</td>
<td>1.1</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Spotswood</td>
<td>14.7 (19.8)</td>
<td>3.5</td>
<td>1.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Shorter Headway Less Reliable</td>
<td>Pascoe Vale</td>
<td>12.3 (16.6)</td>
<td>1.9</td>
<td>1.1</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Parkdale</td>
<td>16.6 (25.3)</td>
<td>2.2</td>
<td>1.6</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Murrum- beena</td>
<td>15.0 (20.7)</td>
<td>2.3</td>
<td>1.1</td>
<td>4.1</td>
</tr>
</tbody>
</table>
It is evident from these results that Alamein (the terminus station) has noticeably unique arrival offset characteristics. As a result, further analysis was carried out on arrival offsets with a particular focus on comparing Alamein to the other stations. In a previous stage, the median arrival offset of each train user was calculated. Figure 4 illustrates the within-sample cumulative distribution of these median arrival offsets for each station, with Alamein highlighted. The x-axis represents the median arrival offset in minutes (in reverse order reaching zero, or the scheduled service time, on the right hand side).

The results in Figure 4 indicate that:

- The distribution of median arrival offset is systematically different for Alamein than all other stations. Whilst the other stations experience differing headways and levels of reliability, the distribution of median arrival offsets for its users is remarkably similar. For Alamein, its users typically turn up 1 to 2 minutes earlier than the equivalent users at other stations.

- The arrival offset behaviour of the commuters is markedly heterogeneous within the station samples. An even distribution of users occurs at around 3 minutes and under. For each minute from 3 minutes before scheduled arrival to the scheduled arrival time, roughly 20% of users will achieve their median arrival time. A selection of users will achieve a median arrival time of 4 minutes or more before scheduled arrival. However, this behaviour is less prevalent. At most, 5% of users will experience their median arrival time per minute.

- There is evidence that ‘late running awareness’ arrival behaviour is conducted by a small proportion of commuters. It was speculated in previous research (Csikos & Currie 2007, Luethi et al 2007, Jolliffe & Hutchinson 1975, Bowman & Turnquist 1981) that some regular commuters became aware of consistently late services and adjusted their arrival times to a point after the timetable suggestion. All stations had a small number of users (less than 10%) who achieved a median arrival time around 30 seconds after the scheduled time. For example, if the 7:50am service typically arrives at least one minute late, then arriving at 7:50.30s will mean the commuter will generally not miss that train.

Overall this analysis suggests that the overriding feature of train commuters’ longitudinal behaviour is heterogeneity. A variety of arrival behaviour types and consistency would exist in all groups of train users. Even a fundamentally different station type, a terminus station, experiences marked heterogeneity. It is clear that one size does not fit all with regards to arrival behaviour.

**ASSESSMENT**

The analysis reveals that public transport passenger arrival behaviour is more complex than simply random and non-random behaviour. When looked at longitudinally, a range of non-random arrival behaviours exist within samples of train users. These behaviours can be broadly categorised into ‘like clockwork’, ‘consistent within a wider window’, and ‘consistent plus outliers’. However, no one arrival behaviour archetype dominated. There was a broadly equal share of arrival behaviours, from ‘like clockwork’ through to ‘largely random’. Defining consistency of arrival behaviour is largely arbitrary. The measure of consistency utilised in this analysis (25th percentile to 75th percentile arrival window) considers the intricacies of the behavioural archetypes.

The peaked aggregate arrival profiles seen in previous research (Csikos & Currie 2007, Luethi et al 2007, Jolliffe & Hutchinson 1975, Bowman & Turnquist 1981) are explained by the combination of arrival behaviours. It is possible for two identically looking ‘peaked’ arrival profiles to be composed of differing proportions of behavioural archetypes. The more ‘peaked’ an arrival profile is, the greater the influence of ‘like clockwork’, ‘consistent within a wider window’, and ‘consistent plus outliers’ behaviours. However, heterogeneity of arrival behaviour is an overriding feature. Observations at stations with differing headways and service reliability saw no systematic difference, except for the terminus station.

The results raise caution regarding any singular ‘arrival offset’ that theoretical models may assume public transport users seek. The range of ‘arrival offsets’ observed highlight that personal factors are involved. Travellers’ differing values of time, aversion of risk, and journey to the station will be involved in forming their ‘arrival offset’. Whilst one may say that ‘timed’ passengers typically arrive 1-3 minutes before
scheduled departure, this by no means characterises all users.

The possibility of 'late running awareness' behaviour observed in previous research (Csikos & Currie 2007, Luethi et al 2007) has been confirmed. There is evidence that a small number of regular commuters (less than 10%) exhibit this as a serial behaviour.

The segmentation of transit users into their behavioural archetypes provides opportunities for the fashioning of traveller information to best suit their needs. Information provision regarding operational reliability would need to consider that users are not all affected by variability in the same way.

CONCLUSIONS

This paper has explored the longitudinal arrival behaviour of morning public transport commuters through an empirical analysis of ticket validations for rail services. Ticket data was used to build a database of passenger arrivals over 4 weeks to undertake the empirical analysis. This provided a large sample of passengers behaviours upon which previous observed arrival patterns could be further investigated. The concept of day-to-day consistency of a passenger’s non-random behaviour, which was largely overlooked in the literature, was explored.

Four main categories of arrival behaviour were defined. ‘Like Clockwork’ arrival behaviour exhibits the strongest regularity. Here, the user arrives in a narrow ‘window’ each day. For most of these users, the time of arrival is 1-3 minutes before the scheduled service. This behaviour is a means of minimising waiting time. ‘Consistent within a wider window’ arrival behaviour exhibits a slightly weaker regularity. The arrival ‘window’ is wider but the spread of arrivals is evenly distributed. ‘Consistent plus outliers’ arrival behaviour occurs where the user exhibits regularity most days, but has their movements to the train station is interrupted some days. ‘Largely random’ arrival behaviour occurs where the user does not ‘time’ their arrival. This behaviour exhibits no significant regularity and arrivals can occur at any number of minutes to the next service.

Within the sample of regular commuters, no one arrival behaviour archetype dominated. Arrival behaviour was mostly a steady transition from ‘like clockwork’ through to ‘largely random’. Therefore defining consistency of arrival behaviour is largely arbitrary. Specifying the bounds of an arrival window that is considered ‘consistent’ is discretionary. Likewise for the number of outliers (arrivals outside of ‘normal’ arrival offset) existing for ‘consistent plus outliers’ arrival behaviour. Nevertheless, the analysis utilised a measure of the difference between a user’s 25th percentile and 75th percentile arrival offset. This represents the central 50% of the user’s arrivals. It excludes the influence of outliers yet still provides a quantification of arrival consistency.

Public transport users exhibiting a greater amount of consistency (that is they have a narrow arrival window) also show other arrival characteristics. These users tended to have a finer ‘arrival offset’ (less time between user’s arrival and the scheduled time of the service), used fewer scheduled services and travelled earlier in the peak.

A range of areas for further analysis are suggested by the research:

- Expansion of the research to public transport services with other headway and reliability characteristics. It is also possible to explore other modes where ticketing data provides a good estimate of a passenger’s stop/station arrival, and AVL provides information on the service’s reliability.
- Expand the investigation to other trip types, such as afternoon/evening commuting and off-peak travel.

Overall the analysis supports the use of electronic ticketing data as a tool to understand passenger behaviour. A large and statistically significant pool of data is readily available from ticketing systems to explore important issues related to public transport service quality and performance.

REFERENCES


Salek MD and Machemehl (1999) ‘Characterising bus transit passenger wait times’ Austin Texas; University of Texas Centre for Transportation Research.


