Analysis of the differences in travel behaviour between Pay As You Go and season ticket holders using smart card data

William Tran

ABSTRACT

This paper establishes several approaches of assessing whether the use of different fare payment options affect passenger travel behaviour. The analysis compares the behavioural traits of season ticket (Travelcard) holders against those who travel on a Pay As You Go (PAYG) basis, on the London Underground over a four-month period. The regularity of passenger journeys is considered through the extent of both their spatial and temporal variation.

Temporal analysis revealed that the journey durations of PAYG users were consistently more variable. Spatial analysis provided mixed indications, but highlighted the contrasts in passenger composition for each fare option. Overall, the results show that the PAYG system is suitable for a wide range of users, and that regular commuters behave indifferently regardless of their fare. Neither pricing structures have been shown to encourage substantially more predictable travel behaviour, but rather that they appeal to different groups of individuals.

1 INTRODUCTION

Since 2000, there has been substantial demand growth on the principal transport networks of London. Two factors have driven this change. Firstly, the population and employment growth have influenced the demand for travel. Transport for London (TfL) reported an 8% increase in population and that 4.8% more jobs were created in London between 2000 and 2010. Secondly, a progressive shift in mode share from private to public transport of 6% has occurred over the same period (TfL, 2011a).

Travelling in London is possible either via the use of a magnetic strip fare card, or an Oyster card. The Oyster card is a contactless Automatic Fare Collection (AFC) system introduced by TfL in 2003. Since 2010, more than 80% of all bus and London Underground (LU) journey payments were made using Oyster cards (TfL, 2010).

Oyster cards are primarily used either as a season ticket (Travelcard), or to pay for travel on a ‘pay as you go’ (PAYG) basis using pre-loaded credit. Travelcards are valid across several transport modes within the TfL network, limited to a spatial and temporal validity. PAYG users are charged for each journey until the daily cumulative fare reaches the price cap.

This research is motivated by the hypothesis that season ticket holders travel with much less regularity as compared to PAYG users. The significance of travel behaviour arises from the value of being able to accurately determine route demand for capacity management purposes.

1.1 Research Objectives

The primary objective of this research is to assess the extent of differences in travel behaviour between Travelcard and PAYG users on the LU network. Additionally, the study aims to quantify the extent of regularity in passenger travel decisions.

To expose the multiple ways in which travel behaviour can differ, the following sub-objectives are set:

• Develop a framework for analysing Oyster card data, including a robust methodology for pre-processing raw data.
• Analyse the regularity of journey duration (temporal variation) for both fare options.
• Analyse the extent of variation in the range of stations used (spatial variation) by both fare options.
• Establish a set of behavioural criteria to group patrons, from which categorical comparisons can be made.

A secondary objective, applicable to all analyses carried out, is to take measures of promoting a fair representation of all individuals.

1.2 Structure

This paper is organised into 8 sections. Section 2 presents key literature relevant to this research. Section 3 highlights the selection of data and associated challenges. Section 4 outlines the methodology of each analysis and the rationale for assumptions made. Section 5 categorises the
findings according to each study. Section 6 discusses the key findings and their implications.

2 LITERATURE REVIEW

This review focuses on the key background elements of this project: the use of smart card data to assess travel behaviour, its value towards service planning, and comparisons made between the Travelcard and PAYG fare options.

2.1 Travel Behaviour

In recent years, research content analysing travel patterns have differentiated user groups either by their trip purpose or by demographics. This is often achieved through generalising the composition of patrons at a macro scale, according to the time of day. Studies of the connection between fare structures and travel behaviour is a relatively under-researched area.

Morency et al. (2007) measured the spatial and temporal variability of transit use for various types of smart cards: senior, student, and 3 categories of adult cards. As part of a related study, the regularity of travel behaviour was shown through three metrics (Morency, et al., 2006):

1. Activity rate: proportion of days where users boarded
2. Average number of boardings per day (by day of week)
3. Different stops used vs. number of days travelled (over 9 months)

2.2 Service Planning

Chapleau and Chu (2007) analysed the variability in passenger boardings on a specific route and their transfer activities to develop a method of coordinating bus and rail schedules.

Paul (2010) combined the use of Oyster card, RODS, and Networking Management Information System (NetMIS) data to produce a method for modelling train allocation. The developed route choice model uses a shortest path algorithm. Despite the usability of this approach, it does not differentiate between the behavioural difference of Travelcard and PAYG users.

Park et al. (2008) uses train line specific analysis of smart card data rather than network averages to improve the accuracy of modelling travel characteristics.

2.3 Season Tickets Vs PAYG

Frumin (2008) considered the choice faced by London Oyster card users, to use either a season ticket or to travel PAYG. His research is the only existing study that compares the two fare options. The analysis studies the financial efficiency of ticket type decisions made by LU patrons. The results show that the likelihood of not breaking even increases with users who purchase greater zone (spatial) coverage season tickets. Overall 33% of Travelcard users do not break even on their purchase. Cost wise, these users would have been better off travelling PAYG.

3 DATASET

All analyses carried out within this research are based upon LU Oyster card data, which is provided by TFL. Four separate months (November 2010 & 2011, February 2011 & 2012) have been selected based on the likelihood to which they reflect normal travel volumes. Bank holidays, school holidays, festive seasons, and special events were taken into consideration. Due to the total volume of journeys made on the LU and constraints due to computational power, a 5% sample of individuals is analysed.

Exactly 28 consecutive days of data are extracted from each month to achieve an equal representation of each period. This enables fair comparisons to be made between each of the periods, and equally weighted results aggregation across the periods.

Every trip made by an Oyster card user induces a number of data fields to be recorded. Data entries where any of the seven fields are missing are omitted from the study. In 99% of cases, the need for data omission results from journeys that are either undocumented at the origin (unstarted) or undocumented at the destination (unfinished). The loss of information because of missing data fields is on average 8.8% across the periods of investigation.

It is possible for a PAYG individual to complete a trip paying a discounted fare. This occurs when the daily price cap is reached, causing the marginal cost of additional trips within the same day to become zero. Such ‘mixed’ journeys amount to less than 3% of LU Oyster card data. Given the uncertainty of their implications upon travel behaviour, these journeys have also been excluded from the dataset. There is a data retention rate of 88% following the omission of all unsuitable LU Oyster card journey records.

The full dataset of LU journeys excluding omitted data contains 15 million journeys, collected from approximately 300 thousand smart cards.

3.1 Additional Fields

To prepare the data for analysis, four additional fields are derived from the extracted data:
• **OD String**: A journey identifier constructed with the combination of the entry station code and exit station code.

• **Journey Duration**: The time taken to complete the journey calculated as the difference between the start and end time.

• **Median time**: The time at the mid-point of the journey based upon the start and end time.

• **Period Classification**: A classification of the period of travel based upon the median time. Classifications segregate trips into groups where journeys have shown to reflect similar characteristics.

Workable datasets for each month is produced separately to allow for independent monthly results to be produced.

### 3.2 Data Limitations

Two main limitations arise from properties of the Oyster card data:

1. **Oyster card Timestamp Truncation**
   Originally identified by Paul (2010), all Oyster card journeys receive an entry and exit timestamp that is registered in absence of the seconds measurement. The consequence is a maximum timestamp error of 59 seconds. Comparisons made between trips of short journey durations are most susceptible to this lack of precision, due to the size of the error as a proportion of the travel time. This inaccuracy also affects the period classification of trips.

2. **Sampling**
   The use of a 5% sample of individuals has resulted in an almost equal split of Travelcard and PAYG journey records. However, due to the fact that Travelcard users make more trips on average, there are over twice as many PAYG users captured in the sampling process. The difference invalidates any comparisons made between absolute travel volumes.

### 4 METHODOLOGY

Passenger travel behaviour is considered through two separate investigations. The evaluation of journey duration assesses the temporal variation of each fare group. Spatial analysis investigates how often individuals make the same journeys, and how often their station choices change.

#### 4.1 Journey Duration

Due to the number of journey attributes, introducing too many variables will result in the over-sorting of data. To avoid comparisons being made between small clusters of individuals that are not representative of their grouping, two separate analytical approaches were developed to assess journey duration.

Table 1 summarises the data fields considered for each approach and their assignment within a Microsoft Access Crosstab Query.

#### Table 1 Data fields considered by each approach

<table>
<thead>
<tr>
<th>Crosstab Assignment</th>
<th>Data Field</th>
<th>Network wide Daily Comparisons</th>
<th>Journey Specific Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Row Headings</strong></td>
<td>Date</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OD String</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Period Classification</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Column Heading</strong></td>
<td>Fare</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Row &amp; Column Intersection</strong></td>
<td>Mean Journey Duration</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Deviation of Journey Duration</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

#### Network wide daily comparisons approach

The network wide daily comparisons approach provides a system-wide assessment of journey duration for a given point in time. The daily timeframe was chosen to optimise between sufficient data sizes, and the preservation of conditions specific to each period.

#### Journey specific comparisons approach

The journey specific comparisons approach disregards the date of travel but introduces extensive data sorting according to the origin and destination. It examines the extent of variation in journey time within each of these unique combinations, for both Travelcard and PAYG users. An additional criterion is set to prevent the inclusion of results derived from small samples. Because of this criterion, 84% of data is omitted. The journey specific comparisons approach is therefore a study of travel behaviour, which focuses upon only the most popular routes.

An exampleOne of the analytical techniques introduced, is the relative comparison metric, which is used to prevent skew towards either fare group:

$$Mean \left\{ \frac{StDev \ (Travelcard)}{StDev \ (Travelcard) + StDev \ (Pay \ As \ You \ Go)} \right\}$$

This limits the feasible range of results between 0% and 100% allowing for meaningful average values to be calculated.
4.2 Extent of Travel

The techniques developed to analyse spatial patterns are sub-divided into three studies.

Study 1: Unique Journeys

The study of unique journeys provides an understanding of the number of different routes that London Underground patrons embark upon within a given period.

Specific For this study, individuals who have made journeys under both fare structures within any given month are further examined. For instances where 90% of an individual's trips are recorded under a single fare structure, only journey records made under their primary choice of fare will be taken into account. For cases where the allocation of trips made under the primary fare is less than 90%, all records for the individual are disregarded. These measures are taken to ensure that results are truly reflective of the fare groups.

Study 2: Maximum trips Vs Unique Journeys

The second study of this Chapter investigates each individual independently based upon two variables: the maximum number of trips on a particular route and the number of unique routes. These independent variables form the axes scales for graphical results. A third variable stores information of the percentage fare group population falling under each category within a 2D matrix array.

The following steps are applicable for both study 2 and study 3.

Individuals are required to have made at least one journey within each of the four months in order to be included within the combined dataset. This ensures a base level of activity and removes a proportion of passengers who have perhaps made a one-off journey on the London Underground. Such data is found to obscure significant behavioural trends that are found in individuals with a more substantial Oyster card usage history.

When generating 3-D passenger distribution graphs, three additional steps have been taken. Firstly, to avoid comparisons between absolute numbers, results are expressed as a proportion of the total fare group population. Secondly, extreme values are truncated to allow a more focused assessment of core trends on graphical outputs. Thirdly, results are plotted on a logarithmic scale to accommodate the huge variation in the percentage distribution of individuals.

Study 3: Average trips Vs Unique Journeys

The third study replaces the maximum trips criteria of the second study with a focus on the average number of trips made across all routes taken. Averages are rounded to one decimal place before passenger distributions are generated.

A Criteria for user classifications have been developed to allow for categorical comparisons. These are defined in Table 2.

<table>
<thead>
<tr>
<th>User classification</th>
<th>Description</th>
<th>Average number of trips made on each route</th>
<th>Number of unique journeys (over 4 months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visitor</td>
<td>Low level of activity</td>
<td>&lt; 2</td>
<td>&lt; 40</td>
</tr>
<tr>
<td>Recurrent</td>
<td>Frequent use of the same routes</td>
<td>≥ 2</td>
<td>&lt; 40</td>
</tr>
<tr>
<td>Varied</td>
<td>Infrequent use of a variety of routes</td>
<td>&lt; 2</td>
<td>≥ 40</td>
</tr>
<tr>
<td>Diverse</td>
<td>Frequent use of a variety of routes</td>
<td>≥ 2</td>
<td>≥ 40</td>
</tr>
</tbody>
</table>

5. RESULTS

5.1 Journey Duration

The network wide daily comparisons approach indicates that the journey duration of PAYG patrons are more variable. Figure 1 presents the daily standard deviations for weekends over the four-month period.

Throughout the 672 periods considered within this study, the number of times that the standard deviation of Travelcard journey durations exceeding the equivalent for PAYG journeys is twice. The standard deviation of PAYG journeys is on average 2.74 minutes greater than Travelcard journeys during weekends and 1.84 minutes greater during weekdays.
Figure 2 presents an extract of the journey specific comparisons results. The feasible range of results will lie between 0% and 100%, where a value of 50% indicates that for exactly half of the OD specific journeys, the standard deviation of PAYG trips exceed that of Travelcard trips. We notice that overall, the aggregated results remain between 45% and 50%.

A relative comparison metric introduced for this approach revealed that: the standard deviation of the time it takes a PAYG user to complete a particular journey, is 3.6% greater than Travelcard users during weekdays, and 2.8% greater on weekends. The closest degree of variation between the two fares occurs during the morning peak period. The greatest difference in variation is observed during the weekday between 10am to 4pm. For this period, the standard deviation of PAYG journeys was on average 10.8% greater than Travelcard journeys on the exact same routes.

5.2 Extent of Travel

Study 1: Unique Journeys

The distribution of PAYG individuals is concentrated towards fewer unique journeys, with over 70% of all users taking five or less unique routes each month. This compares to approximately 45% for Travelcard users. Over 25% of Travelcard users travel on 11 or more unique routes within a given month. Figure 3 shows that for both fare groups, there is an inverse exponential distribution of individuals against the number of unique trips made each month.

The magnitude of the PAYG curve’s gradient is the greater of the two gradients, indicating a stronger inverse exponential relationship. This shows that, by increasing the number of unique journeys by one route, the proportional decline in the number of PAYG users who are able to travel on them, is greater than Travelcard users.

Study 2: Maximum trips Vs Unique Journeys

On a contour plot of maximum trips per route against unique journeys (Figure 4), there is a greater reflection of commuters (higher value on Z-axis) within the Travelcard population as compared to the PAYG population. These are individuals who frequently use the same routes, and make as many as 20 trips on a single route per month. Only 1.0% of Travelcard users and 0.2% of PAYG users exceed 20 trips on a single route per month. 0.9% of PAYG users make more than 15 trips on a single one-way route every month. These individuals do not enter or exit the network from any other stations.

Study 3: Average trips Vs Unique Journeys

51.0% of PAYG users are classified as Visitors, who have a low level of travel activity, compared to 27.7% of Travelcard users. Outside of this classification, PAYG users are likely to either make a high number of trips on few routes (Recurrent), or a low number of trips on many different routes (Varied). 21.8% of Travelcard users are classified Diverse, which is over three times the proportional distribution of PAYG patrons.

Amongst all individuals who travel the volume of a regular commuter (4 to 6 return journeys per week), a PAYG user is as likely to make use of a diverse variety of routes as the Travelcard user.

6. CONCLUSION

Section 1 introduced the hypothesis that season ticket holders travel with less regularity as compared to PAYG users. On the basis of using standard deviation to measure the variability of journey duration, the reverse has been shown to be true. The variability observed within the PAYG user population is greater, regardless of whether the comparison is between all trips made simultaneously at a given point in time, or between
like-for-like journeys originating and arriving at the same stations.

The analyses of this research have shown that the PAYG fare structure caters for a large variety of patrons. Furthermore, 27.7% of the Travelcard population are classified as Visitors. These patrons are likely to have been financially better off traveling PAYG given their low trip volumes.

Frumin (2008) suggested that Travelcard fares (as they exist today) could be abolished on the London Underground. From a behavioural standpoint, had the findings of this research shown that the travel patterns of PAYG users are more predictable, a case could have been made in favour of the simplified ticketing system. However the results of the journey duration analyses demonstrate that Travelcard patrons are more consistent than PAYG users. Separately, the mixed indications presented from spatial analysis neither supports or disagrees with the policy recommendation. The findings of this research further show that a significant proportion the Travelcard population would not alter their route choice even if such a radical fare policy change was implemented.

This paper does not display either London Underground fare structures to be more favourable towards service planning. It is obvious that each fare option appeals to a different user group. However, the findings do show that it is possible for a PAYG structure to cater for patrons currently travelling with Travelcards, and in some cases, more suitable.

7 ACKNOWLEDGEMENTS

This paper is based on work undertaken by William Tran for his undergraduate research at Imperial College London, under the supervision of Dr. Arnab Majumdar. The author wishes to express thanks to Transport for London for allowing access to their data for this research study. All interpretations based on such data are those of the author alone.

8 REFERENCES


