Rail Transit OD Matrix Estimation and Journey Time Reliability Metrics Using Automated Fare Data

by
Joanne Chan

B.S. in Civil and Environmental Engineering
Cornell University, 2005

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Transportation
at the
Massachusetts Institute of Technology
June, 2007

Signature of Author ______________________
Department of Civil and Environmental Engineering
May 25, 2007

Certified by ______________________
Nigel H.M. Wilson
Professor of Civil and Environmental Engineering
Chair, Master of Science in Transportation Program
Thesis Supervisor

Certified by ______________________
John Attanucci
Research Associate for Center for Transportation and Logistics
Thesis Supervisor

Accepted by ______________________
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Rail OD Estimation and Journey Time Reliability Metrics Using Automated Fare Data

by

Joanne Chan

Submitted to the Department of Civil and Environmental Engineering on May 25, 2007, in partial fulfillment of the requirements for the degree of Master of Science in Transportation at the Massachusetts Institute of Technology.

Abstract

The availability of automatic fare collection (AFC) data greatly enhances a transit planner's ability to understand and characterize passenger travel demands which have traditionally been estimated by manual surveys handed out to passengers at stations or on board vehicles. The AFC data also presents an unprecedentedly consistent source of information on passenger travel times in those transit networks which have both entry and exit fare gates. By taking the difference between entry and exit times, AFC transactions can be used to capture the bulk of a passenger's time spent in the system including walking between gates and platforms, platform wait, in-train time, as well as interchange time for multi-vehicle trips.

This research aims at demonstrating the potential value of AFC data in rail transit operations and planning. The applications developed in this thesis provide rail transit operators an easy-to-update management tool that evaluates several dimensions of rail service and demand at near real-time. While the concepts of the applications can be adapted to other transit systems, the detailed configurations and unique characteristics of each transit system require the methodologies to be tailored to solve its needs.

The focus of this research is the London Underground network which adopted the automatic fare collection system, known as the "Oyster Card", in 2003. The Oyster card is now used as the main form of public transport fare payment in all public transport modes within the Greater London area. The two applications developed for the London Underground using Oyster data are (1) estimation of an origin-destination flow matrix that reflects current demand and (2) rail service reliability metrics that capture both excess journey time and variation in journey times at the origin-destination, line segment or line levels.

The Oyster dataset captures travel on more than three times the number of OD pairs in one 4-week AM peak period compared to those OD pairs evident in the RODS database – 57,407 vs. 17,421. The resulting Oyster-based OD matrix shows very similar travel patterns as the RODS matrix at the network and zonal levels. Station level differences are significant at a number of central stations with respect to entries, exits and

interchanges. At the OD level, the differences are the greatest and a significant number of OD pairs in the RODS matrix seem to be erroneous or outdated.

The proposed Excess Journey Time Metric and Journey Time Reliability Metric utilize large continuous streams of Oyster journey time data to support analyses during short time periods. The comparison of the Excess Journey Time Metric and the official Underground Journey Time Metric show significant differences in line level results in terms of both number of excess minutes and relative performance across lines. The differences are mainly due to differences in scheduled journey times and OD demand weightings used in the two methodologies. Considerable differences in excess journey time and reliability results exist between directions on the same line due to the large imbalance of directional demand in the AM peak.

Thesis Supervisor: Nigel H.M. Wilson
Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: John Attanucci
Title: Research Associate, Center for Transportation and Logistics

Acknowledgements

There are many people I need to thank for enriching my MIT/Boston/London experience and helping to make this thesis rewarding.

First of all, I would like to express my deepest gratitude to Professor Nigel Wilson and John Attanucci who have helped bring my greatest potentials I would not even have dreamed of. I thank you for your valuable advice and encouragements and I will miss our weekly meetings very much.

I would also like to thank Mikel Murga, Fred Salvucci, and Haris Koutsopoulos, who have been very inspiring and helpful. Also, Ginny, thank you for all your help.

Special thanks to Fabio Gordillo, who first developed the OD estimation methodology which this thesis is based on and has been dedicatedly answering all my questions and encouraging me. Also big thanks to Zhan Guo and David Uniman who have helped with the technical aspects of this research.

Transport for London has generously made this research possible and I would like to especially thank Lauren Sager Weinstein and Shashi Verma for their continuous support for this research. I would also like to thank Geoffrey Maunder, Gerry Weston, Peter James, Tony Richardson, Mike Collop, Peter Svensson and Mick Bradley at Transport for London who provided valuable help and advice both during and after my internship.

My dear friends at MIT have all been very supportive – Hong Yi, Grace Fattouche, Tanachai Limpaitoon, Vaibhav Rathi, David Uniman, Zhan Guo, Ritesh Warade and many others. My group of Hong Kong friends in Boston has always made me laugh and I am very grateful for knowing each of you – Calvin, Wilson, Raymond, Yeelok and Kwan. I must also express my most heartfelt thanks to my best girl friends – Jacqueline, Lucienne, Wendy, Chris and Erica – who have always been there for me. Also, special thanks to Leon who keeps me motivated.

Finally, thank you to my parents and sister who always love and support me, and to Terence who makes me feel deeply loved.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

With the adoption of automatic fare collection (AFC) systems, many transit agencies across the world now enjoy a wealth of information gathered every day. The number of daily transactions in an AFC system can range from tens of thousands to millions depending on the size of the transit network, the prevalence of system entry and exit controls and the types of transit smart cards in use. The availability of such data greatly enhances a transit planner's ability to understand and characterize passenger travel demands which have traditionally been estimated by manual surveys handed out to passengers at stations or on board vehicles. The AFC data also presents an unprecedentedly consistent source of information on passenger travel times in those transit networks which have both entry and exit fare gates. By taking the difference between entry and exit times, AFC transactions can be used to capture the bulk of a passenger's time spent in the system including walking between gates and platforms, platform wait, in-train time, as well as interchange time for multi-vehicle trips.

This research aims at demonstrating the potential value of AFC data in rail transit operations and planning. The applications developed in this thesis provide rail transit operators an easy-to-update management tool that evaluates several dimensions of rail service and demand at near real-time. While the concepts of the applications can be adapted to other transit systems, the detailed configurations and unique characteristics of each transit system require the methodologies to be tailored to solve its needs.

The focus of this research is the London Underground network. The Underground runs 12 lines serving 273 stations between 5AM and 1AM, carrying over 3 million passengers on an average weekday. In 2003, the Underground became part of Transport for London which adopted the automatic fare collection system, known as the "Oyster Card", in the same year. The Oyster card is now used as the main form of public transport fare payment in all public transport modes within the Greater London area. Chapter 2 provides an overview of Transport for London and transit services it provides, and discusses the Oyster Card.

The two applications developed for the London Underground using Oyster data are (1) estimation of an origin-destination flow matrix that reflects current demand and (2) rail service reliability metrics that capture both excess journey time and variation in journey times at the origin-destination, line segment or line levels. Both applications are capable of representing the rail demand and service performance during any time period and day, the most disaggregate level being a time period on a specific single day.

In addition to reflecting network demand, the OD flow matrix also serves as a building block to all other applications involving passenger demand, such as the reliability metrics mentioned above. By explicitly incorporating passenger demand in routine performance measures, a delay that affects many passengers is recognized to be more disruptive than a similar delay that affects fewer passengers. This inserts a true customer focus into any performance monitoring program that a transit agency intends to develop and maintain.

1.1 Overview of Automatic Fare Collection

From the transit agency's perspective, the major goals of an AFC system include reducing the costs of revenue collection, reducing fare evasion and fraud and speeding up the fare payment and access/egress control process (Hong, 2006). From the passenger's perspective, a transit smart card is easy to use and entitles the card holder universal access to various types of public transit where the card is accepted. This section discusses the physical configurations of AFC systems, and their benefits and capabilities from an operational perspective. A more general overview of the evolution of AFC systems can be found in Section 1.1 of Gordillo (2006) and Chapter 2 of Hong (2006).

1.1.1 *Automatic Fare Collection Systems*

While AFC systems are designed differently to meet the unique needs of each rail transit system, the AFC data collected can be generally categorized by two features of the system – access/egress control and fare validation.

An AFC system can either include or exclude gates at each rail station throughout the network. With gates, a transit agency can easily track entry (and exit) counts at each station because each movement triggers an entry (or exit) record regardless of payment type. The availability of accurate station level entry (and exit) control totals for various time periods of the day is critical to the robustness of the time period OD flow matrix which will be discussed in Section 3.3. On the other hand, transit agencies can still apply the OD matrix estimation methodology to non-gated rail networks if they can develop certain data processing functions to overcome the lack of automated entry (and exit) gate counts (Zhao, 2004; or see Section 3.2).

The fare validation function of an AFC system depends on the fare structure and the fare collection approach. Urban rail systems with a distance-based or zonal fare structure usually require fare validation at both entry and exit gates. Since a transaction record is generated at both ends of a passenger's journey, origin-destination level data can be observed directly. In addition to the stations of entry and exit, the time of each transaction is also recorded and thus the elapsed travel time can be calculated to measure service delivery quality in a general sense. Tables 1-1 and 1-2 summarize the gate control and fare validation requirements for the OD estimation and service reliability methodologies. Gated AFC systems with both entry and exit fare validation controls are easiest for the OD estimation and service reliability metric methodologies presented in this thesis because they rely on observed origin-destination data for at least a sample of passengers to make inferences about network wide characteristics. To estimate origin-destination travel patterns in systems that require fare validation only at entry, refer to Rahbee (2002), Zhao (2004) and Wilson et al (2005).

		Type of Fare Validation	
		Entry Only	**Entry & Exit**
Type of Access Gate Control	**Gated**	Feasible	Ideal
	Non-Gated	Infeasible	Feasible

Table 1-1: AFC System Requirements for OD Matrix Estimation

		Type of Fare Validation	
		Entry Only	**Entry & Exit**
Type of Access Gate Control	**Gated**	Infeasible	Ideal
	Non-Gated	Infeasible	Feasible

Table 1-2: AFC System Requirements for Service Reliability Metric

The London Underground is a partially gated system that requires both entry and exit fare validation due to its zonal fare structure. Section 2.2.2 discusses the characteristics and problems of fare and data collection in the Underground.

Three other types of non-gated fare collection approaches are also commonly used (TCRP, Report 94):

- **Pay On Boarding** – Passengers pay cash or by farecard upon boarding the vehicle in a Pay On Boarding system. This approach is most widely used in light rail, bus and bus rapid transit.
- **Conductor-Validated** – Passengers can purchase tickets before boarding or purchase on board from a conductor. This approach is most commonly used in commuter rail systems.

- **Proof of Payment (POP)** – Passengers are required to carry valid tickets or passes on board and are subject to random inspection in a POP system. It is the newest fare collection approach and has seen increasing use over the past two decades. POP is most prevalent in light rail transit systems but can also be found on every other public transport mode.

Some POP systems equipped with AFC have stand-alone card readers or readers mounted at the vehicle doors for passengers to validate their payment upon boarding (and alighting) the vehicles. Since POP systems are not gated, other types of automatic data collection techniques, such as automatic passenger count systems, are needed to obtain reliable station control totals for OD estimation. While OD estimation can be performed for POP systems that require entry-only validation using destination inference (Zhao, 2004), journey time assessment is only feasible if passengers are required to validate upon both boarding and alighting.

1.1.2 *Advantages of Automatic Fare Collection Data*

In addition to the expected benefits of lower operating costs and faster payment processing provided by AFC systems, transit agencies also recognize the richness of AFC data and its potential value in understanding passenger demand and other attributes of the passenger travel experience.

AFC data can supersede some aspects of traditional manual surveys because its collection at the time of transaction incurs essentially no marginal cost to the transit agency. In contrast, traditional manual surveys are separate undertakings aimed at a specific subset of passenger demand and are generally very costly. Due to limited resources dedicated to the manual surveys, the collected surveys inevitably provide only a small sample of all journeys undertaken in the system. AFC data represent a complete sample of transit journeys made by passengers using smart cards. Therefore, the higher the penetration of smart cards, the more representative the AFC data. Transit agencies with AFC also generally have a strong incentive to increase smart card penetration to streamline the fare payment process and reduce both ticketing and system operating costs.

The amount of post-survey processing effort required to extract AFC data is significantly less than that for manual survey data because the former are collected electronically. Traditional manual surveys typically require labor-intensive processing tasks such as data entry and logical checks, especially when the returned surveys are not completed fully and accurately. After being collected at the point of transaction, AFC records are usually uploaded to the data warehouse at the end of each day where they are stored to facilitate revenue management and support the calculation of travel statistics. The data are also near real-time – a transit planner can, at least in theory, perform analyses as soon as each day's data are uploaded. The timeliness of data extraction greatly enhances a planner's ability to track the impacts of specific events, for example, a large scale disruption in service.

1.2 Motivation

Before the introduction of AFC systems, some transit agencies employed manual surveys to estimate origin-destination level travel patterns. Since these surveys are expensive to conduct and process, the frequency and sample size of each survey are restricted by budget constraints of the transit agencies. As a result, these surveys are only conducted infrequently and may fail to capture changes in travel demand until well after they have occurred. The sample size targeted in each survey is usually calculated such that the results lie within specific error margins. The resulting sample size is usually not large enough to capture a high proportion of specific origin-destination journey patterns on the transit network. Biases in the sample, due to uneven response rates from different types of passengers and journey trip purposes, are likely to negatively affect the accuracy of the resulting OD matrix and associated planning decisions.

With the time period level OD matrix methodology developed in this thesis, planners can more easily monitor changes in travel patterns over time and therefore make informed decisions to adjust services as necessary. The major benefits of having frequent time period level OD estimates are improved service planning and customer segmentation, improved system performance measurement assessment, clearer justifications for capital investments and more informed fare policy decisions. Each of these benefits is described in the following sections.

In particular, Section 1.2.2 discusses the value of an application of two new service performance measures which will be developed in Chapter 5. The Excess Journey Time Metric is a new tool transit planners can utilize to measure actual journey times experienced by passengers by use of time-stamped smart card transactional data. The Journey Time Reliability Metric quantifies service reliability experienced by passengers from the observed journey time distributions captured by Oyster transactional data. These two service performance metrics are also applications of the time period level OD matrix since the results are weighted by passenger demand to reflect the experience of an average passenger on a line, line segment or specific station-to-station origin-destination pair.

1.2.1 Service Planning

There are four components of transit service planning that a frequently updated time period level OD matrix can enhance:

Network and Route Design

Origin-destination level passenger travel patterns are an important input to demand models which are used to evaluate alternative network and route designs. For transit planners to make informed decisions, the demand models need to predict future changes in demand and therefore it is critical that the current OD estimates are accurate. Moreover, accurate OD estimates can also be used to validate demand forecasts given current variables such as price and travel time. On congested rail lines, an examination of specific OD patterns may allow an agency to plan alternative bus services to better serve some of the shorter trips served by the most congested rail links.

Separate Estimates for Weekday and Weekend Travel Patterns

Due to budget constraints, weekend travel patterns are often not surveyed or surveyed at much smaller sample sizes than weekdays. To overcome the lack of weekend data, some transit agencies use weekday data fully or partially to estimate weekend demand patterns. However, weekend travel patterns are very unlikely to be similar to weekday demands because of the different types of activities and destinations passengers travel to. With AFC data, weekend travel patterns can be estimated much more easily and accurately. As a result, weekend transit services and line frequencies can be better tailored to the needs of passengers to access various activity centers during the weekend such as shopping centers and parks.

Frequency Setting

When setting service frequency, transit planners' objective is usually to minimize passengers' generalized cost of travel and provide enough capacity to carry all passengers subject to available resources. The definition of generalized cost of travel usually includes access and egress walk time, platform wait time and on train time. The platform wait time is dependent on service frequency because passenger arrival is considered to be uniform when service is frequent (e.g. at least every 10 to 12 minutes).

The headway also affects the dwell time because the longer the headway, the greater the number of passengers waiting, and therefore the longer the boarding time.

Special Event Planning

With an easy-to-update OD matrix, transit planners can better understand travel patterns before and after special events such as festivals and sport games. Since the travel patterns observed during these events are atypical and infrequent, it is not justifiable to allocate resources to conduct manual surveys to understand them. AFC data can provide large samples of reliably origin-destination data from previous events and allow transit planners to better understand passengers' public transport needs during these events and plan for alternative services if necessary.

1.2.2 *System Performance*

Accurate estimates of origin-destination travel patterns can enhance system performance monitoring in at least three ways.

First, OD level data can provide not only estimates of the volume carried on the system, but also estimates of passenger-miles carried. These two figures are often used to quantify the productivity of a transit system by metrics such as cost per passenger and cost per passenger-mile. For transit agencies that receive public funding, such metrics are often important performance criteria and can be inputs into funding provision.

Second, transit planners rely on OD flow matrices and path choice models to estimate passenger loads on line segments during different times of the day. The line loads are important to make sure that the system has sufficient capacity to allow most passengers to board the first available train and that in-train crowding is acceptable. Line loads are good indicators of crowding levels of different segments in the system.

Third, a time period level OD matrix can also serve as a building block for other performance measures that are customer focused. A customer focused performance measure is one that takes passenger demand into account and captures the experience of passengers in the transit system. It is desirable to incorporate passenger demand so that a service delay affecting a large number of passengers is weighted more heavily than one that affects a smaller number of passengers.

The Excess Journey Time and Journey Time Reliability metrics are customer focused measures of station-to-station journey time experienced by passengers in transit systems with both entry and exit control. Since passengers are required to validate fare payment at both entry and exit, the journey time for every trip made can be calculated from the elapsed time between entry and exit. This is an exceptionally valuable feature of AFC transactions because every journey with properly recorded entry and exit transactions provides a vast and exceptionally accurate set of journey times at a very small incremental data processing cost.

Although the Excess Journey Time Metric measures gate-to-gate travel experience by calculating the elapsed time, it is possible to disaggregate the journey time estimates by supplementing the analysis with train operational data. In rail networks with an automatic train signaling system, train arrival and departure data are collected at every station. Actual platform wait times can be calculated by combining the AFC and train operations data. Station dwell time and station-to-station on-train travel time can also be calculated. By subtracting the three train-related time components from the AFC journey time, walk time can be calculated to estimate station crowding levels that may have impeded passengers' walk speed. Station dwell time is also a proxy for crowding because it takes longer for passengers to alight when the train is crowded and the platform is packed with passengers waiting to board. Boarding time also increases when trains are crowded.

The Journey Time Reliability Metric is also superior to traditional manual surveys due to the ability to estimate journey time distributions for any OD pair which is not feasible with small sample surveys. By establishing journey time distributions, transit agencies can quantify and compare delays between and within time periods and days. With journey time distributions, transit agencies can calculate the mean as well as the variation of journey times. The mean journey time can be compared to the scheduled journey time to give the excess journey time which is a reflection of the gap between expectations and reality. The variation of journey times also provides insight into the extent of delays in passenger experience. While traditional surveys can estimate mean excess journey time at the line level, they are rarely capable of measuring delays for specific OD pairs due to the large number of samples required. Traditional surveys cannot be used to estimate the extent of delays because the surveys often do not capture the extreme cases.

1.2.3 Capital Investment and Demand Management

The time period level OD matrix and service performance metrics both provide a basis to evaluate capital investments that aim at increasing capacity and improving the customer travel experience.

By expressing crowding levels in passengers carried per rail car, line load profiles estimated from the OD matrix can provide the basis for longer trains or more frequent services needed to increase the capacity of the line. Transfer volumes at major interchange stations in a transit system can also be estimated by assigning the OD matrix using a path choice model. These transfer volumes can justify station improvements such as installing more escalators or widening passageways and platforms.

In addition to providing more physical capacity, line load profiles could also help transit agencies free up capacity along the most congested line segments by advising passengers to take less crowded routes. In a complex transit network, there are multiple paths between many OD pairs. Transit agencies can make use of the transit card registration information to notify passengers of faster paths to take or most congested interchange stations to avoid. By linking line load profiles with the JTRM, a notification program can be set up to detect frequent journeys that are delayed and notify passengers of better paths via email or text message.

The journey time distributions from the Journey Time Reliability Metric also allow transit planners to provide "worst case scenario" estimates for passengers' trip planning. Transit agencies nowadays often provide trip planning tools on the internet which usually provide several paths and associated average journey times for specific origin and destination. Passengers can be better informed if some measure of extra journey time is also provided on top of the expected journey time. For example, a trip planning tool can state that while a trip from Station A to Station B is expected to take 25 minutes, a passenger should budget an extra 10 minutes to be 95% certain to arrive on time. This information will allow the passengers who are most time-sensitive to better plan their transit journeys.

1.2.4 *Fare Policy Decisions and Monitoring*

In addition to maximizing ridership and/or revenue, some transit agencies also utilize fare policy as an innovative demand management tool. By establishing fares that spread demand temporally and spatially, transit agencies can provide financial incentives to influence passengers' travel behavior and thus increase efficient utilization of the system's capacity. With accurate estimates of the time period level OD matrix, transit agencies can make informed decisions to alter temporal and geographic demands within the network.

Most transit networks serve demands that are very intense during the AM and PM peak hours. The intensity of demand is usually greatest within a half-hour to one-hour period in each of the peaks. Transit agencies may be able to shift part of the intensity to the shoulders of the peak by adding a premium fare on passengers who enter or leave the network during the peak intervals. For example, a premium fare could be added to all exiting journeys in the central area during the busiest one-hour in the AM peak period when demand is greatest. Similarly, a premium fare could be added to all entering journeys in the central area during the busiest one-hour in the PM peak period when most passengers are commuting home.

Innovative fare policies could also be used to influence passengers' paths between specific origin-destination pairs. Transit agencies could encourage passengers to take alternative paths by compensating them with transit fare credits. This could be done by setting up additional stand-alone transit smart card readers for passengers to validate when they take the alternative paths and thus get credit rewards. Despite initial equipment costs of the additional readers, valuable capacity could potentially be "found" in congested portions of the system.

Lastly, transit agencies could easily and frequently update estimates of OD travel using AFC data to monitor the effects of fare policy changes. Theoretically, transit agencies could iteratively alter fares to achieve optimal network conditions in terms of throughput, crowding and passengers' generalized cost of travel.

1.3 Research Objectives

The primary objective of this research is to explore the potential of using AFC data, in combination with automated gate counts and survey data, to replace manual data collection methods now used for rail planning and operations. In particular, this research aims at demonstrating how AFC data can be used to improve methodologies used in the London Underground to estimate OD flow and monitor service performance developed when Oyster data were unavailable.

The primary objective is achieved through the following sub-objectives:

1. Understanding the characteristics and quality of AFC data and what adjustments are needed to correct biases and deficiencies in the data
2. Discussing how a time period should be defined to support a time period level OD Matrix
3. Demonstrating how an unbiased Time Period Level OD Matrix can be estimated from AFC and survey data (based on a previously developed methodology to estimate a Full Day OD Matrix)
4. Evaluating the AFC data-based OD Matrix against the existing survey-based OD Matrix used by transit agencies by making systematic comparisons
5. Demonstrating how excess journey times and journey time reliability factors can be quantified from AFC transactional data

1.4 Research Approach

This research relies on the availability of large samples of empirical AFC data to observe origin-destination travel patterns and journey time distributions. By understanding the characteristics and potential biases of the Oyster data, methodologies can be developed to take advantage of the strengths of the data and correct the potential biases. The methodologies of the time period level OD estimation and the service performance metrics are evaluated by comparing results with the survey-based methodologies currently used in the Underground, and suggestions can be made based on the comparisons.

1.4.1 *Understanding of AFC Data Quality and Characteristics*

Before using the Oyster data as the methodological basis, it is important to thoroughly understand the data characteristics, strengths and weaknesses. This research studies the following Oyster data characteristics:

- Representativeness of travel demand revealed by Oyster data between gated and non gated stations
- Temporal demand revealed by Oyster data
- Definition of Oyster journey time
- Representativeness of Oyster journey time distributions

1.4.2 *Developing Methodologies based on AFC Data*

When developing the methodologies, the Oyster data characteristics are taken into account and the methodologies aim at correcting the potential data biases and taking advantages of the data strengths. The large same size, ease of manipulation and flexibility of AFC data are also major reasons that make these methodologies attractive.

The time period level OD matrix estimation methodology combines current AFC data and manual surveys, and builds on the full day OD matrix estimation methodology previously developed for the Underground. Refinements needed to adapt the full day methodology to the time period level include:

- Definition of time periods
- Selection of appropriate data that pertain to a given time period
- Revised formulation for station level estimates

The Excess Journey Time and Journey Time Reliability Metrics are based on the recorded elapsed journey times between entry and exit Oyster transactions to infer journey time characteristics for each active OD pair. When developing each of the metrics, the following issues are discussed:

- Sample size requirements
- Formulation for the OD level
- Formulation for the line or line segment level using OD demand weightings

1.4.3 Evaluation of Results & Improvement Suggestions

A thorough evaluation of the methodologies is important for the Underground to understand three aspects regarding its current survey-based methodologies to estimate OD travel patterns and assess journey time performance:

- Robustness of the current methodologies
- How the current methodologies can be improved by incorporating AFC data
- How the existing survey programs can be restructured to supplement the AFC data-based methodologies proposed in this research.

The evaluation of the AFC data-based OD estimation methodology includes both qualitative and systematic comparisons with the survey-based methodology and the resulting OD matrix. The systematic comparisons between the two matrices are made on various levels: network, zonal, station and OD levels. In addition to comparing the numerical results, possible explanations are also given for the discrepancies observed.

The evaluation of the service performance metrics includes a comparison between the excess journey time results from the official Underground reports and the proposed Excess Journey Time Metric. Since there is no equivalent measure of journey time reliability in the Underground, the Journey Time Reliability Metric results will be presented independently.

1.5 Thesis Organization

This thesis is organized into seven chapters, covering methodologies of the time period level origin-destination flow matrix and the service performance metrics:

- Chapter Two provides the background of the institutional organization of Transport for London, the transit services it provides and the AFC system used.
- Chapter Three presents the literature review of previous research on origin-destination flow estimation in transit network, the existing survey-based method currently used by the Underground, and the refinements needed to estimate a time period level OD matrix based on the AFC data-based methodology developed by Gordillo (2006).
- Chapter Four evaluates the time period level OD matrix for the AM Peak estimated by the AFC data-based methodology and compares travel patterns with those revealed in the current survey-based methodology.
- Chapter Five introduces the Excess Journey Time Metric and Journey Time Reliability Metric by presenting prior research on monitoring transit service reliability, discussing the existing methodology used by the London Underground and proposing AFC data-based methods that capture both excess journey time and journey time reliability.
- Chapter Six compares the results between the Excess Journey Time Metric and the existing survey-based methodology, and presents the Journey Time Reliability Metric results independently.
- Chapter Seven summarizes the research findings, discusses possible improvements to enhance current practices and the methodologies, and suggests future research topics.

Chapter 2 Introduction to Transit in London

Transport for London (TfL) was created in 2000 by the Greater London Authority which is a form of citywide government for London consisting of an elected Mayor of London and a 25-member assembly. The London Underground became part of TfL in 2003.

Transport for London is under the direct control of the Mayor and is responsible for delivering the Mayor's Transport Strategy through[1]:

- Managing the London public transport modes, including London Underground, London Buses, Croydon Tramlink, the Docklands Light Railway, and London River Services
- Managing London's road networks and traffic lights
- Regulating taxis and minicabs
- Coordinating alternative transport modes such as Dial-a-Ride for passengers with mobility problems.

In addition to direction over Transport for London, the Mayor also has powers in:

- Setting Transport for London's budget
- Deciding the public transport fare structures in London
- Influencing the National Rail services within Greater London
- Funding new transport services and investing in new transport systems

Section 2.1 provides an overview of transit services in London and Section 2.2 introduces the automatic fare collection system in London, known as the Oyster Card, and discusses issues in Oyster fare and data collection.

[1] http://www.london.gov.uk/gla/transport.jsp, retrieved May 14, 2007.

2.1 Transit Services in London

According to the London Travel Report in 2005, about 27 million journeys are made in Greater London on an average weekday. As shown in Figure 2-1, 42% of journeys are by car, motorcycle or taxi, 35% by transit and 22% by bicycle or on foot.

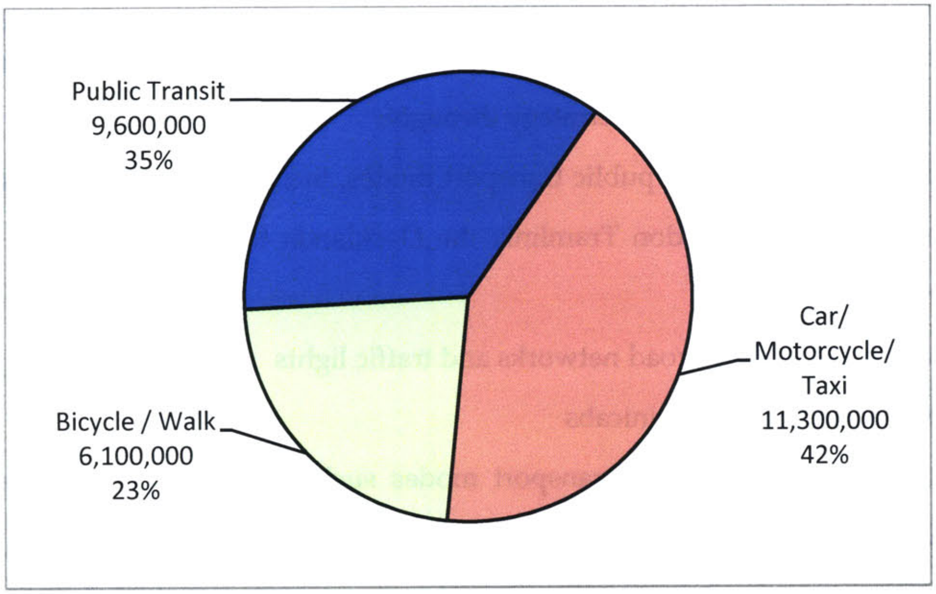

Figure 2-1: Distribution of Daily Journeys in Greater London

The 9.6 million transit journeys on an average weekday are served by five transit modes in London, as shown in Figure 2-2.

- Bus is the most widely used public transit mode in London, carrying half of all passengers on 6,800 scheduled buses and 700 different routes[2]. The bus network has grown significantly in the past several years and bus usage rate is growing at its fastest since 1946. TfL acknowledges that buses are the best option for increasing transit capacity in the short-term and many initiatives are in place to improve bus services.
- The Underground is the second most widely used transit mode, carrying 27% of all transit demand. The Underground network serves 273 stations on 12 lines

[2] http://www.tfl.gov.uk/corporate/modesoftransport/1548.aspx, retrieved May 14, 2007.

and 253 miles of track. Over the next five year, TfL will spend over half of its £10 billion investment budget to improve and expand the Underground infrastructure.

- National Rail journeys constitute 21% of all transit demand and half of these journeys are within Greater London. TfL is committed to deliver high standards and services for London's rail users similar to standards experienced on buses the Underground and provide better integration of rail services with other transit modes.
- The Docklands Light Rail (DLR) serves the redeveloped Docklands area east of central London with 38 stations on 4 routes and 19 miles of track. The DLR carries less that 1% of all transit demand.
- The Tramlink (also Croydon Tramlink) is also part of London's transit network but it is not reported in the London Travel Report due to its low patronage compared to the other transit modes. The Tramlink operates in the Croydon area south of central London which is relatively under-served by the Underground. It serves 39 stops on 3 routes and 18.5 miles of shared street track, dedicated street track and off-street track.

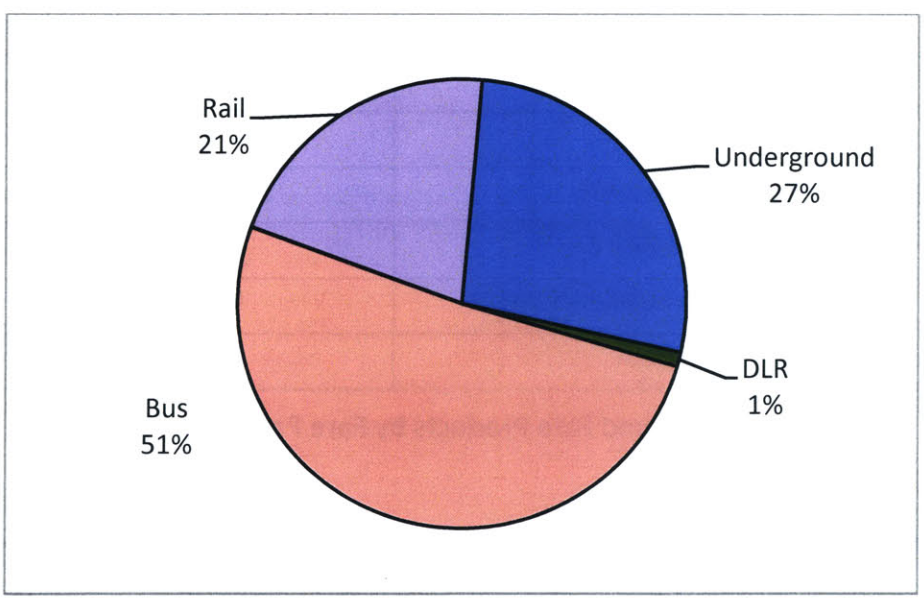

Figure 2-2: Distribution of Daily Transit Journeys in Greater London

2.2 Automatic Fare Collection in London

London's transit smart card, known as the Oyster card, was introduced by Transport for London in 2003 and has since expanded to include many fare types that were previously issued on magnetic stripe tickets. Oyster card allows passengers to store Pay-As-You-Go (PAYG) credits and up to three ticket contracts. PAYG credits are "stored value", from which fares are deducted upon using TfL's services. Ticket contracts are fare products with specific restrictions on spatial and temporal validity, as well as public transport mode. The most common type of ticket contract is the Travelcard that entitles passengers to unlimited usage within specified spatial and temporal bounds. Specific ticket contracts are also offered for the elderly, the disabled and TfL staff.

This section focuses on AFC in the Underground. Table 2-1 illustrates the combinations of fare product and payment media available on the Underground.

Fare Product Type	Fare Media Available
National Rail – Underground "Combo" Tickets	Magnetic
Single and Return Tickets	Magnetic and Oyster PAYG
1-Day Travelcard	Magnetic and Oyster PAYG
3-Day Travelcard	Magnetic
7-Day Travelcard	Oyster Ticket
Monthly Travelcard	Oyster Ticket
Annual Travelcard	Oyster Ticket
Elderly / Disabled Freedom Pass	Oyster Ticket
Staff Pass	Oyster Ticket

Table 2-1: Underground Fare Products by Fare Payment Media Type

All passengers are required to hold a valid ticket while traveling on the Underground. Cash is not accepted at Underground gates so in order to gain access, passengers must purchase a fare medium at the station if they do not already have an Oyster card with sufficient PAYG credits or a valid Travelcard. Passengers may receive a penalty fare or be prosecuted if unable to present a valid fare payment medium.

In order to develop an Oyster-based OD estimation methodology, it is essential to understand the Oyster fare policies pursued by Transport for London to encourage Oyster usage and the characteristics of data available in the Oyster transaction database.

2.2.1 *Oyster Fare Policies*

Transport for London has been actively pursuing fare policies that promote Oyster usage by providing significant monetary incentives. These fare policies have been successful, and as of March 2007, approximately 70% of journeys on the Underground are made using Oyster. Figure 2-3 shows the increase in Oyster use in the Underground since mid 2005. The London Underground operates on a 13-period calendar which starts on April 1st each year.

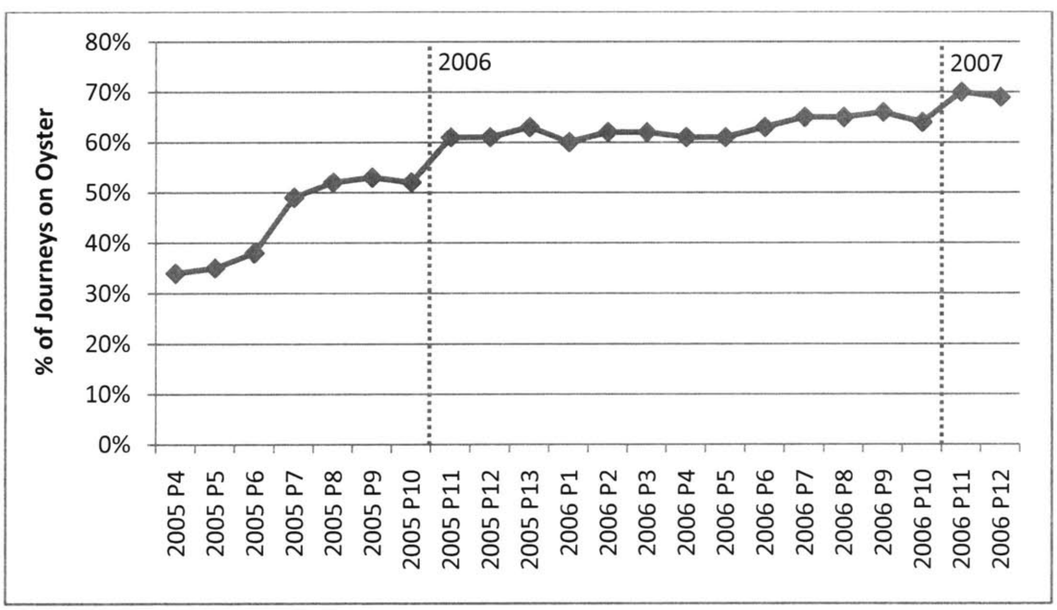

Figure 2-3: Oyster Use in the Underground[3]

[3] Data from Oyster Monthly Report February 2007

Fare Structure

Since January 2006, the single cash fare (i.e. magnetic ticket) has been at least double the equivalent Oyster Pay-As-You-Go fare. The significant differences in Oyster PAYG and cash fares help occasional Underground passengers justify the £3.00 deposit requested for the purchasing of an Oyster PAYG card. Table 2-2 illustrates the contrasting fares between Oyster PAYG and Cash.

	Oyster PAYG	Cash Fare
January 2005	£1.70	£2.00
January 2006	£1.50	£3.00
January 2007	£1.50	£4.00

Table 2-2: Peak Fares for Zone 1 Travel[4]

Figure 2-4 shows the changes in Oyster PAYG and cash fare use as a result of changes in fare structure. After the fare change in January 2006, the use Oyster PAYG increased immediately from 9% to 13%, and continued to increase to 23% by the end of the year. During the same period, the use of the cash fare decreased from 12% to 7% within a month, and continued to decrease to 5% by the end of the year. As of February 2007, 24% of all Underground journeys are Oyster PAYG and 4% cash.

Daily Capping

Oyster PAYG was introduced in the Underground in January 2004 and the world's first daily fare capping scheme was launched in February 2005. Daily capping means that the fare collection system keeps a running total of Oyster PAYG journeys made on a specific Oyster card each day and ensures that when the cap is reached the total fare deducted is always £0.50 less than the equivalent magnetic stripe 1-Day Travelcard. Oyster PAYG users are therefore promised the lowest fares and no longer have to decide on a payment type at the beginning of the day.

[4] Fare information provided by Tony Richardson (Fares & Ticketing, Transport for London)

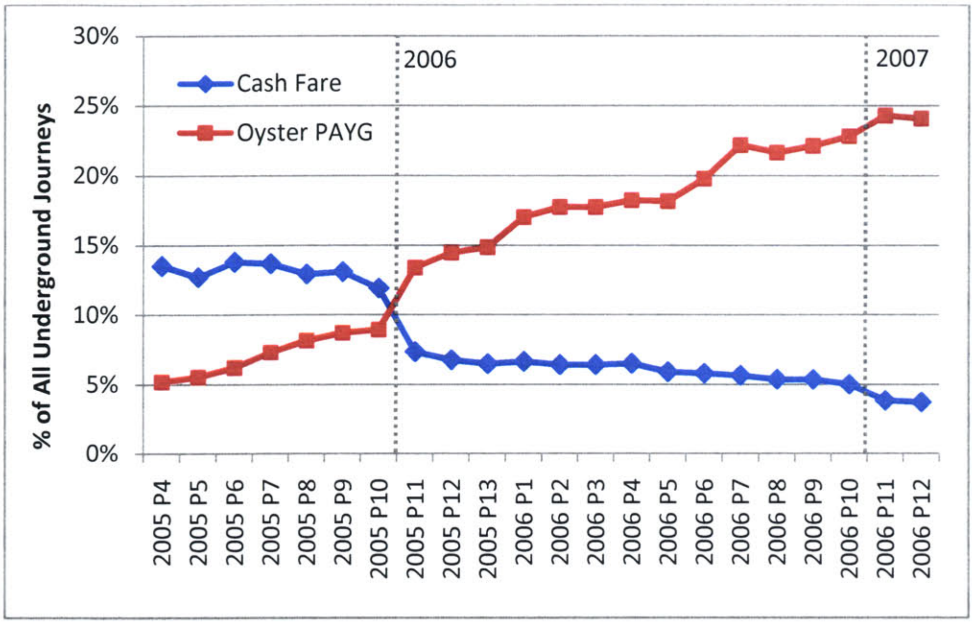

Figure 2-4: Oyster PAYG and Cash Fare Use in the Underground[5]

Oyster PAYG Max Fare

Starting in November 2006, a maximum cash fare is deducted from an Oyster PAYG card if the passenger does not validate the card at both entry and exit. The amount of maximum cash fare is specific to the zone of the validated journey end. For example, if a passenger making a Zone 1-to-Zone 1 journey fails to validate at exit, the maximum cash fare of £4 will be charged instead of the £1.50 Zone 1-to-Zone 1 PAYG fare.

While a passenger using a Travelcard can travel on any National Rail service within the London fare zones for which the Travelcard is valid, only a small number of National Rail services within the Greater London area currently accept Oyster PAYG. For journeys starting at certain Underground stations with National Rail connection where Oyster PAYG is accepted, an entry charge of £4 or £5 will be deducted depending on the station configuration and on exit the card balanced is adjusted to the advertised fare[6]. The £4 entry charge, representing the Oyster maximum PAYG fare, is deducted at gates

[5] Data provided by Tony Richardson (Fare & Ticketing, Transport for London)
[6] http://nationalrail.co.uk/times_fares/oystercard.html, retrieved May 12, 2007.

that lead only to Underground services. At gates that lead to both Underground and National Rail services, the £5 entry charge is deducted representing the average rail fare within Greater London.

In addition to being a higher fare, any maximum cash fare charged will not count towards the Oyster card's daily price cap. Thus the maximum fare will be deducted even if the Oyster card has reached the daily cap if the passenger fails to validate at both entry and exit.

In February 2007, three months after the implementation of Max Fare, it was estimated that an average of 18,000 passengers, or 0.6% of all Underground journeys, were paying the max fare each day.

Availability of Oyster on National Rail

To expand the coverage and usage of Oyster, the mayor of London has offered the National Rail operating companies grants to install Oyster equipment. Oyster availability on National Rail within the Greater London area would allow passengers to carry only one fare medium instead of separate payment types for National Rail and TfL services.

The promotion of Oyster on National Rail was a logical next step after National Rail adopted the Underground fare zones for services within the Greater London area in January 2007. Previously, the 330 National Rail stations affected serve 97,300 station-to-station combinations which had their own set of point-to-point fares. The adoption of Underground fare zones simplified the fare structure to only 21 zone-to-zone combinations. For station-to-station pairs that run on routes shared between the National Rail and the Underground, the Underground fare would be charged as before the adoption.

2.2.2 *Fare and Data Collection*

Although Underground passengers must validate their fare payment media at both entry and exit to be charged the correct zonal fare, some Underground stations do not have gates to enforce payment. Of the 273 Underground stations, 48 are non-gated meaning some or all passengers can enter or exit the Underground platforms without passing through a gate. At these stations, passengers holding magnetic stripe tickets are not required to validate their payment but Oyster cards must be validated at stand-alone card readers. London Underground employs a team of inspectors to control fare evasion by asking passengers on platforms or on board trains to show their validated payment. Passengers failing to present proof may be charged a £20.00 penalty or be prosecuted.

The access gate control situation in the Underground is further complicated by the fact that 63 stations have connections with National Rail. Passengers who regularly interchange between National Rail and the Underground at these stations usually hold a National Rail Travelcard Season Ticket[7] which entitles them to:

1. Travel between a specific National Rail origin or destination station to the boundary of London Travelcard Zone 6, and
2. Unlimited travel within all 6 London Travelcard zones on the London Underground, National Rail services, as well as two other rail services (Docklands Light Railway and Croydon Tramlink), excluding bus.

National Rail Travelcard Season Tickets are available only on magnetic stripe tickets and they constitute a major portion of all magnetic stripe journeys in the Underground. As mentioned previously, passengers with magnetic stripe tickets are not required to validate their payment. This is particularly relevant at non-gated Underground stations

---

[7] http://www.nationalrail.co.uk/times_fares/season_tickets/

with National Rail connections because most of the passengers transferring between National Rail and the Underground carry tickets that are valid on both networks.

Figure 2-5 illustrates possible access control layouts at gated Underground stations with National Rail connection. At these stations, the National Rail section may or may not be separately gated. Since passengers coming from both the National Rail and the street are required to validate their payment at gates, the automated gate counts capture all Underground entry and exit transactions.

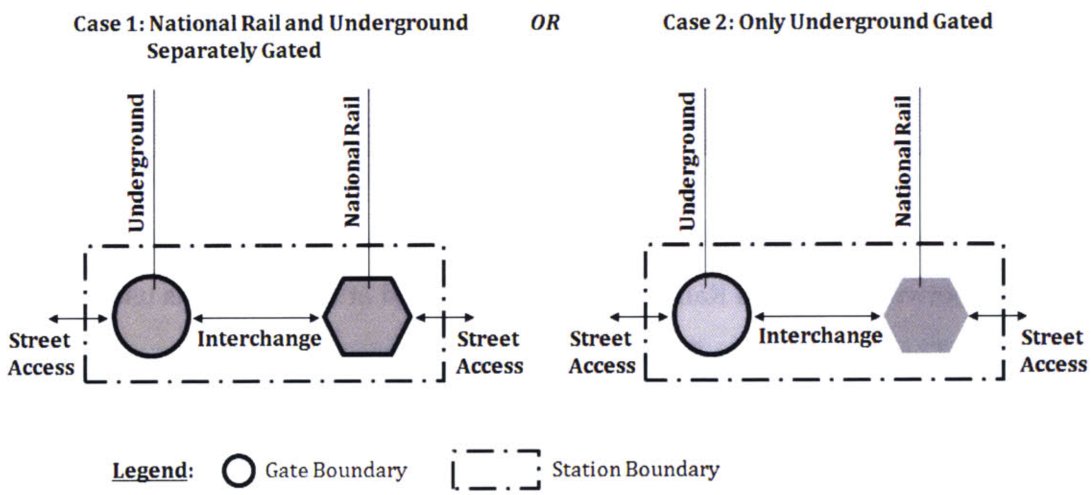

Figure 2-5: Layouts of Gated Stations with National Rail Connection

Figure 2-6 illustrates possible access control layouts at non-gated Underground stations with National Rail connections. At these stations, the National Rail section may not be separately gated from the Underground and the combined sections may or may not be gated from street access. Since passengers transferring between National Rail and the Underground are not required to pass through gates, the automated gate counts at these stations do not capture all Underground entries and exits.

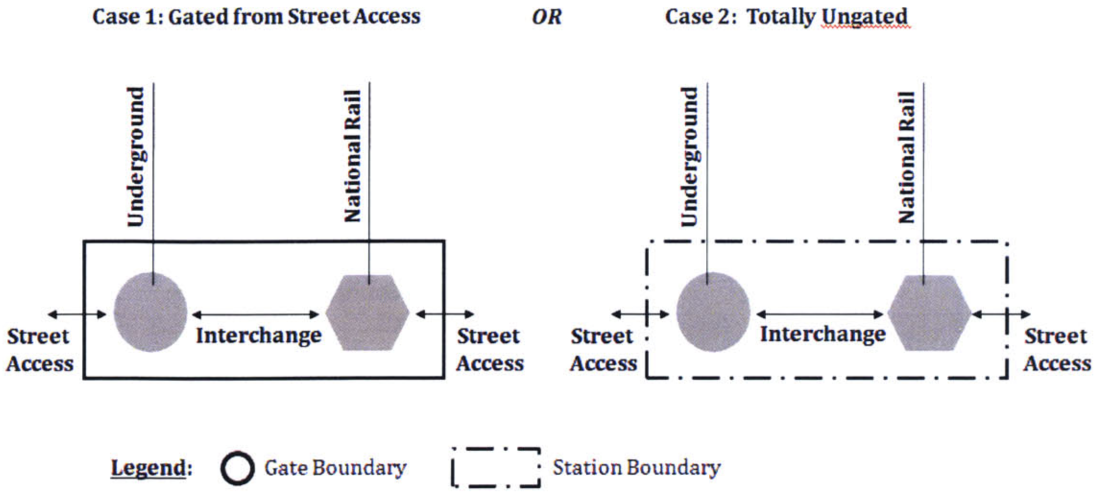

Figure 2-6: Layouts of Non-Gated Stations with National Rail Connection

Of the 63 National Rail / Underground interchange stations, 19 are gated and 44 are non-gated. The interactions of these 63 interchange stations with the Underground access gate control types and station ridership estimate types are summarized in Figure 2-7. Gate counts include both magnetic stripe and Oyster transactions as well as manual counts which are conducted annually. Some gated stations have manual counts because their gate counts are believed to be inaccurate.

In this research, all Underground stations are re-grouped according to the reliance on manual counts in estimating station level ridership:

- **Fully Gated (FG) Stations** – they are gated stations with reliable gate counts and no manual counts. There are 205 FG stations, 19 of which have National Rail connections.
- **Non Fully Gated (NFG) Stations** – they are gated or non gated stations with manual counts. There are 68 NFG stations, 41 of which have National Rail connection.

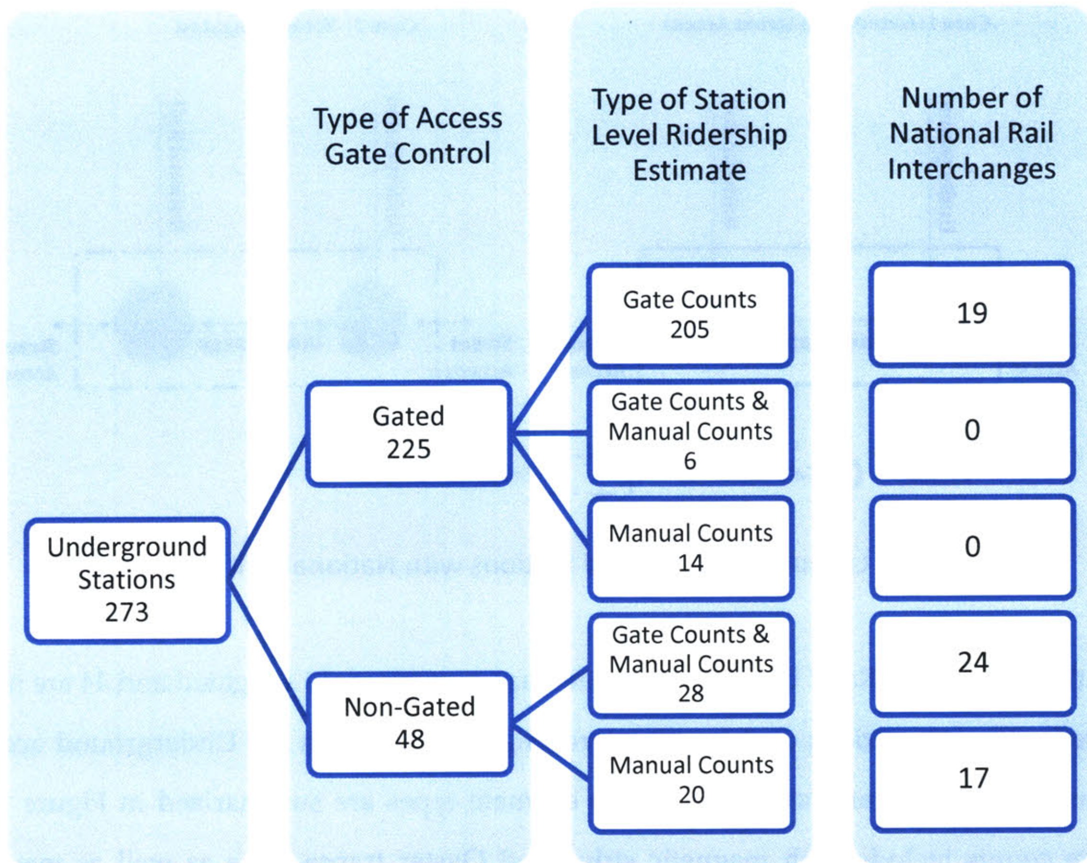

Figure 2-7: Classification of Underground stations by Type of Access Control and Ridership Estimate

2.2.3 *Types of Oyster Transactions*

Ideally, Oyster transactions are recorded at both entry and exit gates. Incomplete Oyster transactions are generated when passengers do not validate their Oyster cards at one end of the journey. The incomplete ends usually occur at non fully gated stations for one, or more, of the following reasons:

- Oyster cards are not validated because there are no gates at the entry (or exit) station
- Gates are left open for passengers to pass through because there is no station staff on duty
- Gates are intentionally left open as a station management tool when there is a sudden influx of passengers, usually at National Rail connections when

hundreds of arriving passengers interchange from National Rail to the Underground

For all practical purposes in the following analysis, four types of Oyster journeys are considered:

1. **Completely Documented (CD) Journeys:** they contain distinct origin and destination stations
2. **No Travel (NT) Journeys:** they start and finish at the same station within a short period of time
3. **Unfinished (UNF) Journeys:** they contain only origin information because there is no exit transaction
4. **Unstarted (UNS) Journeys:** they contain only destination information because there is no entry transaction

Completely Documented Journeys are most common on OD pairs between fully gated stations. No Travel Journeys mostly correspond to staff passes used to let passengers through the gates in cases of gate or card failure. On the other hand, no travel journeys made by non-LU staff are of special interest in understanding station crowding because it is believed that some passengers leave the station and seek alternate modes of transport if they see very crowded conditions at the platform. Lastly, unfinished and unstarted journeys are generally classified as Incompletely Documented (ID) Journeys which result from open gates or lack of gates, and in some cases, fare evasion. Further investigation is needed to understand the proportion of incompletely document journeys due to fare evasion.

Chapter 3 Time Period Level OD Matrix

AFC data have enhanced transit agencies' ability to obtain large samples of current origin-destination information. The goal of this chapter is to provide a description of previous and current research on OD estimation in transit networks in general, and also in the London Underground before and after the availability of AFC data. Section 3.1 introduces the survey method currently used in the London Underground to estimate OD passenger travel. Section 3.2 provides an overview of previous research on OD estimation in public transit networks, including a detailed discussion of the full day OD estimation methodology combining AFC and survey data previously developed by Gordillo (2006) for the London Underground. Section 3.3 extends Gordillo's methodology by refining the process to estimate OD demand at the time period level.

3.1 OD Flow Estimation in London Underground

This section provides an overview of the RODS process focusing on issues that are most closely related to the proposed Time Period Level OD Flow Estimation in Section 3.3. Refer to Section 2.3 of Gordillo (2006) for more general background on RODS.

3.1.1 RODS Process and Outputs

In order to estimate passenger travel patterns in the complex London Underground network, RODS was designed to incorporate passenger surveys from a sample of Underground stations over multiple years and scale results to November entry and exit counts of each year. November is the month of heaviest loading in the Underground and therefore for planning and scheduling purposes it is the basis for estimating the RODS matrix.

RODS surveys cover about 30 to 40 stations in the fall of each year and approximately 8 to 10 years are required to cover all stations in the network. The selection of stations to be surveyed depends on observed or expected changes in station ridership and/or service provision. The number of surveys handed out at each station is based on a target sample size, estimated hourly ridership and expected response rate. On the assigned date, surveys are randomly distributed to passengers entering the station. Passengers return the surveys by mail upon completion of their trips.

The RODS questionnaire, shown in Appendix A, asks passengers to provide details of the trip being made when the questionnaire is distributed. Requested data include access to the station, trip purpose, path taken within the Underground, times of entry and exit, postal codes of final origin and destination, ticket type and various personal characteristics.

The next step is to obtain estimates of station entry and exit counts to serve as control totals in estimating the final matrix. The station estimates rely heavily on automated gate counts (magnetic stripe and Oyster journeys) and a manual passenger flow survey is conducted each year to supplement the gate counts at non-gated stations and selected gated stations where gate count data are believed to be unreliable. Appendix B presents the classification of all Underground stations. Of the 225 gated stations, 205 rely exclusively on gate counts, 14 on manual counts and the remaining 6 on both gate and manual counts. Of the 48 non gated stations, 20 rely exclusively on manual counts while the remaining 28 stations use both gate and manual counts.

By combining the station control totals and RODS survey responses, a seed origin-destination is developed as follows:

1. RODS responses from the current and past years are combined and assigned to the appropriate 15-minute time interval. For intervals with too few or zero

responses, survey records from adjacent intervals at the same entry station are duplicated.

2. Each survey record is multiplied by an expansion factor such that the expanded number of records matches the total station entry counts for each 15-minute interval.
3. An iterative adjustment procedure is used to match the distributions of passenger age and gender, journey purpose and ticket type observed in the Underground User Survey which is an interview of randomly selected passengers on station platforms.

While the RODS surveys are distributed to passengers in the autumn of each year, the results are usually made available internally in October of the following year. Major RODS outputs include:

1. **Origin-Destination Flow Matrices** – specifies the station-to-station passengers flow by 15-minute time interval based on station entry times.
2. **Route Choice** – specifies the flow on each revealed path for each origin-destination pair in the OD Matrix by 15-minute interval.
3. **Station Flow Counts** – specifies directional passenger flow at various points in stations where the manual count survey is conducted including movements that may not be related to LUL journeys, for example, passengers using the Underground stations to access National Rail service.

3.1.2 Assessment of RODS

Seven years after the introduction of RODS in 1998, RODS 2005 became the first origin-destination dataset that contains survey data from all stations, except Euston Square. While the information obtained through RODS is extremely valuable, especially the revealed path choice, several aspects of the process are believed to make the outputs less than perfectly reliable.

The weaknesses of RODS are:

1. **Small Sample Sizes** – Depending on the targeted sampling rate and actual response rate, the expansion factors attached to the survey records vary significantly across stations and by time-of-day. For the AM Peak, 42% of OD pairs in RODS use a single survey response and on average each of the survey response represents 18 journeys in the RODS matrix. Another 40% of OD pairs in RODS have 2 to 5 survey responses in the AM Peak and each survey represents 14 journeys on average. At the 15-minute interval, the expansions are even larger because each response can be assigned to only one 15-minute interval and therefore many intervals are based on a small numbers of survey responses, some of which may be duplicated from an adjacent time interval. Large expansion factors can result in potential inaccuracies in the resulting OD matrix.
2. **Low Response Rate** – The response rate for RODS surveys has been between 20 to 30 percent in recent years, which potentially introduces bias in terms of the travel characteristics of the responding passengers.
3. **Inclusion of Obsolete Travel Patterns** – Due to the rolling nature of the survey program, RODS is incapable of capturing changes in travel patterns over time since most stations are surveyed every 8 to 10 years. In fact, the RODS outputs represent multi-year rolling average travel patterns rather than true annual estimates. Section 4.5 will present some examples of obsolete travel patterns.
4. **Lack of Direct Measurement of Weekend Travel Patterns** – Since RODS surveys are distributed only between 7AM and midnight during weekdays, the Saturday and Sunday matrices are estimated based on the assumption that these travel patterns are similar to those observed during weekdays and so weekday RODS matrices are simply scaled to the corresponding weekend counts. This means that the weekend matrices capture the weekday rush hour characteristics which are not representative of weekend travel patterns.

5. **Inaccurate Station Exit Totals** – The RODS iterative adjustment procedure balances journeys to match station entries in the last step, and therefore exit totals may not be accurate. Section 4.4 will illustrate this problem.

3.1.3 Applications of OD Estimation at London Underground

The major output of the RODS survey program is an annual estimate of origin-destination station level travel demand matrices at 15-minute intervals. In addition to providing an estimate of travel demand, the RODS matrix also supports other modeling, monitoring and planning tools in the London Underground, as illustrated in Figure 3-1.

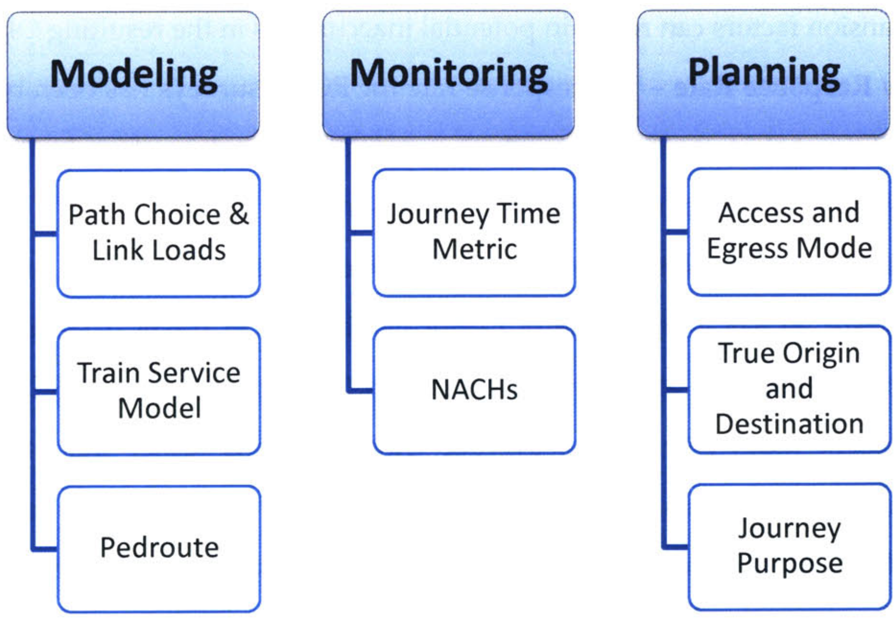

Figure 3-1: Applications of RODS OD Matrix

The RODS OD Matrix enhances three models used by the Underground service planners:

- The RODS survey is the only tool with which to validate path choice models as it provides passenger reports on the specific interchange stations used in their journeys. In a complex network such as the London Underground, an accurate path choice model is critical to the calculation of link loads which helps depict system performance and crowding.

- The Train Service Model is a simulation model that assesses the impacts of disruptions such as delays and cancellations on line operations. The model takes RODS OD demands, train line loads derived from RODS and train schedules as inputs.
- The Pedroute model is a station-based simulation model that represents both passenger and train movements within a station complex. It is primarily used to assess delays and congestion at station passageways by taking into account passenger flows at each station entrance, exit and walkways.

The RODS matrix is also used as an input to two service performance measures:

- The Journey Time Metric (JTM) is an overall Underground service performance measure that emphasizes customers' perspectives by evaluating performance in excess journey time components. The JTM uses the RODS OD demand and link loads to aggregate component level results. The JTM methodology is discussed in detail in Section 5.2.
- The NACHS (Nominally Accumulated Customer Hours) system assesses the impacts of service and mechanical disruptions on passenger journey time. NACHS values are assigned by incident type, duration, location and corresponding RODS passenger demands. The NACHS tables are an integral part of the Public-Private Partnership contract between LUL and train operators to set the basis for service performance assessment.

Finally, the RODS survey also provides valuable information to the Underground planners about certain characteristics of passenger journeys. Such information includes access and egress modes for the Underground journey, actual path taken in the Underground network, the purpose of the journey and its true origin and destination addresses. These questions provide insight into how passengers use the system and how best the Underground network can function in a complex multi-modal network.

3.2 Literature Review

This section focuses on previous research that tackles the problem of origin-destination (OD) estimation for transit networks. Rail and bus networks usually require slightly different estimation methodologies due to differences in data availability and network structure. Section 3.2.1 reviews the Iterative Proportional Fitting technique which can be applied to any type of input data to estimate OD matrices. Section 3.2.2 reviews two works on OD estimation for entry-only rail transit systems using automatically collected data. Lastly, Section 3.2.3 presents in detail an OD estimation methodology for entry-and-exit rail system using AFC and manual data for the London Underground.

3.2.1 Iterative Proportional Fitting

The Iterative Proportional Fitting (IPF) method estimates an overall passenger OD flow matrix from any sources of partial OD data. IPF involves obtaining a seed matrix and scaling it to match entry and exit counts. The AFC data-based OD estimation methodology developed by Gordillo presented in Section 3.2.3 uses the IPF method. Based on Gordillo's work, the time period level OD matrix estimation methodology developed in Section 3.3 of this thesis also uses the IPF method.

The Iterative Proportional Fitting method combines a seed origin-destination matrix with entry and exit counts, usually on a single route level, to estimate the origin-destination matrix. A doubly constrained OD matrix that matches both station entry and exit counts can be obtained by iteratively solving the simultaneous equations [3-1] to [3-3]:

$$\mathbf{T_{ij} = a_i * b_j * t_{ij}} \qquad \forall\, \mathbf{i, j} \qquad \textbf{[3-1]}$$

$$\mathbf{\sum_j T_{ij} = M_i} \qquad \forall\, \mathbf{i} \qquad \textbf{[3-2]}$$

$$\mathbf{\sum_i T_{ij} = N_j} \qquad \forall\, \mathbf{j} \qquad \textbf{[3-3]}$$

Where

T_{ij} is the total passenger flow between origin segment (or stop) i and destination segment (or stop) j,

t_{ij} is the seed matrix flow between i and j,

$a_i \geq 1$ is the row factor,

$b_j \geq 1$ is the column factor,

M_i is the total entry count for segment (or stop) i, and

N_j is the total exit count for segment (or stop) j.

The matrix cells having zero OD flows due to the structure of the matrix are called structural zeros, and those due to the low sampling rate of low OD flows are called non-structural zeros or sampling zeros. If the matrix contains consistent data and no structural zeros, IPF will yield a unique solution (Ben-Akiva, 1987). If the potential matrix has many OD pairs with no observed trips, the IPF will result in non-zero flows for only a portion of the OD pairs which have flow. This probability of producing a matrix with little or no travel on a significant number of potential OD pairs is the main concern in using IPF.

The method requires a representative sample of OD movements to construct the seed matrix which undergoes an iterative procedure to the match the control totals. While the accuracy of the control totals is critical to the resulting OD matrix, the seed sample must also be representative of all journeys made in the system.

Maximum Likelihood Estimation (MLE) is also a popular method to estimate OD flow matrices by making inferences about parameters of the probability distributions of the partial OD data, and the entry and exit counts. For more detailed formulations of the IPF and MLE techniques, refer to Cui (2006, Section 2.1). Other approaches that are based on statistical methods with known properties and other more ad-hoc procedures can be found in Ben-Akiva (1987).

3.2.2 *Entry-only Transit System OD Estimation Using Automated Data*

Since different transit agencies have different data availability and system characteristics, prior research on OD estimation using automated data have had specific objectives and methodologies. There are also differences between rail and bus systems in the ease of determining journey origins and destinations. The location is usually known when a rail passenger trip transaction occurs but it is not necessarily the case for bus. This section focuses on OD estimation methodologies for rail and bus systems which require only entry fare validation.

Destination Inference for Rail Systems using AFC data (Zhao, 2004)

Most rail systems in the US are entry-only, including New York City Transit (NYCT) and Chicago Transit Authority (CTA), and the exit stations of journeys cannot be obtained directly from farecard data. Trip chaining is used to infer the destination of a rail passenger in this kind of system, based on the following three assumptions:

1. The destination station of one trip is the origin station of the following trip, or another station in close proximity. (the "next trip" method)
2. The destination station of the last rail trip of the day is the same as the origin station of the first rail trip of the day. (the "last trip of the day" method)
3. If a bus trip follows a rail trip, the intersection between the bus and rail routes indicates the destination of the rail trip. (Zhao 2004)

The "next trip" and "last trip of the day" methods allowed inference of the destinations of nearly 80% of farecard entries in the CTA study. In order to account for the rest of the farecard journeys as well as non-farecard journeys, the seed matrix obtained from the trip chaining method needs to be scaled to match the station entry and exit totals to estimate a network level OD matrix. The resulting matrix reveals only origin-destination travel patterns without information on path choice. A separate assignment process is needed to estimate which lines and interchange stations passengers use in their journeys.

Further descriptions of the trip chaining method for destination inference can be found in Barry (2002), Rahbee (2002), Zhao (2004) and Zhao, Wilson and Rahbee (2006).

OD Estimation for Bus Systems using AFC, APC and AVL data (Cui, 2006)

This OD estimation methodology requires that the bus network is equipped with multiple automatic data collection systems to obtain three required types of data:

1. Boarding and alighting counts for all stops or segments, known as the marginal values, obtained from Automatic Passenger Count (APC) and Automatic Vehicle Location (AVL) systems
2. A sample of passenger trips with known linked boarding and alighting stops or stop segments, known as the seed matrix, obtained from AFC data
3. Total transfer flows by route and direction obtained from AFC data

Given the required data sources, the general approach consists of three steps:

1. **Obtaining the Marginal Values** – Boarding and alighting counts (or marginal values) can be aggregated from the raw APC data and the locations of the counts can be obtained by linking the transaction times in the APC data with the AVL data. Limitations of data availability may seriously hinder this methodology by introducing bias and inaccuracies in the results. Data availability may be limited by the absence of or partial coverage of AVL or APC systems.

2. **Obtaining the Seed OD Matrix** – Since the transit network requires only entry fare validation, the destination of each unlinked bus trip can be inferred from the AFC data using the trip chaining method described above. The destination of the current trip is assigned differently depending on the mode of the next trip:
    a. If the next trip is a rail trip, the bus stop on the current trip that is closest to the rail station is identified as the alighting stop of the current trip.
    b. If the next trip is a bus trip but on a different route (or if the route travels in different directions), the bus stop on the current trip that is closest to the boarding stop of the next trip is identified as the alighting stop of the current trip.

    The known origins and inferred destination of trips are then aggregated to form the seed OD matrix.
3. **Single Route OD Estimation** – With the marginal values and seed matrix, the OD matrix can be estimated by applying either the Iterative Proportional Fitting or Maximum Likelihood Estimation procedure described in Section 3.1.1.
4. **Network Level Linked Trip OD Estimation**
    a. **Linked Trips** – A linked or transfer trip includes one, or more, unlinked trips, each consisting of boarding and alighting on a single bus, connecting the origin to the ultimate destination. Linked trips seldom contain more than two unlinked trips which are defined to be within a specified time of each other to be defined as the same linked trip. The origin and destination stops of the linked trip are therefore the boarding stop of the first unlinked trip and the alighting stop of the last unlinked trip, respectively.
    b. **Non Transfer Flow OD Matrices** – Total transfer counts can usually be obtained from AFC data. The transfer flows are separated from the single route OD matrices and distributed to eligible origins and destinations (on different routes) over the bus network using a proportional distribution method or a modified iterative proportional fitting method (Cui, 2006).

The result is a system of linked trip segments which can be removed from the interim individual route OD matrices to obtain a set of single route OD matrices that contain only non transfer flows.

c. **Transfer Flow OD Matrices** – In order to obtain the transfer linked trips, the distribution of the origin of the first unlinked trips and the distribution of the destination of the last unlinked trips in the trip series are linked. Once the transfer trips are known, they can be combined with the non transfer single route matrices to obtain the network level OD matrix.

Although the ultimate goal of this methodology is a network level OD matrix for an entry-only bus system, the intermediate products such as the single route OD matrices and the transfer flows also provide significant planning value.

3.2.3 *Entry-and-Exit Rail System OD Estimation Based on AFC Data*

This section discusses Gordillo's methodology to estimate the OD matrix for an average weekday in the London Underground in one TfL accounting period (corresponding to a 4-week period in LUL's calendar), without distinguishing between time periods within the day. Section 3.3 will then discuss the refinements required to adapt the methodology to estimate OD flow for a time period within the day.

The survey-based OD estimation methodology currently used in the Underground, known as RODS, was discussed in Section 3.1. Gordillos' OD flow estimation methodology is based on Oyster (TfL's AFC system) data meaning that observed Oyster origin-destination flow serves as the basis of the matrix augmented by station control totals and measures to eliminate biases in the Oyster dataset. RODS data are used selectively in this method to supplement Oyster data where appropriate.

(a) Estimating Station Entry and Exit Counts

The estimation of station entry and exit counts is straightforward by combining automated gate counts and supplementary manual counts if necessary. Equations [3-4] and [3-5] illustrate how Oyster and magnetic counts constitute total automated gate counts.

$$\mathbf{GATE_{EN(A)} = CDJ_{EN(A)} + UNF_{(A)} + NT_{(A)} + MAG_{EN(A)}} \quad [3\text{-}4]$$

$$\mathbf{GATE_{EX(A)} = CDJ_{EX(A)} + UNS_{(A)} + NT_{(A)} + MAG_{EX(A)}} \quad [3\text{-}5]$$

In words, equation [3-4] states that the total automated gate entries at station A is the sum of all completely documented and unfinished journeys originating from A, as well as no travel journeys and magnetic entries recorded at A. Equation [3-5] is a similar summation for exit counts. Gordillo's analysis revealed that most of the no travel journeys correspond to station staff using their staff passes to allow passengers through the gates in cases of gate or card failures. Since these no travel journeys represent actual entries and exits made by passengers, they should be included in the automated gate count totals. The small fraction of no travel journeys that are made on non-staff passes may indicate passengers deciding to leave the station after seeing severe crowding on the platforms. While there is no way to prove that such a customer did not actually use the Underground and thus should be omitted from the control total, these entries constitute only a very small fraction of all automated gate counts.

When entering or exiting at FG stations, all passengers are required to validate their Oyster or magnetic tickets at the gates. Station entry and exit counts are therefore entirely captured by these two ticket types. The station entry and exit totals are simply the automated gate entry and exit totals, respectively.

On the other hand, passengers can enter or exit NFG stations without validating their travel tickets. In order to estimate the station entry and exit counts accurately, the manual counts from the annual Station Flow Count Survey are used to supplement the recorded Oyster and magnetic entry and exit counts. Based on the following observations, Gordillo proposed an adjustment to include the correct number of manual counts in the station entry and exit totals.

- The manual counts do not differentiate between entries and exits to the Underground and to National Rail at interchange stations, and therefore using them directly would over-estimate Underground ridership.
- The manual counts include passengers who validate their Oyster cards at stand-alone card readers for entry and exit, and therefore adding the manual counts to the AFC counts would result in double counting.
- The entry and exit counts observed at FG stations in April 2006 (TfL 2006/07 Period 1) were very similar to the estimates from RODS based on November 2004 data. Therefore, it was reasonable to assume that the entry and exit counts at NFG stations in RODS would also be very similar to the actual exit counts in April 2006.

The Adjusted Manual Counts (AMC) are therefore formulated to estimate entries and exits that are not recorded by gate transactions and their values will help determine the control totals at NFG stations. Equations [3-6] and [3-7] were developed in Section 3.2.1 of Gordillo to calculate the Adjusted Manual Counts (AMC) for NFG stations which, when added to the current gate counts, result in the best possible approximation to the station level estimates from the most recent RODS OD matrix.

$$\mathbf{AMC_{EN(A)}} = \begin{cases} \mathbf{max\{0, min[RODS_{EN(A)} - GATE_{EN(A)}, MC_{EN(A)}]\}} & \mathbf{if\ A\ is\ NFG} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \quad \text{[3-6]}$$

$$\mathbf{AMC_{EX(A)}} = \begin{cases} \mathbf{max\{0, min[RODS_{EX(A)} - GATE_{EX(A)}, MC_{EX(A)}]\}} & \mathbf{if\ A\ is\ NFG} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \quad \text{[3-7]}$$

By adding the Adjusted Manual Counts to the automated gate counts, the station level entry and exit counts can be obtained, as shown in Equations [3-8] and [3-9].

$$\mathbf{ENTRIES_{(A)} = GATE_{EN(A)} + AMC_{EN(A)}} \qquad [3\text{-}8]$$

$$\mathbf{EXITS_{(A)} = GATE_{EX(A)} + AMC_{EX(A)}} \qquad [3\text{-}9]$$

Table 3-1 summarizes the three types of relation between the Adjusted Manual Count entries and the total entries for NFG stations. The relations for NFG exits are similar. The interactions between RODS counts, Oyster and magnetic gate counts and manual counts relative to the control totals for NFG stations can be summarized into three scenarios. Since MC_{EN} values were surveyed in autumn 2005 and $RODS_{EN}$ values are based on November 2005 counts, they are expected to be different from the $GATE_{EN}$ values which are based on January 2007 Oyster and magnetic stripe gate counts. The entry ratio in Table 3-1 is defined by the number of entries in the Oyster-based matrix divided by that in RODS.

Scenario	AMC_{EN}	ENTRIES	Entry Ratio
$RODS_{EN} - GATE_{EN} \leq MC_{EN}$	$RODS_{EN} - GATE_{EN}$	$RODS_{EN}$	1
$RODS_{EN} - GATE_{EN} > MC_{EN}$	MC_{EN}	$GATE_{EN} + MC_{EN}$	< 1
$GATE_{EN} > RODS_{EN}$	0	$GATE_{EN}$	> 1

Table 3-1: Scenarios for Equation [3-6] for NFG Stations

In the first and second scenarios, the AMC_{EN} values estimate the proportion of entries that do not have ticket transactions and the total number of entries is restricted to be no larger than in RODS. The total entry is therefore equal to $RODS_{EN}$ or the $GATE_{EN}$ + MC_{EN} values. The third scenario is the most interesting because ridership growth is so significant that $GATE_{EN}$ recorded in January 2007 alone exceeds $RODS_{EN}$ which was estimated in November 2005. The total entry is therefore larger than the RODS entry.

(b) Estimating Expansion Factor

The intent of expansion factors is to correct biases in the Oyster dataset, which serves as the basis of the OD flow matrix developed in Gordillo's research, relative to all journeys made in the Underground network. As the name suggests, expansion factors are used to factor up completely documented journeys to match station entry and exit counts and correct biases in the Oyster dataset.

While RODS may not perfectly reflect travel patterns in the Underground network, it provides a reasonable description at a macro-level and is also the only information available on travel patterns of customers using magnetic stripe tickets. A comparison of RODS 2004 and Oyster transactions from 2006 Period 1 revealed that RODS captures travel between just over 29,000 distinct OD pairs whereas a single period of Oyster data captures travel between over 54,000 distinct OD pairs, of which 28,000 pairs are common across both datasets. In terms of journeys, 99% of journeys in RODS and 91% in Oyster are made on the common OD pairs. This means that RODS underestimates the number of journeys made on less popular OD pairs and overestimates those on more popular OD pairs.

Another comparison between RODS 2004 and the Oyster 2006 Period 1 dataset reveals that in the average whole day OD matrix, 29% of all RODS journeys exit at NFG stations whereas only 21% of all Oyster completely documented journeys exit at NFG stations. Section 3.3.3 provides a further breakdown of these proportions for journeys originating from FG and NFG stations to understand the nature of the bias and to assess the need for a formulation that is specific to the gate type at the origin station.

There are two reasons for the underrepresentation of Oyster journeys on OD pairs that are bound to NFG stations:

1. **Lack of fare payment validation** – Passengers do not always pass through gates at NFG stations (either through lack of gates or gates left open by station staff) and therefore journeys involving NFG stations are systematically less likely to be captured in the Oyster CDJ data.

2. **High proportion of Magnetic Stripe Tickets** – As discussed in Section 2.2.2, 41 of 68 NFG stations have National Rail connections at which the proportion of passengers with magnetic stripe tickets is likely to be much higher than the network average because it is the only fare medium available for these interchanging passengers using National Rail Travelcard Season Tickets.

In order to correct for bias in Oyster data, the expansion factors should differentiate between OD pairs that are bound to FG and NFG stations. Although the methodology does not explicitly correct this bias at origin stations, the OD estimation results in Section 4.3.2 show that the entry bias of Oyster data is also corrected. The estimation of bias-corrected expansion factors involves three steps:

Step 1 – Estimating the Number of Unknown-Destination Journeys

Equations [3-10] and [3-11] estimate the number of journeys originating from a given station that are not already captured by Oyster completely documented journeys to FG and NFG stations respectively. These journeys include incompletely documented Oyster journeys and magnetic stripe journey and are collectively called "undocumented journeys". The potential bias of lower-than-expected Oyster journeys bound to NFG stations is corrected by applying the RODS FG-bound proportion.

$$\mathbf{UDJ_{FG(A)}} = \mathbf{min}\left(\frac{\mathbf{RODS_{FG(A)}}}{\mathbf{RODS_{(A)}}} * \mathbf{ENTRIES_{(A)}} - \mathbf{CDJ_{FG(A)}}, \mathbf{ENTRIES_{(A)}} - \mathbf{CDJ_{(A)}}\right) \quad \textbf{[3-10]}$$

$$\mathbf{UDJ_{NFG(A)}} = \mathbf{ENTRIES_{(A)}} - \mathbf{CDJ_{(A)}} - \mathbf{UDJ_{FG(A)}} \quad \textbf{[3-11]}$$

Where

$\mathbf{UDJ_{FG(A)}}$ is the number of undocumented journeys from station A to all FG stations

$\mathbf{UDJ_{NFG(A)}}$ is the number of undocumented journeys from station A to all NFG stations

$\mathbf{RODS_{FG(A)}}$ is the number of journeys from station A to all FG stations in RODS

$\mathbf{RODS_{(A)}}$ is the number of journeys from station A in RODS

$\mathbf{ENTRIES_{(A)}}$ is the number of entries at station A, estimated by Equation [3-8]

$\mathbf{CDJ_{FG(A)}}$ is the number of Oyster CDJ from station A to all FG stations

$\mathbf{CDJ_{(A)}}$ is the number of all Oyster CDJ from station A

Step 2 – Calculating the Expansion Factors

Step 1 splits the unknown-destination journeys between FG and NFG destinations and the Oyster dataset directly splits the completely documented journeys. Equations [3-12] and [3-13] show the expansion factors for an OD pair from origin A to FG and NFG destinations, respectively, knowing the above splits. It will be shown in Section 3.3.3 that the gate type of the origin station does not affect the expansion factor formulation.

$$\mathbf{EXP(A, FG)} = \frac{\mathbf{CDJ_{FG(A)} + UDJ_{FG(A)}}}{\mathbf{CDJ_{FG(A)}}} \qquad \textbf{[3-12]}$$

$$\mathbf{EXP(A, NFG)} = \frac{\mathbf{CDJ_{NFG(A)} + UDJ_{NFG(A)}}}{\mathbf{CDJ_{NFG(A)}}} \qquad \textbf{[3-13]}$$

Step 3 – Applying the Expansion Factors

Depending on whether the destination is FG or NFG, the appropriate expansion factor is multiplied by the CDJ number to obtain the number of journeys for each OD pair in the seed matrix, known as the seed value, as shown in Equations [3-14] and [3-15]. The seed value will be zero if the corresponding CDJ number is zero, meaning there is no Oyster journey recorded for the OD pair.

$$\mathbf{SEED(A, i) = CDJ(A, i) * EXP(A, FG)} \qquad \forall\ \mathbf{i} \in \{\textbf{FG Stations}\} \qquad \textbf{[3-14]}$$

$$\mathbf{SEED(A, i) = CDJ(A, i) * EXP(A, NFG)} \qquad \forall\ \mathbf{i} \in \{\textbf{NFG Stations}\} \qquad \textbf{[3-15]}$$

Where

$\mathbf{CDJ(A, i)}$ is the number of completely document journeys from A to i

The expansion factors are defined such that the sum of all seed values originating from a station equals the estimated number of entries, as shown in Equation [3-16].

$$\sum_{i \in \{\text{Stations}\}} \mathbf{SEED(A, i) = ENTRIES(A)} \qquad \forall \mathbf{A} \qquad \textbf{[3-16]}$$

The resulting seed matrix is a singly constrained OD matrix that matches the estimated station entry totals but not necessarily the exit totals.

(c) Row-Column Balancing

The singly constrained seed matrix calculated from Equations [3-16] can be transformed into a doubly constrained matrix that matches both station entry and exit totals by applying the Iterative Proportional Fitting (IPF) procedure discussed in Section 3.2.1. The IPF method balances the OD flow given by the seed matrix such that the aggregate station level flows in the matrix match the control totals. In this research, the IPF algorithm is implemented in Excel Visual Basic and the code is included in Appendix C for reference. Reference for the IPF algorithm can be found in Navick and Furth (1994).

3.3 Refining OD Matrix Estimation to Time Period Level

Automatic fare collection data not only allow estimates of OD matrices regularly in a timely manner, they also provide the large sample sizes for estimation of OD matrices for shorter time periods within the day. One drawback of RODS is its base 15-minute reporting time unit implies small numbers of survey records available and therefore large expansion factors to match station counts. Oyster has the advantage of providing much larger sample sizes at low marginal costs compared with manual survey methods.

While the methodology developed by Gordillo (2006) constitutes the backbone for the time period level OD matrix estimation discussed in this section, it requires modification of the station level entry and exit counts to estimate the period totals. After addressing the issues associated with time period definitions, the station totals modification will be discussed in two parts: 1) selection of data to match the given time period and 2) use of proportions to estimate the number of exits that correspond to entries during the given time period.

3.3.1 Definition of Time Periods

The primary issue in defining time periods is the use of the resulting time period level OD matrix. As discussed in Section 1.2, a time period level OD matrix can be used to assess crowding by estimating line loads, help redesign routes and timetables by understanding the demands, and serve as a basis of customer-focused performance measures. The AFC-based matrix developed in this thesis can also be used to evaluate the survey-based matrix estimated by the London Underground and vice versa.

A time period is defined by two characteristics – length of time period and definition of qualifying journeys. The intuitive way to define qualifying journeys is by time of entry and/or exit. This method can be simple and makes it possible to avoid double-counting errors if all adjacent time periods are defined by either entry or exit time. There are also cases when using different allocation methods for different time periods of the day is desirable. During the AM Peak, passenger exits are concentrated within a short time interval, especially in central areas where most people work. In contrast, passenger entries are concentrated when people are commuting towards home from the central areas within a short time interval. In this case, the definition of the time period(s) between the AM and PM peaks should avoid double-counting of journeys.

Secondly, the length of a time period depends on how the OD matrix will be manipulated to represent the network demand. To calculate peak line loads, the time period should be defined such that the OD matrix would capture the peak demand and thus the heaviest loads. Such time periods could be defined in consecutive 15-minute intervals so that the network loading can be done sequentially using the resulting OD matrices.

On the other hand, two- to three-hour time periods are usually defined to represent peak flow for planning purposes, depending on the temporal patterns of demand. Figure 3-2 shows the temporal demand on the London Underground using one day of Oyster data aggregated into 15-minute intervals throughout the day. Although the graph shows only Oyster data and does not include magnetic stripe journeys, the overall pattern should be broadly representative of all journeys at the network level as Oyster represents about 70% of the total journeys.

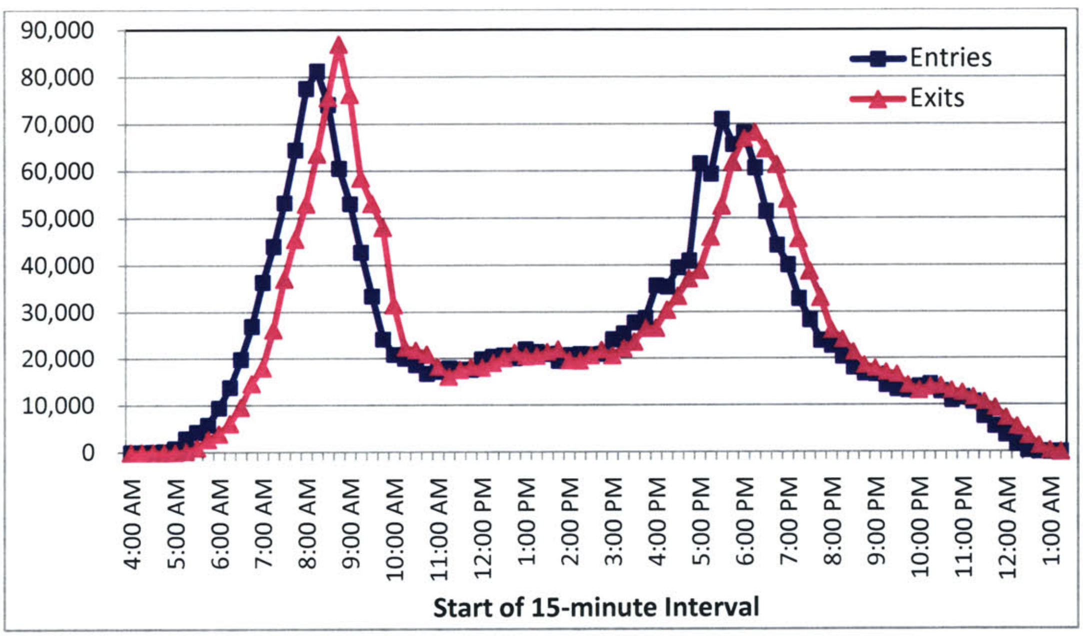

Figure 3-2: Distribution of Temporal Oyster Demand

The highly peaked morning demand reaches its maximum from 8:15 to 8:30AM for entries and 8:45 to 9:00AM for exits. Both the entry and exit distributions during the AM Peak are symmetrical around the respective peak 15-minute intervals. The exit distribution is also consistently shifted to the right of the entry distribution by about 30 minutes.

While the OD estimation methodology developed in this thesis can be used for any consistent definitions of time periods and journey qualifications, the application here is for the AM Peak which includes journeys that start between 7 and 10AM. This time period definition is used for the following reasons:

- To be consistent with the London Underground's definition of the AM peak period.
- The Oyster aggregate data provided by Transport for London at the time the matrix was developed were also defined by LU's definition of the AM Peak and thus they were the most disaggregate level of data available. However, this will no longer be a limiting factor if a full sample of Oyster transaction data is available.

3.3.2 Data Selection

When estimating the daily station totals, entry and exit counts across all hours are aggregated. At the time period level, the aggregation is not direct because of the time which elapses between a passenger's entry and exit. Since the allocation of journeys to time periods is based on entry time, entry counts can be aggregated directly according to the desired time period. The official LU time periods are defined in Table 3-2.

Day	Time Period	Hours
Monday to Friday	Early Morning	5:30 – 6:59AM
Monday to Friday	AM Peak	7:00 – 9:59AM
Monday to Friday	Inter-Peak	10:00AM – 3:59PM
Monday to Friday	PM Peak	4:00 – 6:59PM
Monday to Friday	Evening	7:00 – 9:59PM
Monday to Friday	Late Evening	10:00PM – 12:30AM

Table 3-2: London Underground Weekday Time Periods

Table 3-3 summarizes the base and aggregate data reporting units for each data type used in the OD matrix estimation methodology. The Oyster completely documented, unstarted and unfinished journeys are recorded to the minute whereas magnetic counts are collected every 15 minutes. The RODS OD matrix and manual counts are also reported in 15-minute intervals.

Type of Summary Data	Base Reporting Unit	Aggregate Data Reporting Unit
Oyster Completely Documented Journeys and Unstarted Journeys	Minute	Journey Time Segment (Official TfL Time Period)
Unfinished Journeys	Minute	Time Segment
Magnetic Entry and Exit Counts	15-Minute Interval	Hour
RODS OD Matrix and Manual Counts	15-Minute Interval	15-Minute Interval

Table 3-3: Reporting Units for Different Data Types

Aggregate data are used in this research because the base reporting units will lead to inconveniently large datasets. The "Journey Time Segments" (JTS) definition coincides with the LUL time periods and is used by the Journey Time Metric. The "Time Segments" (TS) definition is used for other analyses and can be slightly shorter than, or equal to, the JTS intervals. The Oyster completely documented journeys and unstarted journeys are aggregated by JTS whereas Oyster unfinished journeys are aggregated by TS because they are not used for journey time calculations. Magnetic entry and exit counts are reported by the hour.

Since all base reporting units shown in Table 3-3 are no longer than the corresponding LU time period, station entry counts from all data types can be aggregated to match a given time period. However, adjustments need to be made to incompletely documented Oyster journeys, magnetic journeys and manual counts to estimate the number of exits at each station which correspond to entries that occur during the specified time period. This adjustment is discussed in the following section.

The second issue in data selection is the number of days to be included if the intent is to represent the typical system demand during the specified time period. The span of data can be any number of days. The application illustrated in this thesis is based on a single LUL Period (4 weeks) of weekday AM Peak data, that is 20 weekdays.

3.3.3 Proportion of Journeys Between Stations of Different Gate Types

Section 3.2.3 showed that when comparing the RODS 2004 and Oyster 2006 Period 1 full day matrices, it was found that 29% of all RODS journeys exit at NFG stations whereas only 21% of all Oyster completely documented journeys exit at NFG stations. The proportions are further analyzed in this study at the time period level using more recent data. Figure 3-3 compares the proportions of journeys between stations of different gate types in the Oyster 2006 Period 11 and RODS 2005 matrices for the AM peak.

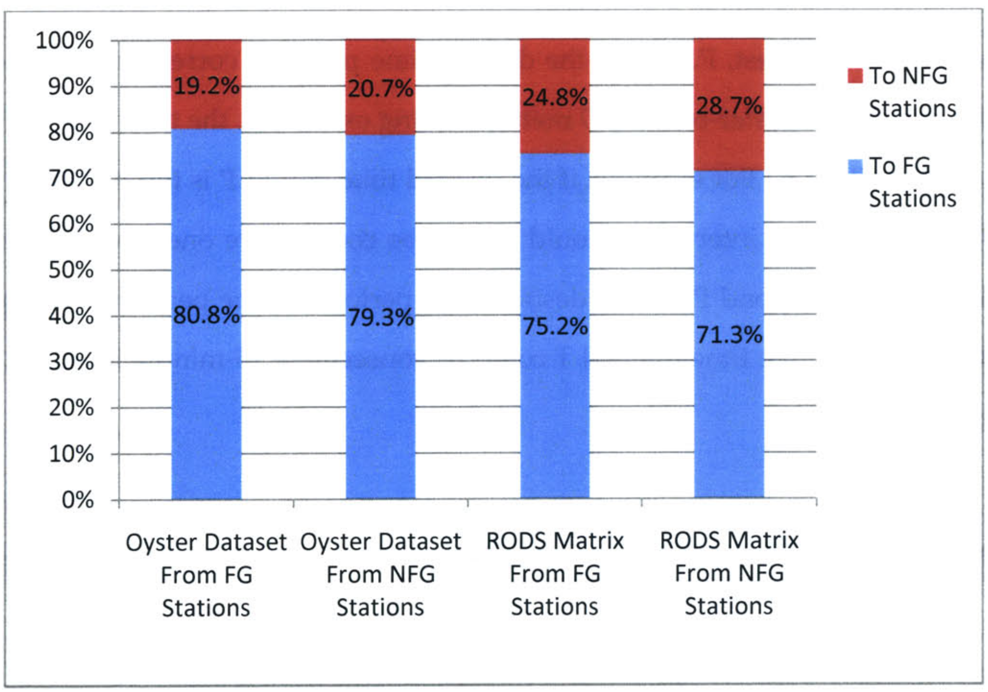

Figure 3-3: Proportion of Journey Between Stations by Gate Type

The proportions of journeys from FG stations to NFG stations are 19.2% and 24.8% in the Oyster dataset and RODS, respectively. The discrepancy is larger for journeys from

NFG stations to NFG station: 20.7% in the Oyster dataset and 28.7% in the RODS. As expected, NFG-to-NFG travel is most weakly captured by Oyster since passengers are not required to validate payment at either end of their journey. The bias of proportion of journeys bound to NFG stations persists in both journeys originating from FG as well as from NFG stations.

Despite this finding, the formulation of expansion factors is not changed because it is found in the resulting matrices that the methodology corrects both entry and exit bias at NFG stations, as will be shown in Section 4.3.2 (Figure 4-3).

3.3.4 *Exit Proportions*

Equation [3-17] defines the Oyster-based exit proportion (EP) which is the proportion of exits during any time interval T at station A that correspond to entries that occur during the time period of interest, P. While the desired time period P corresponds to the time period for which the Oyster-based OD matrix is being estimated, the time interval T may or may not be within P. For example, if the desired time period P is the AM Peak which spans 3 hours, the time interval T should ideally be consecutive one-hour or half-hour intervals extending beyond P. If the desired time period P is the peak one-hour period within the AM Peak, the time interval T could be consecutive 15-minute intervals, again extending beyond P.

$$\mathbf{EP(A,T)} = \frac{\sum_{\text{all X}} \sum_{\text{i} \in \text{P}} \sum_{\text{j} \in \text{T}} \mathbf{CDJ(X,A;i,j)}}{\sum_{\text{all X}} \sum_{\text{all i}} \sum_{\text{j} \in \text{T}} \mathbf{CDJ(X,A;i,j)}} \qquad \text{[3-17]}$$

Where

$\mathbf{CDJ(X,A;i,j)}$ represents completely documented Oyster journeys starting from station X at time i and finishing at station A at time j (i and j are recorded in minutes)

The numerator is the number of Oyster completely documented journeys that start during time period P conditional on ending at station A during time interval T. The denominator is all Oyster completely documented journeys that end at station A during time interval T regardless of start time. The quotient gives the exit proportion that is station- and time interval-specific.

The EP values are then multiplied by non-Oyster counts, such as magnetic stripe exits and manual exit counts, recorded during the corresponding time interval T to estimate the number of exits during T that correspond to entries during time period P. A numerical example is provided at the end of this section for illustration.

The underlying assumption is that on average, completely documented Oyster journeys have similar travel time distributions as incompletely documented journeys and magnetic journeys. Although the proportion of CDJs varies from station to station, the assumption should be robust on an hourly basis especially at high exit volume stations. In the following example for the AM Peak, EPs are calculated for consecutive 1-hour periods starting from 7AM and extending beyond 10AM in order to capture the completion of all trips starting in the AM Peak. It is suggested that EPs be calculated for at least two hours beyond the time period of interest in order to include the longest journeys. EP values before 7AM need not be calculated because all journeys that end before 7AM must also have started before 7AM. The formulation of EP is also capable of filtering out journeys that begin before 7AM but end during the AM Peak. These journeys are filtered out by the EP value for 7 and 8AM by including only journeys that start after 7AM.

Table 3-4 shows a numerical example of the calculation and application of the EP values for one hypothetical station.

	Hour					
	7 - 8	**8 - 9**	**9- 10**	**10 - 11**	**11 - 12**	**Total**
EP Numerator	1,125	2,352	2,200	1,020	250	-
EP Denominator	1,500	2,400	2,200	1,200	1,000	-
EP	0.75	0.98	1	0.85	0.25	-
Recorded Magnetic Stripe Exits	1,300	1,800	1,650	800	400	-
Adjusted Magnetic Stripe Exits	975	1,764	1,650	680	100	5,169
Recorded Manual Count Exits	400	700	950	500	260	-
Adjusted Manual Count Exits	300	686	950	425	65	2,426
Recorded Oyster Exits	14,000					14,000
Total Exits	-					21,595

Table 3-4: Numerical Example to Illustrate the Exit Proportion Concept

The *EP Numerator* and *EP Denominator* values are aggregated from Oyster transactional data according to [3-14]. The *EP Denominator* value of 1,500 is the number of completely documented Oyster journeys finishing at this station during 7 and 8AM. The *EP Numerator* value of 1,125 is the number of CDJ journeys that begin during the AM Peak hours of 7 to 10AM and finish at this station between 7 and 8AM. The remaining 375 exits therefore correspond to entries before 7AM. Similarly, out of the 1,200 CDJ exits between 10 and 11AM, 1,020 correspond to entries during the AM Peak hours and the remaining 180 are entries after 10AM. The resulting *EP* value is simply the quotient of the *EP Numerator* and *EP Denominator* values (EP = 1,125/1,500 = 0.75).

The estimated EP values are then applied to the recorded counts to filter out exits that do not correspond to entries during the AM Peak. The recorded magnetic stripe exits are recorded at the gates and are multiplied by the EP value of the same hour to obtain the number of Adjusted Magnetic Stripe Exits (for 7 to 8AM, Adjusted Magnetic Stripe Exits = Recorded Magnetic Stripe Exits * EP = 1,300*0.75 = 975). Similarly, the Adjusted Manual Count Exits are estimated by multiplying the Recorded Manual Count Exits by the EP value of the same hour. Finally, the value of Recorded Oyster Exits are extracted from the aggregate Oyster data (not transactional data) which are defined by the sum of entries of completely documented journeys, unfinished journeys and no travel journeys.

Summing over all hours, the total number of exits at this station is 21,595. Along with the directly aggregated entry total, the estimated exit total is used in the row-column balancing process to obtain a doubly-constrained OD matrix for the AM Peak.

The EP values shown in Appendix D are calculated using the 5-weekday AM Peak data from February 4 to 8, 2007. Five weekdays of data are used to minimize biases at low exit volume stations. The numerator and denominator in Equation [3-17] are calculated from the sum of the corresponding journeys over the five days.

3.3.5 Revised Station Level Estimates

The equations used to estimate total station entry and exit counts need to be adjusted to incorporate the Exit Proportions. Equations [3-18] to [3-21] below are adjusted versions of Equations [3-4] to [3-7].

$$\mathbf{GATE_{EN}(A,P)} = \textstyle\sum_{t\in P}\mathbf{CDJ_{EN}(A,t)} + \sum_{t\in P}\mathbf{UNF(A,t)} + \sum_{t\in P}\mathbf{NT(A,t)} + \sum_{t\in P}\mathbf{MAG_{EN}(A,t)} \quad \textbf{[3-18]}$$

$$\mathbf{GATE_{EX}(A,P)} = \textstyle\sum_{t\in P}\mathbf{CDJ_{EX}(A,t)} + \sum_{t\in P}\mathbf{UNS(A,t)} * \mathbf{EP(A,t)} + \sum_{t\in P}\mathbf{NT(A,t)} + \sum_{t\in P}\mathbf{MAG_{EX}(A,t)} * \mathbf{EP(A,t)} \quad \textbf{[3-19]}$$

$$\mathbf{AMC_{EN}(A,P)} = \begin{cases} \mathbf{max\{0, min[RODS_{EN}(A,P) - GATE_{EN}(A,P), MC_{EN}(A,P)]\}} & \mathbf{if\ A = \{NFG\}} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \quad \textbf{[3-20]}$$

$$\mathbf{AMC_{EX}(A,P)} = \begin{cases} \mathbf{max\{0, EP(A,p) * min[RODS_{EX}(A,P) - GATE_{EX}(A,P), MC_{EX}(A,P)]\}} & \mathbf{if\ A = \{NFG\}} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \quad \textbf{[3-21]}$$

The calculations for total station entry and exit counts remain unchanged as shown in Equations [3-8] and [3-9].

3.3.6 *Summary of Steps*

The steps needed to produce an Average Weekday OD Matrix for Time Period P are summarized below:

1. Using Oyster transactional data, calculate the exit proportions for each station corresponding to entries during time period P
2. Using AFC and manual count data, estimate station entry totals for time period P
3. Using AFC and manual count data, estimate station exit totals scaled by exit proportions from (1) for time period P
4. Using the RODS OD Matrix, calculate FG station-bound ratios for time period P and calculate the expansion factors
5. Apply the expansion factors from (5) to the time period P Oyster CDJ to obtain the seed matrix
6. Using the seed matrix and station totals as inputs, apply the row-column balancing algorithm until the OD matrix is consistent with both entry and exit totals

3.3.7 *Application to Shorter Time Periods*

The OD matrix estimation methodology discussed above is applicable to single or combined LUL time periods. Theoretically, this methodology could also be applied to estimate the OD matrix for a time interval as small as the longest base reporting unit among all data sources. Given the current base reporting structure, the shortest possible time frame is 15 minutes, limited by the base reporting units of both magnetic counts and RODS data. Since Oyster transactional data are available, they can be aggregated to 15-minute intervals to match the time frame of the other data sources.

Despite the availability of short time interval data, this methodology is not recommended for use to estimate OD matrices for time intervals shorter than one hour. At the one-hour time frame, the expansion factors obtained from hourly RODS OD matrix should be reasonably accurate and therefore the resulting seed matrix should be representative of true travel patterns. For any time frame shorter than an hour, it is questionable how accurate the expansion factors would be because some portion of RODS journeys within each 15-minute interval are duplicated from adjacent intervals due to small numbers of survey responses.

Chapter 4 Evaluation of Oyster-Based OD Matrix

The RODS OD matrix is an input to many LUL modeling and performance tools and it is important to be confident that the matrix reflects actual travel patterns. It would be unwise to replace the RODS OD matrix with the Oyster-based matrix outlined in this thesis without making systematic comparisons between the two. Differences observed at the OD level may or may not be significant enough to impact planning decisions such as timetable setting and capital investment decisions. The systematic comparisons will give planners a better sense of the differences between the RODS OD matrix and the Oyster-based matrix.

This chapter presents qualitative comparisons of the RODS and Oyster-based methodologies and systematic comparisons of the resulting matrices for the AM Peak. The RODS matrix is based on manual surveys collected in the autumns of 1998 to 2005 reconciled to average November 2005 station counts, and is known as "RODS 2005". The Oyster-Based matrix is estimated based on the methodology outlined in Section 3.3 and Oyster and magnetic ticket data from January 2007 (20 weekdays during Period 12 of 2006/07 in the London Underground) and manual survey counts from RODS 2005.

Section 4.1 discusses the qualitative differences between the RODS and Oyster-based methodologies. Section 4.2 outlines the systematic comparisons needed between the two matrices and the actual results are presented in Sections 4.3 to 4.6. Suggestions to improve the origin-destination estimation methodology will be presented in Section 7.1.3.

4.1 Qualitative Comparison of Methodologies

The advantages of RODS over Oyster are:

- **Path Choice Information** – In addition to the access and egress stations, the RODS survey asks about all transfer stations along the passenger's Underground journey on the day of survey. In most cases, the Underground line segments traveled on a journey are determined once the transfer stations are known. Transfer information is particularly important for assigning journeys to the network and thus calculating line loads. In contrast, an Oyster journey defines only the access and egress stations without indication of transfer stations a passenger travels through. However, OD path choice should not change as often as OD level demands, and path choice can be predicted with planning models which can be periodically updated via surveys when the network changes. It is therefore more important to capture variations in demand than path choice within short time periods and Oyster data allow demand variations to be easily monitored.
    - **Access and Egress Mode Information** – The RODS survey inquires about the real origin and destination of a passenger's journey and the access and egress modes taken by the passenger before and after riding the Underground. The access and egress information provides insight into how passengers use the Underground as part of their uni-modal or multi-modal journeys. While Oyster does not explicitly contain such information, the TfL services used for access and egress to and from the Underground could be deduced by the preceding and following journeys on the passengers' Oyster cards within a defined time frame. For example, if an Oyster card records a bus boarding transaction within a reasonable time frame (e.g. 5 to 15 minutes) after a passenger exits from an Underground station, the egress mode can reasonably be assumed to be bus (Zhao 2004).
    - **Fair Representation of All Passengers in RODS** – Oyster systematically under-represents journeys to and from National Rail stations especially during the peak hours because many National Rail commuters use magnetic tickets to access the Underground. Therefore, Oyster provides much less

information than RODS on passengers who tend to access the Underground from the National Rail connections in Zone 1 and usually make shorter journeys. In this respect, RODS is less biased because the surveys are randomly distributed to passengers entering the survey stations and the resulting RODS OD matrix is scaled to match ticket types. This difference is likely to lessen in importance as Oyster PAYG becomes available on National Rail services within Greater London in the near future. It is also expected that National Rail Season Travelcards will also adopt Oyster as the train companies acquire the technology and realize benefits of reduced ticketing costs and smoother fare validation processes.

The advantages of Oyster over RODS are:

- **Sampling Frequency** – While RODS surveys about 30 origin stations each year, Oyster's sampling frequency is virtually every minute when Oyster transactions are made throughout the network.
- **Strong and Increasing Sample Size** – There have been a total of about 240,000 RODS surveys collected over 7 years – compared to over 2 million Oyster journeys made on the Underground every weekday. The number of daily Oyster journeys is expected to grow as Oyster's penetration continues to rise and as it is increasingly accepted on the National Rail network in Greater London.
- **Data Completeness** – Dealing with missing or illogical information on returned RODS surveys poses a challenge to the accuracy of the resulting RODS OD matrix. LU planners sometimes supplement the missing information based on similar surveys in the dataset. On the other hand, Oyster transactional data are accurate and should not include illogical information. Although incompletely documented Oyster journeys have missing information, such journeys are becoming a smaller proportion of all journeys in the network after the Max Fare policy was implemented.

- **Cost Effectiveness** – Oyster data used for OD estimation is virtually a by-product of the smart card revenue collection mechanism and imposes no significant additional cost to LU. RODS is an expensive undertaking involving manual distribution of surveys at stations and intense post-survey processing.

4.2 Outline of Systematic Comparisons

In order to evaluate the resulting Oyster-based matrix and suggest improvements to the RODS methodology, the two matrices need to be systematically compared for LU planners to understand the differences. The systematic comparisons are developed in two respects – the OD matrix itself and travel patterns based on trip assignment.

The OD matrices are compared and evaluated at 4 levels, from top to bottom:

- **Network Level** – compares the number of OD pairs covered and the distribution of journeys on the OD pairs that are common and unique to the two matrices
- **Zonal Level** – compares the distribution of zone-to-zone journeys across LU's fare zones
- **Station Level** – compares the number of entries and exits
- **Origin-Destination Level** – compares estimated number of journeys between specific OD pairs

After assigning the OD matrix to the Underground network by the shortest-path algorithm provided by the TransCAD network analysis program, two comparisons are made:

- **Link Loads** – compares passenger loads assigned to adjacent station-to-station links and evaluates differences in maximum loads
- **Interchange Volumes** – compares the number of transfer passengers at the major interchange stations and provides inputs into station passenger congestion

4.3 Network and Zonal Levels

Since the AFC system captures 100% of journeys made with this fare medium, the Oyster dataset is expected to be much richer than the survey-based RODS dataset in terms of OD travel patterns. While the Oyster-based matrix captures most of the OD pairs that RODS captures, the reverse is not true.

Before the comparisons are presented, it is important to understand the process which leads to the resulting cell values in the final Oyster-based matrix. Table 4-1 shows the different stages of the OD matrix in the estimation process.

Stage of Matrix	Format	Description
Oyster CDJ	Decimal	Daily average number of journeys observed over 20 AM peaks
Seed	Decimal	Oyster CDJ multiplied by expansion factor
Post IPF	Decimal	Result after applying the Iterative Proportional Fitting method to match entry and exit control totals
Final ("Oyster-Based")	Integer	Rounding the Post IPF result to the nearest integer

Table 4-1: Stages of OD Matrix Estimation

The Oyster completely document journeys (CDJ) dataset is the basis of OD travel patterns which are scaled up by the expansion factors to produce the seed matrix. The seed matrix is a singly constrained matrix that matches the station level entry totals but not necessarily the exit totals. By applying the Iterative Proportional Fitting method, a doubly constrained matrix which matches both station entry and exit totals is obtained referred to as the "Post IPF Result". Since the Oyster CDJ is the OD level daily average number of journeys over 20 AM peaks, its value is a decimal in most cases and the decimals are carried through the post IPF matrix. Finally, the values in the post IPF matrix are rounded to the nearest integer because the number of journeys should be discrete. In all subsequent comparisons presented in this chapter, only ODs with a (rounded) daily average of 1 journey or more are included and "Oyster-based" figures present the rounded values unless otherwise noted.

4.3.1 Network Level

Table 4-2 shows the number of OD pairs and corresponding journeys in each OD category network wide. The AFC system records travel on 57,407 OD pairs but 15,506 pairs (daily average of 1,884 journeys) are lost due to rounding the number of journeys to the nearest integer (rounded values being zero). The resulting AM peak Oyster-based matrix has 881,206 journeys, which is about 1% more than the number of journeys in RODS. This difference is probably due to increases in ridership from November 2005 to January 2007.

	# OD Pairs	# Oyster-Based Journeys	# RODS Journeys
Underground Network	74,256 (total possible)		
Post IPF Oyster-Based Matrix Total (before rounding)	57,407	883,090	-
Final Oyster-Based Matrix Total (after rounding)	41,901	881,206	-
RODS Matrix Total	17,421	-	873,100
Common to Both Matrices	16,465	763,456	858,489
Unique to Oyster-Based Matrix (after rounding)	25,436	117,750	-
Unique to RODS Matrix	956	-	14,611

Table 4-2: Network Coverage of OD Pairs

Figure 4-1 illustrates the distribution of OD demand by category. Out of the 74,256 possible OD pairs on the Underground network (273 origins by 272 possible destinations), 42% are not observed in either matrix, 22% are captured by both matrices, 34% are captured only by the Oyster-based matrix and the remaining 1% are captured only by the RODS matrix.

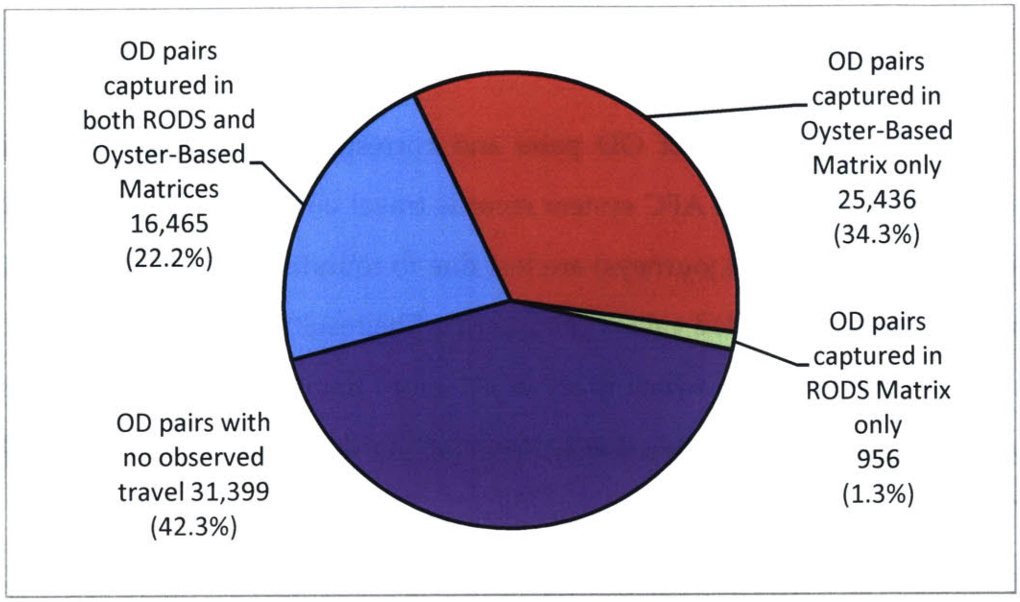

Figure 4-1: Types of OD Pairs

In terms of number of journeys, the common OD pairs carry most of the journeys in both matrices – 87% of the Oyster-based matrix and 98% of the RODS matrix, as illustrated in Figure 4-2. The difference arises from the fact that while both matrices contain about the same number of journeys, RODS captures less than half of the OD pairs in the Oyster-based matrix. Therefore, most of the journeys in RODS are allocated to the more popular OD pairs.

4.3.2 *Journey Proportions Between Stations of Different Gate Types*

The OD estimation methodology corrects the underrepresentation of journeys both from and to non fully gated stations in the Oyster dataset. Oyster data include lower-than-actual proportions of journeys originating from, and destined to non fully gate stations because passengers are not required to validate their fare payment at these stations. Although the OD estimation methodology does not explicitly correct for the bias of journeys originating from NFG stations, the resulting Oyster-based matrix has very similar proportions of journeys as RODS between each gate barrier type, as shown in Figure 4-3.

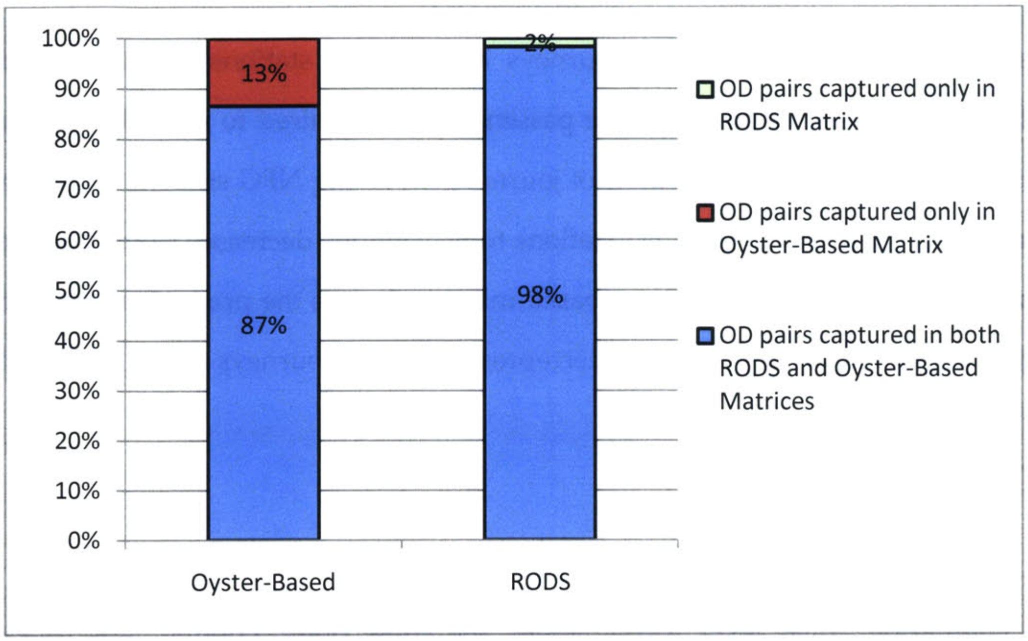

Figure 4-2: Journey Distribution by OD Pair

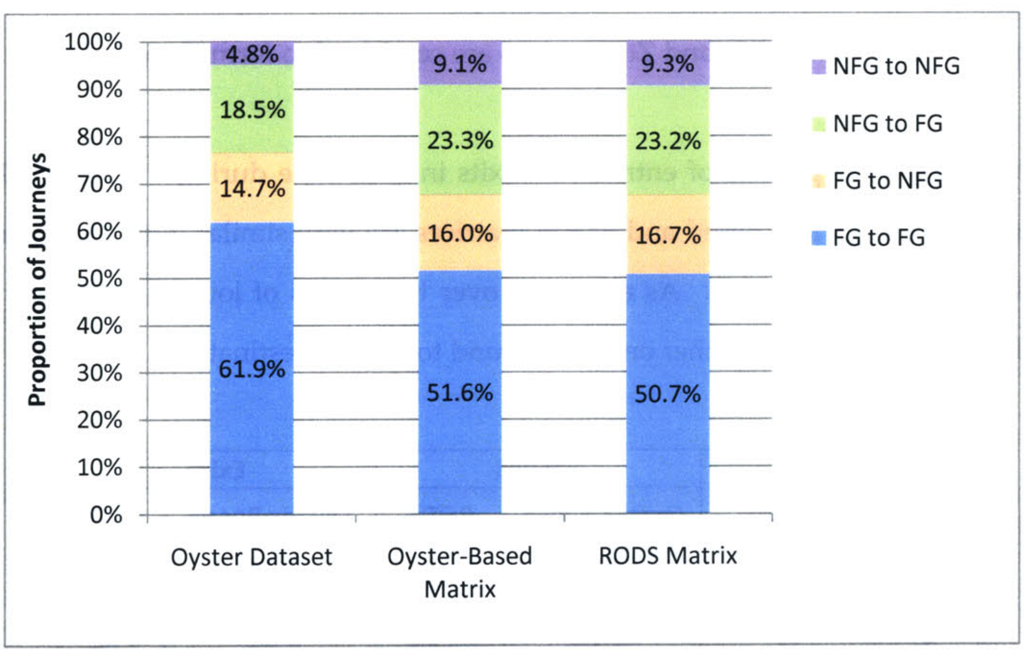

Figure 4-3: Proportion of Journeys Between FG and NFG Stations

Overall, the proportions of journeys originating from and destined to NFG stations are similar between the Oyster-based and RODS matrices because they are simply results of accurate estimates of station control totals.

It is expected that the number of journeys between FG stations is most accurately recorded in the Oyster dataset because passengers are required to validate payment at both ends and much smaller portions of journeys involving NFG stations are recorded. The proportion of journeys from FG stations to FG stations decreases from 61.9% in the Oyster dataset to 51.6% in the Oyster-based matrix whereas the proportions of all three other journey types increase as the underrepresentation of journeys to and/or from NFG stations is corrected.

4.3.3 *Zonal Level*

When comparing zonal level travel below, the three outer London fare zones are combined to one zone because they constitute relatively low demand and may be surveyed less often and provide fewer useable samples in the RODS methodology. Zones 1 to 3 remain the same and Zones 4 to 6 are combined to form Zone 4+.

Table 4-3 shows the percent of entries and exits in each zone during the AM Peak for each matrix. The Oyster-Based and RODS matrices are very similar with discrepancies of less than 0.8% in all cases. As expected, over two-thirds of journeys are bound to Zone 1 destinations, with another one-fifth bound to Zone 2 destinations.

Zone	Number of Stations	Entries		Exits	
		Oyster-Based	RODS	Oyster-Based	RODS
1	62	35.8%	36.2%	68.0%	68.3%
2	73	25.7%	25.7%	19.4%	18.6%
3	45	21.1%	20.8%	6.9%	7.0%
4+	93	17.4%	17.3%	5.8%	6.1%

Table 4-3: Percent of Entries and Exits by Zone

Figure 4-4 illustrates the zone-to-zone demand. Again, the two matrices are very similar with no discrepancies larger than 0.5%. It can be seen that RODS tend to allocate slightly more demand to popular zonal travel such as Z1 to Z1, Z2 to Z2 and Z3 to Z1. This is because RODS captures less than half the number of OD pairs captured in the Oyster-based matrix and therefore RODS inevitably over-allocates journeys to the more popular OD pairs. The Z4+ to Z3 and Z4+ to Z4+ demands are also higher in RODS and the differences are mainly travel on OD pairs that are captured only by RODS. Many of these journeys are implausible and may be reporting and/or processing errors (some examples will be given in Section 4.5.4).

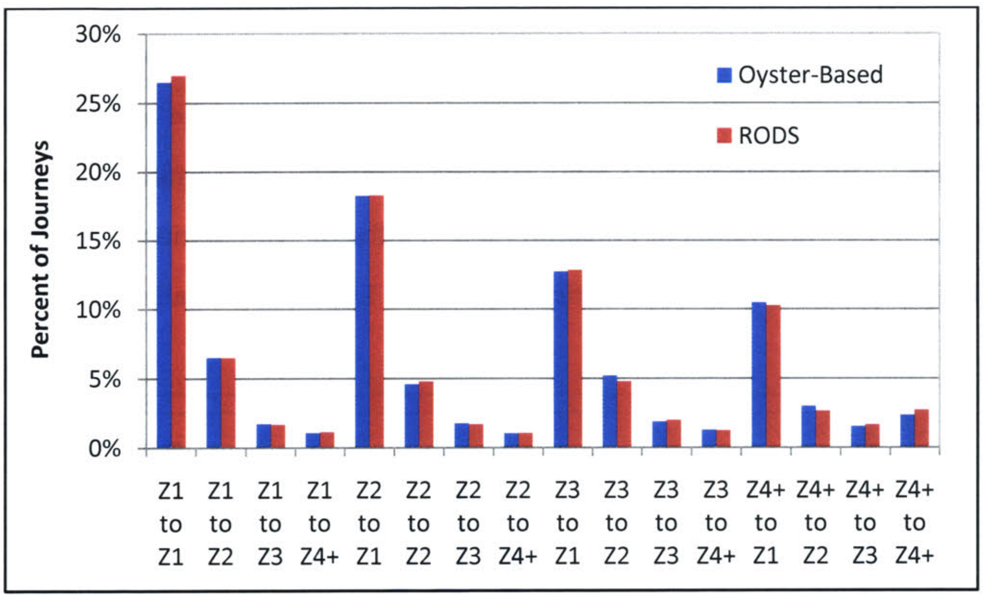

Figure 4-4: Distribution of Zone-to-Zone Journeys

4.4 Station Level

Since the station entry and exit totals are the control totals used in the row-column balancing procedure, the accuracy of the Oyster-based matrix depends greatly on the accuracy of the control totals. The control totals for fully gated (FG) stations are aggregated directly from gate counts and are therefore reliable because all passengers are required to validate their fare payment at gates. On the other hand, the control totals of non fully gated (NFG) stations are estimated from Oyster, magnetic stripe and manual counts, and may not be entirely accurate. The interactions between RODS counts, Oyster and magnetic gate counts and manual counts for NFG stations to estimate the control totals were summarized in Section 3.2.3.

The entry ratio is defined by the number of entries in the Oyster-based matrix divided by that in RODS, and the exit ratio is defined similarly. Appendix E shows a complete list of entry and exit ratios by station. Figure 4-5 shows the distribution of entry ratios for both NFG stations and FG stations. The total number of FG and NFG stations does not add up to 273 because 2 stations were closed in January 2007. The horizontal axis shows the magnitude of entry ratio in 5% intervals. As mentioned before, FG stations have reliable gate counts because all passengers are required to validate fare payment at gates. The distribution of FG stations is fairly symmetrical about the middle interval of 0.975 – 1.025. The distribution of NFG stations is mostly concentrated in the middle interval due to the constraints of Equation [3-3]. Of the 67 NFG stations, 47 fall into the 0.975 – 1.025 interval, 12 have entry ratios larger than 1.025 and 8 smaller than 0.975.

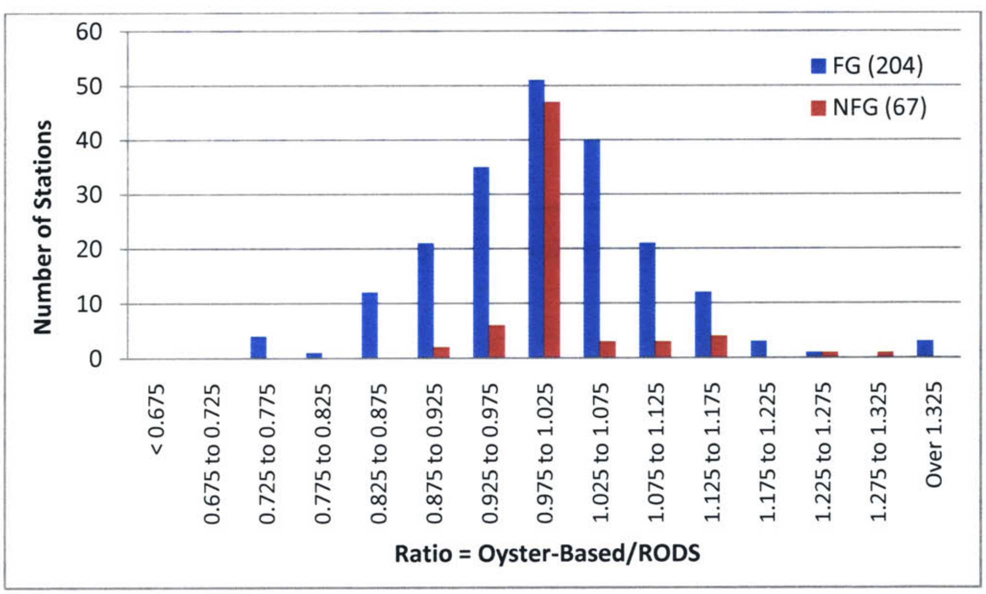

Figure 4-5: Distribution of Entry Ratios

Figure 4-6 shows the distribution of exit ratios. The distributions for both FG and NFG stations are much broader than for entry ratios. There are several possible reasons for this:

- The last step in the RODS iterative process is balancing entries, therefore entries are more accurate than exits.
- In addition to the two closed stations during January 2007, Theydon Bois and Grange Hill stations have no recorded exits in RODS because no surveys indicated these stations as destinations. These two FG stations are assigned an exit ratio of zero.
- The magnetic stripe exit counts and manual counts in the Oyster-based matrix are estimated by multiplying the hourly counts by the exit proportions estimated from Oyster completely documented journeys. It is possible that Oyster CDJs exhibit different journey characteristics, such as distribution of journey length and journey start time, than magnetic and manual count journeys.
- 20 of 26 stations with exit ratios smaller than 0.825 are Zone 4+ stations with very few exits in the AM Peak. A small difference in exits between the Oyster-based and RODS matrices can lead to a low exit ratio.

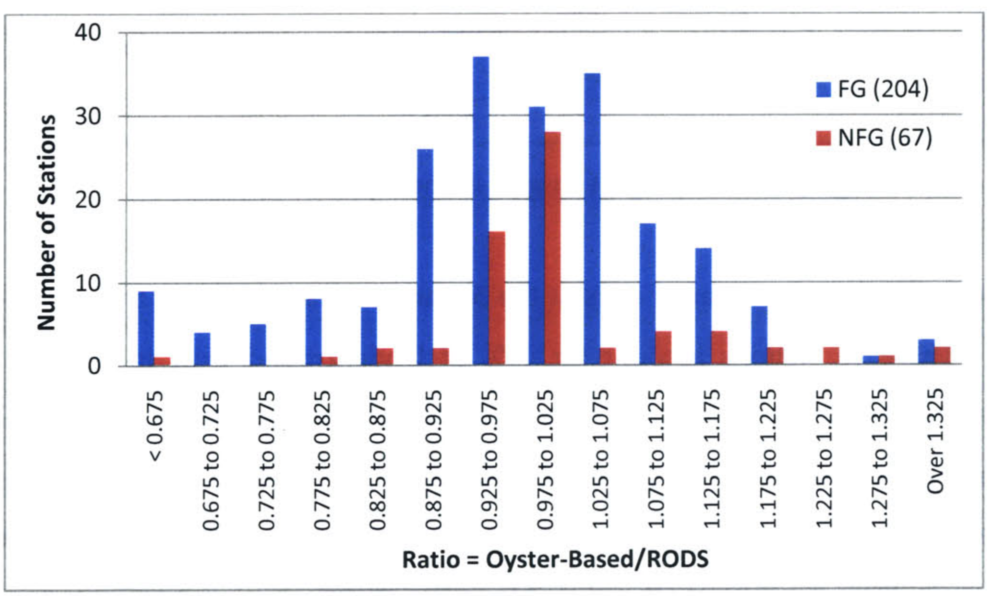

Figure 4-6: Distribution of Exit Ratios

Figure 4-7 shows the number of entries at the 50 largest origins in the Oyster-based matrix. These 50 stations account for 53% of Underground entries in the AM Peak. The two distributions are highly similar with significant differences seen only at Victoria and London Bridge. Victoria is a fully gated station but it is known that gates are sometimes opened to allow passengers transferring from National Rail through. Also, Victoria is occasionally closed for short intervals when station congestion is severe to prevent more passengers from entering the station. Both practices may lead to a lower number of recorded gate entries. London Bridge is also a fully gated station and this discrepancy may be due to ridership increases from November 2005 to January 2007.

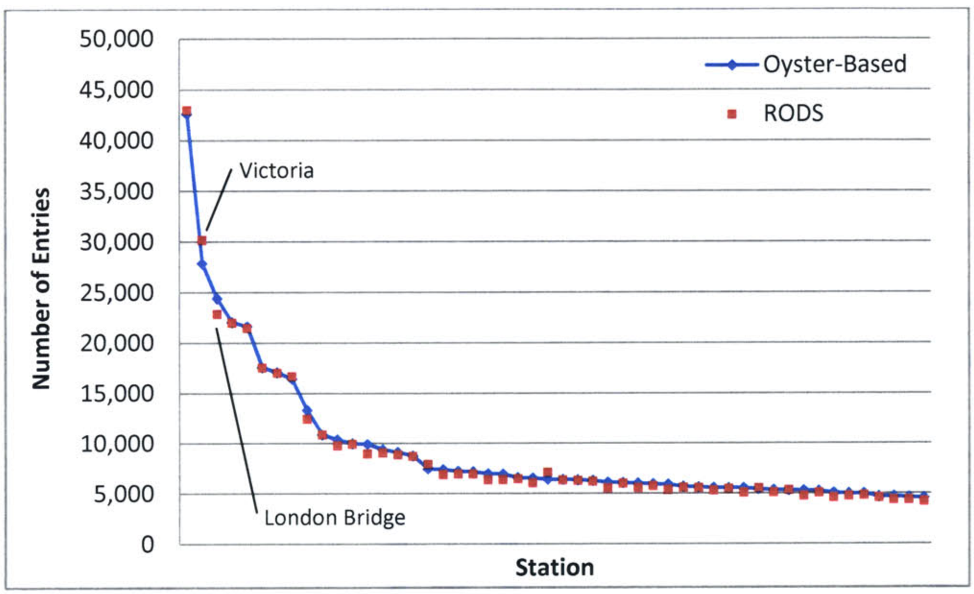

Figure 4-7: 50 Largest Origins in the Oyster-Based Matrix

Figure 4-8 shows the number of exits at the 50 largest destinations in the Oyster-based matrix. These 50 stations account for 43% of Underground exits in the AM Peak. Significant differences from the RODS station totals are seen at 5 stations; all but Canary Wharf are Zone 1 stations with National Rail connections. Other than the explanation that the RODS' iterative process ends with entries balancing and therefore exits may be less accurate, there are several other possible explanations:

- Canary Wharf is a fully gated station and the difference of 6,000 is mostly attributable to ridership growth, especially with the rapid commercial growth near the station.
- Liverpool Street is also a fully gated station and the station exit count is reliable.
- Moorgate and Farringdon are non fully gated stations and the gate counts alone (without manual counts) exceed the numbers of exits in RODS.
- Kings Cross has 30,000 exits in RODS but only 16,500 exits in the Oyster-based matrix. It is possible that RODS overestimates the number of exits due to passengers interchanging between lines at Kings Cross having to pass through two sets of gates, and therefore RODS makes assumptions for the station totals to avoid double counting. After the new ticket hall opened in October 2006, this potential error would have been eliminated because passengers will no longer need to pass through two sets of gates and RODS 2006 station totals at Kings Cross can be directly obtained from gate counts.

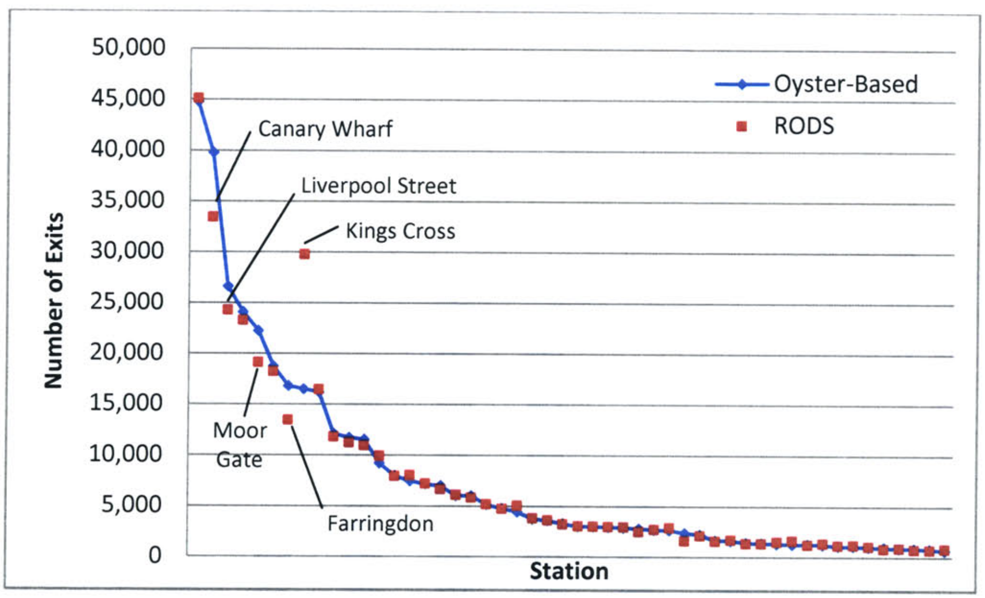

Figure 4-8: 50 Largest Destinations by Oyster-Based Exits

4.5 Origin-Destination Level

By examining differences between the RODS and Oyster-based matrices at the origin-destination level, the methodological differences and possible errors can be discussed at a more detailed level. Interesting findings lead to better understanding of aspects of station layouts that affect the collection of Oyster data and therefore accuracy of the Oyster-based matrix. Some of the following examples also show the dynamics of changing travel patterns and the effects of dated surveys on the resulting RODS matrix.

4.5.1 Destination Coverage

Section 4.3.1 showed that Oyster data capture 57,407 OD pairs whereas RODS captures only 17,421, a difference of almost 40,000 pairs. As a result of such difference, many journeys in RODS are allocated to the more popular OD pairs. On average, an origin station has 90 more destinations in Oyster than in RODS, equivalent to 150% more destinations. Figure 4-9 shows the distribution of the number of "new" destinations in Oyster that are unreported in RODS. The station with the largest number of new destinations is Upton Park with an increase of 176 from 64 to 240.

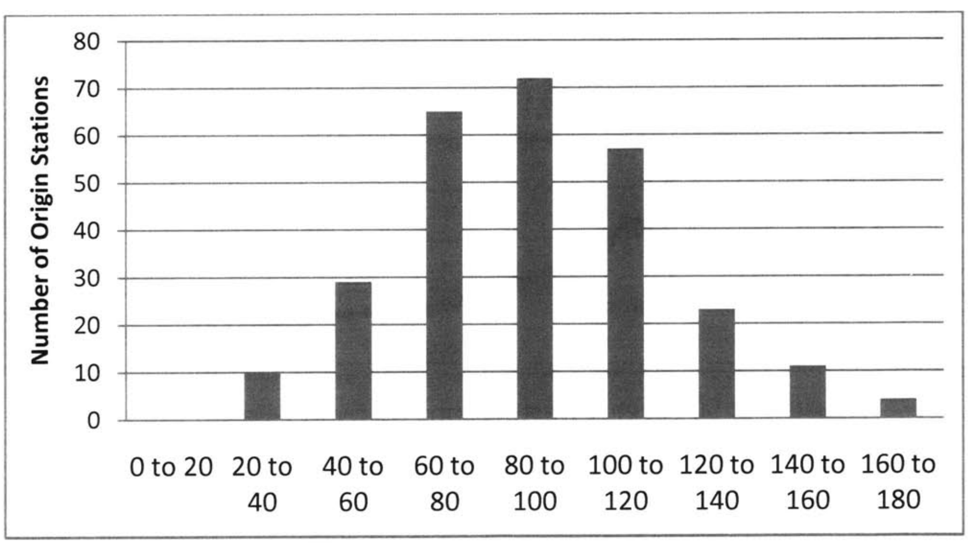

Figure 4-9: Distribution of the Number of "New" Destinations Captured by Oyster

Figure 4-10 shows the distribution of "destination ratios" which is defined for each origin station as the number of destinations in Oyster divided by that in RODS. The largest destination ratio of 6.0 is observed at Heathrow Terminal 4, with 120 destinations in Oyster and 20 destinations in RODS. This evidence shows that passengers use Oyster to access six times the number of destinations in Oyster reported in RODS.

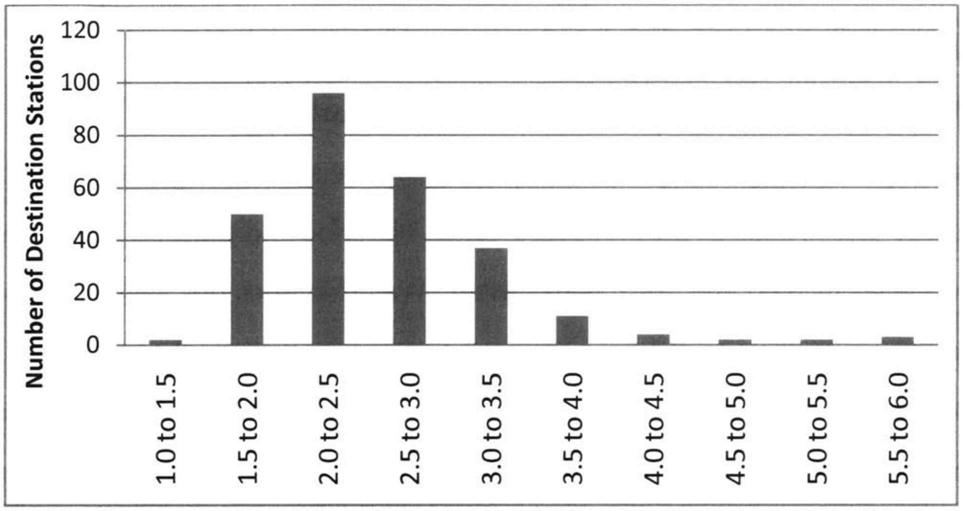

Figure 4-10: Distribution of Destination Ratios

4.5.2 *Largest Origin-Destination Pairs*

Figure 4-11 shows the numbers of journeys assigned to the top 50 origin-destination pairs in the Oyster-based matrix, with the corresponding number of journeys in RODS. These 50 OD pairs carry 10% of AM peak network demand. Appendix F lists all 50 OD pairs with the respective numbers of journeys in each matrix.

Possible explanations of the observed differences are[8]:

- **Waterloo to Bank & Monument** – Waterloo and Bank & Monument are uniquely connected by the Waterloo & City Line where passengers are not required to pass through gates when entering at the Waterloo end. These journeys will be recorded as unstarted journeys that finish at Bank & Monument and do not

[8] Geoffrey Maunder, London Underground

contribute to the seed matrix. Secondly, RODS does not capture the current travel patterns originating from Waterloo since it was last surveyed in 2000. Since the opening of the Jubilee Line Extension in 1999 between Westminster and Stratford, there are more destinations from Waterloo that were not captured in the 2000 survey. As a result, RODS assigned more journeys to Bank & Monument because it is a major destination from Waterloo, especially in the absence of the new destinations.

- **Waterloo and London Bridge to Canary Wharf** – Canary Wharf has grown rapidly in the past few years and is currently the second largest destination with nearly 40,000 exits in the AM Peak. Since Waterloo and London Bridge were last surveyed in 2000 and 2005, respectively, Waterloo does not have enough RODS responses to Canary Wharf compared to London Bridge. In addition, London Bridge is a fully gated station whereas Waterloo is non fully gated, the seed matrix definitely captures a more accurate profile of outgoing journeys from London Bridge than Waterloo. Therefore, many fewer journeys are assigned from Waterloo to Canary Wharf in RODS despite the fact that Waterloo is the largest origin with over 42,000 entries whereas London Bridge is the third largest origin with 24,000 entries.
- **Bank & Monument to Liverpool Street** – The large number of journeys from Bank & Monument to Liverpool Street in the Oyster-based matrix may be due to the lack of Waterloo & City Line gates at Waterloo. Passengers traveling from Waterloo to Central Line destinations via Waterloo & City Line need to pass through gates at Bank & Monument when transferring between the two lines. This might create the illusion that these passengers start their journeys at Bank & Monument instead of Waterloo. It is possible that this illusion also happens at other Central line destinations but the discrepancy is most significant at Liverpool Street because it is the largest destination on Central Line (also serving 3 other lines) and the third largest destination in the network.

- **Victoria to Kings Cross** – The Oyster-based matrix has only half the number of journeys from Victoria to Kings Cross as RODS because the number of exits at Kings Cross in the Oyster-based matrix is only half RODS, as shown in Figure 4-9. Other OD pairs that are possibly affected by the low number of exits at Kings Cross are journeys from Liverpool Street, Finsbury Park, Seven Sisters, Bank & Monument and other stations. However, the correct number of exits at Kings Cross is more plausibly closer to the 16,000 estimated in the Oyster-based matrix than the 26,000 in RODS.

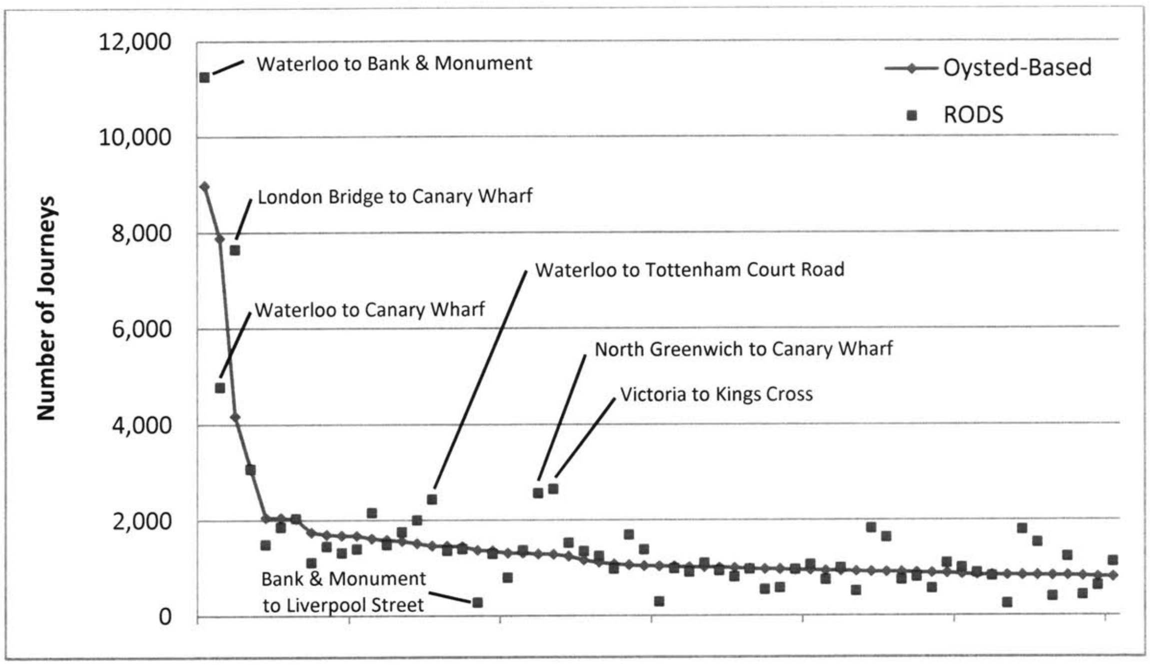

Figure 4-11: The 50 Largest Origin-Destination Pairs in the Oyster-Based Matrix

4.5.3 *Largest OD Flow Differences*

Figures 4-12 to 4-14 depict graphically the largest differences in OD journeys between the two matrices on the Underground network. The magnitude of the difference is represented by the thickness of the dotted line connecting the origin and destination stations. The Circle line is also shown (yellow solid line) to indicate the relative locations of the differences, since the Circle line covers most of Zone 1. Figure 4-12 shows the OD pairs that have at least 500 more journeys in the Oyster-based matrix than in RODS. Figure 4-13 shows the OD pairs that have at least 500 more journeys in RODS than in the Oyster-based matrix, and Figure 4-14 is a closer view of Figure 4-13 showing Zone 1 in greater detail.

The directions and magnitudes of flow differences shown in the visualizations are also shown in Tables 4-4 and 4-5, ordered by the decreasing magnitudes of the difference. The two and three largest differences in Tables 4-4 and 4-5, respectively, were discussed in Section 4.5.2 as they are also some of the largest OD pairs in the network.

From Station	To Station	Oyster-Based	RODS	Difference
Waterloo	Canary Wharf	7,884	4,775	3,109
Bank & Monument	Liverpool Street	1,370	265	1,105
Waterloo	Paddington	1,036	276	760
Kings Cross	Victoria	1,744	1,103	641
Victoria	Bank & Monument	843	237	606
Victoria	St James's Park	658	84	574
Stratford	Canary Wharf	2,054	1,481	573
London Bridge	Green Park	1,311	781	530
Liverpool Street	Bank & Monument	734	216	518

Table 4-4: OD pairs with at least 500 more journeys in Oyster-Based Matrix than RODS

While there are only 8 OD pairs with 500 more journeys in the Oyster-based matrix than in RODS, there are 28 OD pairs where the reverse is true. Since RODS captures only one-third of the OD pairs in Oyster, many journeys are over-allocated to the popular OD pairs in RODS. For example, RODS over-allocated many journeys to OD pairs ending at Kings Cross since RODS has twice the number of exits at Kings Cross than the Oyster-based matrix.

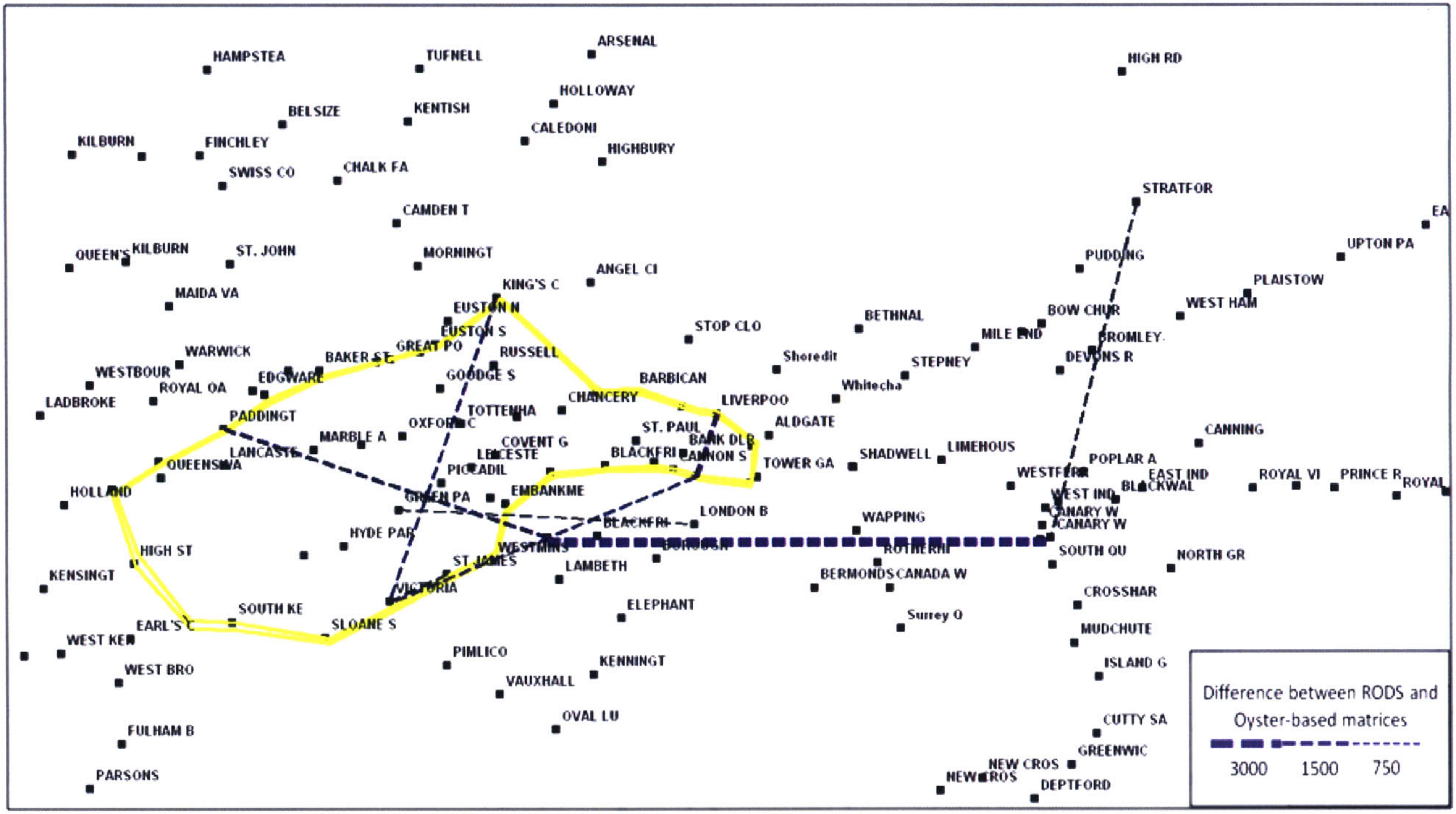

Figure 4-12: Visualization of OD pairs with at least 500 more journeys in Oyster-Based Matrix than RODS

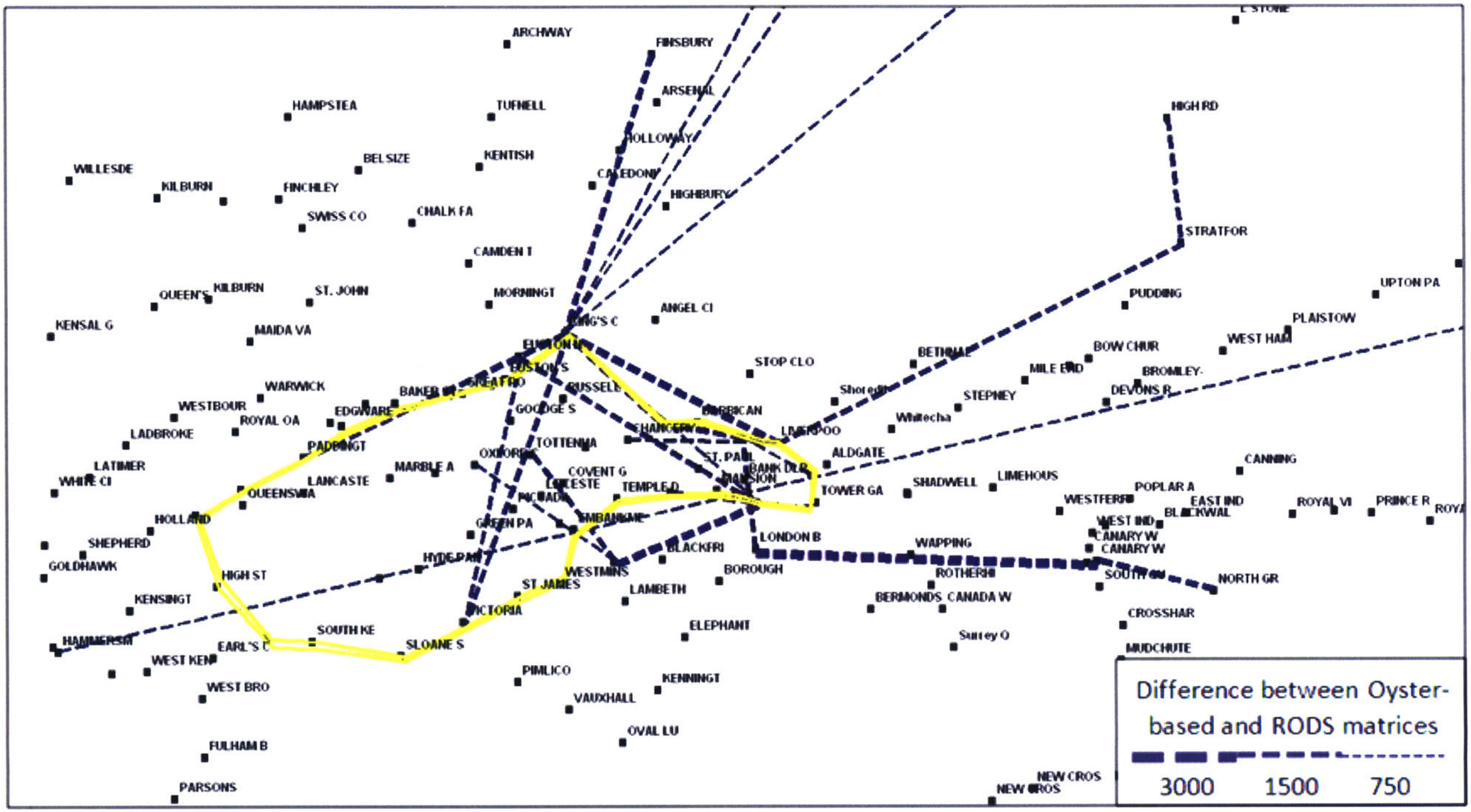

Figure 4-13: Visualization of OD pairs with at least 500 more journeys in RODS than Oyster-Based Matrix

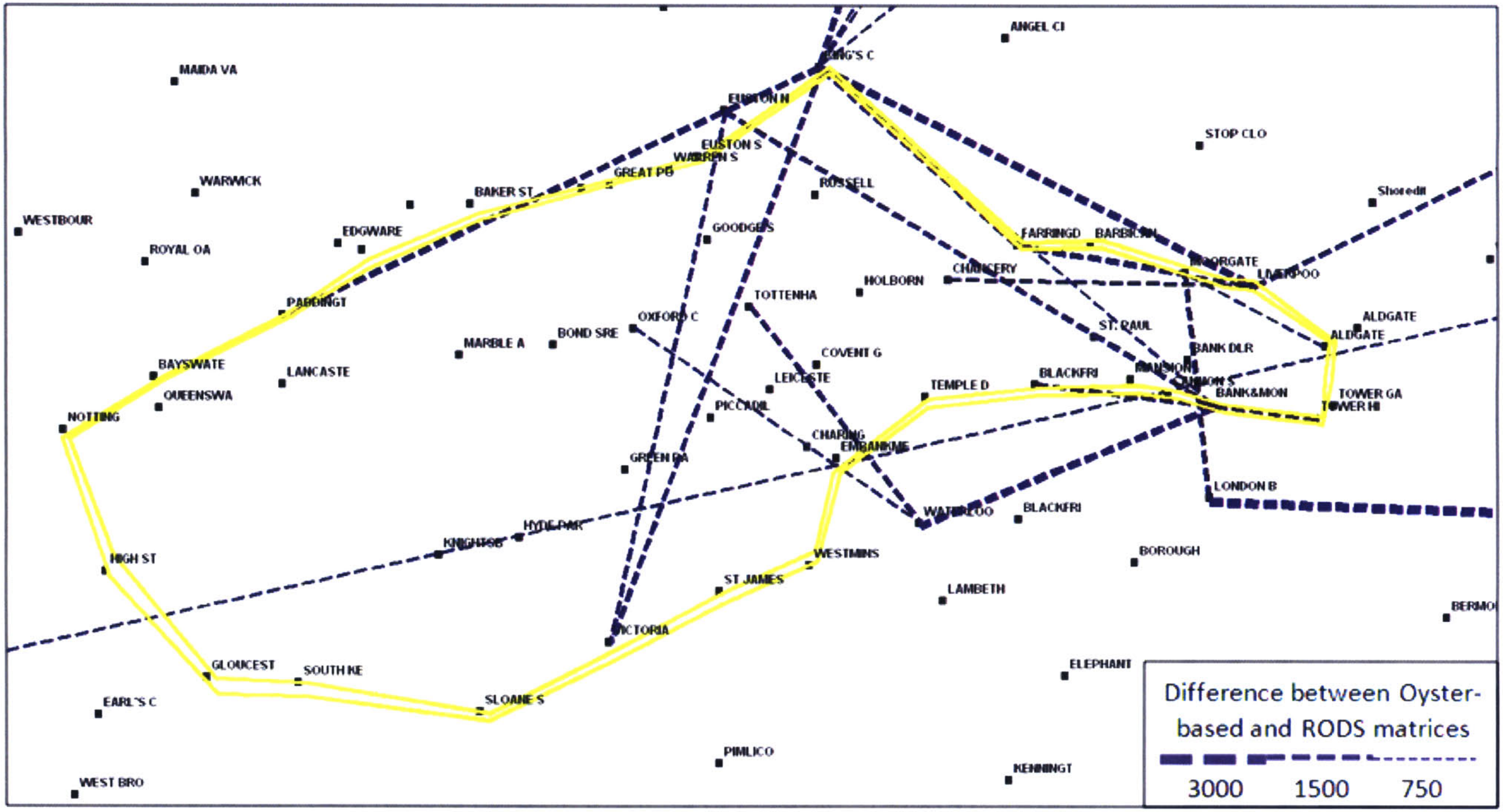

Figure 4-14: Closer View of Figure 4-13

There are significant differences in OD pairs ending at Canary Wharf, which has grown rapidly in the past few years. The RODS database does not have enough records to Canary Wharf to accurately reflect its current demands, especially in the AM peak. Other stations that show significant OD differences are Victoria, Bank & Monument, London Bridge and Waterloo.

From Station	To Station	Oyster-Based	RODS	Difference
London Bridge	Canary Wharf	4,172	7,641	3,469
Waterloo	Bank & Monument	8,981	11,256	2,275
Victoria	Kings Cross	1,282	2,648	1,366
North Greenwich	Canary Wharf	1,288	2,555	1,267
Paddington	Kings Cross	117	1,113	996
Moorgate	London Bridge	628	1,597	969
Waterloo	Tottenham Court Rd	1,461	2,429	968
Farringdon	Moorgate	841	1,788	947
Kings Cross	Farringdon	351	1,264	913
Euston	Bank & Monument	916	1,815	899
Liverpool Street	Stratford	126	880	754
Liverpool Street	Kings Cross	74	796	722
Victoria	Euston	914	1,625	711
Finsbury Park	Kings Cross	841	1,516	675
Kings Cross	Liverpool Street	97	766	669
Tower Hill	Blackfriars	732	1,376	644
Kings Cross	Paddington	138	775	637
Farringdon	Liverpool Street	770	1,402	632
Liverpool Street	Chancery Lane	1,057	1,684	627
Upminster	Hammersmith D	10	609	599
Seven Sisters	Kings Cross	565	1,163	598
Bank & Monument	Kings Cross	189	747	558
Moorgate	Aldgate	309	864	555
Leyton	Stratford	416	947	531
Waterloo	Oxford Circus	1,614	2,145	531
Moorgate	Farringdon	495	1,022	527
Walthamstow Central	Kings Cross	676	1,187	511
Tottenham Hale	Kings Cross	329	835	506

Table 4-5: OD pairs with at least 500 more journeys in RODS than Oyster-Based Matrix

4.5.4 *RODS Only OD Pairs and Obsolete Travel Patterns*

Earlier in this chapter, it was shown that 956 of the 17,421 OD pairs in the RODS matrix are travel patterns not captured in the Oyster-Based Matrix. These "RODS only" ODs account for about 1.7% of demand in the RODS matrix. Of those 956 ODs, 127 are to or from stations that were closed during January 2007, and therefore no journeys were possible. Table 4-6 shows the zonal distribution of the remaining 829 "RODS only" OD pairs between open stations. Travel from an inner zone to Zone 4+ in the AM Peak is uncommon but the table shows that 240 of the 829 "RODS only" ODs are of this type.

Of the 829 ODs, 611 have recorded Oyster journeys during January 2007 but the numbers are so small that they become zero after rounding to the nearest integer, leaving 218 truly "RODS only" OD pairs.

From Zone	To Zone				
	1	2	3	4+	Sub Total
1	10	63	38	143	254
2	10	48	46	76	180
3	15	33	24	21	93
4+	54	114	66	68	302
Sub Total	89	258	174	308	829

Table 4-6: Zonal Distribution of "RODS Only" OD Pairs

Table 4-7 shows several examples of "RODS only" OD pairs and the number of journeys in both the RODS and Post IPF Oyster-based matrix (before rounding). It is important to note that many of these OD pairs involve travel from an inner zone to Zone 4+ via the central zones which is uncommon in the AM Peak. The year of most recent RODS survey at the origin station may also be indicative of obsolete travel patterns. Harlesden, Amersham and Stonebridge Park were last surveyed in 1998, and it is very likely that those OD pairs may now be inactive if no Oyster journeys are recorded in 2007. The number of RODS surveys represents the total number of RODS responses collected since the beginning of the RODS program. All of the OD pairs shown in Table 4-6 have only 1 or 2 RODS surveys for the AM Peak, corresponding to expansion factors ranging from 28 to 166. The numbers of journeys on such OD pairs are likely to be over-allocated in the RODS iterative fit adjustment procedure to match OD flows to station control totals.

Some of the OD pairs involve very lengthy journeys and inconvenient transfers. For example, to go from Hounslow West to Upney, a passenger goes from Piccadilly Line in Zone 5 west of Zone 1 to District Line in Zone 4 east of Zone 1 and the journey takes approximately one and a half hours. It is rather unlikely that as many as 68 passengers would be making such a difficult journey, especially in the AM Peak.

From Station	To Station	From Zone	To Zone	Most Recent RODS Survey	RODS Surveys	RODS Journeys	Oyster-Based Journeys (before rounding)
Tower Hill	Hillingdon	1	4+	2004	1	166	0.12
Northwood	Chorleywood	4+	4+	2003	1	138	0.31
Harlesden	Hammersmith D	3	2	1998	1	93	0.18
North Greenwich	Harrow Wealdstone	2	4+	2004	1	91	0.24
Golders Green	Kew Gardens	3	3	2002	2	81	0.48
Southfields	Brixton	3	2	2004	2	66	0.19
Stonebridge Park	Leicester Square	3	1	1998	1	60	0.36
Green Park	Ladbroke Grove	1	2	2000	1	59	0.43
Snaresbrook	Royal Oak	4+	2	2002	1	57	0.07
Richmond	Balham	4+	3	2004	2	57	0.16
Canning Town	West Acton	3	3	2004	1	57	0.09
Rayners Lane	Bow Road	4+	2	2004	1	56	0.16
Bow Road	Seven Sisters	2	3	2003	1	54	0.48
Gunnersbury	Harrow On The Hill	3	4+	2004	1	52	0.28
Sudbury Hill	Kilburn	4+	2	1999	1	52	0.39
Willesden Junction	Upton Park	3	3	2004	1	52	0.04

Table 4-7: Examples of "RODS Only" OD Pairs for the AM Peak

4.6 OD Matrix Assignment

By assigning the Oyster-based and RODS AM Peak matrices onto the Underground network, differences in link loads and interchange volumes can be examined. The results in this section are obtained from assigning the OD matrices using TransCAD's Pathfinder method. TransCAD is a comprehensive, integrated package for demand forecasting that combines demand modeling procedures and tools with a built-in geographic information system (GIS) for transportation. Pathfinder is one of the six transit assignment methods provided in TransCAD capable of computing specific walk paths for access and egress for any OD pairs and solving for paths that provide the minimum generalized cost of travel using specified penalty factors. Descriptions of TransCAD and Pathfinder can be found in Caliper (2005).

The OD assignment results shown in this section did not follow a calibration exercise because the purpose is only to demonstrate the differences between the Oyster-based and RODS matrices in terms of critical links loads and interchange volumes. The weight factors used to compute the generalized cost of travel are 2.0 for platform wait time and 3.0 for access, egress and interchange walk times. Three interchange penalties were used – 3, 5 and 8 minutes per interchange – to compare the differences between the assigned matrices and the official RODS figures. The results of the 5-minute penalty are closest to the official RODS figures and are shown in this section.

For each link and interchange station, three sources of volumes are compared: 1) volumes based on the TransCAD assignment of the Oyster-based matrix, 2) volumes based on the TransCAD assignment of the RODS 2005 matrix, and 3) TfL-scaled volumes revealed by passengers in the path choice questions in the RODS survey.

Two levels of comparisons can be made from the three sources of volumes. First, the volumes obtained from the TransCAD assignment of the Oyster-based and RODS matrices should show differences resulting from OD flow differences because the same assignment algorithm is used. Second, the volumes obtained from the RODS assignments can also be compared with the RODS path choice volumes because the two are from the same underlying OD matrix but assigned using different path choice methodologies. The first comparison is the focus of this section because the TransCAD assignment exercise has not calibrated to match the official RODS figures.

Sections 4.6.1 and 4.6.2 present the results of preliminary analyses of link loads and interchange volumes, respectively, using a 5-minute interchange penalty factor in Pathfinder. As the graphs show, some significant differences between the Oyster-based and RODS matrices exist.

4.6.1 Link Loads

Figure 4-15 shows the 15 most congested links in the Underground obtained from the "official" RODS AM Peak link loads and the corresponding values from the RODS and Oyster-based matrices assignment. The heaviest links shown are concentrated on two lines: 8 are on the Central Line westbound, and 5 and 2 links are on the Victoria Line northbound and southbound, respectively, as shown in Table 4-8.

The link loads from the matrices assignments are systematically lower than the RODS official loads in all but 1 of 15 links, with an average discrepancy of approximately 5%. The differences are significant in terms of passengers' in-train discomfort and resulting delays. The systematic differences could possibly be due to the specified penalty factors used to compute the minimum generalized cost of travel in TransCAD, leading to chosen paths different from those in the RODS survey.

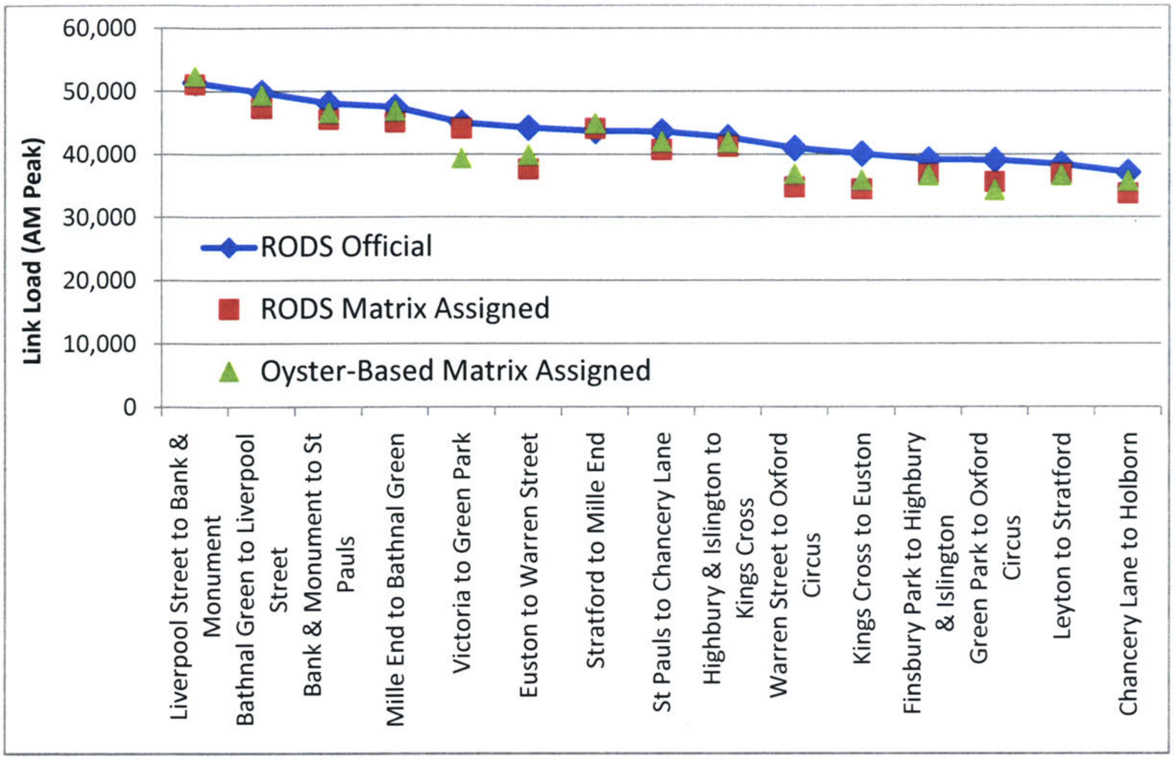

Figure 4-15: Largest 15 Link Loads

OD Pair	Line	Direction	RODS Official	RODS Matrix Assigned	Oyster-Based Matrix Assigned
Liverpool Street to Bank & Monument	Central	WB	51,280	50,928	52,266
Bathnal Green to Liverpool Street	Central	WB	49,738	47,165	49,231
Bank & Monument to St Pauls	Central	WB	47,969	45,441	46,467
Mile End to Bathnal Green	Central	WB	47,436	44,968	46,789
Victoria to Green Park	Victoria	NB	44,899	44,002	39,325
Euston to Warren Street	Victoria	SB	44,143	37,545	39,796
Stratford to Mile End	Central	WB	43,606	44,002	44,837
St Pauls to Chancery Lane	Central	WB	43,584	40,672	41,939
Highbury & Islington to Kings Cross	Victoria	SB	42,622	41,179	41,878
Warren Street to Oxford Circus	Victoria	SB	40,926	34,738	36,861
Kings Cross to Euston	Victoria	SB	40,044	34,402	35,897
Finsbury Park to Highbury & Islington	Victoria	SB	39,093	36,758	36,622
Green Park to Oxford Circus	Victoria	NB	39,034	35,564	34,312
Leyton to Stratford	Central	WB	38,357	36,847	36,647
Chancery Lane to Holborn	Central	WB	37,094	33,751	35,729

Table 4-8: Largest 15 Link Loads

The link loads in the assigned Oyster-based and RODS matrices are very close, with an average discrepancy of less than 2% among the largest 15 links. Since the two matrices are assigned using the same algorithm and factors, the small differences indicate that the OD flow differences do not lead to large differences in the critical link loads. While the average discrepancy is 2%, the maximum is 11% between Victoria and Green Park on the Victoria line northbound. These differences are likely to be even larger for shorter time periods and lead to different assessments of in-train discomfort and journey time delays.

4.6.2 *Interchange Volumes*

The station level interchange volumes are the total platform-to-platform movements made by passengers transferring between lines within a station complex. The three sources of interchange volumes show different numbers of stations with interchange movements, as shown in Table 4-9. Each interchange is counted separately and therefore one journey can contribute multiple interchange movements. The interchange volumes in RODS Path Choice are calculated by scaling the path choice information revealed in the RODS surveys to match station level entry and exit counts, and the interchange volumes are obtained once the number of journeys on each OD pair and path are estimated. The RODS Path Choice has the largest number of interchange stations (93) because it includes passengers who transfer on the same line due to short-turning of trains. For example, a passenger may board a short-turn train and transfer to another train that runs full service to the end of the line. On the other hand, the TransCAD assignment assumes that passengers wait at the origin station for a train that serves the intended destinations instead of boarding a short turning train and then transferring after it terminates.

	Number of Interchange Stations (AM Peak)	Total Interchange Movements (AM Peak)
Oyster-Based Matrix Assignment	87	333,927
RODS Matrix Assignment	75	304,935
RODS Path Choice	93	338,098

Table 4-9: Three Sources of Interchange Volumes

While the Oyster-based matrix contains 1% more journeys than the RODS matrix, it has 10% more interchange movements. This is probably because the Oyster-based matrix captures more OD pairs that are less heavily traveled, longer and require more interchanges. On the other hand, the interchange volumes obtained from the RODS matrix and the RODS Path Choice are based on the same OD matrix, yet the RODS Path Choice contains 11% more interchange movements. The discrepancy could result from differences in assignment rules, most notably the assumption in the TransCAD assignment that passengers wait for full service trains.

Figures 4-16 to 4-18 show the differences of interchange volumes among the three sources by volume categories. When comparing the two interchange volumes obtained from matrix assignments, it can be seen that RODS systematically under-estimates interchange volumes at both high- and medium-volume stations. The differences can also be significant – 15% at Oxford Circus, 14% at Bank & Monument, 17% at Westminster and 27% at Holborn. The discrepancy at Kings Cross is largest at 41% but this is due to the problem of gate volume double counting until November 2006. Section 4.4 points out that at Kings Cross the RODS matrix has 13,500 more exits than the Oyster-based matrix, but 6,000 fewer interchanges as shown in Figure 4-16. RODS surveys that are conducted after November 2006 should provide more accurate estimates of station level entry and exit counts, as well as interchange volumes, because passengers are no longer required to pass through two sets of gates to transfer between certain lines and the double counting should be resolved.

For all three volume categories, significant differences in interchange volume exist between the RODS matrix assignment and the RODS Path Choice. Some of these differences are at stations which are known for passengers interchanging between services on the same line due to short-turning of trains, such as Wembley Park on the Jubilee line and Camden Town on the Northern line.

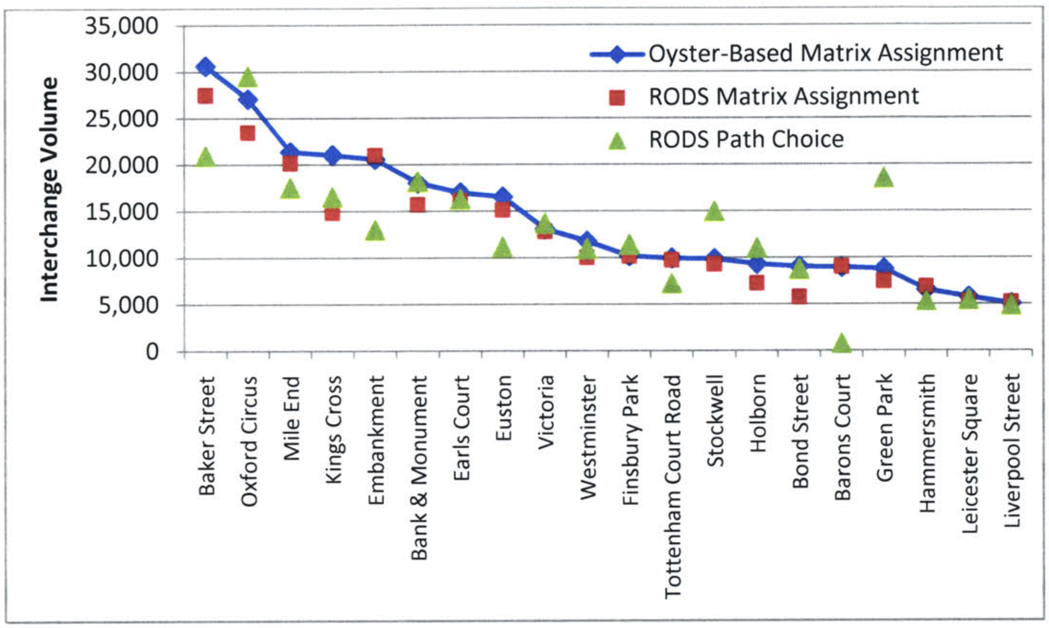

Figure 4-16: High Volume Interchange Stations

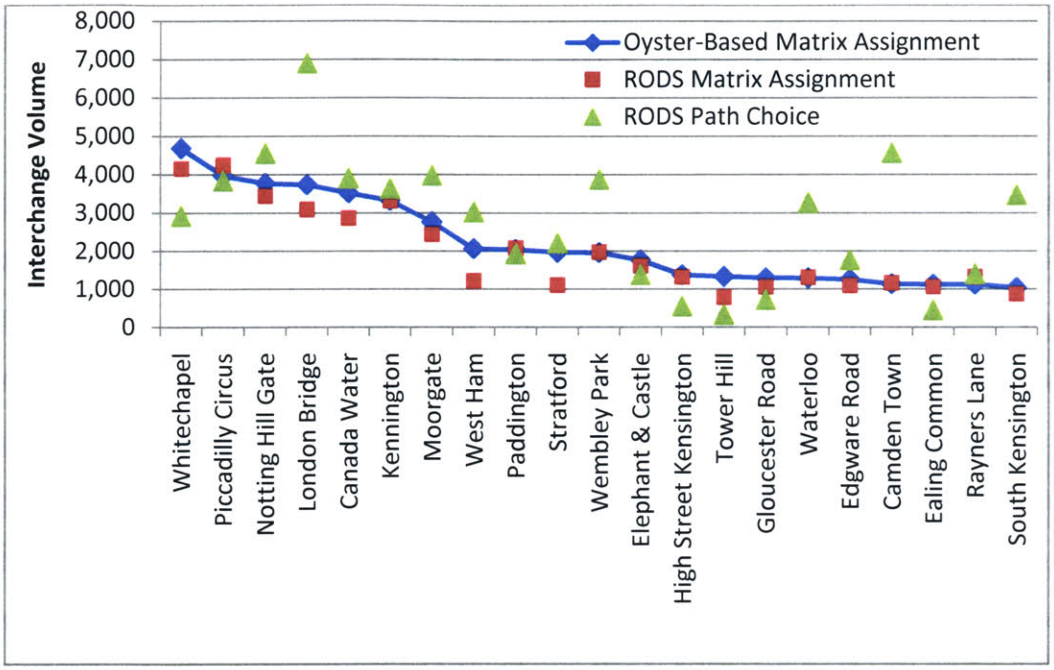

Figure 4-17: Medium Volume Interchange Stations

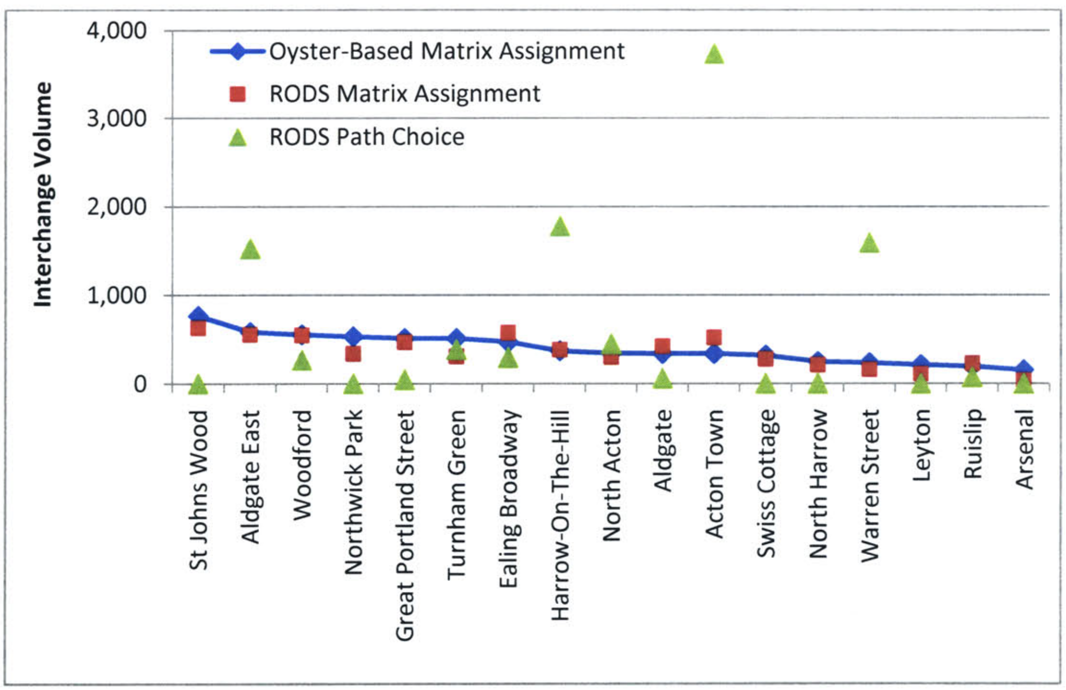

Figure 4-18: Low Volume Interchange Stations

Chapter 5 Journey Time Metrics

In transit systems that have both entry and exit fare payment control, AFC transactional data have enhanced transit agencies' ability to obtain large samples of origin-destination journey time information in close to real time. The goal of this chapter is to provide a complete overview of journey time estimation before and after the availability of AFC data in the London Underground. Section 5.1 provides an overview of previous research on measuring transit service reliability. Section 5.2 introduces the survey method currently used in the London Underground to estimate journey time performance. Section 5.3 discusses the characteristics of Oyster transactional data in relation to measuring elapsed journey time between entry and exit. Section 5.4 proposes the Excess Journey Time Metric which assesses excess OD journey time by selecting OD pairs on the same line. Section 5.5 presents the Journey Time Reliability Metric which quantifies service reliability, again by selecting OD pairs on the same line.

5.1 Literature Review

Considerable research has been conducted on four aspects of transit service reliability:

- the effects of reliability on passengers and transit agencies
- the selection of reliability measures
- the identification of the causes of unreliability
- the application of strategies to improve reliability

This section focuses on the selection of reliability measures which have been revolutionized by the development of automatic data collection in recent years. Further descriptions of the other three aspects can be found in the Abkowitz et al. (1978), and Cham (2005) provides a summary of findings in that study.

Abkowitz et al. (1978) evaluated the typical service reliability measures prior to 1978 and revealed three major weaknesses:

- The use of published schedule to compute measures of delays and deviations may have skewed the results due to schedule inefficiencies such as inadequate running times and recovery times.
- Some measures are incapable of capturing the effects of unreliability from the passenger's perspective.
- Data collection limitations lead to the inability to capture daily and seasonal variability, as well as across time periods.

The authors recommended the following reliability measures, based on the identified weaknesses of previous reliability measures, to address time-of-day and day-to-day variability and reflect both the perspectives of passengers and transit agencies. These measures should enhance transit agencies' ability to consider the interrelationships between different measures and the development of mathematical relationships to better understand their interactions and the effects of service attributes. (Cham, 2006)

- Distributions of travel time (total travel, in-vehicle and wait) – mean, coefficient of variation and percent of observations N minutes greater than the mean value
- Schedule adherence at any point along the route – average deviation from schedule, coefficient of variation for average deviation and percent of arrivals N minutes later than average deviation from schedule
- Distribution of headways – mean, coefficient of variation, percent of headways greater than X (≥ 1) percent of average or scheduled headways and lower than Y (≤ 1) percent of average or scheduled headways
- Seat availability measured by passenger loads and capacity

Cham (2006) recognized that schedule deviations at terminals, passenger loads, running times, environmental factors, operator behavior and the inter-relationships between these causes are the most significant contributors to reliability problems. Cham proposed a practical framework serving as a comprehensive guide for transit agencies to analyze large amounts of ADC data to evaluate performance and implement efficient strategies to improve service planning, operations monitoring and management procedures. The framework consists of three blocks: 1) characterizing service reliability through service measures and performance reports; 2) identifying the causes of reliability problems; and 3) selecting strategies to improve service by targeting critical causes of unreliability.

In particular, Cham suggested four service measures:

- **Distribution of Service Attribute** – The distribution of a specific service attribute reflects variability that is often overlooked by simple statistical measures. Service attributes of interest include running times, deviations from scheduled times, and deviations from scheduled headways.
- **Wait Times** – Wait times can be assessed in terms of expected wait time, excess passenger wait time and percent of excess time. The expected wait time is a function of scheduled headway whereas the excess wait time is the difference between actual expected wait time and the scheduled expected wait time to reflect the impacts of poor headway adherence.
- **Overcrowding** – Overcrowding is a measure of high passenger loads where the number of passengers on board exceeds a threshold value which affect passenger comfort. It is related to reliability because high loads increase dwell times and possibly lead to poor headway adherence, and thus increased wait times.
- **Percent of Unreliable Trips** – Whether a trip is reliable can be defined with a given threshold value for the service attribute concerned. Service attributes of interest include late/early departures, late arrivals, bunches of vehicles, large gaps between vehicle arrivals, late garage pull-out and late relief.

Threshold values should be decided for each of the service measures to classify a service as reliable or unreliable. The thresholds or ranges of thresholds should be based on the level of service the transit agency can cost effectively deliver and may vary by mode, route type and time of day.

5.2 Existing Journey Time Metric at London Underground

The Journey Time Metric (JTM) was developed by London Underground Marketing and Planning in 1997 to monitor overall Underground service performance. By measuring a journey by its time components and incorporating passenger demand, the JTM is a customer-focused performance measure that emphasizes the customer experience rather than simply train operations. The JTM is reported every period at the line and network levels.

Each journey is broken into stages as a passenger would experience them: ***access*** from station entrance to the gate area and then on to platform, ***ticket queuing and purchase, platform wait, on train time, interchange*** between platforms and ***egress*** from platform to station exit. Each time component has a scheduled value which represents the amount of time a passenger should normally expect to take to finish the stage. Actual journey times are measured by component in each period, the difference between the measured and scheduled times is then the indication of service performance expressed as average passenger excess journey time.

In addition to the average excess journey times at the line and network levels for the entire period, the monthly JTM report also disaggregates the excess journey time by day and timeband to highlight the major contributors to delay. This breakdown provides a good tool for management to track improvements made by line and station managers.

5.2.1 Scheduled and Actual Journey Time Components

The scheduled journey time components form the basis of comparison with each period's measured times. The scheduled times are usually stable over time but it is important to maintain up-to-date scheduled times to reflect the effects of both capital investment and non-capital improvements. Capital improvements range from station re-design to changes in rolling stock, and non-capital improvements include tighter management of train dwell times and faster customer service at ticket booths.

On the other hand, measured journey times tend to fluctuate significantly from period to period and are highly sensitive to train performance and travel demand. The periodic measures rely heavily on both manual surveys at stations and train signal data as well as models to estimate the component times.

Access, Egress and Interchange (AEI)

The AEI walk times are measured by manual surveys or a station simulation model depending on station volume. The definitions of the walk times are the same in both cases:

- Access – from station entrance(s) to midpoint of platform(s)
- Interchange – from midpoint of arriving platform to midpoint of departing platform, assuming the interchange walk starts immediately after the train arrives at the platform
- Egress – Midpoint of platform(s) to station exit(s), assuming the egress walk starts immediately after the train arrives

The manual AEI surveys are conducted at 27 major stations which together account for 46% of network demand. These stations are surveyed at least 12 times a period during the busiest timebands between 7AM to 7PM on weekdays. Data are collected by survey staff walking predefined routes that cover every possible walk route within the stations.

The PEDS model is used to assess the congestion at all stations, including those with AEI surveys. The model incorporates recorded events such as lift and escalator failures and demand fluctuations to calculate the excess AEI times at each station. PEDS supplements the surveyed AEI data for the 27 major stations and is solely responsible for the AEI results for the remaining 246 stations.

The scheduled AEI times are assessed as above but assume free flow conditions in the station under which passengers are able to walk unimpeded.

Ticket Purchase Time (TPT)

The ticket purchase time component is measured by the Time In Queue Survey (TIQS) which consists of two parts: queuing time and transaction time. The queuing time is based on the number of people standing in ticket queues during a selected interval. The transaction time is taken at staffed ticket office windows at all stations and passenger operated ticket machines at stations where queues are common. The frequency of TIQS depends on how busy a station is – "busy" stations are surveyed every period, "medium" stations are surveyed once every quarter and "quiet" stations are surveyed once a year.

In order to reflect the impacts of LUL initiatives that encourage off-site and less frequent purchases, the number of ticket sales in each period is normalized by the total number of entries. This ratio is used to calculate the average ticket purchase time for all passengers entering the station regardless of whether or not they purchase tickets.

The scheduled ticket purchase time is based on 90% of average transaction times observed during the previous year for each station. The scheduled time does not include any expected queuing time.

Platform Wait Time (PWT) and Left Behinds

The platform wait time is defined as the elapsed time after the end of access and before the beginning of on-train travel, i.e. PWT is the time from passenger arrival at midpoint of platform to the wheel start of the boarded train. Passengers are assumed to board the first train to their destination or a convenient interchange point.

For all lines except the District, the PWT is derived from the 100% line signaling data that contain train movement information to the minute. For the District Line, PWT is measured by the Time on Platform Survey which employs staff to record departure times and destinations of trains.

Due to the complications of line branching and train short-turning, each line is subdivided into segments that have the same service frequency. The average platform wait time from each origin segment to each destination segment is calculated by the sum of each headway squared divided by twice the sum of all headways.

The PWT also takes into account additional wait time incurred by passengers who are unable to board the first train due to crowding. The additional time is known as "left behinds" and is calculated based on current excess PWT results and RODS link loads scaled by current station entries and exits.

The scheduled platform wait times are derived from current train timetables (by taking one-half of the scheduled headway). The scheduled "left behinds" is zero.

On-Train Time (OTT)

As for platform wait times, the on-train time is also defined by line segments which share the same service frequency. Train information in recorded at the start and end of each line segment. For lines with 100% sampling from the train signaling system, the OTT is calculated directly from train data. For the other lines, OTT is sampled from signal cabin box sheets. These lines include the District and Piccadilly lines, and portions of the Metropolitan and Hammersmith & City lines.

The scheduled on-train times are derived from current train timetables.

Closures

In addition to the above journey time components normally experienced by passengers, JTM also captures excess journey times incurred by short and long term service disruptions. This component is particularly important to JTM calculations during periods of disruptions when empirical data may not be available. There are three types of closures:

- **Unplanned, Short Term** – The Nominally Accumulated Customer Hours (NACHs) system uses information from the Contract Performance Information Database (CuPID) system to calculate excess journey times incurred by short term line or station disruptions that exceed 30 minutes. The CuPID database documents all station and train malfunctions and delays by type and duration. Examples of short-term closures include track failures, and lift and escalator failures.
- **Planned, Short Term** – LUL are aware of planned, short term closures sufficiently in advance to inform passengers. The impacts of these closures are therefore somewhat lessoned by providing alternatives to passengers such as bus replacement services. NACHs is also used to calculate the impacts of these closures.
- **Planned, Long Term** – Planned, long term closures usually last four weeks or longer and significantly affect passenger travel during the disruption. LUL is required to make passengers aware of the closure at least two weeks in advance. The Train Service Model (TSM) is used to estimate the additional journey time which becomes part of the scheduled journey time during the period of planned, long term closure. Therefore the additional estimated journey times will only be included in the actual journey time results but not in the excess journey time results.

5.2.2 *Value-of-Time Weightings*

As a customer focused measure, the JTM uses a value of time (VOT) weighting for each journey time component according to how negatively passengers perceive the journey stage. The overall weighted journey time is known as the generalized, or perceived, time. The VOT weights for each journey element are shown in Table 5-1.

Journey Element	Value-of-Time Weighting
Walking horizontally	2
Walking up stairs	4
Walking down stairs	2.5
Walking up and down stairs	3.25
Taking Escalator or Lift	1.5
Queuing for ticket	3.4
Purchasing ticket	2.5
Waiting on platform	2
Being left behind on platform	3
Traveling on train	Varies between 1 and 2.48 depending on crowding level of train

Table 5-1: Value-of-Time Weightings

The breakdowns of journey time components differ significantly between unweighted and weighted journey times. An average journey in the Underground is 28 minutes and, the unweighted and weighted component times are shown in Figures 5-1 and 5-2, respectively.

5.2.3 *Data Aggregation*

At the end of each period the collected journey time component data are input into the Line Aggregator which calculates the line level average journey time. The demand weightings applied to the disaggregate journey time data are calculated from the RODS

2002 line usage data by timeband. Since 2003, the RODS 2002 line usage data has been used to keep demand constant in each period so that changes in JTM results are solely due to changes in service performance and not demand fluctuations. However, changes in demand influence journey times because congestion can result in delays when trains are running perfectly on schedule.

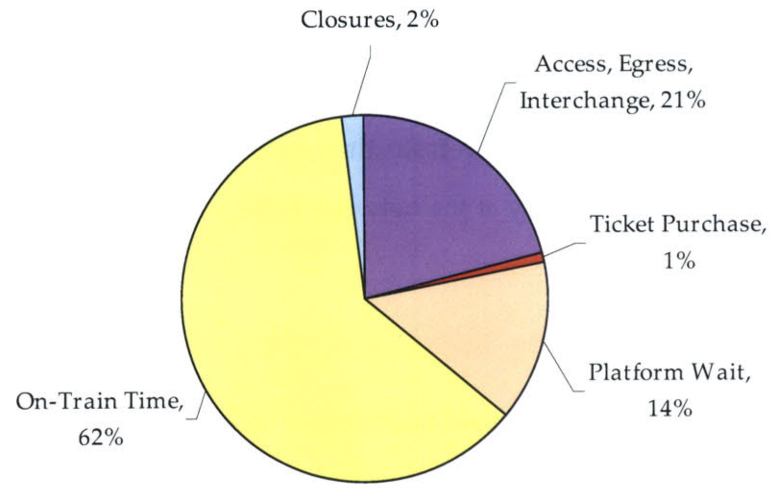

Figure 5-1: Total Unweighted Journey Time Components

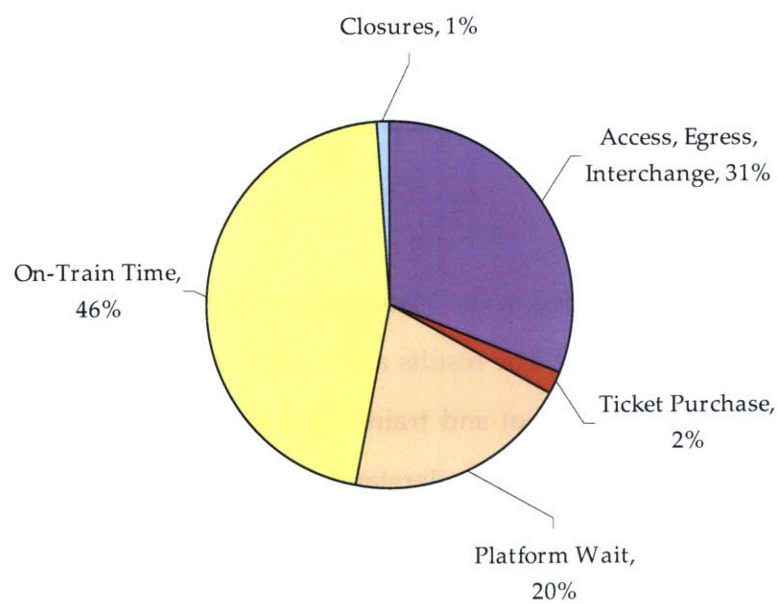

Figure 5-2: Total Weighted Journey Time Components

The line level results are in turn used to calculate the overall network average journey time. Each time component contributed by interchange passengers are captured in the line level results. The Journey Leg Factor is applied to both scheduled and actual journey times to combine the line level results and acknowledges the fact that approximately 40% of all journeys in the Underground involve at least one interchange. The line level results assume that all passengers wait for their desired service and no unnecessary interchanges occur. On the other hand, interchanging passengers experience platform wait time and on-train time more than once in their journeys and these extra times are captured only at the network level.

5.2.4 *JTM Outputs*

The JTM Performance Report published each period includes the following sections:

1. Highlights of major events and disruptions
2. Average actual journey times, both weighted and unweighted, compared to the previous 13 periods as trend analysis
3. Updated scheduled journey times, both weighted and unweighted, as a result of service improvements or alterations
4. Excess journey time, both weighted and unweighted, contribution by time component

The JTM results are also published in the Management Information Brochure for regular review by the line management. The results are separated into two parts, station based time (AEI and ticket purchase time) and train based time (platform wait and on-train time), for line management to better understand the causes of excess journey time in their respective domains.

5.2.5 *Assessment of JTM*

The Journey Time Metric is a well established mechanism and, as with all mechanisms, it has both strengths and weaknesses.

The strengths of JTM are:

- **Effective Management Tool** – The JTM is a single, simple measure for each line and the full network that captures the complete Underground station entry to exit passenger experience. It provides the Underground with an effective management diagnostic by summarizing the overall Underground performance in just a few figures.
- **Truly Customer Focused** – By measuring the distinct journey time components, the perceived journey time (weighted by value-of-time) can be best improved by tackling components that have large excesses and high value-of-time weightings. Also, performance at key stations and line segments are emphasized by weighing the results by estimated demand, although the recent growth in ridership has resulted in an increase in excess journey time on most Underground lines.
- **Reliable Train Data** – all lines except the District provide a complete sample from the train signaling system and thus the measure of on-train time is accurate.

The weaknesses of JTM are:

- **Attribution of Station AEI Times to One Line Only** – The access, egress and interchange times at each station are attributed to the line which manages the station. For example, the Oxford Circus station is under the Bakerloo line management and thus all AEI times within the station are attributed to the Bakerloo line although the station also serves the Victoria and Central lines. JTM argues that the line management should be responsible for station performance as managed by its station staff. However, such attribution potentially leads to unfair comparison of performance between lines because some lines include more busy stations and are thus prone to higher levels of excess AEI times.

- **Use of Static Demand Data** – By keeping the OD travel pattern and line usage information at static RODS 2002 level, the JTM does not reflect the excess journey times for current travel patterns. Average journey lengths and demand will have changed in the past 5 years. For example, RODS 2002 contains only limited information on travel patterns involving stations along the Jubilee Line Extension which was opened in 1999 and significant growth has occurred since then. Also, while JTM includes weekend measurements, RODS 2002 weekend demand is estimated from weekday survey data which are most likely to be different from the weekend.
- **Unreliable and Costly Survey Data** – The AEI and ticket purchase time measurements are partly and wholly survey based, respectively. The AEI and TPT surveys are costly and provide only a small sample for each period's measurements. Small samples are prone to inaccuracies and may misrepresent actual station performance.

5.3 Overview of Oyster Transactional Data

Before utilizing Oyster data for journey time calculations, it is important to have a thorough understanding of the definitions and characteristics of the dataset. Since Oyster usage is not uniform across OD pairs, it is also important to examine how representative Oyster journeys are of all journeys in terms of journey time profiles.

The methodologies to be presented in Sections 5.4 and 5.5 rely on both Oyster transactional data to estimate journey times and the OD flow matrix to weigh the OD level calculations to give line level results. In this and the following chapters, the following data sources are used to illustrate the Journey Time Reliability Metric methodology unless otherwise noted:

- **Oyster Transactional Data from the AM Peak Period of LU Period 12 (20 weekdays in February 2007)** – The Oyster transactional data contain information on station and time for both entry and exit. Journey times can be calculated for each of the 11.6 billion completely documented journeys in the dataset.
- **Oyster-Based OD Flow Matrix for the AM Peak Period of LU Period 11 (January 2007)** – This is the average AM Peak OD flow matrix discussed in Chapter 3. The OD matrix is used to aggregate the OD level journey time results to the line level. Note that the OD matrix is derived from data one period before the Oyster Transactional Data.

The use of data sources from two different LU periods is due to time constraints on data availability and completion of research. While not ideal, the use of an OD demand matrix from the previous LU period is unlikely to significantly alter results at the line level because the differences in individual OD pairs will be reduced by presence of a large number of OD pairs on each line. When aggregating OD level journey time calculations to line or line segment level, the Oyster-based estimated OD demand from the same period should be used.

5.3.1 Definition of Oyster Journey Time

An Oyster transaction is generated every time a passenger taps the Oyster card at a gate or a stand-alone reader. Each transaction records the station, time, fare product and movement direction. Movement direction is either entry or exit and a completely documented journey is recorded if an entry transaction is followed by an exit transaction. However, if two consecutive transactions are of the same movement direction, one of the two transactions will automatically become an incompletely documented journey. With two consecutive entry transactions, the first transaction will be recorded as an unfinished journey since no exit information is recorded. With two consecutive exit transactions, the second transaction will be recorded as an unstarted journey since no entry information is recorded.

Oyster journey times can be calculated from completely documented journeys by taking the difference between the entry and exit transaction times. Oyster transaction times are recorded to the minute, truncating all seconds after the minute. For example, all transactions made between 06:00:00 and 06:00:59 have recorded transaction times of 06:00. Therefore, the value of an Oyster journey time is subject to an error of +/- 59 seconds, as illustrated in Figure 5-3.

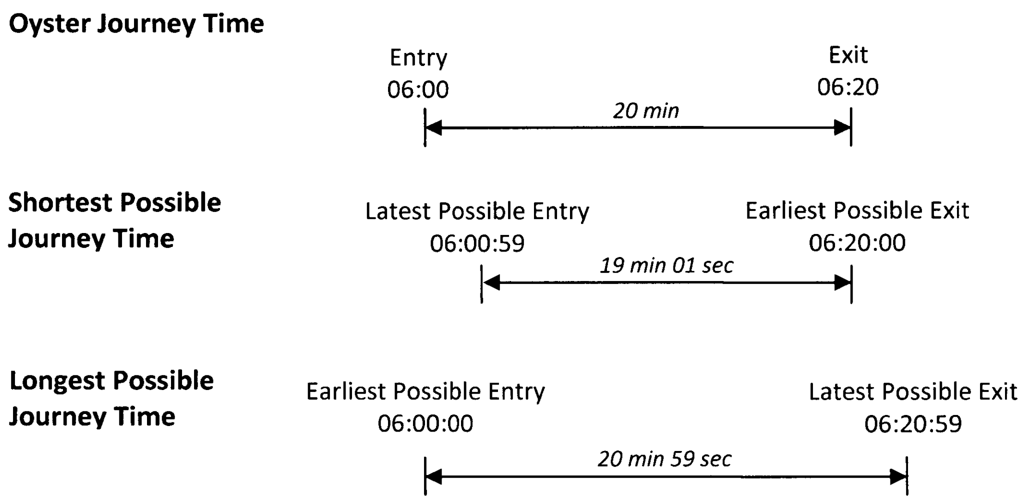

Figure 5-3: Oyster Journey Time Errors

5.3.2 *Oyster Journey Lengths*

Figures 5-4 and 5-5 show the distribution of Oyster journey time observed in the Underground network during 20 weekday AM Peaks in January 2007. There are a total of 11.7 billion Oyster journeys, corresponding to approximately 584,000 journeys in an average weekday AM Peak period. Around 66% of all Underground journeys are Oyster completely documented journeys, although the overall Oyster penetration rate is higher due to unfinished and unstarted journeys. The long tail on the right of Figure 5-4 is indicative of unusual delays, and some unusually lengthy journeys and perhaps some passenger waiting behavior (e.g. for friends) within stations that is independent of the real journey time.

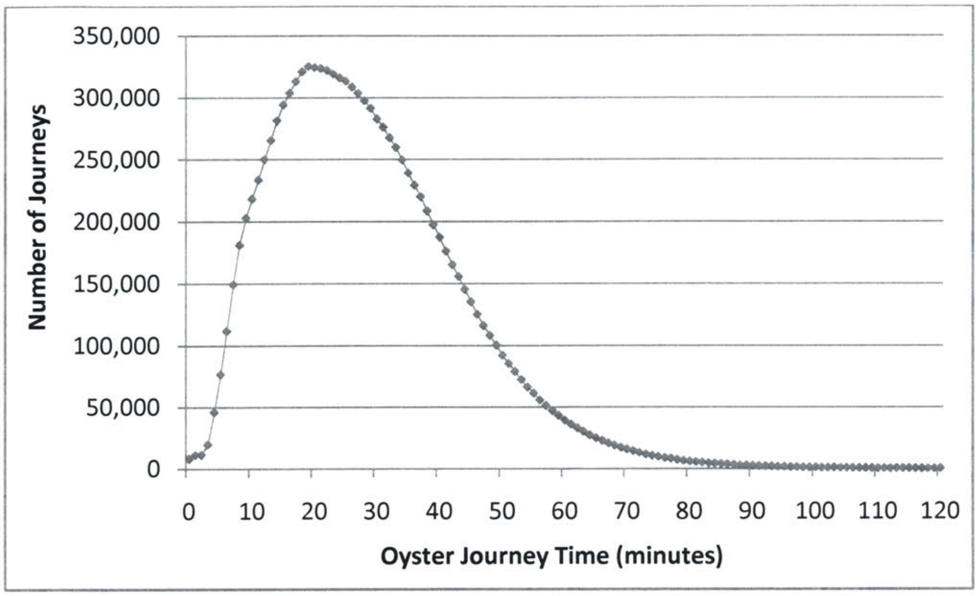

Figure 5-4: Distribution of Oyster Journey Time (Continuous)

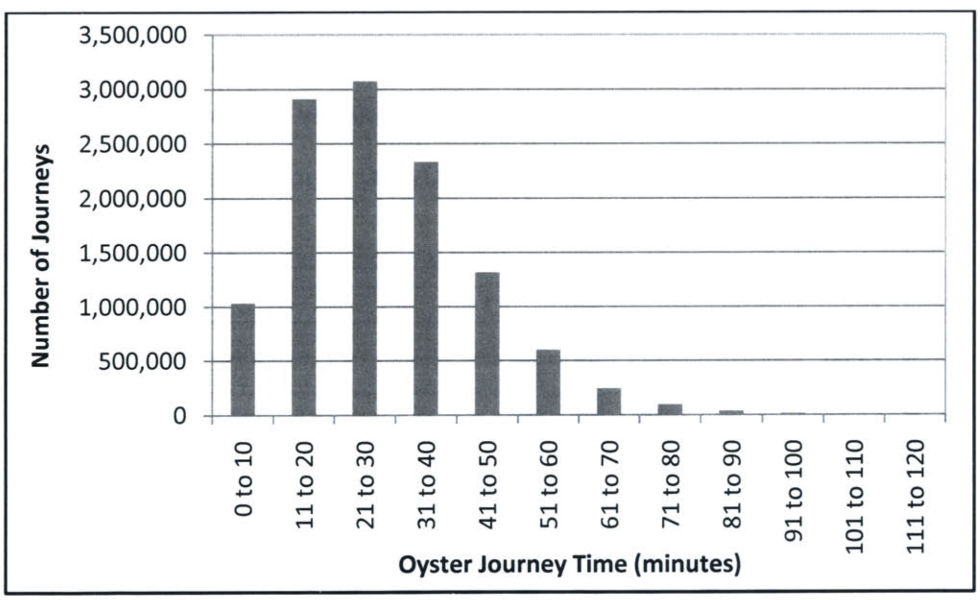

Figure 5-5: Distribution of Oyster Journey Time (Interval)

While the peak of the journey time distribution is at 19 minutes, the median and average journey times are 26.0 and 28.7 minutes, respectively. Half of the journeys range between 11 and 30 minutes, with another one-third of journeys between 31 and 50 minutes.

Figures 5-6 and 5-7 show the distribution of median OD journey times. About 40% of OD pairs have median journey times between 31 and 50 minutes, and 75% between 21 and 60 minutes. The long tail at the right of Figure 5-6 indicates very infrequent and very lengthy journeys.

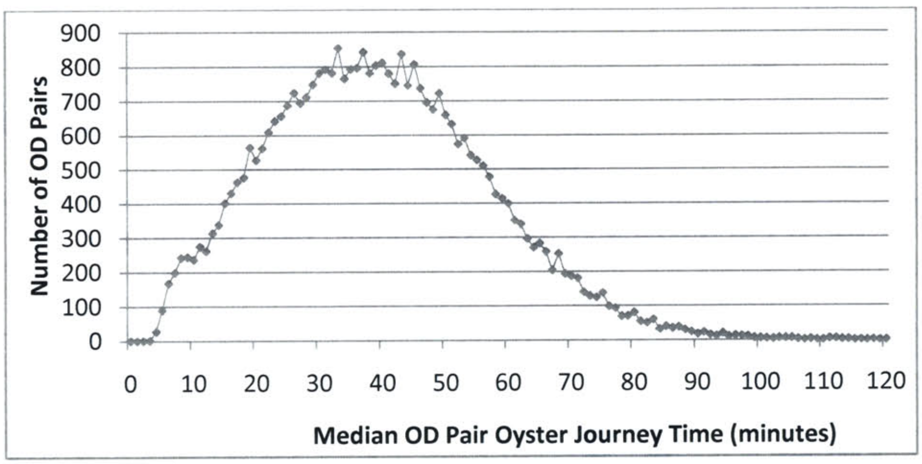

Figure 5-6: Distribution of Median OD Pair Oyster Journey Time (Continuous)

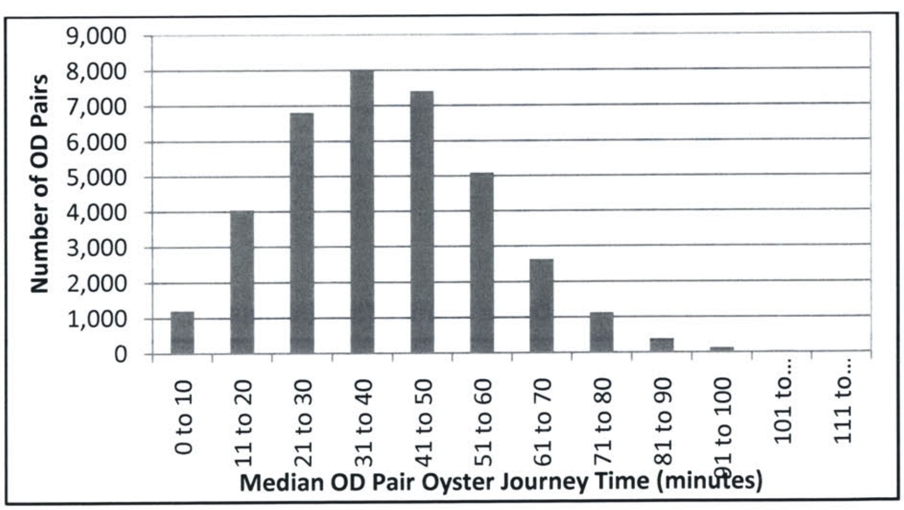

Figure 5-7: Distribution of Median OD Pair Oyster Journey Time (Interval)

5.3.3 Oyster Penetration and Representativeness

The Oyster penetration rate of an OD pair is defined by the number of completely documented journeys divided by the total number of journeys. In Figure 5-8, the numbers of Oyster CDJ are from LU Period 12 and the corresponding total numbers of journeys are from the Oyster-based matrix of LU Period 11 (as explained earlier in this chapter). Only OD pairs with at least 20 journeys in the Oyster-based matrix are included in the figure because infrequently traveled journeys are likely to have low Oyster penetration and therefore little information on journey time is available.

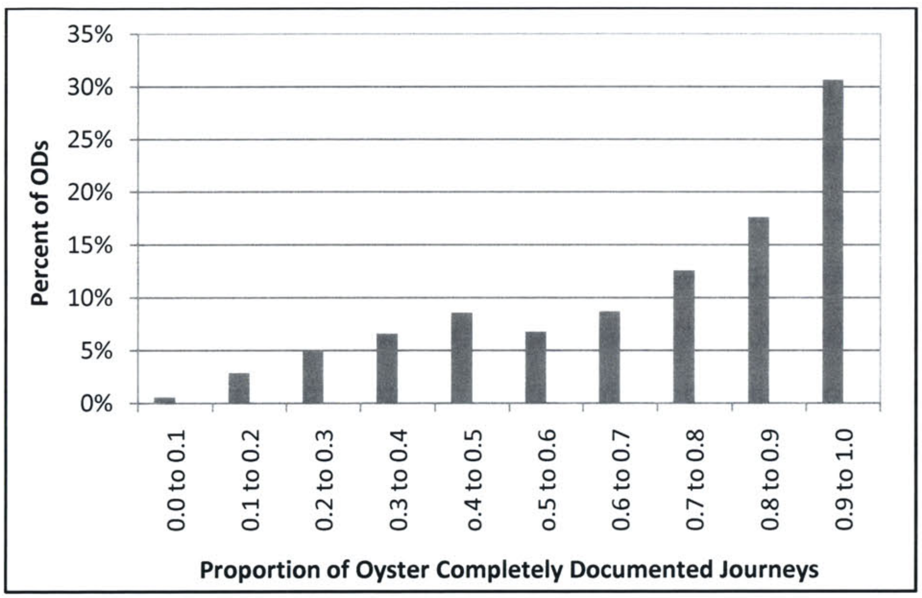

Figure 5-8: Oyster Penetration

Figure 5-8 includes a total of 8,487 OD pairs. Oyster completely documented journeys make up at least 90% of demand for 30% of OD pairs and 50% of demand for 75% of OD pairs. Such high levels of Oyster penetration are favorable for the methodologies developed in the following sections.

It may be argued that Oyster may provide a biased sample of journey times because Oyster card holders are likely to be regular users of the Underground and familiar with the network. In contrast, magnetic ticket users (excluding National Rail Season Travelcard users) are likely to be infrequent travelers who need to spend extra time navigating around the system. The average time spent by an Oyster user inside the system is therefore likely to be shorter than a magnetic ticket user. As regular travelers, Oyster users are also likely to walk faster than magnetic ticket users, thus spending even shorter times in the system. This potential bias currently involves only about 24% of infrequent users who still use single, return or Day Travelcard magnetic tickets[9] and there is already strong evidence that even infrequent Underground users are adopting Oyster Pay-As-You-Go because of the very large price differential.

Passengers who take both the Underground and National Rail on their regular commute are likely to have shorter Underground journeys than Underground-only passengers because once the former arrive at Zone 1 on National Rail, only a short Underground transfer is needed to connect them to their final destinations. Since the National Rail Season Tickets (which entitle passengers to unlimited travel on the Underground) are only available on magnetic stripe tickets, shorter Underground journeys made by National Rail passengers are unlikely to be captured by Oyster. For these OD pairs, the Oyster journey time distributions should still be representative of both types of passengers because National Rail passengers are also regular travelers whose behavior should be similar to other regular Oyster users. This problem will become less relevant as the National Rail Season Tickets are expected to be available on Oyster in the next several years.

5.4 Oyster-Based Excess Journey Time Calculation

The superiority of Oyster data over manual survey journey time data is the former's abundance and ability to capture variations. By comparing each completely documented Oyster record with the appropriate scheduled time for the same OD pair,

---

[9] Based on journey data provided by Tony Richardson, Fares & Ticketing, Transport for London.

the value of excess journey time for each record can be calculated. This section discusses the characteristics of journey time captured by Oyster entry and exit transactions, proposes the definition of comparable scheduled journey time, and develops a formulation to compute excess journey time at the line level.

For illustration, this analysis will only include journeys that start and end on the same line. Journeys that involve travel on more than one line require more consideration due to the possibility of various transfer stations and different paths being taken. Therefore, a path choice model is needed to assign journeys to paths before excess journey time can be calculated for journeys that involve a transfer.

5.4.1 Definition of Scheduled Journey Time

The simplest type of OD pair is one that involves only one on-train time component, and thus is made without a transfer. The corresponding Oyster journey time between entry and exit includes: access walk time from entry gate to platform, wait on platform, on-train travel time and egress walk time from platform to exit gate. Figure 5-9 illustrates the difference between actual Oyster and JTM journey time definitions. JTM also measures time components between station entrance and gates, capturing time spent in purchasing tickets and queuing at gates.

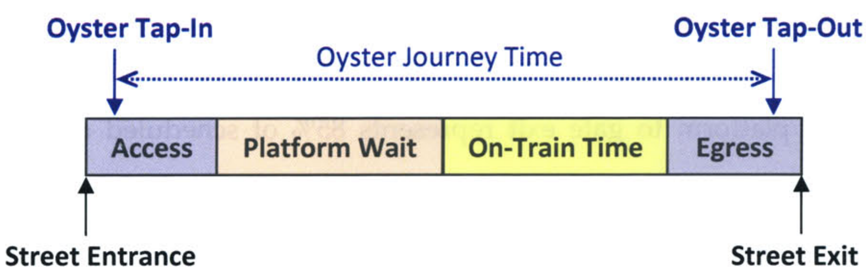

Figure 5-9: Comparisons of Oyster Journey Time with JTM

The scheduled journey time components used in this analysis are taken from the JTM schedule for the following reasons:

- The JTM is the only available source of schedule data.
- The JTM schedule is reliable since it was developed carefully and is updated periodically using manual surveys and simulation models.
- By using the same schedule as the JTM results, the Oyster-based excess journey time results are comparable with JTM. Comparability is important to both validate the new methodology as well as the JTM results.

In order to calculate excess journey time from Oyster data, the schedule journey time should reflect only the components included in Oyster. Three components are measured in JTM but not captured by Oyster in this analysis: (1) ticket purchase time since it occurs outside the gates, (2) interchange walk time (only same-line journeys are considered), and (3) closures because no journeys would have been made in closed parts of the network. A few modifications to the scheduled times are also needed for comparability with the elapsed Oyster journey times. The modified scheduled journey time components include:

- **Access (ACC) and Egress (EGR)** – Since the JTM access and egress scheduled times for each station include time spent within the station but outside the gated area, they should be factored to include only time inside the gate. It is estimated that walk time from gate entry to departing platform represents about 85% of the scheduled access time at the origin station, and similarly, walk time from arriving platform to gate exit represents 85% of scheduled egress time at the destination station[10].
- **Platform Wait Time (PWT)** – The JTM scheduled platform wait is calculated as half of the effective headway by station, line and direction. Since the Underground service is frequent, passengers are assumed to arrive at platforms randomly and therefore half of the passengers are expected to wait longer than

[10] Peter James, London Underground

the scheduled platform wait. This results in excess platform wait time even when all trains adhere to the schedule perfectly because half of the passengers will always experience platform wait. To provide a "headway-neutral" measure that is comparable between lines and to capture only "excess" time beyond any scheduled time, the scheduled platform wait in the proposed Oyster formulation is the full effective headway by station, line and direction.

- **On-Train Time (OTT)** – The Oyster-based methodology uses the same scheduled on-train time as the JTM except that Oyster considers every OD pair whereas JTM calculates on-train time for line segments that aggregate several adjacent station-to-station links between transfer stations. The scheduled on-train time is presented as departure time between two adjacent stations and therefore includes dwell time at the second station. The scheduled on-train time between adjacent stations are summed to give the scheduled on-train time between an OD pair on the same line.

In summary, the scheduled journey time for an OD pair from origin O to destination D consists of four JTM schedule components, as shown in Equation [5-1].

$$\mathbf{Scheduled\ JT}_{\mathbf{OD}} = \mathbf{0.85} * \mathbf{ACC_O} + \mathbf{2} * \mathbf{PWT}_{\mathbf{OD}} + \mathbf{OTT}_{\mathbf{OD}} + \mathbf{0.85} * \mathbf{EGR_D} \qquad \textbf{[5-1]}$$

5.4.2 OD Level Excess Journey Time

Since Oyster transactions record time elapsed between entry and exit, the inferred journey time can be affected by passenger behavior such as walk speed and getting lost in the network. While utilizing the easy availability and large sample sizes of Oyster data, it is important to include only data that represent system performance and not individual behavior.

Figure 5-10 shows the journey time distributions for five randomly selected OD pairs from the February 2007 AM Peak transactional Oyster data. The five OD pairs have different sample sizes (shown in parentheses on the graph) and median journey times but a common profile with three stages – a slight rise from the minimum value to the 10^{th} percentile journey time, a steady and slight slope from the 10^{th} to 90^{th} percentile and a very steep increase from the 90^{th} percentile to the maximum value.

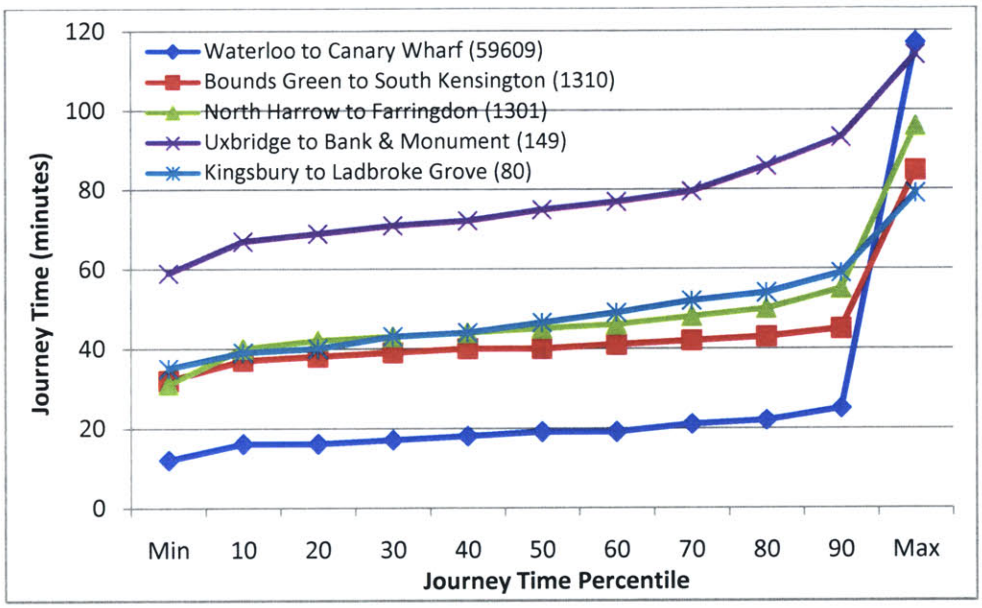

Figure 5-10: Full Journey Time Distribution for Selected OD Pairs

Figure 5-11 zooms in on the portion from the 95^{th} percentile to the maximum value. Three of the OD pairs show a steady slope between the 95^{th} and 99.5^{th} percentiles before a significant increase to the maximum value whereas the remaining two OD pairs show a steady slope over the entire distribution.

The significant increase beyond the 99.5^{th} percentile is likely to represent individual behavior such as taking the wrong train or waiting for friends within a station. It is also possible that these extremely long journeys are two incompletely documented journeys mistaken to be one journey if a passenger fails to validate both the exit at the end of the first journey and the new entry at the beginning of the second journey. To eliminate the effects of individual and extreme behavior, only data within the 99th percentile range for each OD pair are included in the excess journey time calculation.

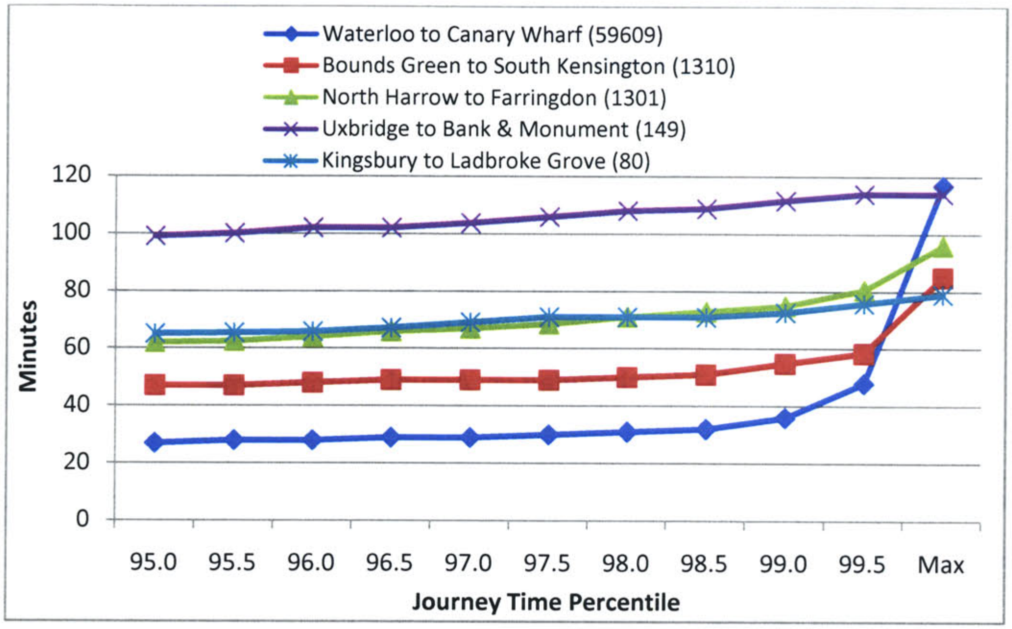

Figure 5-11: Top End of Journey Time Distribution for Selected OD Pairs

Figure 5-12 shows the value of the scheduled journey time with respect to a hypothetical but typical OD journey time distribution from the minimum value to the 99th percentile. If the scheduled journey time included only half the headway, it should be expected that approximately half of the passengers would experience journeys longer than the schedule. However, since the scheduled journey time defined in this methodology (Equation [5-1]) includes a full headway, it is expected that less than half of the passengers would experience journeys longer than the schedule, and thus positive excess journey time. By definition, passengers who experience shorter journeys than the schedule would not contribute to excess time.

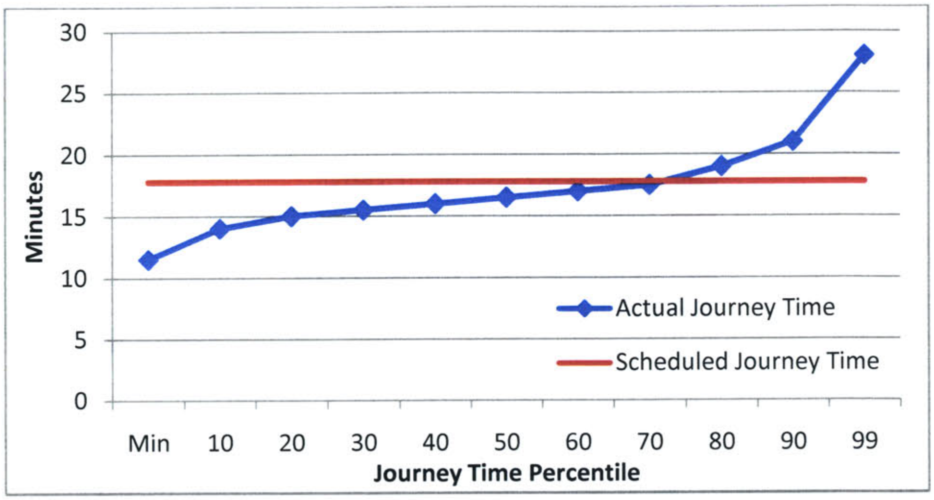

Figure 5-12: Scheduled vs. Actual Journey Time

Equation [5-2] expresses the excess journey time for each transaction for an OD pair and ensures that only journeys longer than the schedule contribute to positive excess.

$$\textbf{Excess JT}_{\textbf{OD,Transaction}} = \textbf{MAX}(\textbf{Actual JT}_{\textbf{OD,Transaction}} - \textbf{Scheduled JT}_{\textbf{OD}}, \textbf{0}) \quad \textbf{[5-2]}$$

The average excess journey time of an OD pair is simply the sum of positive excess journey time divided by the number of journeys that experience journey time between the minimum value and the 99th percentile, as shown in Equation [5-3].

$$\textbf{Average Excess JT}_{\textbf{OD}} = \frac{\sum_{\textbf{Transaction}} \textbf{Excess JT}_{\textbf{OD,Transaction}}}{\textbf{SampleSize}_{\textbf{OD}}} \quad \textbf{[5-3]}$$

5.4.3 Line Level Excess Journey Time

To aggregate the OD level excess journey times to the line or line segment level, the OD average excess journey times need to be weighted by their respective demands. Since the Oyster transactional sample is not entirely representative of all passenger travel patterns, the demand weightings should be taken from the estimated Oyster-based OD flow matrix discussed in Section 3.3 (that is developed using both current Oyster data and RODS data).

By selecting OD pairs that start and end on the same line, a demand weighted average excess journey time at the line level can be calculated as shown in Equation [5-4]. The formulation should accurately reflect the average experience of passengers on one line with no transfers.

$$\textbf{Average Excess JT}_{\textbf{Line}} = \frac{\sum_{\mathbf{OD \in Line}} \textbf{Average Excess JT}_{\mathbf{OD}} * \textbf{Demand}_{\mathbf{OD}}}{\sum_{\mathbf{OD \in Line}} \textbf{Demand}_{\mathbf{OD}}} \qquad \textbf{[5-4]}$$

This formulation does not take into account passengers who travel on one line then transfer to another. While passengers who travel from station A to station C via transfer station B should experience the same journey time on the first leg of their journeys as passengers traveling only from A to B, only the latter passengers are included in the demand weightings in Equation [5-4].

In the London Underground, approximately 40% of journeys involve at least one transfer. In order to quantify excess journey time at the network level, a better understanding of travel behavior between lines using data or models of path choice is needed. The extension of excess journey time calculation to the network level is left for further research.

The average excess journey time can be calculated for any time period and over any number of days. The excess journey time results presented in Section 6.1 are based on AM peak data over 4 weeks in February 2007. Excess journey times for each AM peak can also be calculated to analyze day-to-day fluctuations.

5.5 Oyster-Based Journey Time Reliability Metric

The quantification of journey time reliability is a new concept that has not been viable using traditional survey methods to assess journey time performance. Traditional survey methods collect samples to estimate the average journey time taken by passengers during different time periods. However, the survey samples are usually too small to allow transit agencies to assess variability in journey time performance. With AFC data, especially in transit networks with high use of AFC media and both entry and exit control, journey time distributions can be obtained on almost every active OD pair. Transit agencies can obtain a better understanding of journey time performance by assessing both excess journey time and variability in journey time.

Recalling the definition of elapsed journey time between entry and exit in the Oyster transactional dataset, passengers' time spent inside the gates include access and egress walk time, platform wait time and on-train time. For journeys that require one (or more) interchange, interchange walk time and additional platform wait are also included.

5.5.1 Characteristics of Journey Time Reliability Factor

There are various ways to quantify the journey time reliability metric from journey time distributions depending on the objectives of the methodology and the intended uses of the results. In this research, the following characteristics of the methodology and results are desired:

- **Straightforward Interpretation** – An easily understood interpretation of the methodology and results will enable the metric to be effectively used by a transit agency as a management tool. This research suggests that a metric represented in time units is most straightforward and meaningful, the same representation as the Journey Time Metric currently used in the London Underground. Both transit planners and passengers can easily understand service performance expressed in terms of minutes.

- **Representativeness of Service Performance** – When computing service reliability, only data that are representative of the state of service for most passengers should be included. The inclusion of all data in the calculation may reflect individual behavior as well as service performance. A minimum sample size should also be required for each OD pair on the line to eliminate potential errors associated with small samples. For OD pairs that meet the minimum sample size, an upper limit for journey time percentile should also be imposed to exclude long journeys that are the result of individual behavior.
- **Complementary to Excess Journey Time Metric** – The excess journey time and reliability calculations should be complementary and consistent with each other to provide a better understanding of service reliability. Since the excess journey time calculations already show results relative to the scheduled values, the reliability factor would be best defined to be independent of the schedule. For example, the reliability factor methodology should be able to distinguish between OD pairs with wider journey time variations but similar excess journey times.
- **Comparability of Results Across Lines** – Factors such as headway and journey length affect the observed journey time variations for different OD pairs and lines. Lines with longer headways are prone to larger journey time variations because on platform wait times are longer and the impacts of a cancelled service are greater. On the other hand, both short and long journeys can lead to large variations in journey time. Short journeys have a larger proportion of access and egress walk times and the journey time distributions may be dominated by passengers' walk speed. Long journeys have higher chances of experiencing delays than short journeys, especially journeys that involve interchanges. To make the comparison of results across lines fair and meaningful, the methodology should emphasize the effects of service unreliability and minimize the effects of headway and journey length.

5.5.2 *Definition of OD Level Journey Time Reliability Factor*

The inverse cumulative density functions in Figure 5-10 illustrate that between any origin and destination, there is a minimum amount of time needed to travel but there is no equivalent upper bound. The lower ends of the distributions represent passengers who walk fast and experience short, or no, platform wait times, and thus very short journey times. In contrast, the upper ends of the distributions represent passengers who walk slowly, experience long platform wait times, or long in-train travel times, and thus exceptionally long journey times.

The journey time reliability for an OD pair can be generally defined by quantifying the spread of the journey time distribution. The tighter the distribution, the more reliable the service. This methodology quantifies the spread of journey times, known as the Reliability Factor, as the difference between an upper threshold, the N^{th} percentile journey time, and the median journey time. The N^{th} percentile should be defined such that journey times beyond which are considered to be highly undesirable for passengers. Transit planners can see how long these undesirable journeys are compared to the typical journey (median journey time) on the same OD pair. For example, if an OD pair has a median journey time of 20 minutes and a 95^{th} percentile ($N = 95$) journey time of 36 minutes, the resulting reliability factor of 16 minutes means that the bad 5% ($1 - N$) of journeys take 16 minutes longer than typical journeys. An alternative interpretation is that passengers, and transit planners, can be 95% confident that a journey on the OD pair can be traveled within 36 minutes, whereas a typical journey takes 20 minutes.

The value of N should depend on three factors: 1) the observed OD journey time distributions, 2) the desired sensitivity of the reliability results, and 3) the service standards of the transit agency. The value of N should be large enough to be representative of most journeys on the OD pair and reflects only service performance. Section 5.4.2 showed that for the Underground, journey times beyond the 99^{th} percentile are likely to be influenced by individual behavior rather than service performance. If N is set to equal 99, the difference between the 99th percentile and median journey times represents the extra time (on top of the median journey time) experienced by 1 out of 100 journeys on the OD pair. While N = 99 is a valid and reasonable parameter, passengers may not be sensitive to additional journeys times with only a 1% chance of occurring.

For the Underground, the reliability factor for an OD pair is defined as the difference between the 95th percentile journey time and median journey time. The value of the reliability factor is therefore the amount of extra journey time experienced by 1 out of 20 journeys on the OD pair. Passengers can take into account the reported values of reliability factors when planning their travel. Depending on their time sensitivity, passengers should allow more, or less, time to travel the OD pair.

In this analysis, 100% of Oyster transactional data with properly recorded entry and exit times are used to calculate the elapsed time and thus a journey time distribution for each OD pair. For an OD pair with a sufficiently large sample, the journey time distribution should be smooth and thus the 95th percentile can be correctly calculated by interpolation of data points. However, for OD pairs with small samples, the interpolated 95th percentile may be affected by the extreme values. It is suggested that for OD pairs with sample sizes of between 20 and 200 the 95th percentile journey time be replaced by the second maximum recorded journey time if this is smaller. Lastly, OD pairs with few than 20 journeys should not be included in the reliability calculation because the results would be prone to small sample biases and thus have the potential to misrepresent service performance. Equation [5-5] summarizes the calculation of the OD level unreliability factor for different sample sizes. The larger the value of the reliability factor, the less reliable the service between the OD pair.

$$\textbf{Reliability Factor}_{OD} = \begin{cases} \textbf{95th Percentile JT}_{OD} - \textbf{Median JT}_{OD} & \textbf{if SampleSize}_{OD} \geq \textbf{200} \\ \textbf{Max(95th Percentile JT}_{OD}\textbf{, Second Max JT}_{OD}\textbf{)} - \textbf{Median JT}_{OD} & \textbf{if 20} \leq \textbf{SampleSize}_{OD} < 200 \end{cases} \quad \textbf{[5-5]}$$

5.5.3 *Definition of Line Level Journey Time Reliability Factor*

To obtain the line level reliability factor, same-line OD pairs can be aggregated by demand weightings to represent the performance reliability faced by an average passenger on the line. Since each line has different characteristics in terms of service provision and demand patterns, the definition of the reliability factor at the line level requires consideration of what the results will be used for. If the results are to be compared among different lines in the network, it is important that they are fundamentally comparable in terms of the types of OD pairs included in the calculation. It is believed that increasing headway and journey length variations affect reliability negatively because they introduce more uncertainties in the journeys.

Two factors affect the headways along the same line: short-turning of trains and branching of the line. The common element of these two characteristics is the shorter effective headways in the "trunk" portion of the line where all trains pass (i.e. before the short-turn or branching point). In this analysis, the portion of the line with the most frequent service is termed the "trunk" portion and the remaining portion the "non-trunk". Trunk OD pairs are defined as those that have both origin and destination on the trunk whereas non trunk OD pairs involve at least one end of the journey at a non trunk station. A non-trunk OD pair can be a branch-to-branch movement which requires an interchange and involves more than one platform wait time.

In the Underground, the trunk portions usually have very similar headways across different lines although the lengths of the trunks vary. Table 5-2 presents relevant characteristics for the five Underground lines analyzed in this research. While all five lines use short-turning, only two have branches.

Line	Branching	Total Number of Stations	Number of Trunk Stations	Trunk Definition	Trunk AM Peak Headway (minutes)	Non Trunk Maximum AM Peak Headway (minutes)
Bakerloo	No	25	16	Queens Park to Elephant &Castle	2.8	12.1
Central	Yes	49	20	White City to Leytonstone	2.2	21.2
Jubilee	No	27	17	Willesden Green to North Greenwich	2.7	5.9
Piccadilly	Yes	51	24	Acton Town to Arnos Grove	2.5	15.2
Victoria	No	16	13	Seven Sisters to Brixton	2.3	4.0

Table 5-2: Characteristics of Underground Lines Analyzed

Some train services on a line may short-turn at an intermediate station rather than running to the terminal station in order to provide more frequent service on the middle portion of the line with the heaviest demands. The frequencies of trains going to the short-turn and terminal stations depend on demand, rolling stock and the physical design of the tracks and stations on the line.

Table 5-3 shows the mean and standard deviation of reliability factor for trunk and non trunk ODs on the Bakerloo line classified by median journey time. For the same median journey time and thus similar journey length, non trunk OD pairs have consistently higher reliability factors than trunk OD pairs. This is because the non trunk OD pairs are served by lower frequencies and thus have larger variations of observed journey times. Non trunk OD pairs also have higher proportions of longer journeys because they are not restricted to the middle section of the line with the most frequent service. For the non trunk OD pairs, longer journey lengths are also generally associated with higher reliability factors which is not the case for trunk OD pairs.

Median Journey Time Range (minutes)	Trunk ODs			Non Trunk ODs		
	Mean RF (minutes)	Standard Deviation of RF (minutes)	Number of ODs	Mean RF (minutes)	Standard Deviation of RF (minutes)	Number of ODs
5-10	7.36	5.26	50	8.51	2.45	33
10-15	8.70	6.29	51	8.75	2.29	32
15-20	6.34	2.39	41	10.84	4.24	34
20-25	7.86	5.87	32	12.28	5.18	30
25-30	8.51	5.48	16	12.22	5.18	30
30-35	7.01	2.31	7	13.64	10.24	30
35-40	-	-	0	12.20	5.64	29
Over 40	-	-	0	13.15	5.30	50

Table 5-3: OD Reliability Factors of Bakerloo Line

Another example is the Piccadilly line which has both short-turning and branching. Branching of lines also increases the reliability factor because service frequencies in the branches are lower than in the trunk portion where all trains provide service. Table 5-4 shows the variation of reliability factors for trunk and non-trunk OD pairs on the Piccadilly line. There is a clear relationship in Table 5-4 that non-trunk OD pairs systematically have higher reliability factors than trunk OD pairs in the same median journey time range, except for short journeys with median journey time less than 10 minutes.

To ensure comparability of reliability factors across lines with different designs and headways, the line level reliability factor should include only OD pairs on the trunk portion, as illustrated in Equation [5-6].

$$\textbf{Reliability Factor}_{\textbf{Line}} = \frac{\sum_{\textbf{OD}\in\textbf{Trunk of Line}} \textbf{Reliability Factor}_{\textbf{OD}} * \textbf{Demand}_{\textbf{OD}}}{\sum_{\textbf{OD}\in\textbf{Trunk of Line}} \textbf{Demand}_{\textbf{OD}}} \quad \textbf{[5-6]}$$

	Trunk ODs			Non Trunk ODs		
Median Journey Time Range (minutes)	Mean RF (minutes)	Standard Deviation of RF (minutes)	Number of ODs	Mean RF (minutes)	Standard Deviation of RF (minutes)	Number of ODs
5-10	9.14	9.26	7	8.06	2.44	10
10-15	6.19	3.52	92	7.52	3.60	89
15-20	6.15	3.52	96	7.91	1.89	86
20-25	5.92	2.04	92	8.00	1.78	91
25-30	7.04	2.67	68	8.82	2.19	93
30-35	7.65	2.85	61	10.23	3.13	107
35-40	9.05	4.30	47	10.45	3.72	114
Over 40	9.31	5.26	29	11.81	4.59	115

Table 5-4: OD Reliability Factors of Piccadilly Line

Similar to the excess journey time results, the reliability factor can be calculated for any time period and any number of days. The reliability factor results presented in Section 6.2 are based on AM peak data over 4 weeks in February 2007. To monitor day-to-day variations of a line's reliability factor, the period (e.g. 4 weeks) median journey time, instead of the daily median journey time, should be used as the basis of comparison. This ensures that the daily reliability factor reflects service performance with respect to typical journey times throughout the period but not daily median journey times which could differ significantly as a result of unreliability on a particular day.

Chapter 6 Evaluation of Journey Time Metrics

The full sample of Oyster transactional data collected during one Underground period, corresponding to four weeks, is used to compute the excess journey time and reliability factor results shown in this chapter. Section 6.1 presents the Excess Journey Time Metric for five Underground lines and compares them with the official JTM results. Section 6.2 presents the Journey Time Reliability Metric and discusses potential service conditions reflected in the results. The line level results are also divided into directional results to evaluate the effects of journey length and link loads on delays and reliability.

6.1 Excess Journey Time Results

The results of the excess journey time calculations can be analyzed in several dimensions given the large sample sizes of the Oyster transactional data. The following sections evaluate service performance of the Underground by comparing the observed Oyster journey times against the scheduled OD journey times. Due to issues in computing the OD level schedules without detailed service pattern information, only 5 of the 12 lines in the London Underground network are analyzed to demonstrate the Oyster-based methodology.

In the following sections, the results of each line are comprised of only same-line journeys which begin and end on the same line and include branch-to-branch movements that require interchanges. Branch-to-branch movements have two platform wait times but the scheduled journey times include only the larger of the two branch headways. Although this may potentially lead to higher-than-actual excess journey time results, branch-to-branch movements are usually small and should not significantly affect the overall line level results. For example, less than 2% of same-line journeys on the Piccadilly line are branch-to-branch movements.

Path choice is not considered when applying demand weightings to aggregate the OD level excess journey time results. This means that same-line OD pairs that can be traveled on two or more lines will be reported for each of the possible lines using the same OD demand weighting. For example, all passengers traveling from Finsbury Park to Kings Cross are counted separately for both Victoria and Piccadilly lines whereas in reality more passengers would likely travel on the Victoria line as it involves fewer stations.

6.1.1 *Proportion of Delayed Journeys*

Table 6-1 shows the sample from the 100% Oyster transactional data used in this analysis. The Oyster data are taken from 20 AM Peaks in February 2007 and the average daily demand represents the AM Peak same-line demand obtained from the Oyster-based OD matrix for January 2007 and used to scale the OD level excess journey time results.

	Same-Line Journeys (AM Peak)		
Line	Oyster Transactional Observations (20 days)	Percent of Observations with Excess Journey Time	Average Daily Demand
Bakerloo	280,428	22.78%	29,379
Central	1,171,203	36.30%	84,429
Jubilee	845,138	40.40%	70,110
Piccadilly	844,615	37.18%	57,091
Victoria	633,014	32.12%	60,046

Table 6-1: Oyster Transactional Data Sample Size

Since the scheduled OD journey time used here includes a full headway as the platform wait time, it is expected that less than half of all passengers would experience excess journey time because the median journey time corresponds to a journey that on average experiences platform wait time of one-half the headway. Among the five lines, Bakerloo has the smallest percentage of same-line journeys with excess time of 23% whereas Jubilee has the largest percentage with 40% of same-line journeys experiencing delays.

Figure 6-1 shows the proportion of journeys experiencing delays in each direction. For the Bakerloo, Central and Piccadilly lines, the directional deviations are within 4% of the average line-level proportions. For the Jubilee line, 45% of eastbound journeys experience delays compared with 34% of westbound journeys. This is likely to be due to imbalanced demand in the AM Peak as 57% of same-line journeys on the Jubilee line are eastbound towards the large destinations such as Canary Wharf. The Victoria line has the largest discrepancy in proportion of delayed journeys between the two directions – 23% northbound versus 41% southbound – while the same-line demand is balanced between the two directions. Section 6.2.2 investigates the directional differences for the Victoria line.

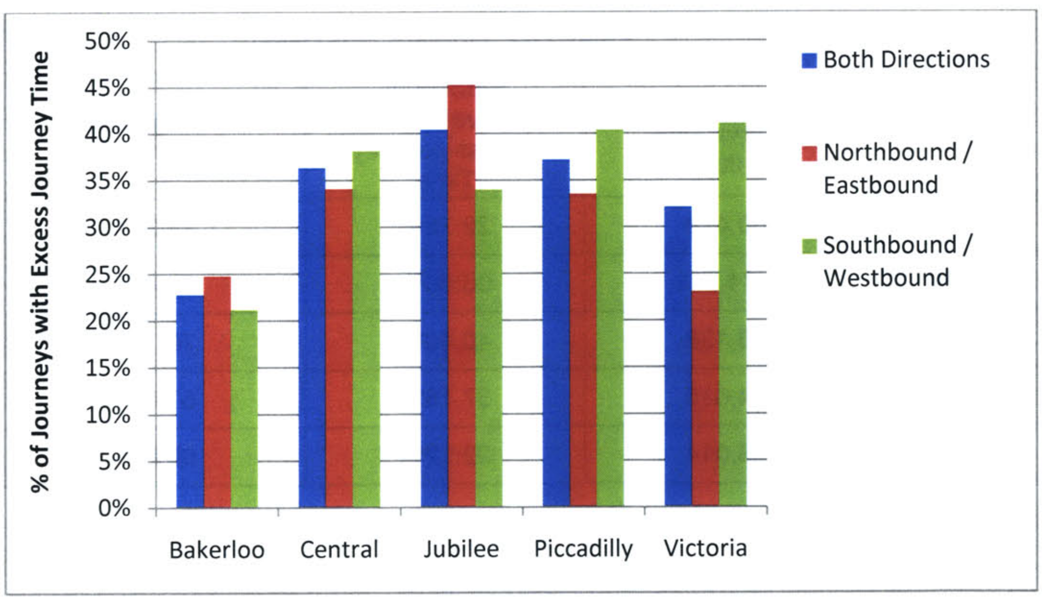

Figure 6-1: Proportion of Oyster Journeys with Excess Journey Time

6.1.2 *Line Level Excess Journey Time Results*

The line level excess journey time is aggregated from the lower 99% of Oyster journey times in each OD pair and weighted by OD demand obtained from the Oyster-based flow matrix. Figure 6-2 shows the excess journey time results for each line and by direction.

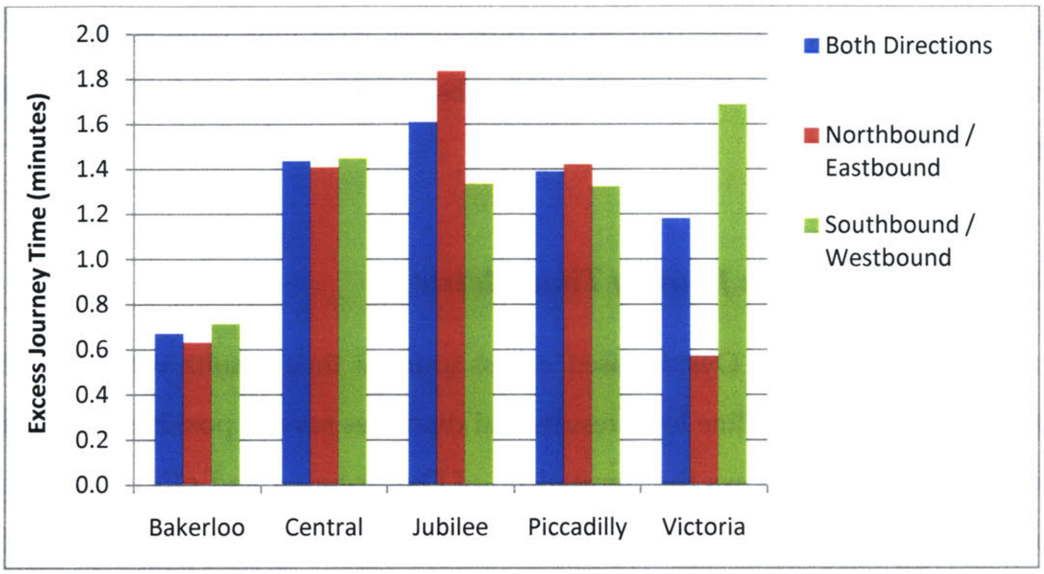

Figure 6-2: Oyster-Based Line Level Excess Journey Time

On the line level, Bakerloo has the lowest excess journey time of 0.7 minutes, followed by Victoria with 1.2 minutes. Bakerloo and Victoria are the shortest lines among the five, serving 25 and 16 stations, respectively. While Victoria has fewer stations than Bakerloo, it serves some of the busiest stations in the network including Victoria, Kings Cross and Oxford Circus. The Piccadilly and Central lines have an average excess journey time of 1.4 minutes and Jubilee has the largest excess at 1.6 minutes.

On the directional level, the Bakerloo, Central and Piccadilly lines have directional discrepancies within 0.1 minute of the respective line average. Similar to the directional discrepancies in the proportion of delayed journeys, the Victoria and Jubilee lines also have the largest directional discrepancies in average excess journey time.

It is somewhat counterintuitive that for the Bakerloo and Piccadilly lines, the direction with a higher proportion of delayed journeys (in the Oyster transactional data) results in a smaller average excess journey time. This means that while a direction has a higher proportion of delayed journeys, the extent of delays is smaller and thus the average excess journey time is lower than the other direction. The OD demand weightings used to scale the OD excess journey time to the directional level may have also shifted the relative delays between the Oyster transaction observations and the demand weighted results.

6.1.3 *Comparison with Journey Time Metric*

This section compares the Oyster-based excess journey time results with the official Journey Time Metric (JTM) line level results and discusses several possible causes of the differences. Figure 6-3 shows the comparison of Oyster-based and official JTM excess journey times for February 2007 and Table 6-2 shows the corresponding values. Although the total JTM excess time shown in this section includes only time components inside the gates that are captured by Oyster – access, egress, platform wait, left behind and on-train time – the two results are not expected to be totally comparable due to the differences in calculating the scheduled and actual journey times.

The JTM excess times range from 2.1 to 4.6 minutes and are significantly higher than the Oyster-based excess time that range from 0.7 to 1.6 minutes for all five lines analyzed. The major difference between the Oyster-based and JTM excess journey time is the inclusion of full headway and half headway, respectively, in the scheduled journey times. The "Adjusted JTM Excess" columns in Figure 6-3 show the comparable JTM excess times by adding the average half headway to the Total JTM excess for each line. These columns are expected to be more similar to the JTM excess journey times as both have approximately the same scheduled platform wait. However, the line level differences are not fully resolved, and the differences between the Oyster-based excess and the adjusted JTM excess range from -0.3 to 1.4 minutes. As shown in Table 6-2, 3 and 2 lines have positive and negative differences, respectively, and this shows that the Oyster-based excess journey times are not systematically biased.

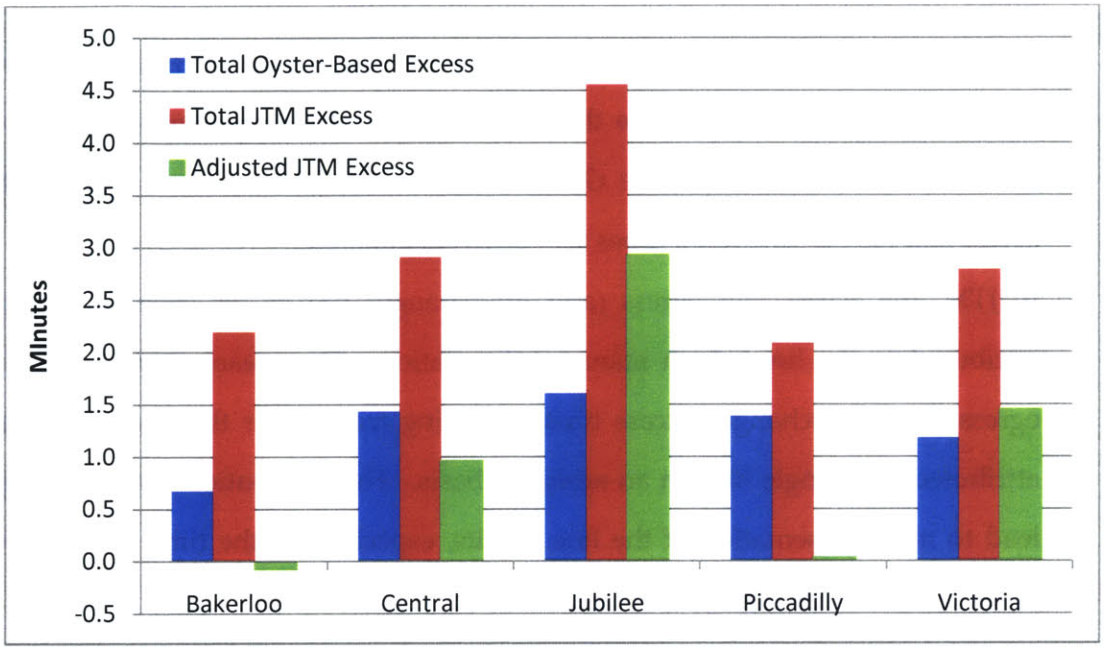

Figure 6-3: Oyster-Based and JTM Excess Journey Time

Line	Total Oyster-Based Excess	Total JTM Excess	Average Half Headway	Adjusted JTM Excess	Difference between Oyster-based Excess and Adjusted JTM Excess
Bakerloo	0.67	2.19	2.268	-0.07	0.75
Central	1.44	2.90	1.944	0.96	0.47
Jubilee	1.61	4.55	1.621	2.93	-1.32
Piccadilly	1.39	2.08	2.048	0.04	1.35
Victoria	1.18	2.79	1.333	1.45	-0.27

Table 6-2: Total Excess Journey Time in Minutes

There are three possible explanations for the observed line level discrepancies:

1. The JTM access and egress times capture passenger walk time between station street entrance/exit and platform whereas the Oyster-based times do not include

the walk between station street entrance/exit and the fare gates. The station access and egress scheduled times used to calculate the Oyster-based excess times are adjusted for this difference and a factor of 0.85 is applied to all JTM station scheduled times to allow for comparable scheduled and actual journey times. The access/egress factor of 0.85 may be too high, meaning that over 0.15 of access and egress walk times are spent between the station street entrance/exit and the fare gates. In this case the higher-than-actual access/egress factor will lead to underestimation of the Oyster-based excess journey time because the access and egress scheduled times will be overestimated.

2. In JTM, the access and egress (and interchange) results for each station are attributed to the line which manages the station. The means that access and egress (and interchange) excess times resulting from other than lines are also attributed to a single line on an arbitrary basis. This attribution can potentially lead to misrepresentation of the line results, especially at the time period level when demands are highly directional.
3. The line level results are scaled by different demand matrices from different years. The JTM is using the RODS OD matrix from 2002 as the basis for demand weightings whereas the Oyster-based results are weighted by current 2007 demand. Another difference is that JTM incorporates demands incurred by all journeys with or without transfers at the line level whereas the Oyster-based results are solely based on no-transfer journeys. This means that if severe delays occur on a line segment with a high proportion of transfer journeys, the JTM results would capture more excess than the Oyster-based results.

The JTM methodology allows the results to be evaluated by time component and improvements can be targeted at the components with high levels of excess. Figure 6-4 shows the breakdown of JTM excess minutes into time components and Figure 6-5 shows the same results but normalized by total excess minutes for each line. The major causes of excess time in each line differ, for example, excess access times are significant for the Jubilee and Victoria lines whereas excess egress times are significant only for the Jubilee line. The large excess egress times in the Jubilee line are likely to be due to extended escalator or elevator failures at one or more major destinations on the line, such as Waterloo and Canary Wharf.

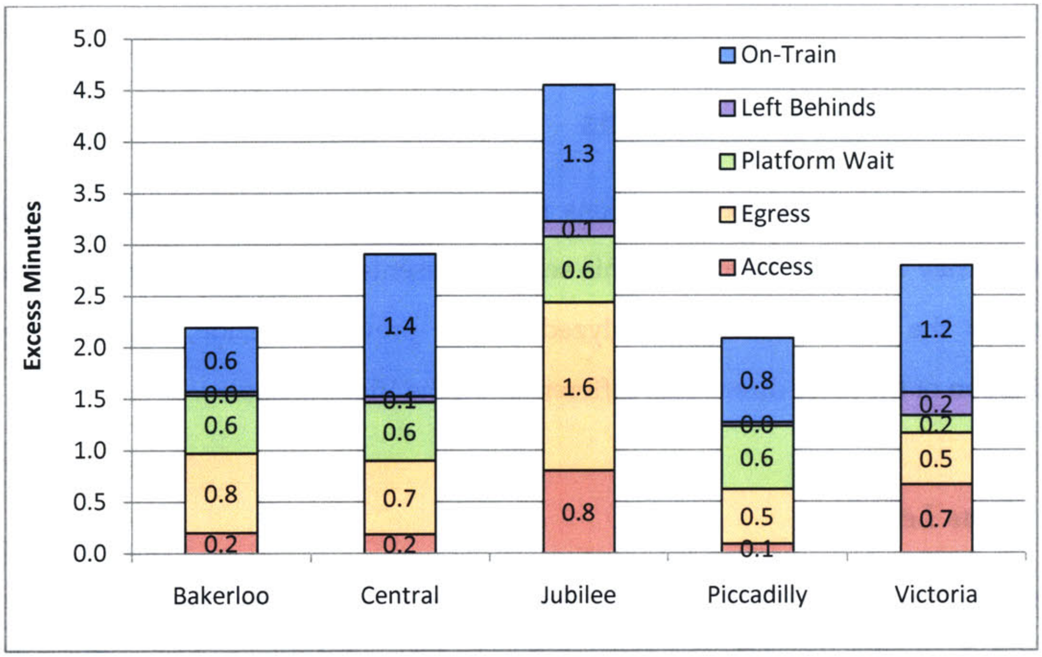

Figure 6-4: Breakdown of JTM Excess Journey Time (Actual)

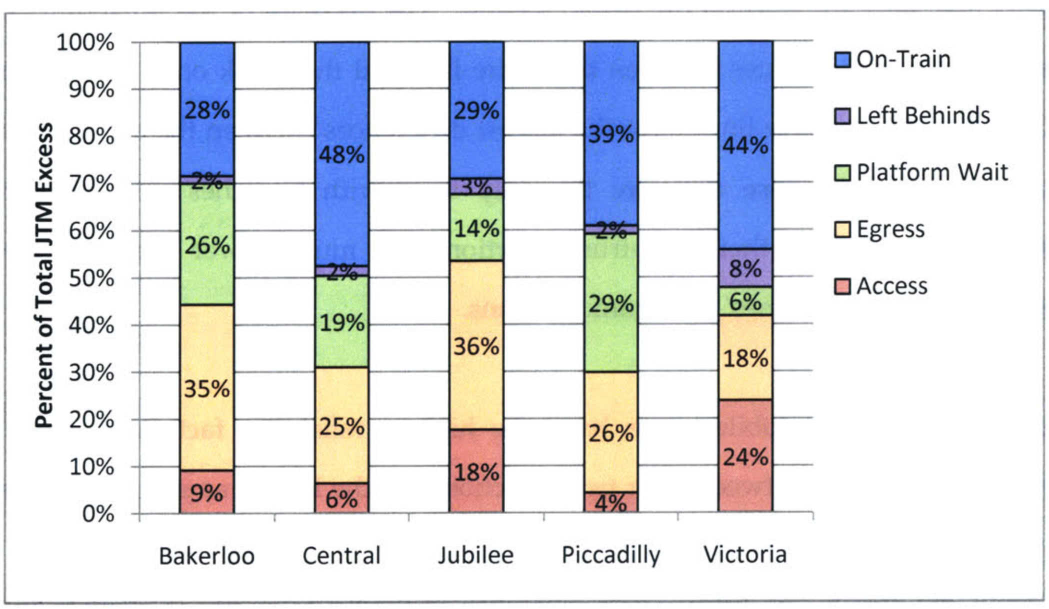

Figure 6-5: Breakdown of JTM Excess Journey Time (Normalized)

It is interesting to note that the Victoria line has the lowest excess platform wait but the highest left behinds. The left behinds in JTM are calculated based on RODS link loads and current excess platform wait results which may not be accurately estimated from train operations data (for all lines other than the District). Therefore, the distinction between excess platform wait times and left behinds is unclear in the JTM methodology.

6.2 Reliability Factor Results

The reliability factor captures variations in journey time that cannot be captured by excess journey time results alone. This section presents the line level reliability factor results for the same five lines analyzed in the previous section, followed by an examination of the large directional differences on the Victoria line.

6.2.1 Line Level

Figure 6-6 shows the reliability factors for the five Underground lines analyzed, and Table 6-3 shows the actual figures for each line and the reliability rankings. The column "Entire Line" includes all OD pairs on each line whereas "Trunk Only" represents only OD pairs that start and end on the trunk portion. The decrease of the reliability factor ranges from 0.5 to 2 minutes between the entire line and the trunk only for all 5 lines. The Central and Piccadilly lines have the largest differences between the entire line and trunk only results because they are the only lines with branches and the service frequencies on some of their non-trunk portions are much lower than the lowest frequencies on the other lines' non trunk portions.

While the Central and Jubilee lines have the highest reliability factors of about 8.5 minutes, the variations between their two directions are the lowest among the five lines. This means the service performance is balanced between the two directions. On the other hand, the directional difference in reliability on the Bakerloo line is likely to be due to an imbalance in directional demands. Since the Bakerloo line has its southern terminus within Zone 1, there are no northbound journeys coming into Zone 1 to balance the heavy southbound demand from the north, and thus the southbound direction has a much higher reliability factor due to crowding.

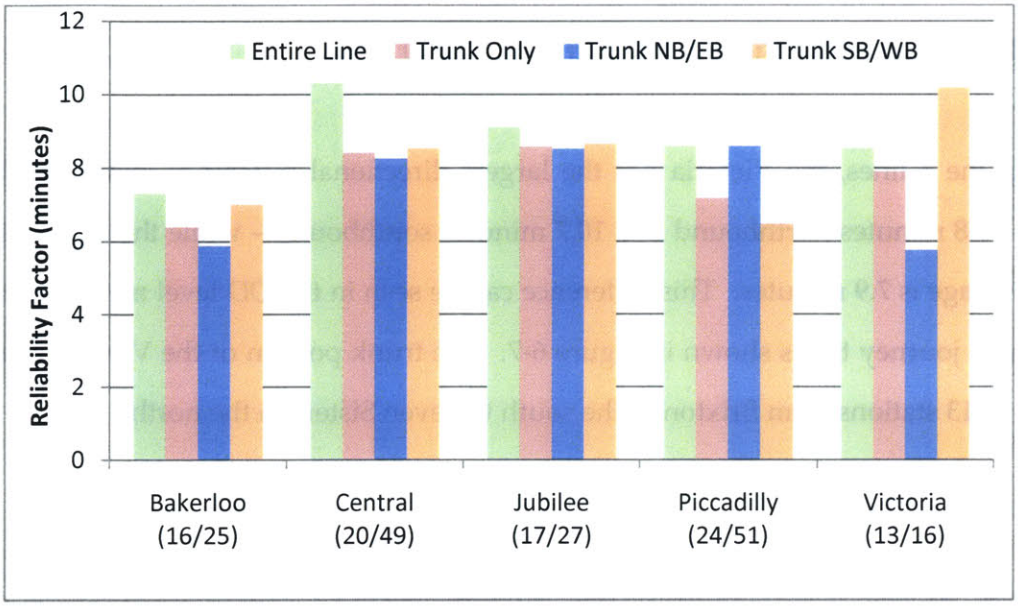

Figure 6-6: Line Reliability Factors

	Reliability Factor (minutes)				Line Rank	
	Entire Line	Trunk Only	Trunk NB/EB	Trunk SB/WB	Entire Line	Trunk Only
Bakerloo	7.30	6.37	5.87	6.99	1	1
Central	10.30	8.41	8.25	8.52	5	4
Jubilee	9.12	8.57	8.52	8.65	4	5
Piccadilly	8.59	7.17	8.59	6.46	3	2
Victoria	8.55	7.89	5.76	10.18	2	3

Table 6-3: Line Reliability Factors and Ranks

6.2.2 *Victoria Line*

The Victoria line is of special interest in this analysis because of its strategic station locations connecting National Rail stations to the heart of Zone 1. It is the shortest line among the five lines in this analysis with 16 stations, of which 13 are in the trunk portion. The Victoria line serves 7 of the 15 highest demand origins in the AM peak, all of which are connected to the National Rail network. In addition, Victoria also serves 6 of the 15 busiest interchange stations in the AM peak, the busiest being Oxford Circus with an estimated 27,000 AM Peak interchanges. Oxford Circus is also the largest AM peak destination on the Victoria line with over 37,000 exits.

Among the 5 lines, the Victoria has the largest directional difference in the reliability factor – 5.8 minutes northbound and 10.7 minutes southbound – while the overall trunk only average is 7.9 minutes. This difference can be seen in the OD level median and 95th percentile journey times shown in Figure 6-7. The trunk portion of the Victoria line has a total of 13 stations from Brixton in the south to Seven Sisters in the north.

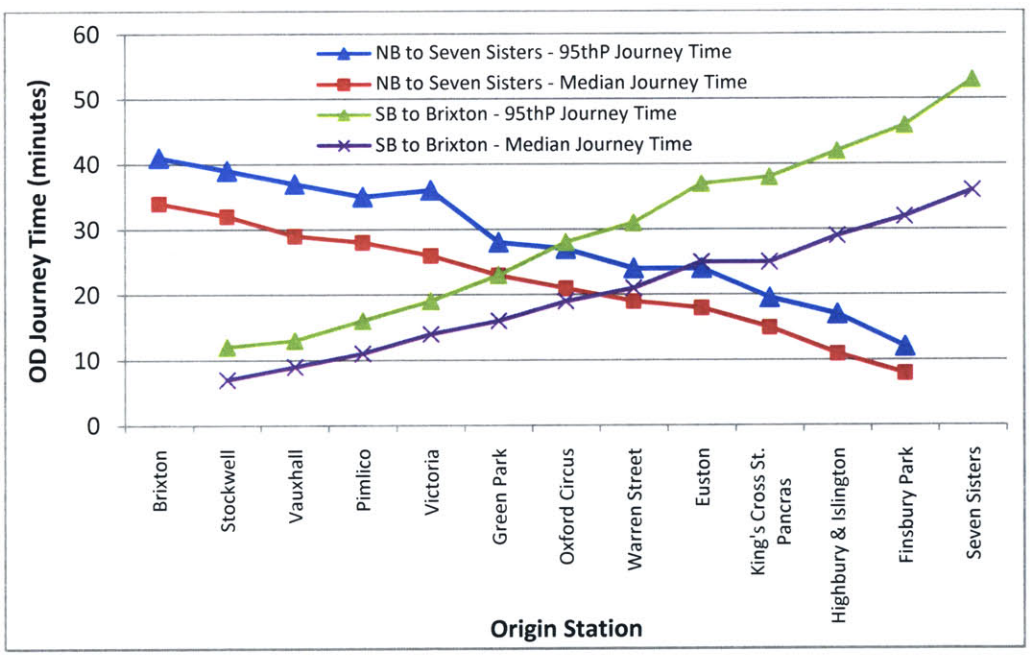

Figure 6-7: Victoria Line OD Level Reliability Factors

The two downward sloping lines in the figure show the northbound journey times from each station to Seven Sisters. The difference between the 95th percentile and median journey times generally decreases with journey length. An exception occurs at Victoria where the 95th percentile journey time is higher than at the previous station. The large variation of journey times originating from Victoria highlights the problem that many passengers are unable to board the first northbound train due to high levels of crowding on the Victoria platforms and perhaps on the trains arriving at Victoria. As a result, many passengers at Victoria have to wait for the second or even third train before they can board.

Figures 6-8 and 6-9 show the Victoria line OD level reliability factors together with link loads at the boarding station northbound and southbound, respectively. These figures also provide insight into the relationship between the reliability factor and journey length.

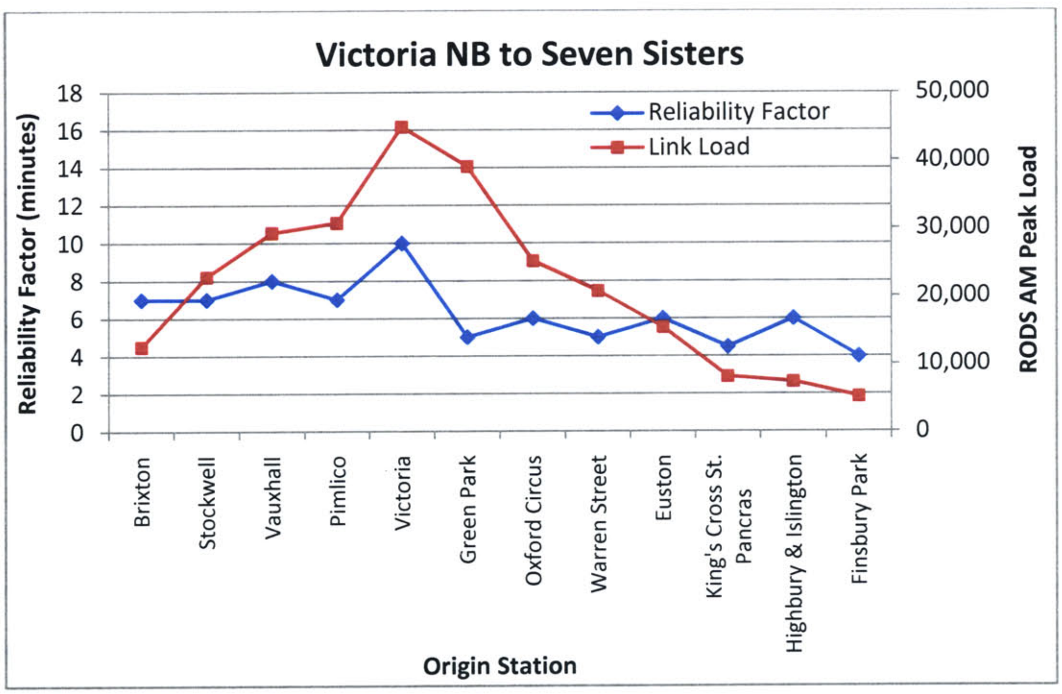

Figure 6-8: Victoria Line Northbound Reliability Factors and Link Loads

In the northbound direction, the reliability factors are only slightly and non-linearly affected by journey length, decreasing from 7.5 minutes at Brixton to 4 minutes at Finsbury Park. The trend of reliability factors clearly follows the variation in link load along the line. The link load represents the total number of passengers traveling on a single station-to-station link during the AM peak. For example, the heaviest link in the northbound direction is between Victoria and Green Park, on which over 45,000 passengers travel during the AM peak. Serious in-train crowding builds up as the trains pick up high volumes of passengers, and as a result, trains are already crowded when they arrive at Victoria where a huge platform demand awaits. The northbound reliability factor peaks at Victoria and decreases with the link load.

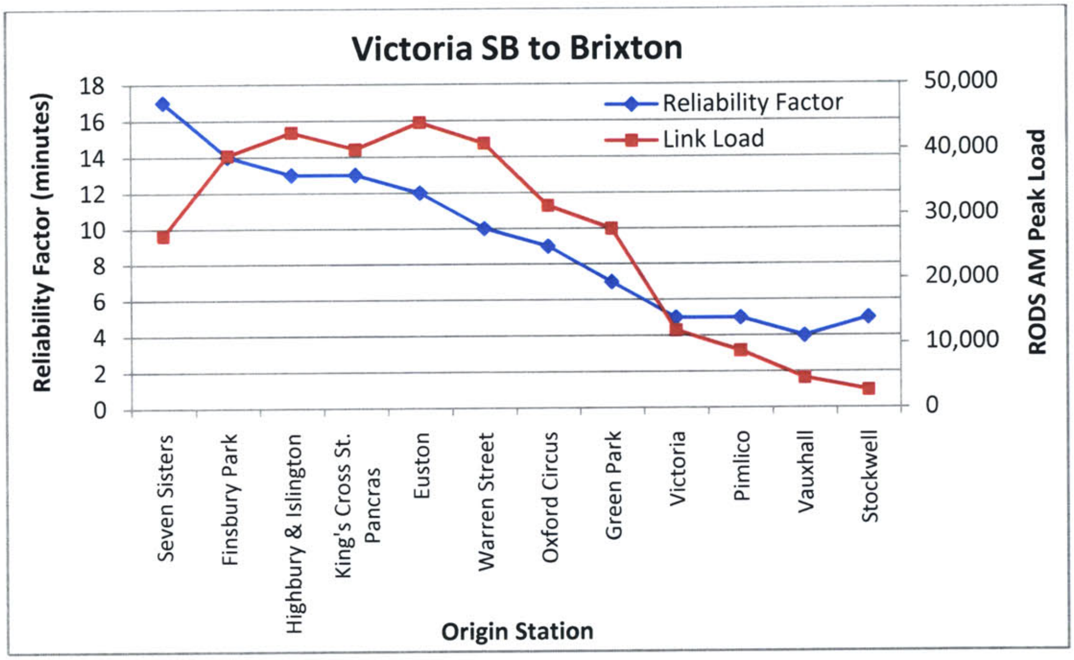

Figure 6-9: Victoria Line Southbound Reliability Factors and Link Loads

On the other hand, the southbound reliability factor decreases with both journey length and link load. The reliability factor decreases from 17 minutes at Seven Sisters to 4.5 minutes at Stockwell. Given the low link load at Seven Sisters, the reliability factor of 17 minutes for the OD pair from Seven Sisters to Brixton is unusually high. This poor level of reliability could be attributed to delays incurred by passengers when the trains pass through the most congested area between Finsbury Park and Warren Street where the link loads are consistently over 40,000.

Overall, the large differences in reliability factors between the northbound and southbound directions are mainly attributable to imbalance in directional link loads and demands. It should also be noted that since the OD level results are weighed by demand on only same-line OD pairs, the demand incurred by interchange passengers who travel on the Victoria line as a part of their journeys are not taken into account. This may lead to an unfair representation of the directional on-train volumes since the Victoria line serves very large numbers of interchange passengers.

Chapter 7 Summary and Conclusions

Transit agencies equipped with AFC systems can now make use of the large samples of AFC data to perform easy-to-update, more detailed and accurate OD estimation and journey time analyses which were more difficult to develop and update using manual survey data. This thesis shows that AFC data-based methodologies can provide cost-effective solutions to obtain better understanding of passenger demand and service quality.

This thesis presented two planning applications developed for the London Underground using Oyster data: 1) estimation of a time period level origin-destination flow matrix that reflects current demand, and 2) rail service reliability metrics that capture both excess journey time and variation in journey times at the origin-destination, line segment or line levels. The reliability metrics use the OD demand obtained from the first application to scale the OD level journey time results to the line segment or line levels.

Sections 7.1 and 7.2 summarize the research findings, discuss its limitations and suggest improvements for the time period level OD estimation and journey time reliability methodologies, respectively. Section 7.3 proposes future research directions related to this thesis.

7.1 Time Period Level OD Estimation

This research refines the origin-destination matrix estimation methodology for the full day to the time period level for a transit network equipped with both entry and exit controls for most stations. This research suggests that OD estimation should be based primarily on the AFC data and be supplemented with survey data when needed. The time period level OD estimation methodology presents a valuable opportunity for the

Underground to transform the current survey-based methodology to a more cost-effective and easy-to-update methodology that enhances the ability of transit planners to estimate OD demands for any time period and any time of the year.

7.1.1 *Overview of Research Findings*

The proposed Oyster-based methodology enables transit agencies to make use of the extensive AFC dataset as the basis to understand passenger travel behavior. The Oyster dataset captures travel on more than three times the number of OD pairs in one 4-week AM peak period compared to those OD pairs evident in the RODS database - 57,407 vs. 17,421. Comparison of the Oyster-based and RODS matrices yield the following conclusions:

- Demands are very similar at the network and zonal levels.
- At the station level there are significant differences for a number of central stations with respect to entries, exits and interchanges.
- At the OD level, the differences are greatest and a significant number of OD pairs in the RODS matrix seem to be erroneous or outdated.

While the differences in travel patterns between the Oyster-based and RODS matrices have not been shown here to be great enough to influence operations plans for the Underground, a more accurate matrix that captures active travel on many more OD pairs is beneficial for long term planning, such as new route design and capital investment assessment. In addition to a better understanding of demand in the network, the Oyster-based methodology is easier-to-update and much more cost-effective that the survey-based methodology. The flexibility of the Oyster-based methodology also allows transit agencies to easily assess the impacts of short-term policy and operational changes on passenger travel patterns.

7.1.2 *Limitations*

The Oyster-based OD estimation methodology has two main limitations. First, the methodology relies on the accuracy of station entry and exit counts which is currently dependent on the use of annual manual surveys. Since the manual count survey is conducted in November and the RODS OD matrix is scaled to average November station counts, the Oyster-based OD matrix is potentially skewed by combining daily changes in gate counts and annual manual counts, which may also vary significantly on a daily basis. This could be tested at the larger stations that currently include significant manual counts.

Second, the Oyster-based methodology does not advise the estimation of an OD matrix shorter than one hour because the RODS proportions used to calculate the expansion factors may not be accurate for such short periods. More detailed study on the OD estimation methodology is needed to estimate OD matrices for time periods shorter than one hour. Short time period OD matrices are needed to assess crowding levels on the critical links and interchange stations during the most congested half- or one-hour period of the AM and PM peaks.

7.1.3 *Suggested Improvements*

In the short run, the annual RODS survey program can be redesigned to more effectively complement the weaknesses of the Oyster-based method. The format of RODS can be modified and resources reallocated in the following ways:

- **Increase Response Rate** – The response rate to the RODS survey can be increased by shortening the questionnaire to make it easier for passengers to complete it. The RODS questionnaire can be simplified to obtain only the necessary information for Underground service and operations planning, including origin and destination stations, path and ticket type. Other information such as linked trips, journey purpose and passenger background can be obtained from the London Area Transport Survey (LATS) and the London Travel Demand Survey (LTDS), or perhaps by new internet surveys of Oyster users. The surveys provide information about Londoners' travel patterns and modal share for a wide range of users by interviewing selected households across London[11].

- **Selection of Stations** – Instead of surveying the Underground stations cyclically, RODS can target stations where Oyster data is weakest, especially non fully gated stations and stations with high proportions of magnetic stripe ticket use. Survey responses from these stations can improve the accuracy of the resulting OD matrix by refining our knowledge of non-Oyster users at these stations.
- **Ticket Type Differentiation** – RODS can emphasize surveying passengers using magnetic stripe tickets because their travel patterns are not captured in the Oyster data. It is believed that magnetic ticket users tend to make shorter trips within Zone 1 after transferring from the National Rail. By obtaining a better understanding of travel patterns of magnetic ticket users, the accuracy of the resulting OD matrix can be improved.

In the long run, however, it is worthwhile to re-evaluate the need for manual survey programs as Oyster usage and coverage continue to increase. More innovative strategies combining automated data collection systems with targeted manual surveys can help transit agencies estimate OD travel patterns more accurately and cost effectively.

- **Manual Count Surveys** – To improve the accuracy of station entry and exit counts and to more accurately estimate OD travel, manual counts at non fully gated stations need to be conducted more frequently to capture seasonality and day-to-day variation. Other technological advances, such as video-based automatic passenger counting technology, can also be used to estimate station control totals.

[11] http://www.tfl.gov.uk/assets/downloads/LTDS-research-supplement.pdf, retrieved May 17, 2007.

- **Path Choice Information** – While Oyster data capture more extensive travel patterns than manual surveys, they do not include information about the actual paths taken by passengers. Other sources are needed to obtain path choice information which is critical to understanding passenger behavior in general and system link loads in particular. Although the RODS surveys collect path choice information, they are not a cost effective solution to obtain OD travel and path choice information in the long run. Three alternative approaches are presented below:
  - **Extensive Path Choice Survey** – Since path choice behavior is unlikely to change as rapidly as OD travel, a large-scale path choice survey could be conducted to obtain a thorough understanding of passenger preference. Such a survey should target only stations serving more than one line to capture the maximum amount of complex choice information. Path choice is also likely to differ across time periods due to differences in time sensitivity of passengers at different times of the day. The results from the path choice survey should be representative of travel in the network until major changes in service occur and can be used for path choice model formulation and estimation for use in the longer term planning context.
  - **Automated Path Choice Data Collection** – Path choice data collection can also be automated by installing stand-alone Oyster card readers at major interchange stations for passengers to validate while walking from one platform to another. Financial incentives can be provided to increase passengers' likelihood to provide such information. For example, a certain number of passengers who reveal their paths can be randomly selected at each interchange station to be rewarded with Oyster card credits towards the next PAYG or Travelcard purchase. A cost benefit analysis is needed to assess the feasibility of this automated path choice data collection method.

- **Internet-Based Survey** – Surveys to obtain path choice information can be conducted online when Oyster card users manage their online accounts. Simple one-question surveys can be prompted when Oyster uses access their account to ask about the path taken on their previous transfer trip. The implementation cost of this kind of survey is low but further study is needed to understand its potential bias and limitations.

- **Reduction of Oyster Data Bias** – As Oyster PAYG becomes available on the National Rail, more information about travel from non fully gated stations with National Rail connections can be obtained. It is expected that National Rail Season Travelcard will also switch from magnetic stripe tickets to Oyster in the future.

7.2 Journey Time Reliability Metrics

AFC transactional data have enhanced transit agencies' ability to obtain large samples of origin-destination journey time information in transit systems that have both entry and exit fare payment control. This research proposes an AFC data-based methodology to assess journey time performance by selecting OD pairs on the same line. The Excess Journey Time Metric and Journey Time Reliability Metric can be effective management tools for assessing system performance using simple figures that have straightforward interpretations.

7.2.1 Overview of Research Findings

The proposed Excess Journey Time Metric and Journey Time Reliability Metric utilize large continuous streams of data to support analyses during short time periods. The methodology is very flexible since it considers different levels of origin and destination stations or line segments. Detailed differences in journey time distributions across different time periods can also be captured. Several ways to evaluate the proposed metrics are suggested:

- The comparison of the Excess Journey Time Metric and the official Underground Journey Time Metric show significant differences in line level results in terms of both number of minutes and relative performance across lines. The differences are due to differences in scheduled journey times and OD demand weightings used in the two methodologies.
- By comparing link load data and the reliability results, the effects of crowding on unreliability can be better understood.
- The excess journey time and reliability results generally agree with each other – a line with high excess journey times usually exhibits a high level of unreliability for the same period.
- Considerable differences in excess journey time and reliability results exist between directions on some lines. This is due to the highly imbalanced directional demand on some lines in the AM peak. The differences are likely to be smaller for the full day.

7.2.2 *Limitations and Challenges*

While Oyster data present a consistent and indisputable source of journey time information, there are three limitations of the Oyster-based methodology:

- Oyster journey time is captured between entry and exit and it is difficult to attribute the excess journey time and reliability results to various causes of delays. Without knowing the causes of delays and unreliability, operations planners cannot target specific aspects of passenger journeys to improve their Underground experience.
- The methodology assumes that passengers using magnetic tickets have the same OD journey time distributions as Oyster users. Passengers using single or return magnetic tickets are likely to be less familiar with the Underground and thus may need to spend more time to navigate within the system. The interactions between the journey time distributions of Oyster and magnetic ticket users

require more study. Although the journey time results should be representative of all users, the performance rating should not be penalized by passengers' individual behaviors that lead to longer journey times.

- The journey time reliability methodology considers only "trunk" journeys in order to eliminate the discrepancy between line level results due to the length of headways. Since the non-trunk portions are generally served by much longer headways than the trunk, the variations of non-trunk headways will introduce variations in the reliability results. However, on longer lines such as the Central and Piccadilly, over 40% of same-line journeys are non-trunk, and a truly customer focused metric needs to consider such a large number of passengers. The impacts of headway variations across the non-trunk portions of different lines can possibly be eliminated by expressing reliability as a fraction of median journey time (e.g. RF = 95^{th} percentile journey time / median journey time instead of taking the difference between the two). More detailed study is needed to assess the implications of such a normalized reliability factor.

The Journey Time Reliability Metric also presents a challenge for transit planners to implement directly. Due to the large amount of transactional data required to construct accurate journey time distributions, immediate processing of the data is required to prevent the loss of data storage capacity. Alternatively, the Oyster data collection unit can generate summary OD level journey time reports for transit planners and minimize data storage and processing costs.

7.2.3 *Suggested Improvements*

This thesis presents a powerful method to assess journey time performance using elapsed journey time between AFC transactions. Below are four suggestions to overcome some of the methodological limitations discussed in the previous section:

- **More Accurate Scheduled Access and Egress Times** – The scheduled access and egress times between the gates and the platforms should be carefully measured for each station. The use of a universal factor of 0.85 to transform the JTM scheduled access and egress times in this methodology is only a rough estimate and both the excess journey time and reliability results could be skewed. Shorter journeys are particularly prone to erroneous scheduled access and egress times because they constitute a larger proportion of total journey time on short journeys.
- **Extension to Non Trunk Journeys (for Reliability Metric only)** – Since over 40% of same-line journeys in the longer lines, such as the Central and Piccadilly, involve the non-trunk portion for one or both end, journey time performance for non-trunk journeys is also important for the Underground to truly understand the passenger experience.
- **Network Level Results** – Since approximately 40% of journeys in the Underground involve at least one interchange, it is important to assess journey time performance for journeys involving interchanges. An accurate path choice model is needed to assign the interchange journeys to the Underground network and adapt the same-line journey formulations presented in this thesis to the network level.
- **Separation of In-Train and Out-of-Train Time** – Although the Oyster-based excess journey time methodology cannot break the total journey time into components, train operations data can be used to separate the in-train time and the out-of-train time. The out-of-train time consists of access walk, platform wait including left behind, interchange walk and egress walk. Each of the walk times can be estimated from station flows and crowding levels in each passageway. By subtracting the walk times from the out-of-train time, the actual platform wait time can be estimated.

7.3 Future Research Directions

In addition to the suggested improvements discussed in Sections 7.1.3 and 7.2.3, several future research directions are listed below:

Research to further improve the understanding of travel patterns:

1. **Calibration of Transit Assignment Method** – By combining the OD matrix estimation and journey time results, the assignment procedure used to allocate OD demand onto the transit network can be more accurately calibrated since journey time is an important element of the path choice decision. Both the crowding and journey time reliability levels for an OD pair, or a line, are likely to affect passengers' path choices.

2. **Multimodal Travel Analysis** – The multimodal travel behavior of passengers is captured by the AFC system as the same transit smart card is accepted on most, if not all, transit modes. By analyzing passengers who take the Underground with one or more other modes as part of their linked trips, the interactions among modes can be better understood.

3. **Personal Data** – With personal data such as home addresses and demographic characteristics, transit agencies can better understand mode choice and path choice given the available transit services in the vicinity of a home address. More concentrated analyses of OD travel behavior for transit users in a particular geographic area can help evaluate the existing services and potential alternatives.

Research to enhance service performance monitoring:

1. **Crowding Analysis** – The most crowded state of a transit network using different spans of one half or one-hour durations during the peak periods. A

shorter journey time analysis time frame, such as during the peak half hour, is needed to assess and quantify the critical crowding levels.

2. **Separation of Train and Non-Train Journey Time** – The in-train time component can be separated from the total Oyster-based journey time using train operations data. By analyzing the out-of-train time components, it is possible to obtain distributions of platform wait time which is most negatively impacted by crowding.

Appendix A Sample RODS Questionnaire

Please answer the questions below about the journey you were making when you were handed this questionnaire at Tower Hill.

Please either tick the relevant box ☑ or write in the appropriate answer e.g. [9.37] AM

Section 1: Your journey to the Underground

1

Where have you just come from?

Please tick ONE box only

- Home ☐
- Normal workplace ☐
- Other workplace/business meeting ☐
- Visiting friends/relatives/on holiday ☐
- Theatre/cinema/concert etc ☐
- Sporting activity/event ☐
- Museum/exhibition ☐
- Other social (e.g. restaurant, pub) ☐
- Shopping ☐
- School/college/university (as student) ☐
- School/college (accompanying pupil) ☐
- Taking someone to airport, station, hotel etc ☐
- Meeting someone at airport, station, hotel etc ☐
- Personal business (e.g. doctor, hospital, bank) ☐
- Sightseeing ☐
- Hotel/guest house etc ☐
- Other (please tick and write in) ☐

2

It would help us if you were willing to enter the address of the place where you **started this journey.** This information is used to plan station entrances and exits and will not be used for marketing purposes.

Please give us as much information as possible

Name of shop/hotel etc. (if appropriate)

Street & number

District/Town

Postcode

3

At what time did you set out on this journey?

AM

PM

4

At what time did you actually reach Tower Hill Underground station?

AM

PM

5

How did you get to Tower Hill Underground station from the place mentioned in Q2?

Please complete ONE box only

If you used more than one type of transport please complete the MAIN method used

National Rail (⇌) - Please give origin station

Docklands Light Railway - Please give origin station

Bus - Please give bus route number

Tram - Please give origin station

Another Underground train (⊖) - Please give origin station

- Car/van - parked at /near station ☐
- Car/van - dropped off ☐
- Coach/workbus ☐
- Motorcycle ☐
- Bicycle ☐
- Air ☐
- Taxi/minicab ☐
- Walked all the way from the start ☐
- Boat ☐
- Other (please tick and write in) ☐

Section 2: Your journey from Tower Hill Underground station

6a

On this journey which train service did you use from Tower Hill station?

Please tick ONE box only and continue to Q6b unless otherwise indicated

London Underground: District Line ☐

London Underground: Circle Line ☐

None - I left Tower Hill through the exit to the street ☐ **go to Q8**

6b

At which National Rail, DLR or Underground station did you finish your journey?

Please write in the name of the Underground (⊖), DLR or National Rail (⇌) station where you ended your journey

7

Please write in the name(s) of all the Underground (⊖), DLR and National Rail (⇌) station(s) where you changed trains during your journey.

Please leave blank if you did not make any changes

First change at	Second change at	Third change at	Fourth change at

8

When you arrived at your destination station, how did you complete the journey to your destination address?

Please complete ONE box only

If you used more than one type of transport please complete the MAIN method used

Bus - Please give bus route number

Trams - Please give destination station

Car/van - parked at /near station ☐

Car/van - picked up ☐

Coach ☐

Motorcycle ☐

Bicycle ☐

Air ☐

Taxi/minicab ☐

Walked all the way from the station ☐

Boat ☐

Other (please tick and write in) ☐

9

Why were you travelling to this place/destination?

Please tick ONE box only

Going home ☐

Going to normal workplace ☐

Going to other workplace/business meeting ☐

Visiting friends/relatives/on holiday ☐

Going to the theatre/cinema/concert etc ☐

Going to a sporting activity/event ☐

Going to a museum/exhibition ☐

Other social (e.g. restaurant, pub) ☐

Going shopping ☐

Going to school/college/university (as student) ☐

Accompanying pupil to/from school/college ☐

Taking someone to airport/station/hotel etc ☐

Meeting someone at airport, station, hotel etc ☐

Personal business (e.g. doctor, hospital, bank) ☐

Going sightseeing ☐

Going to hotel/ guest house etc ☐

Other (please tick and write in) ☐

10

It would help us if you were willing to enter the address of the place you were **travelling to.** This information is used to plan station entrances and exits and will not be used for marketing purposes.

Please give us as much information as possible

Name of shop/hotel etc. (if appropriate)

Street & number

District/Town

Postcode

Section 3: Ticket type

11 What type of ticket, Oyster card or pass did you use for the Underground part of this journey?
Please tick ONE box only

Tickets

- Single ticket ☐
- Return ticket ☐
- Carnet ☐
- Extension ticket ☐
- Pre Pay Oyster ☐

Passes/Permits

- Elderly Persons Permit/Freedom Pass ☐
- Disabled Persons Permit/Freedom Pass ☐
- LT/NR staff pass ☐
- Police pass ☐

Travelcards/LT Cards

- One Day Travelcard peak ☐
- One Day Travelcard off peak ☐
- Weekly Travelcard/Oyster card ☐
- Monthly Travelcard/Oyster card ☐
- Annual Travelcard/Oyster card ☐
- Other Length Travelcard/Oyster card ☐
- Weekend Travelcard ☐
- Family Travelcard/Group Day ☐
- Visitor Travelcard ☐
- Daily LT Card (Bus/Underground) ☐
- 16-17 youth Travelcard ☐
- Student Travelcard ☐

Other (please write in)

12 **Travelcard/LT Card users only**, otherwise go to **Q13**
Please tick ALL zones covered by your Travelcard/LT Card

Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Outer zone station

Section 4: Background information

We need to finish this section of the questionnaire by collecting some background information about you.

13 How often do you make this particular journey?
Please tick ONE box only

5 or more days a week	1-4 days a week	Once a fortnight	Once a month	Less than once a month	First time ever

14 And are you

- Male ☐
- Female ☐

15 What age were you on your last birthday?
Please tick ONE box only

Under 16	16 - 19	20 - 24	25 - 34	35 - 44	45 - 59	60 - 64	65 - 70	Over 70

16 How many cars or private vans does your household have regularly available for use (including company cars?)
Please tick ONE box only

0	1	2	3	more than 3

17 **If the journey you are answering about does not start or finish at home, please write in where you normally live or the postcode.** This information will not be used for marketing purposes.

Street & number

District/Town

If outside UK, please give country

Postcode

Thank you for taking the time to complete this questionnaire.

Appendix B Classification of Underground Stations

Station Name	National Location Code	Gated	Non Gated	National Rail Connections	RODS 2005 Counting Method	Grouping in this research
Acton Town	500	YES	-	-	AFC counts	FG
Aldgate	502	YES	-	-	AFC counts	FG
Aldgate East	503	YES	-	-	AFC counts	FG
Alperton	505	YES	-	-	AFC counts	FG
Amersham	506	-	YES	YES	Manual + AFC counts	NFG
Angel	507	YES	-	-	AFC counts	FG
Archway	508	YES	-	-	AFC counts	FG
Arnos Grove	509	YES	-	-	AFC counts	FG
Arsenal	510	YES	-	-	AFC counts	FG
Baker Street	511	YES	-	-	AFC counts	FG
Balham	512	YES	-	YES	AFC counts	FG
Bank & Monument	513	-	YES	-	Manual + AFC counts	NFG
Barbican	501	-	YES	YES	Manual + AFC counts	NFG
Barking	514	-	YES	YES	Manual + AFC counts	NFG
Barkingside	515	YES	-	-	AFC counts	FG
Barons Court	516	YES	-	-	AFC counts	FG
Bayswater	517	YES	-	-	AFC counts	FG
Becontree	518	YES	-	-	Manual counts	NFG
Belsize Park	519	YES	-	-	AFC counts	FG
Bermondsey	787	YES	-	-	AFC counts	FG
Bethnal Green	520	YES	-	-	AFC counts	FG
Blackfriars	521	YES	-	YES	AFC counts	FG
Blackhorse Road	522	-	YES	YES	Manual + AFC counts	NFG
Bond Street	524	YES	-	-	AFC counts	FG
Borough	525	YES	-	-	AFC counts	FG
Boston Manor	526	YES	-	-	AFC counts	FG
Bounds Green	527	YES	-	-	AFC counts	FG
Bow Road	528	YES	-	-	AFC counts	FG
Brent Cross	529	YES	-	-	Manual counts	NFG
Brixton	778	YES	-	YES	AFC counts	FG
Bromley-by-Bow	530	YES	-	-	AFC counts	FG
Buckhurst Hill	531	YES	-	-	AFC counts	FG
Burnt Oak	532	YES	-	-	Manual counts	NFG
Caledonian Road	534	YES	-	-	AFC counts	FG
Camden Town	535	YES	-	-	AFC counts	FG
Canada Water	788	YES	-	-	AFC counts	FG
Canary Wharf	852	YES	-	-	AFC counts	FG
Canning Town	884	-	YES	YES	Manual + AFC counts	NFG
Cannon Street	536	YES	-	YES	AFC counts	FG
Canons Park	537	YES	-	-	AFC counts	FG

Chalfont & Latimer	539	-	YES	YES	Manual + AFC counts	NFG
Chalk Farm	540	YES	-	-	AFC counts	FG
Chancery Lane	541	YES	-	-	AFC counts	FG
Charing Cross	718	YES	-	YES	AFC counts	FG
Chesham	543	YES	-	-	AFC counts	FG
Chigwell	544	YES	-	-	AFC counts	FG
Chiswick Park	545	YES	-	-	AFC counts	FG
Chorleywood	546	-	YES	YES	Manual counts	NFG
Clapham Common	547	YES	-	-	AFC counts	FG
Clapham North	548	YES	-	YES	AFC counts	FG
Clapham South	549	YES	-	-	AFC counts	FG
Cockfosters	550	YES	-	-	AFC counts	FG
Colindale	551	YES	-	-	Manual counts	NFG
Colliers Wood	552	YES	-	-	AFC counts	FG
Covent Garden	553	YES	-	-	AFC counts	FG
Croxley	554	YES	-	-	AFC counts	FG
Dagenham East	555	YES	-	-	AFC counts	FG
Dagenham Heathway	556	YES	-	-	Manual counts	NFG
Debden	557	YES	-	-	AFC counts	FG
Dollis Hill	558	YES	-	-	AFC counts	FG
Ealing Broadway	560	-	YES	YES	Manual + AFC counts	NFG
Ealing Common	561	YES	-	-	AFC counts	FG
Earl's Court	562	YES	-	-	AFC counts	FG
East Acton	563	YES	-	-	AFC counts	FG
East Finchley	565	YES	-	-	AFC counts	FG
East Ham	566	YES	-	-	AFC counts	FG
East Putney	567	YES	-	-	AFC counts	FG
Eastcote	564	YES	-	-	AFC counts	FG
Edgware	568	YES	-	-	Manual counts	NFG
Edgware Road (Bak)	774	YES	-	-	AFC counts	FG
Edgware Road (Cir)	569	YES	-	-	AFC counts	FG
Elephant & Castle	570	YES	-	YES	AFC counts	FG
Elm Park	571	YES	-	-	AFC counts	FG
Embankment	542	YES	-	-	AFC counts	FG
Epping	572	YES	-	-	AFC counts	FG
Euston	574	YES	-	YES	AFC counts	FG
Euston Square	575	YES	-	-	AFC counts	FG
Fairlop	576	YES	-	-	AFC counts	FG
Farringdon	577	-	YES	YES	Manual + AFC counts	NFG
Finchley Central	578	YES	-	-	Manual counts	NFG
Finchley Road	579	YES	-	-	AFC counts	FG
Finsbury Park	580	-	YES	YES	Manual counts	NFG
Fulham Broadway	581	YES	-	-	AFC counts	FG
Gants Hill	582	YES	-	-	AFC counts	FG
Gloucester Road	583	YES	-	-	AFC counts	FG
Golders Green	584	YES	-	-	AFC counts	FG
Goldhawk Road	585	YES	-	-	AFC counts	FG

Goodge Street	586	YES	-	-	AFC counts	FG
Grange Hill	587	YES	-	-	AFC counts	FG
Great Portland Street	588	YES	-	-	AFC counts	FG
Green Park	590	YES	-	-	AFC counts	FG
Greenford	589	-	YES	YES	Manual + AFC counts	NFG
Gunnersbury	591	-	YES	YES	Manual + AFC counts	NFG
Hainault	592	YES	-	-	AFC counts	FG
Hammersmith (Dis)	593	YES	-	-	AFC counts	FG
Hammersmith (H&C)	773	YES	-	-	AFC counts	FG
Hampstead	594	YES	-	-	AFC counts	FG
Hanger Lane	595	YES	-	-	AFC counts	FG
Harlesden	596	-	YES	YES	Manual counts	NFG
Harrow & Wealdstone	597	-	YES	YES	Manual counts	NFG
Harrow-on-the-Hill	598	-	YES	YES	Manual + AFC counts	NFG
Hatton Cross	779	YES	-	-	AFC counts	FG
Heathrow Terminal 4	781	YES	-	-	AFC counts	FG
Heathrow Terminals 123	780	YES	-	-	Manual counts	NFG
Hendon Central	601	YES	-	-	Manual counts	NFG
High Barnet	602	YES	-	-	Manual counts	NFG
High Street Kensington	605	YES	-	-	AFC counts	FG
Highbury & Islington	603	-	YES	YES	Manual + AFC counts	NFG
Highgate	604	YES	-	-	AFC counts	FG
Hillingdon	606	YES	-	-	AFC counts	FG
Holborn	607	YES	-	-	AFC counts	FG
Holland Park	608	YES	-	-	AFC counts	FG
Holloway Road	609	YES	-	-	AFC counts	FG
Hornchurch	610	YES	-	-	AFC counts	FG
Hounslow Central	611	YES	-	-	AFC counts	FG
Hounslow East	612	YES	-	-	Manual + AFC counts	NFG
Hounslow West	613	YES	-	-	AFC counts	FG
Hyde Park Corner	614	YES	-	-	AFC counts	FG
Ickenham	615	YES	-	-	AFC counts	FG
Kennington	616	YES	-	-	AFC counts	FG
Kensal Green	617	-	YES	YES	Manual + AFC counts	NFG
Kensington (Olympia)	618	-	YES	YES	Manual counts	NFG
Kentish Town	619	-	YES	YES	Manual + AFC counts	NFG
Kenton	620	-	YES	YES	Manual counts	NFG
Kew Gardens	621	-	YES	YES	Manual + AFC counts	NFG
Kilburn	622	YES	-	-	AFC counts	FG
Kilburn Park	623	YES	-	-	AFC counts	FG
King's Cross St. Pancras	625	-	YES	YES	Manual + AFC counts	NFG
Kingsbury	624	YES	-	-	AFC counts	FG
Knightsbridge	626	YES	-	-	AFC counts	FG
Ladbroke Grove	627	YES	-	-	AFC counts	FG
Lambeth North	628	YES	-	-	AFC counts	FG
Lancaster Gate	629	YES	-	-	AFC counts	FG

Latimer Road	630	YES	-	-	AFC counts	FG
Leicester Square	631	YES	-	-	AFC counts	FG
Leyton	632	YES	-	-	AFC counts	FG
Leytonstone	633	YES	-	-	AFC counts	FG
Liverpool Street	634	YES	-	YES	AFC counts	FG
London Bridge	635	YES	-	YES	AFC counts	FG
Loughton	636	YES	-	-	AFC counts	FG
Maida Vale	637	YES	-	-	AFC counts	FG
Manor House	638	YES	-	-	AFC counts	FG
Mansion House	639	YES	-	-	AFC counts	FG
Marble Arch	640	YES	-	-	AFC counts	FG
Marylebone	641	YES	-	YES	AFC counts	FG
Mile End	642	YES	-	-	AFC counts	FG
Mill Hill East	643	-	YES	-	Manual counts	NFG
Moor Park	646	YES	-	-	AFC counts	FG
Moorgate	645	-	YES	YES	Manual + AFC counts	NFG
Morden	647	YES	-	-	AFC counts	FG
Mornington Crescent	648	YES	-	-	AFC counts	FG
Neasden	649	YES	-	-	AFC counts	FG
New Cross	651	-	YES	YES	Manual counts	NFG
New Cross Gate	652	-	YES	YES	Manual counts	NFG
Newbury Park	650	YES	-	-	AFC counts	FG
North Acton	653	YES	-	-	AFC counts	FG
North Ealing	654	YES	-	-	AFC counts	FG
North Greenwich	789	YES	-	-	AFC counts	FG
North Harrow	656	YES	-	-	AFC counts	FG
North Wembley	659	-	YES	YES	Manual counts	NFG
Northfields	655	YES	-	-	AFC counts	FG
Northolt	657	YES	-	-	AFC counts	FG
Northwick Park	660	YES	-	-	AFC counts	FG
Northwood	661	YES	-	-	AFC counts	FG
Northwood Hills	662	YES	-	-	AFC counts	FG
Notting Hill Gate	663	YES	-	-	AFC counts	FG
Oakwood	664	YES	-	-	AFC counts	FG
Old Street	665	-	YES	YES	Manual + AFC counts	NFG
Osterley	667	YES	-	-	AFC counts	FG
Oval	668	YES	-	-	AFC counts	FG
Oxford Circus	669	YES	-	-	AFC counts	FG
Paddington	670	-	YES	YES	Manual + AFC counts	NFG
Park Royal	671	YES	-	-	AFC counts	FG
Parsons Green	672	YES	-	-	AFC counts	FG
Perivale	673	YES	-	-	AFC counts	FG
Piccadilly Circus	674	YES	-	-	AFC counts	FG
Pimlico	776	YES	-	-	AFC counts	FG
Pinner	675	YES	-	-	Manual + AFC counts	NFG
Plaistow	676	YES	-	-	AFC counts	FG
Preston Road	677	YES	-	-	AFC counts	FG

Putney Bridge	678	YES	-	-	AFC counts	FG
Queen's Park	680	-	YES	YES	Manual + AFC counts	NFG
Queensbury	679	YES	-	-	AFC counts	FG
Queensway	681	YES	-	-	AFC counts	FG
Ravenscourt Park	682	YES	-	-	AFC counts	FG
Rayners Lane	683	YES	-	-	AFC counts	FG
Redbridge	684	YES	-	-	AFC counts	FG
Regent's Park	685	YES	-	-	AFC counts	FG
Richmond	686	-	YES	YES	Manual counts	NFG
Rickmansworth	687	-	YES	YES	Manual + AFC counts	NFG
Roding Valley	688	-	YES	-	Manual counts	NFG
Rotherhithe	689	YES	-	-	AFC counts	FG
Royal Oak	690	YES	-	-	AFC counts	FG
Ruislip	691	YES	-	-	AFC counts	FG
Ruislip Gardens	692	YES	-	-	AFC counts	FG
Ruislip Manor	693	YES	-	-	AFC counts	FG
Russell Square	694	YES	-	-	AFC counts	FG
Seven Sisters	698	-	YES	YES	Manual + AFC counts	NFG
Shadwell	699	YES	-	-	AFC counts	FG
Shepherd's Bush (Cen)	700	YES	-	-	AFC counts	FG
Shepherd's Bush (H&C)	775	YES	-	-	AFC counts	FG
Shoreditch	701	-	YES	-	Manual counts	NFG
Sloane Square	702	YES	-	-	AFC counts	FG
Snaresbrook	703	YES	-	-	Manual + AFC counts	NFG
South Ealing	704	YES	-	-	AFC counts	FG
South Harrow	707	YES	-	-	AFC counts	FG
South Kensington	708	YES	-	-	AFC counts	FG
South Kenton	709	-	YES	YES	Manual counts	NFG
South Ruislip	710	-	YES	YES	Manual + AFC counts	NFG
South Wimbledon	711	YES	-	-	AFC counts	FG
South Woodford	712	YES	-	-	AFC counts	FG
Southfields	705	YES	-	-	AFC counts	FG
Southgate	706	YES	-	-	AFC counts	FG
Southwark	784	YES	-	-	AFC counts	FG
St. James's Park	695	YES	-	-	AFC counts	FG
St. John's Wood	696	YES	-	-	AFC counts	FG
St. Paul's	697	YES	-	-	AFC counts	FG
Stamford Brook	713	YES	-	-	AFC counts	FG
Stanmore	714	YES	-	-	AFC counts	FG
Stepney Green	715	YES	-	-	AFC counts	FG
Stockwell	716	YES	-	-	AFC counts	FG
Stonebridge Park	717	-	YES	YES	Manual counts	NFG
Stratford	719	-	YES	YES	Manual + AFC counts	NFG
Sudbury Hill	720	YES	-	YES	AFC counts	FG
Sudbury Town	721	YES	-	-	Manual + AFC counts	NFG
Surrey Quays	722	YES	-	-	AFC counts	FG
Swiss Cottage	723	YES	-	-	AFC counts	FG

Temple	724	YES	-	-	AFC counts	FG
Theydon Bois	725	YES	-	-	AFC counts	FG
Tooting Bec	726	YES	-	-	AFC counts	FG
Tooting Broadway	727	YES	-	-	AFC counts	FG
Tottenham Court Road	728	YES	-	-	AFC counts	FG
Tottenham Hale	729	YES	-	YES	AFC counts	FG
Totteridge & Whetstone	730	YES	-	-	Manual counts	NFG
Tower Hill	731	YES	-	YES	AFC counts	FG
Tufnell Park	733	YES	-	-	AFC counts	FG
Turnham Green	734	YES	-	-	AFC counts	FG
Turnpike Lane	735	YES	-	-	AFC counts	FG
Upminster	736	-	YES	YES	Manual counts	NFG
Upminster Bridge	737	YES	-	-	AFC counts	FG
Upney	738	YES	-	-	Manual counts	NFG
Upton Park	739	YES	-	-	AFC counts	FG
Uxbridge	740	YES	-	-	AFC counts	FG
Vauxhall	777	YES	-	YES	AFC counts	FG
Victoria	741	YES	-	YES	AFC counts	FG
Walthamstow Central	742	YES	-	YES	AFC counts	FG
Wanstead	743	YES	-	-	AFC counts	FG
Wapping	744	YES	-	-	AFC counts	FG
Warren Street	745	YES	-	-	AFC counts	FG
Warwick Avenue	746	YES	-	-	AFC counts	FG
Waterloo	747	-	YES	YES	Manual + AFC counts	NFG
Watford	748	YES	-	-	AFC counts	FG
Wembley Central	751	-	YES	YES	Manual counts	NFG
Wembley Park	752	YES	-	-	AFC counts	FG
West Acton	753	YES	-	-	AFC counts	FG
West Brompton	755	-	YES	YES	Manual + AFC counts	NFG
West Finchley	756	YES	-	-	Manual counts	NFG
West Ham	757	-	YES	YES	Manual + AFC counts	NFG
West Hampstead	758	YES	-	YES	AFC counts	FG
West Harrow	759	YES	-	-	Manual + AFC counts	NFG
West Kensington	760	YES	-	-	Manual + AFC counts	NFG
West Ruislip	762	-	YES	YES	Manual counts	NFG
Westbourne Park	754	YES	-	-	AFC counts	FG
Westminster	761	YES	-	YES	AFC counts	FG
White City	764	YES	-	-	AFC counts	FG
Whitechapel	763	YES	-	-	AFC counts	FG
Willesden Green	765	YES	-	-	AFC counts	FG
Willesden Junction	766	-	YES	YES	Manual counts	NFG
Wimbledon	767	-	YES	YES	Manual counts	NFG
Wimbledon Park	768	YES	-	-	AFC counts	FG
Wood Green	770	YES	-	-	AFC counts	FG
Woodford	769	YES	-	-	AFC counts	FG
Woodside Park	771	YES	-	-	Manual counts	NFG

Appendix C Visual Basic Code of Iterative Proportional Fitting

```
Sub IPF()

ScaledSheet = "Matrix"
FirstRow = 3
LastRow = 57409
FinalColumn = 9
FactorENColumn = 10
FactorEXColumn = 11

a = 0
While a < 5

   a = a + 1

     'Balance rows
     Worksheets(ScaledSheet).Range(Cells(FirstRow, FactorEXColumn), Cells(LastRow, FactorEXColumn)).Select
     Selection.Copy
     Worksheets(ScaledSheet).Range(Cells(FirstRow, FinalColumn), Cells(LastRow, FinalColumn)).Select
     Selection.PasteSpecial Paste:=xlValues, Operation:=xlMultiply, SkipBlanks _
     :=False, Transpose:=False

     'Balance columns
     Worksheets(ScaledSheet).Range(Cells(FirstRow, FactorENColumn), Cells(LastRow, FactorENColumn)).Select
     Selection.Copy
     Worksheets(ScaledSheet).Range(Cells(FirstRow, FinalColumn), Cells(LastRow, FinalColumn)).Select
     Selection.PasteSpecial Paste:=xlValues, Operation:=xlMultiply, SkipBlanks _
     :=False, Transpose:=False

Wend

End Sub
```

Appendix D Station Exit Proportions

Station	National Location Code	7 - 8 AM	8 - 9 AM	9 - 10 AM	10 - 11 AM	11AM - 12PM
Acton Town	500	0.55	0.99	1.00	0.50	0.01
Barbican	501	0.61	1.00	1.00	0.58	0.01
Aldgate	502	0.50	0.99	1.00	0.70	0.02
Aldgate East	503	0.63	1.00	1.00	0.61	0.02
Alperton	505	0.65	0.99	1.00	0.51	0.04
Amersham	506	0.59	0.96	1.00	0.76	0.25
Angel	507	0.68	1.00	1.00	0.53	0.00
Archway	508	0.64	0.96	1.00	0.58	0.01
Arnos Grove	509	0.67	0.99	1.00	0.44	0.02
Arsenal	510	0.50	0.97	1.00	0.43	0.01
Baker Street	511	0.68	1.00	1.00	0.54	0.00
Balham	512	0.79	0.99	1.00	0.36	0.01
Bank & Monument	513	0.66	1.00	1.00	0.52	0.01
Barking	514	0.70	0.99	1.00	0.39	0.01
Barkingside	515	0.69	0.98	1.00	0.54	0.05
Barons Court	516	0.68	1.00	1.00	0.49	0.01
Bayswater	517	0.71	1.00	1.00	0.42	0.01
Becontree	518	0.68	0.98	1.00	0.32	0.03
Belsize Park	519	0.68	0.99	1.00	0.55	0.01
Bethnal Green	520	0.67	1.00	1.00	0.42	0.00
Blackfriars	521	0.67	1.00	1.00	0.54	0.00
Blackhorse Road	522	0.72	0.99	1.00	0.39	0.01
Bond Street	524	0.71	1.00	1.00	0.47	0.00
Borough	525	0.59	1.00	1.00	0.61	0.01
Boston Manor	526	0.63	0.98	1.00	0.47	0.01
Bounds Green	527	0.63	0.99	1.00	0.40	0.01
Bow Road	528	0.62	0.99	1.00	0.37	0.01
Brent Cross	529	0.70	1.00	1.00	0.52	0.02
Bromley By Bow	530	0.64	1.00	1.00	0.49	0.01
Buckhurst Hill	531	0.60	0.97	1.00	0.44	0.00
Burnt Oak	532	0.52	0.96	1.00	0.52	0.04
Caledonian Road	534	0.64	0.99	1.00	0.54	0.01
Camden Town	535	0.70	1.00	1.00	0.52	0.01
Cannon Street	536	0.62	1.00	1.00	0.52	0.01
Canons Park	537	0.66	0.99	1.00	0.50	0.03
Chalfont & Latimer	539	0.66	0.97	1.00	0.44	0.05
Chalk Farm	540	0.71	1.00	1.00	0.54	0.00
Chancery Lane	541	0.66	1.00	1.00	0.54	0.01
Embankment	542	0.70	1.00	1.00	0.53	0.01
Chesham	543	0.28	0.87	1.00	0.86	0.42
Chigwell	544	0.83	0.97	1.00	0.70	0.12

Chiswick Park	545	0.66	0.99	1.00	0.56	0.03
Chorleywood	546	0.54	0.97	1.00	0.51	0.00
Clapham Common	547	0.67	0.99	1.00	0.45	0.01
Clapham North	548	0.75	0.99	1.00	0.45	0.01
Clapham South	549	0.71	0.99	1.00	0.30	0.01
Cockfosters	550	0.50	0.95	1.00	0.58	0.04
Colindale	551	0.50	0.94	1.00	0.62	0.04
Colliers Wood	552	0.59	0.98	1.00	0.39	0.01
Covent Garden	553	0.70	1.00	1.00	0.54	0.01
Croxley	554	0.62	0.99	1.00	0.41	0.10
Dagenham East	555	0.71	0.98	1.00	0.31	0.02
Dagenham Heathway	556	0.69	0.99	1.00	0.39	0.01
Debden	557	0.70	1.00	1.00	0.61	0.01
Dollis Hill	558	0.73	0.99	1.00	0.48	0.00
Ealing Broadway	560	0.55	0.99	1.00	0.58	0.03
Ealing Common	561	0.63	0.99	1.00	0.51	0.01
Earls Court	562	0.64	1.00	1.00	0.52	0.01
East Acton	563	0.58	0.99	1.00	0.51	0.01
Eastcote	564	0.60	0.99	1.00	0.39	0.01
East Finchley	565	0.52	0.98	1.00	0.52	0.01
East Ham	566	0.67	0.98	1.00	0.47	0.02
East Putney	567	0.53	0.99	1.00	0.60	0.02
Edgware	568	0.62	0.97	1.00	0.52	0.03
Edgware Road M	569	0.66	0.99	1.00	0.56	0.01
Elephant & Castle	570	0.61	0.99	1.00	0.57	0.01
Elm Park	571	0.63	0.97	1.00	0.49	0.04
Epping	572	0.50	0.98	1.00	0.64	0.04
Euston	574	0.70	1.00	1.00	0.44	0.00
Euston Square	575	0.64	1.00	1.00	0.55	0.01
Fairlop	576	0.72	0.99	1.00	0.43	0.07
Farringdon	577	0.69	1.00	1.00	0.62	0.01
Finchley Central	578	0.63	0.92	1.00	0.54	0.03
Finchley Road	579	0.72	1.00	1.00	0.47	0.00
Finsbury Park	580	0.72	0.99	1.00	0.42	0.01
Fulham Broadway	581	0.61	1.00	1.00	0.57	0.00
Gants Hill	582	0.66	0.98	1.00	0.48	0.01
Gloucester Road	583	0.62	1.00	1.00	0.48	0.01
Golders Green	584	0.66	0.99	1.00	0.44	0.01
Goldhawk Road	585	0.61	0.99	1.00	0.46	0.01
Goodge Street	586	0.70	1.00	1.00	0.54	0.01
Grange Hill	587	0.57	0.94	1.00	0.58	0.05
Great Portland St	588	0.65	1.00	1.00	0.56	0.00
Greenford	589	0.50	0.99	1.00	0.56	0.01
Green Park	590	0.75	1.00	1.00	0.44	0.01
Gunnersbury	591	0.53	0.99	1.00	0.68	0.02
Hainault	592	0.49	0.98	1.00	0.52	0.02
Hammersmith D	593	0.65	1.00	1.00	0.54	0.02
Hampstead	594	0.72	0.99	1.00	0.47	0.01

Hanger Lane	595	0.57	0.99	1.00	0.54	0.02
Harlesden	596	0.62	0.97	1.00	0.57	0.05
Harrow Wealdstone	597	0.68	0.99	1.00	0.69	0.03
Harrow On The Hill	598	0.73	0.99	1.00	0.44	0.01
Hendon Central	601	0.66	0.99	1.00	0.54	0.05
High Barnet	602	0.51	0.98	1.00	0.68	0.04
Highbury	603	0.67	1.00	1.00	0.50	0.00
Highgate	604	0.58	0.99	1.00	0.56	0.01
High Street Kens	605	0.61	0.99	1.00	0.57	0.01
Hillingdon	606	0.47	0.97	1.00	0.50	0.04
Holborn	607	0.68	1.00	1.00	0.52	0.00
Holland Park	608	0.69	1.00	1.00	0.55	0.00
Holloway Road	609	0.62	0.99	1.00	0.58	0.01
Hornchurch	610	0.61	0.98	1.00	0.51	0.01
Hounslow Central	611	0.47	0.96	1.00	0.64	0.06
Hounslow East	612	0.49	0.96	1.00	0.61	0.05
Hounslow West	613	0.47	0.94	1.00	0.61	0.03
Hyde Park Corner	614	0.62	0.99	1.00	0.49	0.01
Ickenham	615	0.74	1.00	1.00	0.34	0.02
Kennington	616	0.66	0.99	1.00	0.52	0.01
Kensal Green	617	0.63	0.98	1.00	0.54	0.03
Kensington Olympia	618	0.45	0.95	1.00	0.78	0.04
Kentish Town	619	0.67	0.99	1.00	0.59	0.01
Kenton	620	0.76	0.99	1.00	0.45	0.01
Kew Gardens	621	0.44	0.98	1.00	0.59	0.05
Kilburn	622	0.66	0.99	1.00	0.45	0.02
Kilburn Park	623	0.62	0.99	1.00	0.45	0.01
Kingsbury	624	0.71	0.98	1.00	0.44	0.02
Kings Cross	625	0.66	0.99	1.00	0.47	0.00
Knightsbridge	626	0.63	0.99	1.00	0.52	0.00
Ladbroke Grove	627	0.59	0.98	1.00	0.60	0.02
Lambeth North	628	0.60	1.00	1.00	0.60	0.00
Lancaster Gate	629	0.63	1.00	1.00	0.48	0.00
Latimer Road	630	0.61	0.98	1.00	0.58	0.00
Leicester Square	631	0.71	1.00	1.00	0.51	0.00
Leyton	632	0.68	1.00	1.00	0.41	0.01
Leytonstone	633	0.68	0.99	1.00	0.42	0.02
Liverpool Street	634	0.61	1.00	1.00	0.54	0.00
London Bridge	635	0.63	1.00	1.00	0.51	0.00
Loughton	636	0.63	0.99	1.00	0.46	0.02
Maida Vale	637	0.72	0.99	1.00	0.46	0.01
Manor House	638	0.71	0.99	1.00	0.40	0.01
Mansion House	639	0.67	1.00	1.00	0.50	0.01
Marble Arch	640	0.66	1.00	1.00	0.48	0.00
Marylebone	641	0.58	1.00	1.00	0.47	0.00
Mile End	642	0.72	1.00	1.00	0.44	0.01
Mill Hill East	643	0.57	0.98	1.00	0.54	0.04
Moorgate	645	0.63	0.99	1.00	0.58	0.01

Moor Park	646	0.80	0.99	1.00	0.56	0.05
Morden	647	0.57	0.99	1.00	0.46	0.03
Mornington Crescent	648	0.73	0.99	1.00	0.55	0.00
Neasden	649	0.62	0.98	1.00	0.51	0.00
Newbury Park	650	0.60	0.99	1.00	0.56	0.03
New Cross	651	0.56	0.99	1.00	0.58	0.01
New Cross Gate	652	0.52	1.00	1.00	0.59	0.02
North Acton	653	0.68	1.00	1.00	0.59	0.01
North Ealing	654	0.70	1.00	1.00	0.38	0.00
Northfields	655	0.76	0.99	1.00	0.55	0.03
North Harrow	656	0.66	1.00	1.00	0.37	0.01
Northolt	657	0.65	0.96	1.00	0.46	0.02
North Wembley	659	0.52	0.98	1.00	0.51	0.03
Northwick Park	660	0.65	0.99	1.00	0.66	0.03
Northwood	661	0.63	0.98	1.00	0.51	0.04
Northwood Hills	662	0.67	0.98	1.00	0.38	0.02
Notting Hill Gate	663	0.65	1.00	1.00	0.48	0.00
Oakwood	664	0.66	0.97	1.00	0.51	0.03
Old Street	665	0.63	1.00	1.00	0.61	0.01
Osterley	667	0.41	0.98	1.00	0.56	0.05
Oval	668	0.67	0.99	1.00	0.51	0.01
Oxford Circus	669	0.75	1.00	1.00	0.47	0.00
Paddington	670	0.63	0.99	1.00	0.49	0.01
Park Royal	671	0.61	0.99	1.00	0.64	0.04
Parsons Green	672	0.64	0.99	1.00	0.55	0.01
Perivale	673	0.64	0.98	1.00	0.47	0.02
Piccadilly Circus	674	0.69	1.00	1.00	0.48	0.01
Pinner	675	0.59	0.98	1.00	0.41	0.05
Plaistow	676	0.74	0.99	1.00	0.36	0.01
Preston Road	677	0.69	0.99	1.00	0.40	0.01
Putney Bridge	678	0.51	0.99	1.00	0.60	0.01
Queensbury	679	0.59	0.97	1.00	0.44	0.02
Queens Park	680	0.68	0.99	1.00	0.61	0.00
Queensway	681	0.58	1.00	1.00	0.46	0.00
Ravenscourt Park	682	0.71	1.00	1.00	0.58	0.01
Rayners Lane	683	0.64	0.99	1.00	0.47	0.01
Redbridge	684	0.73	0.99	1.00	0.47	0.04
Regents Park	685	0.00	0.00	0.00	0.00	0.00
Richmond	686	0.50	0.98	1.00	0.62	0.03
Rickmansworth	687	0.56	0.97	1.00	0.51	0.04
Roding Valley	688	0.40	1.00	1.00	0.25	0.00
Rotherhithe	689	0.67	1.00	1.00	0.42	0.00
Royal Oak	690	0.50	0.99	1.00	0.59	0.02
Ruislip	691	0.73	0.98	1.00	0.41	0.03
Ruislip Gardens	692	0.45	0.96	1.00	0.33	0.02
Ruislip Manor	693	0.62	1.00	1.00	0.31	0.03
Russell Square	694	0.62	0.99	1.00	0.55	0.01
St James's Park	695	0.66	1.00	1.00	0.58	0.00

St Johns Wood	696	0.68	0.99	1.00	0.47	0.00
St Pauls	697	0.62	1.00	1.00	0.51	0.00
Seven Sisters	698	0.63	0.99	1.00	0.48	0.01
Shadwell	699	0.76	1.00	1.00	0.38	0.00
Shepherds Bush Ctl	700	0.60	1.00	1.00	0.59	0.00
Shoreditch	701	0.00	0.00	0.00	0.00	0.00
Sloane Square	702	0.66	0.99	1.00	0.52	0.00
Snaresbrook	703	0.71	0.99	1.00	0.65	0.03
South Ealing	704	0.59	0.99	1.00	0.49	0.01
Southfields	705	0.50	0.97	1.00	0.44	0.01
Southgate	706	0.64	0.98	1.00	0.50	0.02
South Harrow	707	0.66	0.98	1.00	0.44	0.04
South Kensington	708	0.62	1.00	1.00	0.51	0.00
South Kenton	709	0.66	0.99	1.00	0.40	0.00
South Ruislip	710	0.53	0.97	1.00	0.43	0.02
South Wimbledon	711	0.66	0.99	1.00	0.45	0.01
South Woodford	712	0.70	0.99	1.00	0.39	0.02
Stamford Brook	713	0.60	0.99	1.00	0.50	0.01
Stanmore	714	0.54	0.96	1.00	0.46	0.04
Stepney Green	715	0.71	1.00	1.00	0.50	0.01
Stockwell	716	0.71	0.99	1.00	0.42	0.00
Stonebridge Park	717	0.41	0.97	1.00	0.67	0.07
Charing Cross	718	0.67	1.00	1.00	0.50	0.00
Stratford	719	0.69	0.99	1.00	0.43	0.01
Sudbury Hill	720	0.58	0.97	1.00	0.50	0.03
Sudbury Town	721	0.71	0.98	1.00	0.43	0.00
Surrey Quays	722	0.57	0.99	1.00	0.37	0.00
Swiss Cottage	723	0.73	1.00	1.00	0.48	0.00
Temple	724	0.73	1.00	1.00	0.48	0.00
Theydon Bois	725	0.69	0.97	1.00	0.54	0.02
Tooting Bec	726	0.70	0.99	1.00	0.41	0.01
Tooting Broadway	727	0.67	0.99	1.00	0.44	0.02
Tottenham Court Rd	728	0.69	1.00	1.00	0.53	0.00
Tottenham Hale	729	0.70	1.00	1.00	0.41	0.01
Totteridge	730	0.55	0.96	1.00	0.76	0.03
Tower Hill	731	0.62	1.00	1.00	0.54	0.01
Tufnell Park	733	0.68	0.99	1.00	0.55	0.01
Turnham Green	734	0.66	0.99	1.00	0.57	0.02
Turnpike Lane	735	0.65	0.99	1.00	0.40	0.01
Upminster	736	0.65	0.99	1.00	0.47	0.04
Upminster Bridge	737	0.69	1.00	1.00	0.55	0.03
Upney	738	0.75	0.99	1.00	0.36	0.01
Upton Park	739	0.72	0.99	1.00	0.38	0.02
Uxbridge	740	0.49	0.97	1.00	0.64	0.06
Victoria	741	0.66	1.00	1.00	0.51	0.00
Walthamstow Central	742	0.65	0.99	1.00	0.44	0.01
Wanstead	743	0.78	0.99	1.00	0.41	0.00
Wapping	744	0.64	1.00	1.00	0.47	0.01

Warren Street	745	0.75	1.00	1.00	0.50	0.00
Warwick Avenue	746	0.62	0.99	1.00	0.55	0.00
Waterloo	747	0.63	1.00	1.00	0.52	0.00
Watford Met	748	0.67	0.98	1.00	0.69	0.08
Wembley Central	751	0.64	0.98	1.00	0.47	0.02
Wembley Park	752	0.64	0.98	1.00	0.51	0.01
West Acton	753	0.67	0.99	1.00	0.44	0.00
Westbourne Park	754	0.69	0.98	1.00	0.59	0.01
West Brompton	755	0.64	0.99	1.00	0.55	0.01
West Finchley	756	0.48	0.96	1.00	0.45	0.02
West Ham	757	0.67	0.99	1.00	0.39	0.01
West Hampstead	758	0.74	0.99	1.00	0.44	0.00
West Harrow	759	0.66	1.00	1.00	0.48	0.05
West Kensington	760	0.62	0.99	1.00	0.55	0.01
Westminster	761	0.64	1.00	1.00	0.58	0.01
West Ruislip	762	0.72	0.98	1.00	0.47	0.04
Whitechapel	763	0.70	1.00	1.00	0.47	0.01
White City	764	0.53	0.99	1.00	0.73	0.01
Willesden Green	765	0.62	0.99	1.00	0.42	0.00
Willesden Junction	766	0.48	0.98	1.00	0.55	0.01
Wimbledon	767	0.58	0.97	1.00	0.53	0.05
Wimbledon Park	768	0.61	0.96	1.00	0.38	0.01
Woodford	769	0.70	0.99	1.00	0.40	0.01
Wood Green	770	0.66	0.99	1.00	0.47	0.01
Woodside Park	771	0.67	0.99	1.00	0.64	0.01
Hammersmith M	773	0.74	0.99	1.00	0.46	0.00
Edgware Road B	774	0.64	1.00	1.00	0.51	0.00
Shepherds Bush Met	775	0.62	0.99	1.00	0.47	0.01
Pimlico	776	0.69	1.00	1.00	0.50	0.00
Vauxhall	777	0.65	1.00	1.00	0.50	0.00
Brixton	778	0.55	0.99	1.00	0.52	0.01
Hatton Cross	779	0.36	0.95	1.00	0.61	0.05
Heathrow Terms 123	780	0.30	0.88	1.00	0.83	0.17
Heathrow Term 4	781	0.32	0.89	1.00	0.85	0.12
Southwark	784	0.67	0.99	1.00	0.55	0.01
Bermondsey	787	0.68	0.99	1.00	0.41	0.01
Canada Water	788	0.69	0.99	1.00	0.38	0.01
North Greenwich	789	0.59	0.98	1.00	0.44	0.01
Canary Wharf	852	0.62	0.99	1.00	0.58	0.01
Canning Town	884	0.71	0.99	1.00	0.45	0.01

Appendix E Station Entry and Exit Ratios

Station	National Location Code	Group	RODS Entries	Oyster-Based Estimated Entries	Entry Ratio	RODS Exits	Oyster-Based Estimated Exits	Exit Ratio
Acton Town	500	FG	3,175	3,137	0.99	1,499	1,592	1.07
Barbican	501	NFG	1,119	1,119	1.00	6,616	7,190	1.09
Aldgate	502	FG	1,400	1,339	0.96	4,296	4,215	0.99
Aldgate East	503	FG	1,538	1,542	1.00	4,529	4,257	0.95
Alperton	505	FG	1,413	1,395	0.99	786	645	0.83
Amersham	506	NFG	1,235	1,410	1.14	445	512	1.16
Angel	507	FG	3,345	3,440	1.03	6,562	5,824	0.89
Archway	508	FG	3,702	3,801	1.03	1,913	1,867	0.98
Arnos Grove	509	FG	2,518	2,492	0.99	612	677	1.11
Arsenal	510	FG	1,472	1,361	0.92	244	422	1.74
Baker Street	511	FG	4,924	5,013	1.02	10,915	11,526	1.06
Balham	512	FG	7,010	7,233	1.03	2,100	2,206	1.06
Bank & Monument	513	NFG	10,851	10,851	1.00	45,124	44,804	1.00
Barking	514	NFG	5,639	5,639	1.00	3,602	3,576	1.00
Barkingside	515	FG	518	616	1.19	209	135	0.65
Barons Court	516	FG	3,432	3,521	1.03	2,962	3,007	1.02
Bayswater	517	FG	1,867	1,395	0.75	989	904	0.92
Becontree	518	NFG	1,658	1,658	1.00	170	228	1.35
Belsize Park	519	FG	2,381	2,336	0.98	1,946	2,027	1.05
Bethnal Green	520	FG	5,653	6,006	1.06	3,221	3,256	1.02
Blackfriars	521	FG	4,253	4,273	1.00	9,456	9,704	1.03
Blackhorse Road	522	NFG	4,046	4,191	1.04	657	761	1.17
Bond Street	524	FG	1,757	1,893	1.08	15,597	16,106	1.04
Borough	525	FG	769	870	1.13	2,143	2,010	0.94
Boston Manor	526	FG	1,023	1,011	0.99	453	339	0.75
Bounds Green	527	FG	3,753	3,948	1.05	472	557	1.19
Bow Road	528	FG	2,392	2,502	1.05	951	859	0.91
Brent Cross	529	NFG	1,357	1,357	1.00	468	465	1.00
Bromley By Bow	530	FG	1,000	1,114	1.11	1,055	731	0.70

Buckhurst Hill	531	FG	1,448	1,484	1.02	191	158	0.83
Burnt Oak	532	NFG	2,518	2,518	1.00	325	372	1.15
Caledonian Road	534	FG	2,014	2,131	1.06	1,598	1,824	1.15
Camden Town	535	FG	2,353	2,571	1.09	5,352	5,743	1.08
Cannon Street	536	FG	3,042	3,070	1.01	2,903	2,704	0.94
Canons Park	537	FG	1,209	1,259	1.04	172	180	1.05
Chalfont & Latimer	539	NFG	862	862	1.00	299	297	1.00
Chalk Farm	540	FG	1,711	1,819	1.06	881	1,009	1.15
Chancery Lane	541	FG	387	873	2.26	12,996	13,506	1.05
Embankment	542	FG	4,154	4,121	0.99	8,104	8,183	1.02
Chesham	543	FG	390	348	0.89	97	93	0.97
Chigwell	544	FG	237	271	1.14	107	108	1.02
Chiswick Park	545	FG	1,157	992	0.86	738	854	1.17
Chorleywood	546	NFG	1,043	1,043	1.00	159	158	1.00
Clapham Common	547	FG	4,335	4,386	1.01	1,598	1,653	1.04
Clapham North	548	FG	3,105	3,074	0.99	854	780	0.92
Clapham South	549	FG	5,686	5,716	1.01	1,498	1,308	0.88
Cockfosters	550	FG	840	833	0.99	278	259	0.94
Colindale	551	NFG	2,214	2,214	1.00	1,275	1,266	1.00
Colliers Wood	552	FG	3,600	3,886	1.08	559	500	0.90
Covent Garden	553	FG	360	311	0.86	4,285	4,448	1.05
Croxley	554	FG	579	603	1.04	236	221	0.94
Dagenham East	555	FG	1,343	1,338	1.00	449	328	0.73
Dagenham Heathway	556	NFG	2,005	2,005	1.00	872	866	1.00
Debden	557	FG	1,010	1,157	1.15	810	660	0.82
Dollis Hill	558	FG	2,171	2,196	1.01	776	781	1.01
Ealing Broadway	560	NFG	9,031	9,910	1.10	3,765	3,738	1.00
Ealing Common	561	FG	1,817	1,925	1.06	947	1,021	1.09
Earls Court	562	FG	7,919	7,450	0.94	5,797	6,013	1.04
East Acton	563	FG	1,743	1,615	0.93	1,176	1,093	0.94
Eastcote	564	FG	1,813	1,716	0.95	397	390	0.99
East Finchley	565	FG	4,303	4,155	0.97	770	716	0.94
East Ham	566	FG	5,840	5,956	1.02	1,685	1,592	0.95
East Putney	567	FG	3,723	3,560	0.96	1,660	1,604	0.97
Edgware	568	NFG	2,512	2,512	1.00	581	577	1.00
Edgware Road M	569	FG	1,250	1,193	0.95	2,849	2,966	1.05

Elephant & Castle	570	FG	5,169	5,557	1.08	4,973	4,435	0.90
Elm Park	571	FG	2,105	2,105	1.00	161	160	1.00
Epping	572	FG	1,824	2,076	1.14	370	285	0.78
Euston	574	FG	9,827	10,359	1.05	9,925	9,237	0.94
Euston Square	575	FG	2,614	3,184	1.22	5,737	5,779	1.01
Fairlop	576	FG	395	466	1.18	98	118	1.21
Farringdon	577	NFG	6,099	6,099	1.00	13,448	16,792	1.26
Finchley Central	578	NFG	3,082	3,082	1.00	872	866	1.00
Finchley Road	579	FG	4,506	4,430	0.98	2,180	2,344	1.08
Finsbury Park	580	NFG	17,530	17,530	1.00	5,158	5,121	1.00
Fulham Broadway	581	FG	3,332	3,078	0.92	3,359	3,972	1.19
Gants Hill	582	FG	3,556	3,479	0.98	470	451	0.97
Gloucester Road	583	FG	4,810	4,331	0.90	4,254	3,967	0.94
Golders Green	584	FG	3,748	3,752	1.00	1,263	1,358	1.08
Goldhawk Road	585	FG	805	729	0.91	343	354	1.04
Goodge Street	586	FG	405	350	0.86	5,050	5,162	1.03
Grange Hill	587	FG	310	360	1.16	0	34	#DIV/0!
Great Portland St	588	FG	957	919	0.96	4,467	4,854	1.09
Greenford	589	NFG	2,146	2,407	1.12	663	658	1.00
Green Park	590	FG	1,613	1,724	1.07	20,505	22,196	1.09
Gunnersbury	591	NFG	1,717	1,717	1.00	2,016	2,121	1.06
Hainault	592	FG	1,786	1,901	1.06	269	247	0.93
Hammersmith D	593	FG	7,027	7,175	1.02	11,795	12,079	1.03
Hampstead	594	FG	1,656	1,521	0.92	1,409	1,339	0.96
Hanger Lane	595	FG	1,446	1,672	1.16	801	831	1.04
Harlesden	596	NFG	1,210	1,210	1.00	827	821	1.00
Harrow Wealdstone	597	NFG	1,745	1,745	1.00	946	939	1.00
Harrow On The Hill	598	NFG	4,106	4,106	1.00	2,738	2,719	1.00
Hendon Central	601	NFG	3,610	3,610	1.00	1,338	1,329	1.00
High Barnet	602	NFG	1,649	1,666	1.01	575	541	0.95
Highbury	603	NFG	8,762	8,762	1.00	6,070	6,027	1.00
Highgate	604	FG	3,686	3,569	0.97	497	464	0.94
High Street Kens	605	FG	2,074	1,910	0.92	5,126	5,017	0.99
Hillingdon	606	FG	927	894	0.96	278	213	0.77
Holborn	607	FG	1,873	1,703	0.91	20,135	20,806	1.04
Holland Park	608	FG	1,607	1,554	0.97	1,380	1,475	1.08

Holloway Road	609	FG	1,825	2,047	1.12	3,137	2,080	0.67
Hornchurch	610	FG	1,420	1,420	1.00	355	352	1.00
Hounslow Central	611	FG	1,427	1,670	1.17	871	870	1.01
Hounslow East	612	NFG	1,928	1,999	1.04	734	712	0.98
Hounslow West	613	FG	1,590	1,646	1.04	230	287	1.26
Hyde Park Corner	614	FG	295	292	0.99	2,342	2,586	1.11
Ickenham	615	FG	606	596	0.98	360	349	0.98
Kennington	616	FG	1,745	1,910	1.09	824	852	1.04
Kensal Green	617	NFG	1,530	1,530	1.00	219	343	1.58
Kensington Olympia	618	NFG	365	365	1.00	352	350	1.00
Kentish Town	619	NFG	2,956	2,956	1.00	2,892	2,872	1.00
Kenton	620	NFG	554	554	1.00	478	475	1.00
Kew Gardens	621	NFG	1,802	1,692	0.94	827	786	0.96
Kilburn	622	FG	4,485	4,724	1.05	1,033	982	0.96
Kilburn Park	623	FG	1,469	1,503	1.02	646	600	0.93
Kingsbury	624	FG	1,844	1,992	1.08	606	579	0.96
Kings Cross	625	NFG	22,002	22,002	1.00	29,788	16,497	0.56
Knightsbridge	626	FG	1,000	950	0.95	6,208	6,374	1.03
Ladbroke Grove	627	FG	1,696	1,526	0.90	1,458	1,434	0.99
Lambeth North	628	FG	643	703	1.09	1,066	946	0.89
Lancaster Gate	629	FG	3,094	2,508	0.81	957	827	0.87
Latimer Road	630	FG	773	643	0.83	665	659	1.00
Leicester Square	631	FG	875	771	0.88	7,259	7,974	1.11
Leyton	632	FG	6,526	6,588	1.01	1,186	1,082	0.92
Leytonstone	633	FG	5,552	5,583	1.01	1,615	1,243	0.78
Liverpool Street	634	FG	21,459	21,589	1.01	24,266	26,609	1.10
London Bridge	635	FG	22,801	24,354	1.07	18,223	18,755	1.04
Loughton	636	FG	1,886	1,968	1.04	510	416	0.82
Maida Vale	637	FG	1,983	1,933	0.97	711	671	0.95
Manor House	638	FG	4,311	4,561	1.06	761	865	1.14
Mansion House	639	FG	430	383	0.89	4,657	5,131	1.11
Marble Arch	640	FG	1,843	1,748	0.95	4,538	4,692	1.04
Marylebone	641	FG	3,670	4,180	1.14	3,719	3,620	0.98
Mile End	642	FG	4,492	4,465	0.99	3,536	3,080	0.88
Mill Hill East	643	NFG	542	542	1.00	172	171	1.00
Moorgate	645	NFG	6,298	6,298	1.00	19,114	22,238	1.17

Moor Park	646	FG	521	531	1.02	194	208	1.08
Morden	647	FG	3,957	4,409	1.11	1,173	925	0.79
Mornington Crescent	648	FG	537	468	0.87	1,710	1,782	1.05
Neasden	649	FG	1,601	1,974	1.23	585	569	0.98
Newbury Park	650	FG	2,678	3,113	1.16	435	415	0.96
New Cross	651	NFG	2,095	2,095	1.00	903	897	1.00
New Cross Gate	652	NFG	2,636	2,636	1.00	726	721	1.00
North Acton	653	FG	1,544	1,701	1.10	2,525	2,582	1.03
North Ealing	654	FG	602	516	0.86	278	299	1.08
Northfields	655	FG	2,749	2,862	1.04	783	886	1.14
North Harrow	656	FG	1,283	1,294	1.01	159	149	0.94
Northolt	657	FG	2,204	2,316	1.05	440	428	0.98
North Wembley	659	NFG	685	685	1.00	288	286	1.00
Northwick Park	660	FG	1,681	1,675	1.00	1,185	947	0.81
Northwood	661	FG	1,350	1,317	0.98	407	499	1.24
Northwood Hills	662	FG	990	962	0.97	311	237	0.77
Notting Hill Gate	663	FG	5,214	5,370	1.03	4,690	4,735	1.02
Oakwood	664	FG	1,302	1,296	1.00	457	295	0.65
Old Street	665	NFG	2,496	2,924	1.17	9,483	10,598	1.13
Osterley	667	FG	1,309	1,296	0.99	404	366	0.91
Oval	668	FG	2,456	2,520	1.03	1,212	1,240	1.03
Oxford Circus	669	FG	1,881	2,194	1.17	36,051	37,459	1.05
Paddington	670	NFG	16,388	16,388	1.00	11,210	11,728	1.05
Park Royal	671	FG	482	491	1.02	750	709	0.95
Parsons Green	672	FG	2,984	2,773	0.93	2,048	1,994	0.98
Perivale	673	FG	1,207	1,276	1.06	473	358	0.76
Piccadilly Circus	674	FG	1,032	981	0.95	12,111	13,047	1.08
Pinner	675	NFG	1,952	2,004	1.03	200	199	1.00
Plaistow	676	FG	3,170	3,037	0.96	1,040	868	0.84
Preston Road	677	FG	2,299	2,338	1.02	263	272	1.04
Putney Bridge	678	FG	2,406	2,150	0.89	1,719	1,627	0.95
Queensbury	679	FG	2,290	2,475	1.08	281	273	0.98
Queens Park	680	NFG	3,496	3,496	1.00	868	988	1.15
Queensway	681	FG	2,213	2,220	1.00	1,113	1,261	1.14
Ravenscourt Park	682	FG	1,110	972	0.88	1,639	1,468	0.90
Rayners Lane	683	FG	2,680	2,627	0.98	550	548	1.00

Redbridge	684	FG	1,624	1,662	1.02	235	143	0.61
Regents Park	685	FG	0	0		0	0	
Richmond	686	NFG	6,404	6,404	1.00	2,706	2,687	1.00
Rickmansworth	687	NFG	1,022	1,168	1.14	595	711	1.20
Roding Valley	688	NFG	249	249	1.00	11	11	1.00
Rotherhithe	689	FG	896	929	1.04	144	126	0.88
Royal Oak	690	FG	714	664	0.93	442	491	1.12
Ruislip	691	FG	1,272	961	0.76	311	281	0.91
Ruislip Gardens	692	FG	698	758	1.09	153	107	0.71
Ruislip Manor	693	FG	1,140	1,021	0.90	365	224	0.62
Russell Square	694	FG	1,671	1,229	0.74	4,776	4,309	0.91
St James's Park	695	FG	1,295	1,647	1.27	11,940	11,850	1.00
St Johns Wood	696	FG	2,638	2,670	1.01	2,234	2,498	1.13
St Pauls	697	FG	728	752	1.03	10,302	11,201	1.10
Seven Sisters	698	NFG	9,992	9,992	1.00	2,941	2,920	1.00
Shadwell	699	FG	497	454	0.91	619	668	1.09
Shepherds Bush Ctl	700	FG	5,608	6,132	1.09	2,988	2,964	1.00
Shoreditch	701	NFG	0	0		0	0	
Sloane Square	702	FG	3,555	3,718	1.05	5,855	5,951	1.02
Snaresbrook	703	NFG	2,174	2,158	0.99	453	401	0.89
South Ealing	704	FG	2,194	2,146	0.98	581	523	0.91
Southfields	705	FG	4,475	4,642	1.04	823	687	0.84
Southgate	706	FG	2,853	3,040	1.07	870	823	0.95
South Harrow	707	FG	1,253	1,149	0.92	443	407	0.93
South Kensington	708	FG	4,256	4,225	0.99	12,419	11,902	0.97
South Kenton	709	NFG	474	474	1.00	44	44	1.00
South Ruislip	710	NFG	1,222	1,222	1.00	143	285	2.01
South Wimbledon	711	FG	2,132	2,412	1.13	884	1,049	1.19
South Woodford	712	FG	2,983	2,541	0.85	493	430	0.88
Stamford Brook	713	FG	1,707	1,567	0.92	647	713	1.11
Stanmore	714	FG	1,679	1,675	1.00	537	627	1.18
Stepney Green	715	FG	1,467	1,357	0.93	1,354	1,235	0.92
Stockwell	716	FG	5,459	5,911	1.08	1,332	1,209	0.91
Stonebridge Park	717	NFG	941	941	1.00	535	531	1.00
Charing Cross	718	FG	5,833	6,542	1.12	6,628	6,975	1.06
Stratford	719	NFG	17,024	17,024	1.00	7,900	7,973	1.02

Sudbury Hill	720	FG	1,417	1,413	1.00	407	471	1.16
Sudbury Town	721	NFG	1,379	1,566	1.14	216	227	1.06
Surrey Quays	722	FG	1,296	1,309	1.01	337	323	0.96
Swiss Cottage	723	FG	3,010	3,239	1.08	2,085	2,021	0.98
Temple	724	FG	330	353	1.07	6,151	5,782	0.95
Theydon Bois	725	FG	538	574	1.07	0	64	#DIV/0!
Tooting Bec	726	FG	5,185	5,258	1.01	823	952	1.17
Tooting Broadway	727	FG	6,387	6,994	1.10	2,485	2,767	1.12
Tottenham Court Rd	728	FG	1,320	1,105	0.84	14,466	14,559	1.01
Tottenham Hale	729	FG	4,703	4,713	1.00	1,331	1,411	1.07
Totteridge	730	NFG	1,157	1,182	1.02	228	275	1.21
Tower Hill	731	FG	7,179	6,415	0.89	8,019	7,462	0.94
Tufnell Park	733	FG	2,059	2,070	1.01	449	476	1.07
Turnham Green	734	FG	3,376	3,176	0.94	1,318	1,404	1.07
Turnpike Lane	735	FG	4,680	5,081	1.09	890	879	0.99
Upminster	736	NFG	2,327	2,327	1.00	568	564	1.00
Upminster Bridge	737	FG	646	646	1.00	308	338	1.10
Upney	738	NFG	1,059	1,164	1.10	469	466	1.00
Upton Park	739	FG	4,123	4,475	1.09	1,029	1,089	1.07
Uxbridge	740	FG	1,621	1,454	0.90	3,003	2,898	0.97
Victoria	741	FG	30,027	27,819	0.93	23,284	24,074	1.04
Walthamstow Central	742	FG	8,904	9,120	1.02	1,341	1,323	0.99
Wanstead	743	FG	1,580	1,525	0.97	379	263	0.70
Wapping	744	FG	788	759	0.96	325	332	1.03
Warren Street	745	FG	1,165	1,131	0.97	11,504	12,445	1.09
Warwick Avenue	746	FG	2,507	2,620	1.05	800	958	1.21
Waterloo	747	NFG	42,575	42,575	1.00	16,315	16,199	1.00
Watford Met	748	FG	795	940	1.18	583	568	0.98
Wembley Central	751	NFG	867	1,121	1.29	953	946	1.00
Wembley Park	752	FG	3,489	3,723	1.07	2,882	2,274	0.79
West Acton	753	FG	1,178	1,076	0.91	325	293	0.91
Westbourne Park	754	FG	1,419	1,339	0.94	805	800	1.00
West Brompton	755	NFG	1,263	1,263	1.00	1,511	1,814	1.21
West Finchley	756	NFG	1,180	1,180	1.00	55	80	1.47
West Ham	757	NFG	3,584	3,584	1.00	1,343	1,333	1.00
West Hampstead	758	FG	5,379	5,577	1.04	1,198	1,258	1.06

West Harrow	759	NFG	957	957	1.00	76	75	1.00
West Kensington	760	NFG	2,491	2,292	0.92	1,453	1,443	1.00
Westminster	761	FG	1,073	1,252	1.17	10,276	10,441	1.02
West Ruislip	762	NFG	1,012	1,012	1.00	171	170	1.00
Whitechapel	763	FG	2,261	2,373	1.05	4,954	5,468	1.11
White City	764	FG	1,265	1,386	1.10	4,936	5,736	1.17
Willesden Green	765	FG	5,421	5,325	0.98	728	784	1.09
Willesden Junction	766	NFG	1,341	1,341	1.00	1,514	1,503	1.00
Wimbledon	767	NFG	6,942	7,386	1.06	3,027	3,006	1.00
Wimbledon Park	768	FG	1,481	1,434	0.97	407	419	1.04
Woodford	769	FG	3,513	3,594	1.02	372	388	1.05
Wood Green	770	FG	4,848	5,301	1.09	1,569	1,635	1.05
Woodside Park	771	NFG	1,522	1,899	1.25	515	511	1.00
Hammersmith M	773	FG	1,429	1,911	1.34	1,609	1,952	1.22
Edgware Road B	774	FG	666	602	0.90	1,414	1,686	1.20
Shepherds Bush Met	775	FG	1,296	1,230	0.95	601	688	1.15
Pimlico	776	FG	2,581	2,509	0.97	4,285	4,104	0.96
Vauxhall	777	FG	9,084	9,373	1.03	7,181	7,161	1.00
Brixton	778	FG	12,392	13,295	1.07	2,857	2,630	0.93
Hatton Cross	779	FG	675	821	1.22	763	1,252	1.65
Heathrow Terms 123	780	NFG	1,637	1,637	1.00	2,472	2,454	1.00
Heathrow Term 4	781	FG	234	260	1.11	240	323	1.35
Southwark	784	FG	2,541	3,764	1.48	4,668	5,277	1.14
Bermondsey	787	FG	2,901	3,201	1.10	1,622	1,831	1.14
Canada Water	788	FG	4,757	5,024	1.06	1,123	1,080	0.97
North Greenwich	789	FG	6,436	6,944	1.08	1,567	2,354	1.51
Canary Wharf	852	FG	5,521	5,447	0.99	33,409	39,784	1.20
Canning Town	884	NFG	6,377	6,377	1.00	2,907	2,886	1.00

Appendix F 50 Largest Origin-Destination Pairs

From Station	To Station	From Zone	To Zone	RODS Journeys	Oyster-Based Journeys
Waterloo	Bank & Monument	1	1	11,256	8,981
Waterloo	Canary Wharf	1	2	4,775	7,884
London Bridge	Canary Wharf	1	2	7,641	4,172
Victoria	Oxford Circus	1	1	3,062	3,088
Stratford	Canary Wharf	3	2	1,481	2,054
Liverpool Street	Farringdon	1	1	1,845	2,048
Waterloo	Piccadilly Circus	1	1	2,027	2,021
Kings Cross	Victoria	1	1	1,103	1,744
Vauxhall	Oxford Circus	1	1	1,435	1,696
Kings Cross	Bank & Monument	1	1	1,302	1,677
London Bridge	Bond Street	1	1	1,383	1,672
Waterloo	Oxford Circus	1	1	2,145	1,614
Finsbury Park	Oxford Circus	2	1	1,469	1,583
Brixton	Oxford Circus	2	1	1,742	1,558
Victoria	Warren Street	1	1	1,989	1,514
Waterloo	Tottenham Court Rd	1	1	2,429	1,461
Waterloo	London Bridge	1	1	1,349	1,459
Waterloo	Bond Street	1	1	1,381	1,448
Bank & Monument	Liverpool Street	1	1	265	1,370
Kings Cross	Oxford Circus	1	1	1,282	1,354
London Bridge	Green Park	1	1	781	1,311
Victoria	South Kensington	1	1	1,359	1,308
North Greenwich	Canary Wharf	2	2	2,555	1,288
Victoria	Kings Cross	1	1	2,648	1,282
Liverpool Street	Holborn	1	1	1,513	1,242
Vauxhall	Green Park	1	1	1,335	1,161
Victoria	Green Park	1	1	1,233	1,111
Liverpool Street	Euston Square	1	1	966	1,079
Liverpool Street	Chancery Lane	1	1	1,684	1,057
Finsbury Park	Victoria	2	1	1,371	1,038
Waterloo	Paddington	1	1	276	1,036
Liverpool Street	Oxford Circus	1	1	973	1,030
Vauxhall	Kings Cross	1	1	897	1,016
Waterloo	Goodge Street	1	1	1,084	1,014
Kings Cross	Holborn	1	1	923	1,006

Highbury	Oxford Circus	2	1	799	998
Blackfriars	Bank & Monument	1	1	958	974
Charing Cross	Oxford Circus	1	1	531	972
Richmond	Bank & Monument	4+	1	566	969
Paddington	Oxford Circus	1	1	954	955
Liverpool Street	Tottenham Court Rd	1	1	1,052	952
Kings Cross	Old Street	1	1	739	938
London Bridge	Kings Cross	1	1	983	935
Liverpool Street	Barbican	1	1	498	926
Euston	Bank & Monument	1	1	1,815	916
Victoria	Euston	1	1	1,625	914
Euston	Victoria	1	1	736	909
London Bridge	Angel	1	1	796	897
Victoria	Blackfriars	1	1	551	889
Finsbury Park	Holborn	2	1	1,088	888
Brixton	Vauxhall	2	1	991	866
London Bridge	Moorgate	1	1	883	848
Kings Cross	Green Park	1	1	815	843
Victoria	Bank & Monument	1	1	237	843
Farringdon	Moorgate	1	1	1,788	841
Finsbury Park	Kings Cross	2	1	1,516	841
Vauxhall	Warren Street	1	1	387	835
Waterloo	Green Park	1	1	1,222	835
Kings Cross	London Bridge	1	1	422	825
London Bridge	Old Street	1	1	616	809
Highbury	Victoria	2	1	1,117	805

Appendix G Bibliography

Abkowitz, M., Slavin, H, Waksman, R., Englisher, L. and Wilson, N. (1978) *Transit Service Reliability, Report UMTA-MA-06-0049-78-1*. USDOT Transportation Systems Center. Cambridge, MA.

Barry, J, Newhouser, R., Rahbee, A. and Sayeda, S. (2002) *Origin and Destination Estimation in New York City Using Automated Fare System Data*. Transportation Research Record 1817, pp183-187.

Ben-Akiva, M. (1987) *Methods to Combine Different Data Sources and Estimate Origin-Destination Matrices*, Transportation and Traffic Theory, Gartner, N. and N.H.M. Wilson (Eds.), Elsevier, New York: Proceedings of the 10th International Symposium on Transportation and Traffic Theory, MIT, Cambridge, MA, July 9-10, 1987.

Cham, L. (2005) *Understanding Bus Service Reliability: A Practical Framework Using AVL/APC Data*. Thesis, Master of Science in Transportation, MIT.

Cui, A. (2006) *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*. Thesis, Master of Science in Transportation, MIT.

Gordillo, F. (2006) *The Value of Automated Fare Collection Data for Transit Planning: An Example of Rail Transit OD Matrix Estimation*. Thesis, Master of Science in Transportation and Master of Science in Technology and Policy, MIT.

Hong, Y. (2006) *Transition to Smart Card Technology: How Transit Operators Can Encourage the Take-Up of Smart Card Technology*. Thesis, Master of Science in Transportation, MIT.

Navick, D.S. and Furth, P. (1994) *Distance-Based Model for Estimating a Bus Route Origin-Destination Matrix*. Transportation Research Record 1433, pp16-23.

Rahbee, A. and Czerwinski, D. (2002) *Using Entry-Only Automatic Fare Collection Data to Estimate Rail Transit Passenger Flows*. CTA, Proceedings of the 2002 Transport Chicago Conference.

Transit Cooperative Research Program (2003) *TCRP Report 94: Fare policies, structures and technologies: Update* on http://www.tcrponline.org

Wilson, N., Zhao, J. and Rahbee, A. (2005) *The Potential Impact of Automated Data Collection Systems on Urban Public Transport Planning*. TRAIL Research School, Delft, May 2005

Transport for London (2006), *London Travel Report 2005* on http://www.tfl.gov.uk

Zhao, J. (2004) *The Planning and Analysis Implications of Automated Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modelling Examples*. Thesis, Master of City Planning/Master of Science in Transportation, MIT.

Zhao, J., Wilson, N. and Rahbee, A. (2006) *Estimating Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection.* Computer-Aided Civil and Infrastructure Engineering, to be published.